STAT 593
Smoothing

Joseph Salmon
http://josephsalmon.eu

Télécom Paristech, Institut Mines-Télécom
&
University of Washington, Department of Statistics
(Visiting Assistant Professor)

Outline

Inf convolution

Moreau envelope

Table of Contents

Inf convolution

Moreau envelope

Inf convolution / epigraph addition

Definition

The **inf-convolution** of two functions $f_1, f_2 : \mathbb{R}^d \mapsto \mathbb{R}$ is denoted $f_1 \,\square\, f_2$, where

$$(f_1 \,\square\, f_2)(x) = \inf_{(u_1,u_2)\in\mathbb{R}^d\times\mathbb{R}^d} \{f_1(u_1) + f_2(u_2) : u_1 + u_2 = x\}$$

$$(f_1 \,\square\, f_2)(x) = \inf_{u\in\mathbb{R}^d} \{f_1(u) + f_2(x-u)\}$$

- $f_1 \,\square\, f_2 = f_2 \,\square\, f_1$
- For closed convex f_1 and f_2, provided they share a common affine minorant, then $f_1 \,\square\, f_2$ is a convex closed function s.t.

$$\mathrm{epi}(f_1 \,\square\, f_2) = \mathrm{epi}(f_1) + \mathrm{epi}(f_2)$$

Properties

<u>Notation</u>: $d_{\mathcal{C}}(x) = \inf_{c \in \mathcal{C}} \|x - c\|, \quad \iota_{\mathcal{C}}(x) = \begin{cases} 0 & \text{if } x \in \mathcal{C} \\ +\infty & \text{if } x \notin \mathcal{C} \end{cases}$

f_1	f_2	$f_1 \square f_2$
f	0	$\inf_{x \in \mathbb{R}^d} f(x)$
$\iota_{\mathcal{C}}$	$\| \cdot \|$	$d_{\mathcal{C}}$
$\iota_{\mathcal{C}}$	$\iota_{\mathcal{D}}$	$\iota_{\mathcal{C}+\mathcal{D}}$
f	$\iota_{\{x\}}$	$f(\cdot - x)$
f	$\langle s , \cdot \rangle$	$\langle s , \cdot \rangle - f^*(s)$
f	f	$2f(\frac{\cdot}{2})$ (f convex)

Fenchel and inf-convolution[1]

Theorem

For any function f and g, one has

$$(f \,\Box\, g)^* = f^* + g^*$$

Proof: simply write the definition

[1] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. New York: Springer, 2011, pp. xvi+468.

Table of Contents

Inf convolution

Moreau envelope

Moreau envelope[2]

Definition

The Moreau envelope of the function f of parameter $\gamma > 0$ is

$$^{\gamma}f := f \,\square \left(\frac{1}{2\gamma} \| \cdot \|^2 \right)$$

i.e.,

$$^{\gamma}f(x) := \inf_{u \in \mathbb{R}^d} \{ f(u) + \frac{1}{2\gamma} \|x - u\|^2 \}$$

Rem: when f is closed and convex the $\inf$ is a $\min$

Rem: γ has the role of a smoothing parameter

[2] J.-J. Moreau. "Fonctions convexes duales et points proximaux dans un espace hilbertien". In: *C. R. Acad. Sci. Paris* 255 (1962), pp. 2897–2899.

Pinball case

For $s_1 \leq 0 \leq s_2$ let us define the general pinball loss :

$$\ell_{s_1,s_2}(x) = \begin{cases} s_1 x & \text{if } x \leq 0 \\ s_2 x & \text{if } x \geq 0 \end{cases}$$

