Signal Processing and Networks Optimization
Part IV: Proximity operator

Pierre Borgnat[1], Jean-Christophe Pesquet[2], Nelly Pustelnik[1]

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

[2] LIGM – Univ. Paris-Est – CNRS UMR 8049
jean-christophe.pesquet@univ-paris-est.fr

Motivation

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

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

- Gradient descent algorithm

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

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

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

- **Alternative algorithm**

  Set $\gamma \in \,]0,+\infty[$ and $x_0 \in \mathcal{H}$.
  For $n = 0, 1 \ldots$

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

Questions:

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

Proximity operator: definition

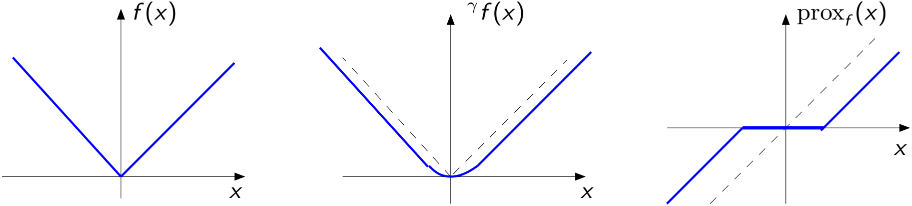
Let $\mathcal{H}$ be a Hilbert space. Let $f \in \Gamma_0(\mathcal{H})$.

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

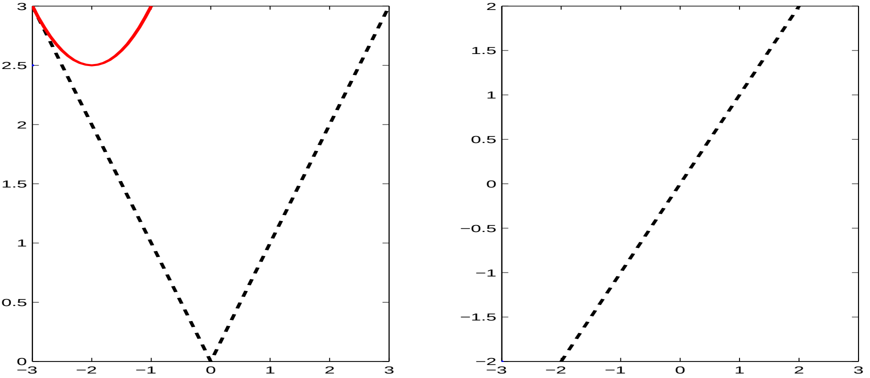
$$^{\gamma}f \colon \mathcal{H} \to \mathbb{R} \colon x \mapsto \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma}\|y - x\|^2.$$

- The **proximity operator** of f is

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



Proximity operator: definition



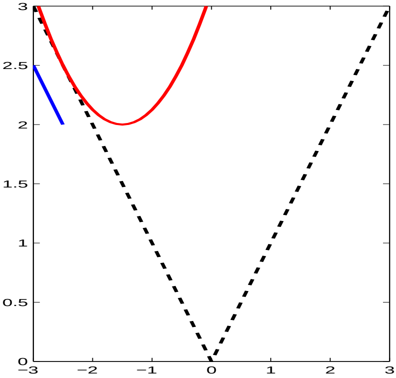
Moreau envelope

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

Proximity operator

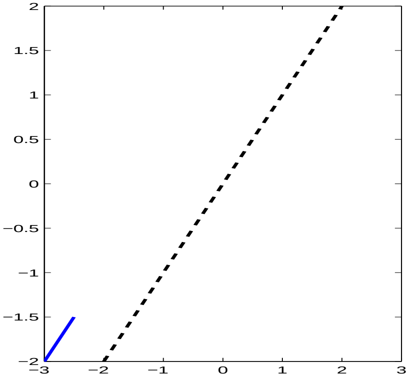
$\mathrm{prox}_f(x) = \underset{y\in\mathcal{H}}{\operatorname{argmin}}\; f(y) + \frac{1}{2}\|y-x\|^2$