Then, writing $f = \ell_{s_1,s_2}$ and $g = \frac{|\cdot|^2}{2\gamma}$, the Moreau envelope ${}^\gamma f$ is defined for any $x \in \mathbb{R}$ by

$${}^\gamma f(x) = (f \,\square\, g)(x) = = \quad \begin{cases} s_1 x - \gamma \frac{s_1^2}{2}, & \text{if } x < s_1 \\ \frac{1}{2\gamma} x^2, & \text{if } x \in [\gamma s_1, \gamma s_2] \\ s_2 x - \gamma \frac{s_2^2}{2}, & \text{if } x > s_2 \end{cases}$$

Rem: a classical example is the Huber function, when one considers the absolution value function $|\cdot| = \ell_{-1,1}$:

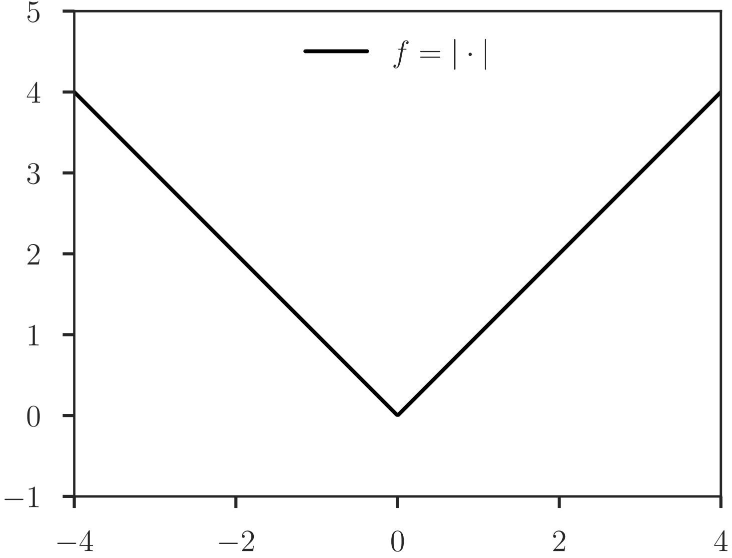
Intermission

Movie!

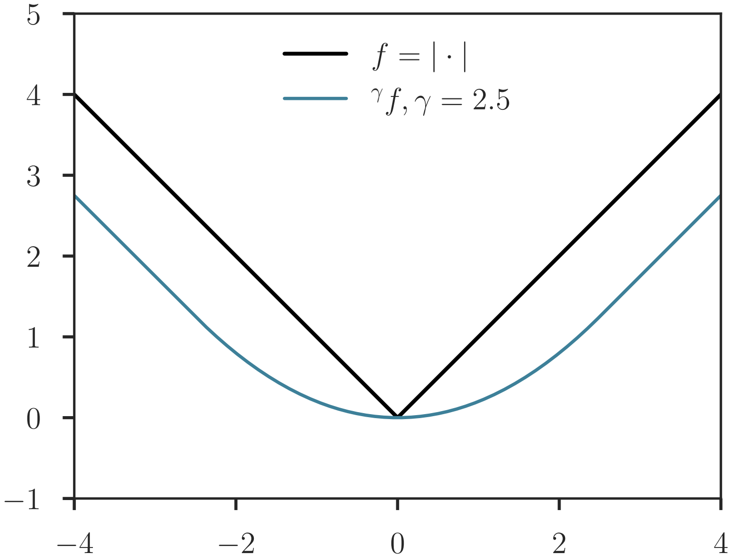
Proof

- if $u < 0$: the Fermat rule leads to $u = x - \gamma s_1$ (and $x < \gamma s_1$) so $(f \,\square\, g)(x) = s_1(x - \gamma s_1) + \gamma \frac{s_1^2}{2} = s_1 x - \gamma \frac{s_1^2}{2}$
- if $u > 0$: the Fermat rule leads to $u = x - \gamma s_2$ (and $x > \gamma s_2$) $(f \,\square\, g)(x) = s_2(x - \gamma s_2) + \frac{\gamma s_2^2}{2} = s_2 x - \gamma \frac{s_2^2}{2}$
- if $u = 0$: the Fermat rule and noticing that $\partial \ell_{s1,s_2}(0) = [s_1, s_2]$ leads to $\frac{x-u}{\gamma} = \frac{x}{\gamma} \in [s_1, s_2]$.

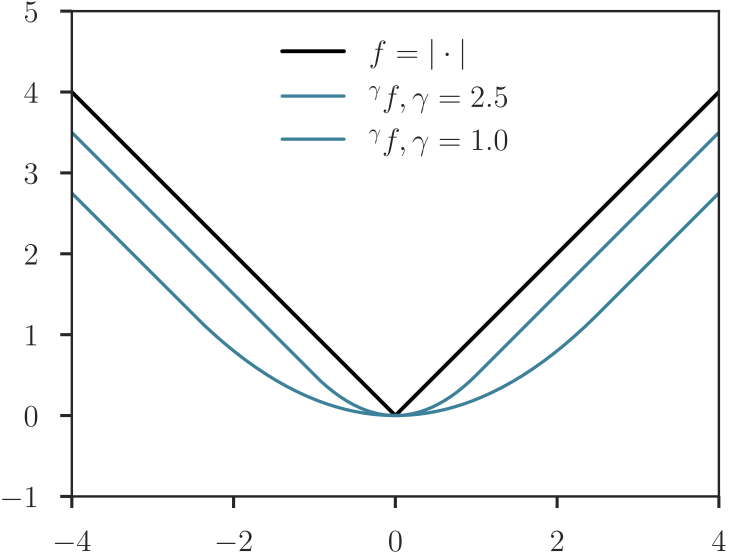
Influence of the smoothing parameter γ



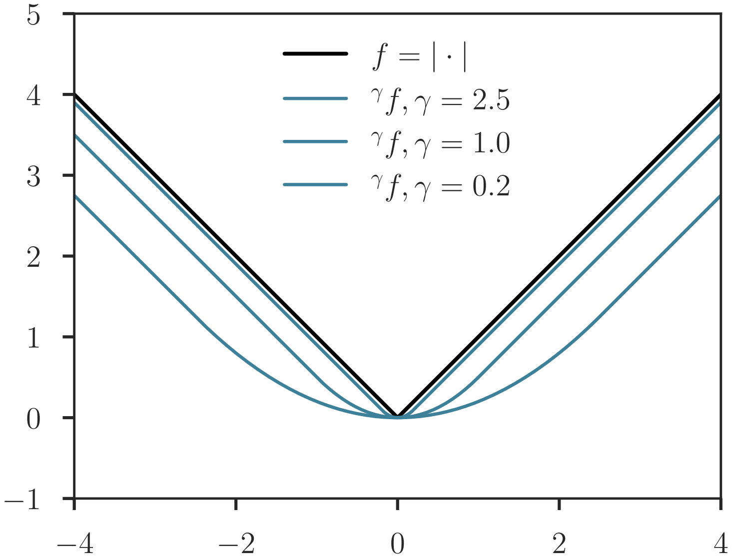
Influence of the smoothing parameter γ



Influence of the smoothing parameter γ



Influence of the smoothing parameter γ



Sharing inf/minimum

Theorem

A function f and its Moreau envelopes share the same minima:

$$\forall \gamma > 0, \quad \inf_x \left(({}^{\gamma} f)(x) \right) = \inf_x f(x)$$

Proof:

$$\begin{aligned} \inf_x \left(({}^{\gamma} f)(x) \right) &= \inf_x \inf_y \left\{ f(y) + \frac{1}{2\gamma} \|x - y\|^2 \right\} \\ &= \inf_y \inf_x \left\{ f(y) + \frac{1}{2\gamma} \|x - y\|^2 \right\} \\ &= \inf_y f(y) \end{aligned}$$