Proximity operator: definition



Moreau envelope

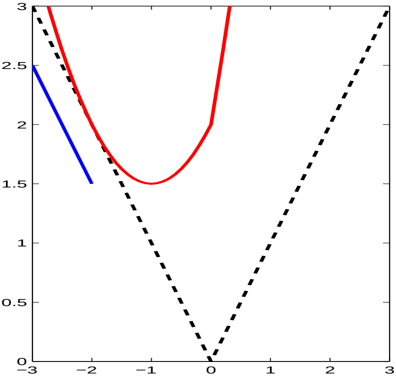
$^{\gamma}f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma}\|y - x\|^2$



Proximity operator

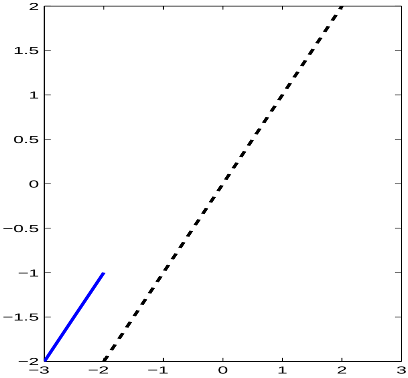
$\mathrm{prox}_f(x) = \underset{y \in \mathcal{H}}{\mathrm{argmin}} \;\; f(y) + \frac{1}{2}\|y - x\|^2$

Proximity operator: definition



Moreau envelope

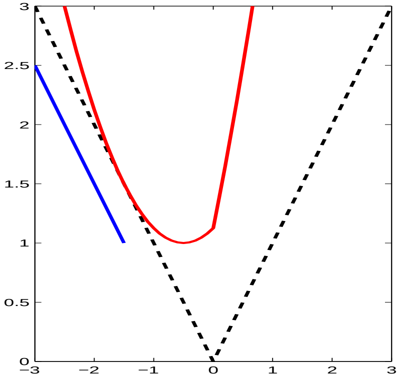
$${}^{\gamma}f(x) = \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|y-x\|^2$$



Proximity operator

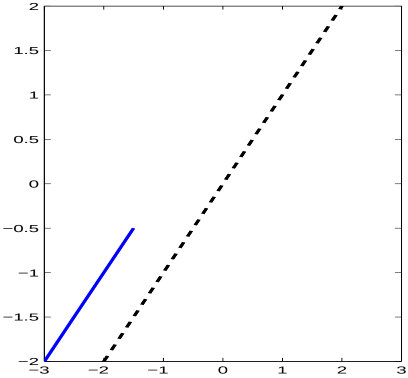
$$\mathrm{prox}_f(x) = \operatorname*{argmin}_{y\in\mathcal{H}} \; f(y) + \frac{1}{2}\|y-x\|^2$$

Proximity operator: definition



Moreau envelope

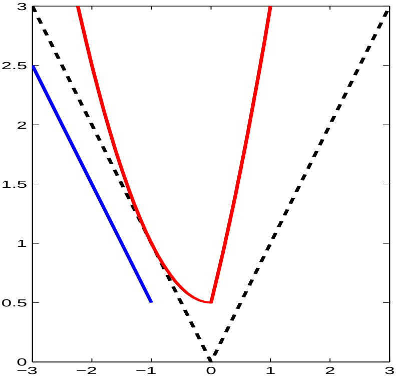
$^{\gamma}f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma}\|y - x\|^2$



Proximity operator

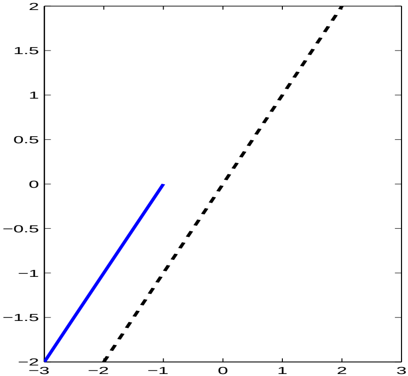
$$\mathrm{prox}_f(x) = \underset{y \in \mathcal{H}}{\mathrm{argmin}} \; f(y) + \frac{1}{2}\|y - x\|^2$$