□

Proximal / Moreau

Recall for the convex case:

$$
\begin{aligned}
{}^{\gamma}f(x) &:= \min_{u \in \mathbb{R}^d} \left\{ f(u) + \frac{1}{2\gamma} \left\| x - u \right\|^2 \right\} \\
\operatorname{prox}_{\gamma f}(x) &:= \operatorname*{arg\,min}_{u \in \mathbb{R}^d} \left\{ f(u) + \frac{1}{2\gamma} \left\| x - u \right\|^2 \right\}
\end{aligned}
$$

Moreau decomposition: $\mathrm{Id} = \operatorname{prox}_f + \operatorname{prox}_{f^*}$

Link proximal operators / Moreau envelopes

Note that by rearranging terms

$$
\begin{aligned}
{}^{\gamma}f(x) &= \frac{1}{2\gamma}\|x\|^2 - \frac{1}{\gamma}\sup_y\left(\langle\, x\,,\, y\,\rangle - \gamma f(y) - \frac{1}{2}\|y\|^2\right) \\
&= \frac{1}{2\gamma}\|x\|^2 - \frac{1}{\gamma}\left(\gamma f + \frac{1}{2}\|\cdot\|^2\right)^*(x)
\end{aligned}
$$

Hint: remind that

$s \in \arg\max_{t\in\mathbb{R}^d}\langle t, x\rangle - f^*(t) \iff s \in \partial f(x) \iff x \in \partial f^*(s)$

$$
\begin{aligned}
\nabla\, {}^{\gamma}f(x) &= \frac{x}{\gamma} - \frac{1}{\gamma}\arg\max_y\left(\langle\, x\,,\, y\,\rangle - \gamma f(y) - \frac{1}{2}\|y\|^2\right) \\
&= \frac{1}{\gamma}(x - \operatorname{prox}_{\gamma f}(x))
\end{aligned}
$$

Consequence: ${}^{\gamma}f$ is $\frac{1}{\gamma}$-Lipschitz !

Smoothing generalization
Nesterov (2005), Beck and Teboulle (2012)

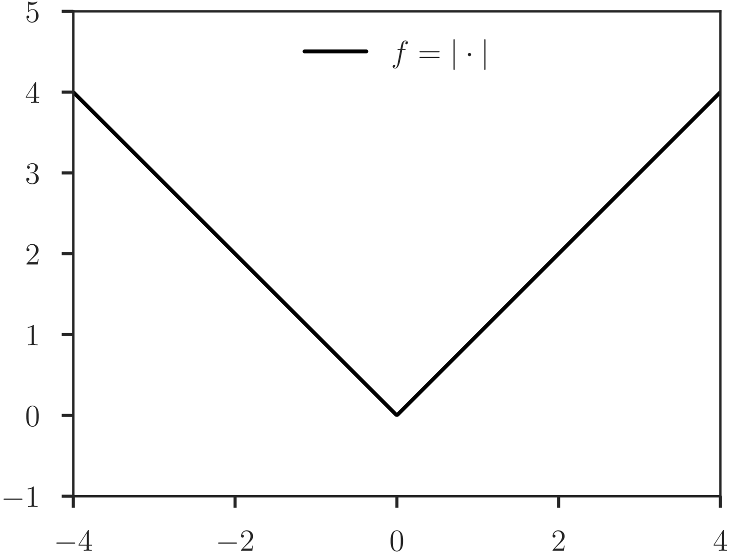
Motivation: smooth a non-smooth function f to ease optimization

Smoothing step: for $\gamma > 0$, a "smoothed" version of f is ${}^{\gamma}_{\omega}f$

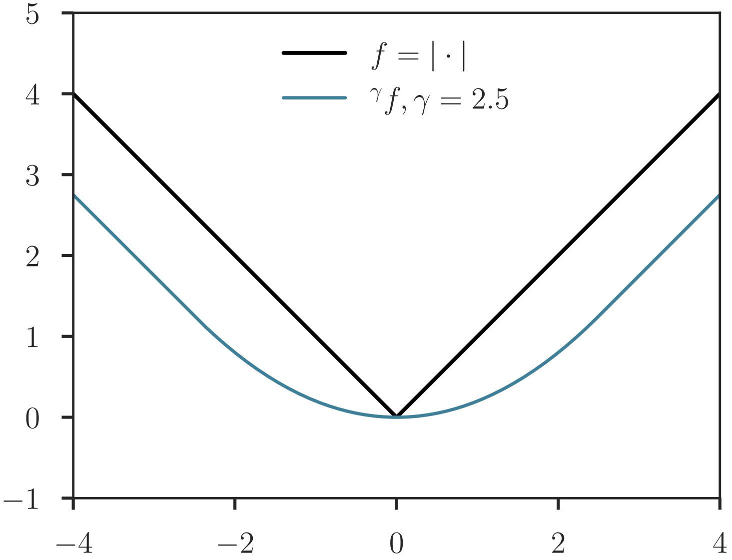
$$
\begin{aligned}
{}^{\gamma}_{\omega}f &= \gamma\omega\left(\frac{\cdot}{\gamma}\right) \Box f \\
&= (f^* + \gamma w^*)^*
\end{aligned}
$$

- $\omega = \frac{1}{2}\|\cdot\|^2$ recovers the Moreau envelop and we can drop the index ω
- ω is a predefined smooth function (s.t. $\nabla\omega$ is L_w Lipschitz)

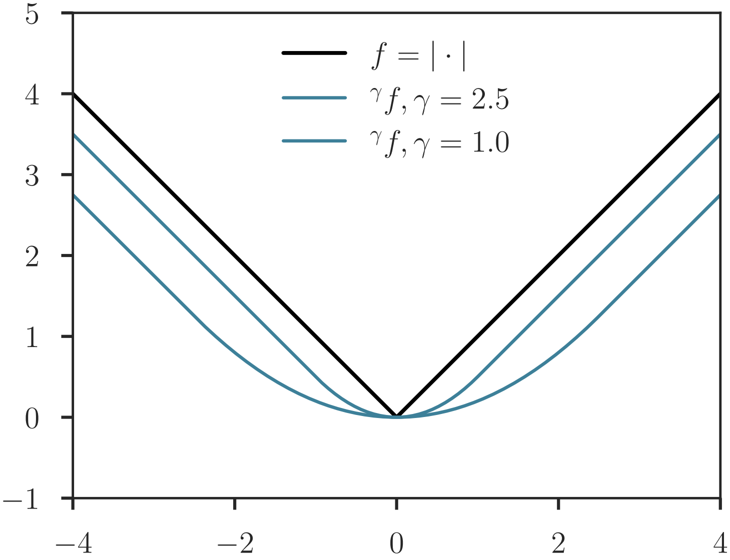
Huber function: $\omega(t) = \frac{t^2}{2}$



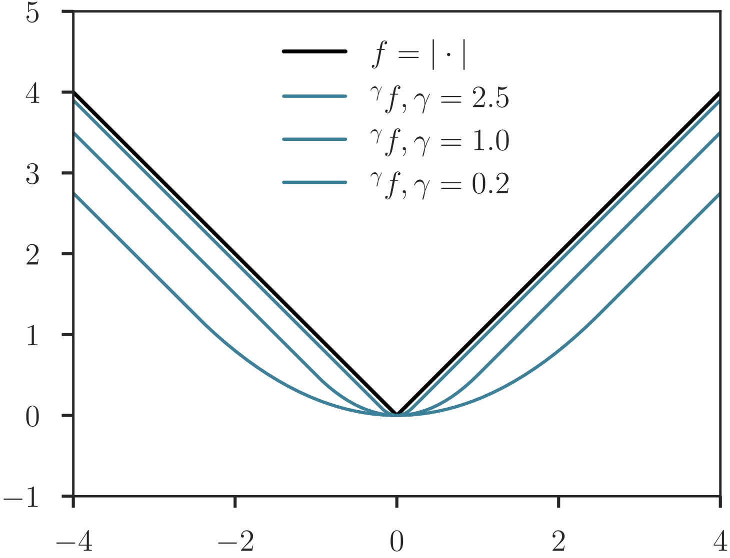
Huber function: $\omega(t) = \frac{t^2}{2}$