Proximity operator: definition



Moreau envelope

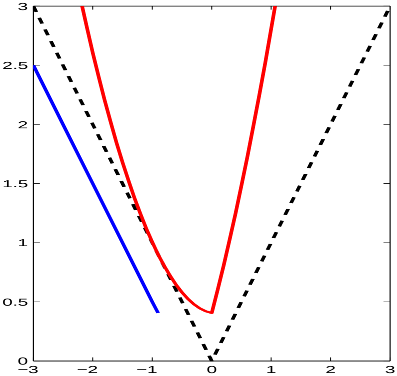
$^{\gamma}f(x) = \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|y-x\|^2$



Proximity operator

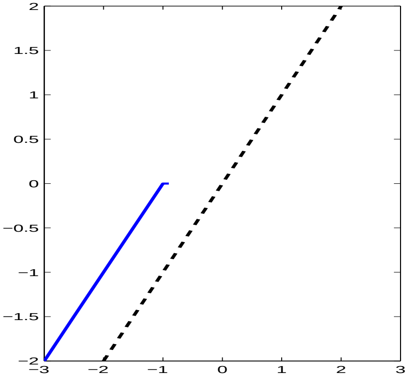
$\mathrm{prox}_f(x) = \underset{y\in\mathcal{H}}{\text{argmin}}\ f(y) + \frac{1}{2}\|y-x\|^2$

Proximity operator: definition



Moreau envelope

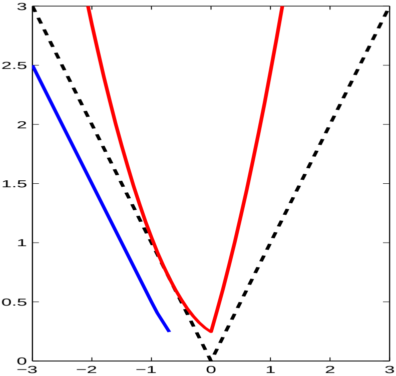
$${}^{\gamma}f(x) = \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|y-x\|^2$$



Proximity operator

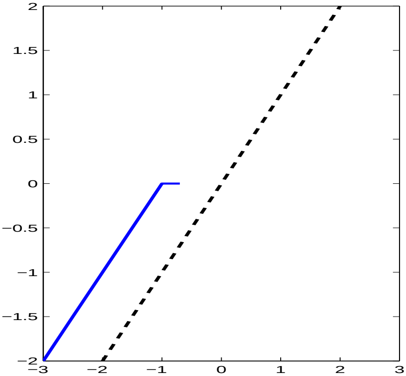
$$\mathrm{prox}_f(x) = \underset{y\in\mathcal{H}}{\operatorname{argmin}}\ f(y) + \frac{1}{2}\|y-x\|^2$$

Proximity operator: definition



Moreau envelope

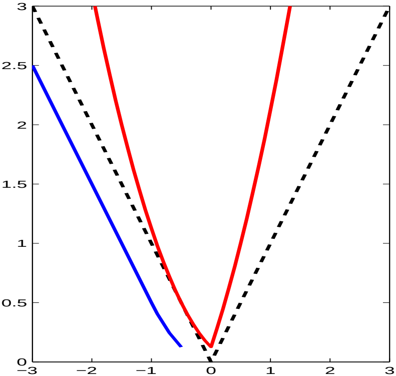
$$^{\gamma}f(x) = \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|y-x\|^2$$



Proximity operator

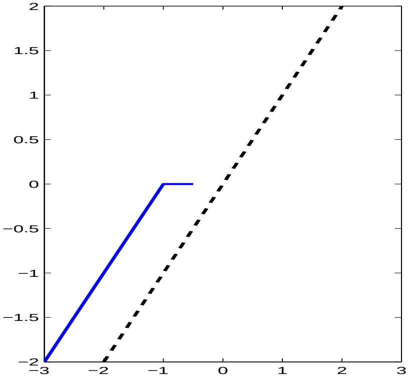
$$\mathrm{prox}_f(x) = \underset{y\in\mathcal{H}}{\text{argmin}}\ \ f(y) + \frac{1}{2}\|y-x\|^2$$