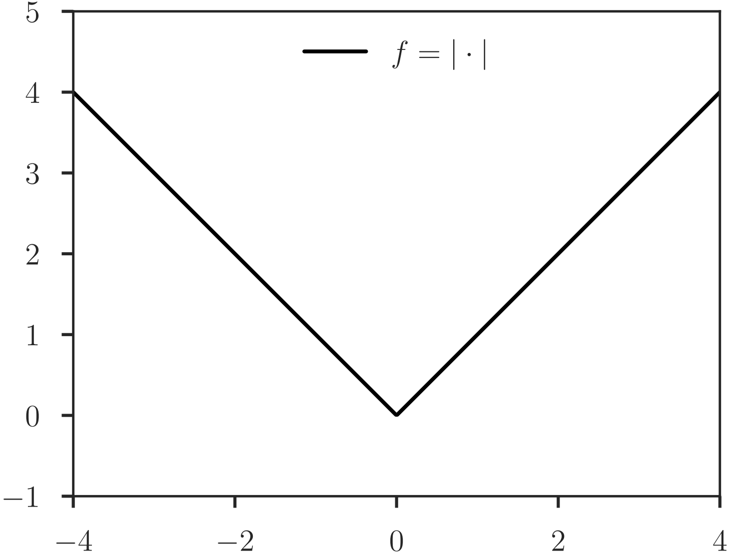
Huber function: $\omega(t) = \frac{t^2}{2}$



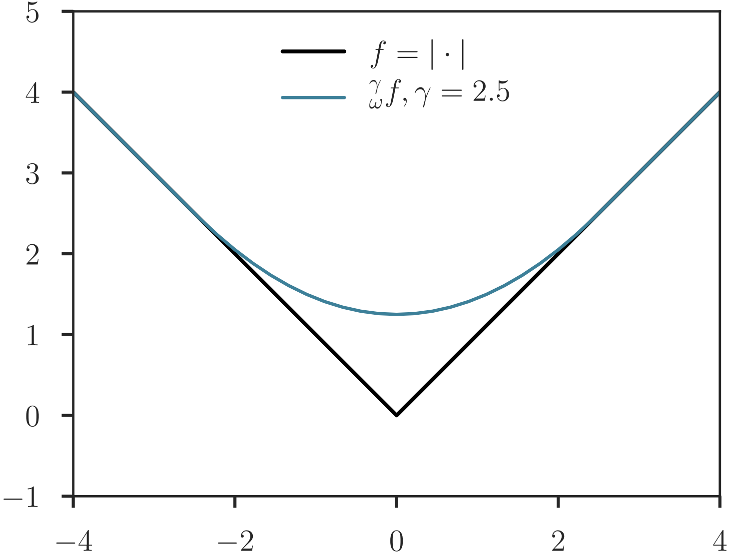
Huber function: $\omega(t) = \frac{t^2}{2}$