Proximity operator: definition



Moreau envelope

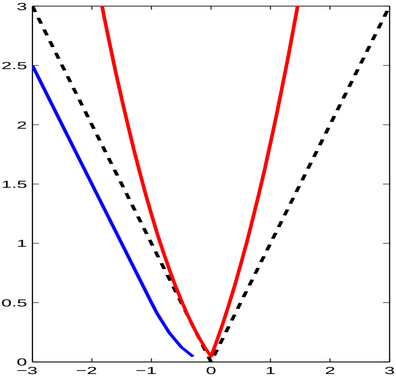
$^{\gamma}f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma}\|y - x\|^2$



Proximity operator

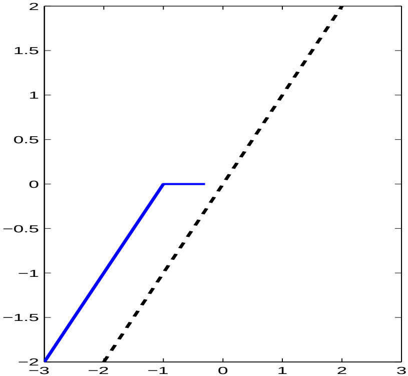
$\mathrm{prox}_f(x) = \underset{y \in \mathcal{H}}{\mathrm{argmin}} \; f(y) + \frac{1}{2}\|y - x\|^2$

Proximity operator: definition



Moreau envelope

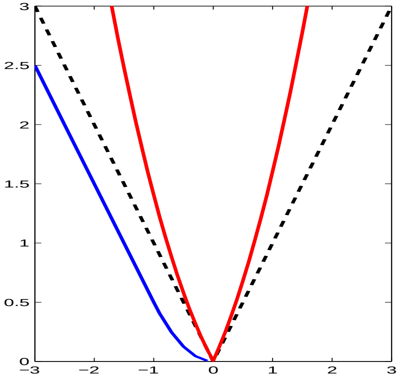
$$^{\gamma}f(x) = \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|y-x\|^2$$



Proximity operator

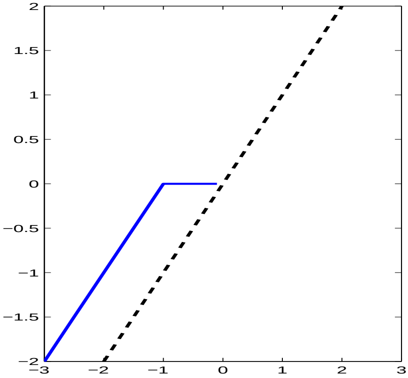
$$\mathrm{prox}_f(x) = \underset{y\in\mathcal{H}}{\operatorname{argmin}}\ f(y) + \frac{1}{2}\|y-x\|^2$$

Proximity operator: definition



Moreau envelope

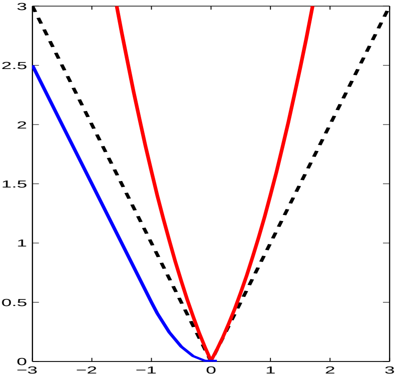
$^{\gamma}f(x) = \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|y-x\|^2$



Proximity operator

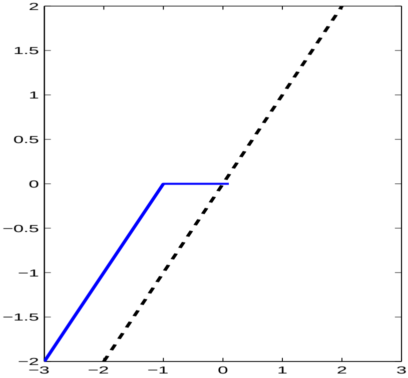
$\mathrm{prox}_f(x) = \underset{y\in\mathcal{H}}{\mathrm{argmin}}\ \ f(y) + \frac{1}{2}\|y-x\|^2$

Proximity operator: definition



Moreau envelope

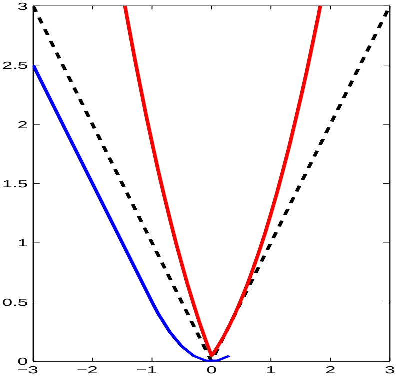
$${}^{\gamma}f(x) = \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|y-x\|^2$$



Proximity operator

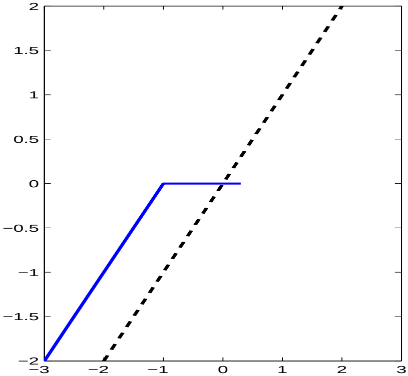
$$\mathrm{prox}_f(x) = \operatorname*{argmin}_{y\in\mathcal{H}} \; f(y) + \frac{1}{2}\|y-x\|^2$$

Proximity operator: definition



Moreau envelope

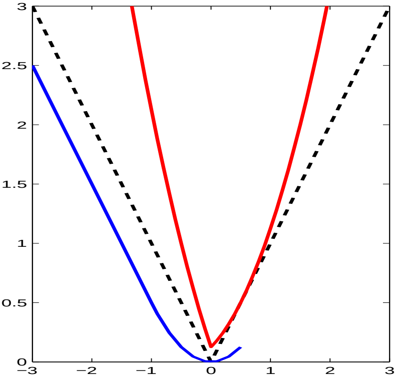
$^{\gamma}f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma}\|y - x\|^2$



Proximity operator

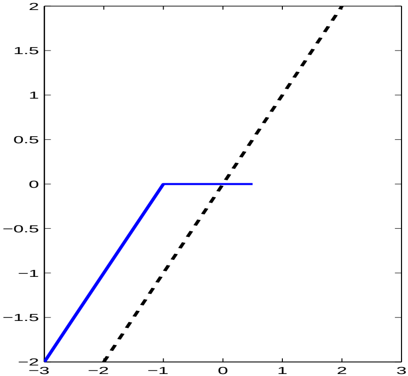
$\mathrm{prox}_f(x) = \underset{y \in \mathcal{H}}{\mathrm{argmin}} \; f(y) + \frac{1}{2}\|y - x\|^2$

Proximity operator: definition



Moreau envelope

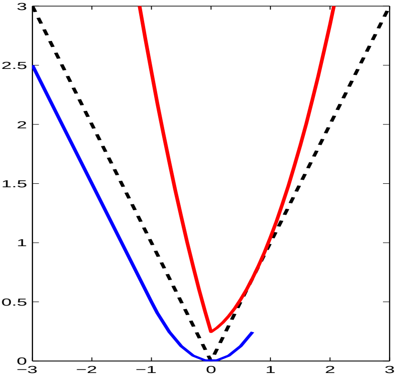
$^{\gamma}f(x) = \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|y-x\|^2$



Proximity operator

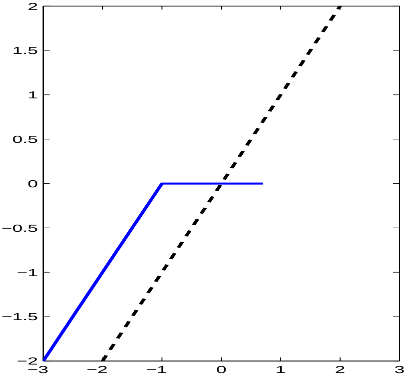
$\mathrm{prox}_f(x) = \underset{y\in\mathcal{H}}{\mathrm{argmin}}\ f(y) + \frac{1}{2}\|y-x\|^2$

Proximity operator: definition



Moreau envelope

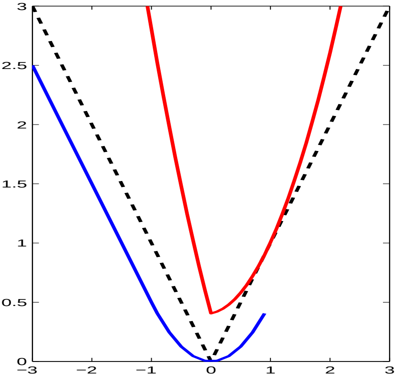
$^{\gamma}f(x) = \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|y - x\|^2$



Proximity operator

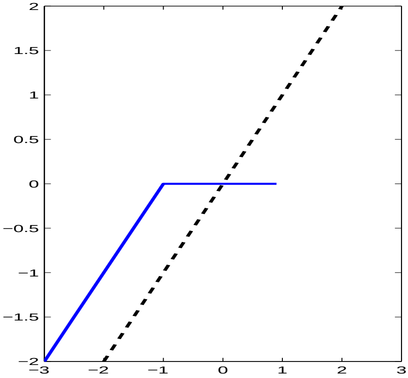
$\operatorname{prox}_f(x) = \underset{y\in\mathcal{H}}{\operatorname{argmin}} \; f(y) + \frac{1}{2}\|y - x\|^2$

Proximity operator: definition



Moreau envelope

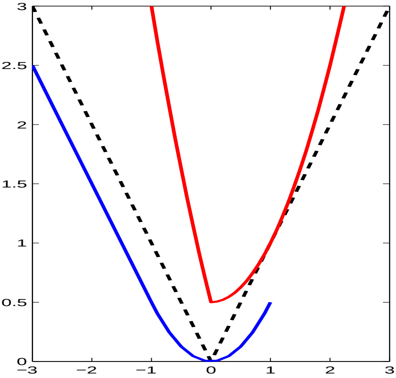
$^{\gamma}f(x) = \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|y-x\|^2$



Proximity operator

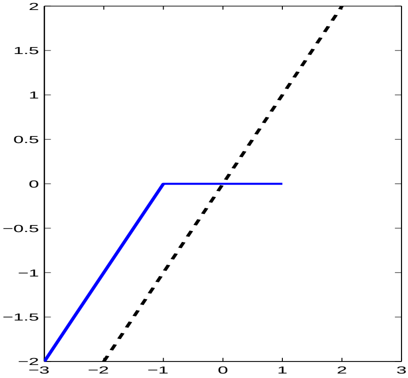
$\mathrm{prox}_f(x) = \underset{y\in\mathcal{H}}{\operatorname{argmin}} \; f(y) + \frac{1}{2}\|y-x\|^2$

Proximity operator: definition



Moreau envelope

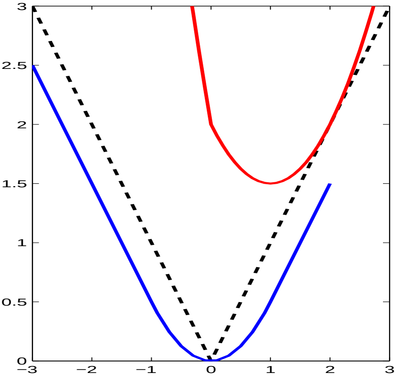
$$^{\gamma}f(x) = \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|y-x\|^2$$



Proximity operator

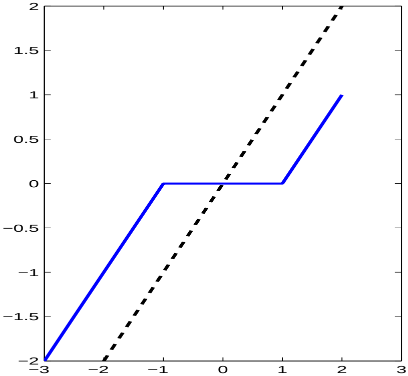
$$\mathrm{prox}_f(x) = \underset{y\in\mathcal{H}}{\mathrm{argmin}}\;\; f(y) + \frac{1}{2}\|y-x\|^2$$

Proximity operator: definition



Moreau envelope

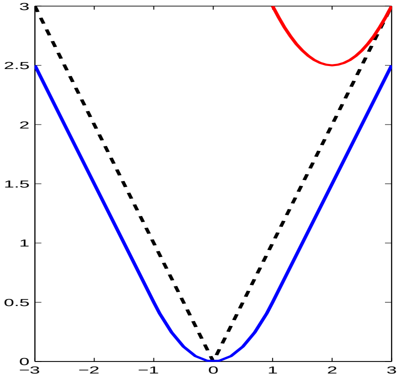
$^{\gamma}f(x) = \inf_{y\in\mathcal{H}} f(y) + \frac{1}{2\gamma}\|y-x\|^2$



Proximity operator

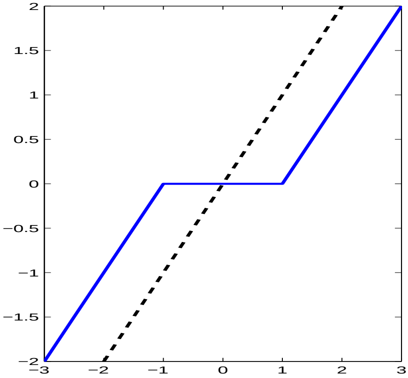
$\mathrm{prox}_f(x) = \underset{y\in\mathcal{H}}{\operatorname{argmin}}\; f(y) + \frac{1}{2}\|y-x\|^2$

Proximity operator: definition



Moreau envelope

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



Proximity operator

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

Proximity operator: definition

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

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

Proximity operator: definition

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

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

Proof: By using Fermat's rule, for every $x \in \mathcal{H}$, if and only if

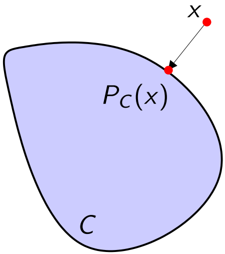
$$\begin{aligned}
& p = \underset{y \in \mathcal{H}}{\arg\min}\; f(y) + \frac{1}{2}\|y - x\|^2 \\
\Leftrightarrow \quad & 0 \in \partial\Big(f + \frac{1}{2}\|\cdot - x\|^2\Big)(p) \\
\Leftrightarrow \quad & 0 \in \partial f(p) + p - x \\
\Leftrightarrow \quad & x \in (\mathrm{Id} + \partial f)(p).
\end{aligned}$$

Proximity operator: examples

Projection :

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

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



Proximity operator: examples

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

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

Remark :

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

Proximity operator: examples

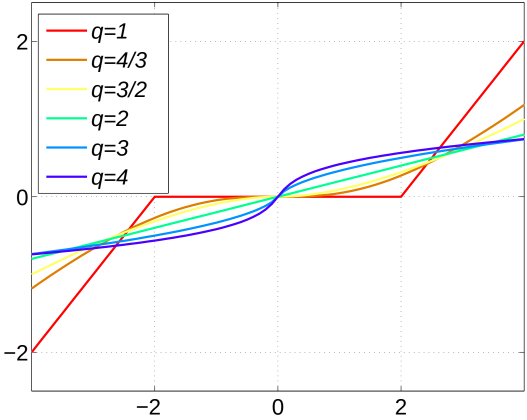
Power q function with $q \geq 1$:

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

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

Proximity operator: examples

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



Proximity operator: examples

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

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

- Exercice : Prove this property.

Proximity operator: properties

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

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

Exercice

Establish all the properties in the previous table.

Proximity operator: properties

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

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

then

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

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

Proximity operator: properties

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

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

then

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

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

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

Proximity operator: properties

Moreau decomposition formula

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

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

Proof:

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

Proximity operator: properties

Moreau decomposition formula

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

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

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

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

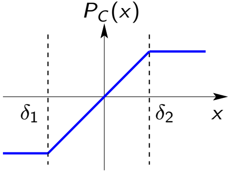
Proximity operator: examples

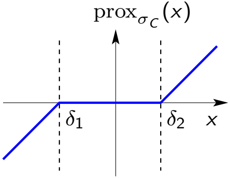
Support function :
Let $\mathcal{H}$ be a Hilbert space and $C \subset \mathcal{H}$ be nonempty closed convex.

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

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

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





Proximity operator: properties

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

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

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 $\operatorname{ran} L = \mathcal{H}$. Then

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

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

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

Proximity operator: properties

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

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