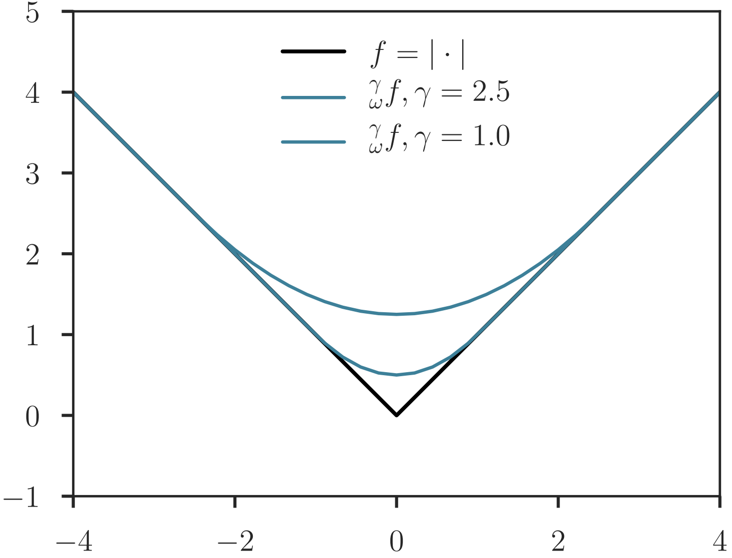
Huber function (bis): $\omega(t) = \frac{t^2}{2} + \frac{1}{2}$



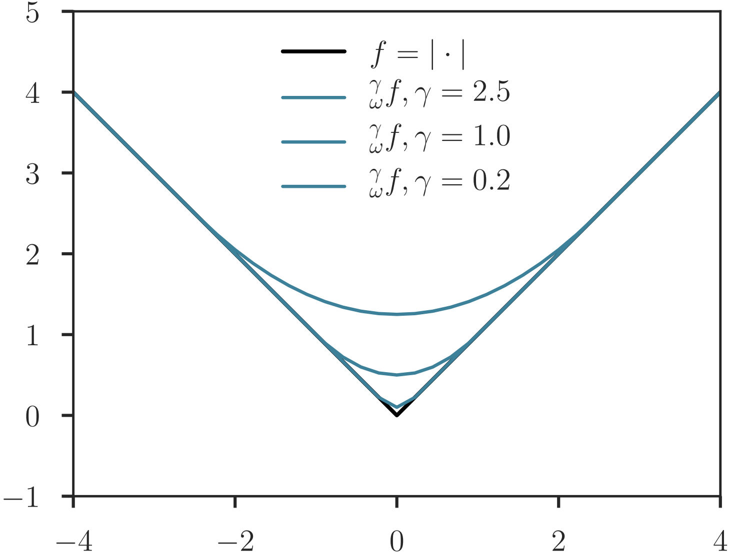
Huber function (bis): $\omega(t) = \frac{t^2}{2} + \frac{1}{2}$



Huber function (bis): $\omega(t) = \frac{t^2}{2} + \frac{1}{2}$



Huber function (bis): $\omega(t) = \frac{t^2}{2} + \frac{1}{2}$



Link: Fourier / Legendre transforms and kernel smoothing

Kernel smoothing analogy:

Fourier/Laplace : $\mathcal{F}(f)$	Fenchel/Legendre: f^*
convolution: $\star$	inf-convolution: $\square$
$\mathcal{F}(f \star g) = \mathcal{F}(f) \cdot \mathcal{F}(g)$	$(f \,\square\, g)^* = f^* + g^*$
Gaussian : $\mathcal{F}(g) = g$	$\omega = \frac{\|\cdot\|^2}{2}$: $\quad \omega^* = \omega$
$f_h = \frac{1}{h} g\left(\frac{\cdot}{h}\right) \star f$	${}^{\gamma}_{\omega}f = \gamma\omega\left(\frac{\cdot}{\gamma}\right) \square\, f$

Smoothing and approximation[3]

Theorem

Let f be a closed function with:

$$ {}^{\gamma}_{\omega}f = \gamma\omega\left(\frac{\cdot}{\gamma}\right) \square f $$

Then, for any $\gamma > 0$ and $x \in \mathbb{R}^d$, one has

$$ f(x) - \gamma\omega^*(g_x) \leq {}^{\gamma}_{\omega}f(x) \leq f(x) + \gamma\omega(0) $$

where $g_x \in \partial f(x)$.

[3] A. Beck and M. Teboulle. "Smoothing and first order methods: A unified framework". In: *SIAM J. Optim.* 22.2 (2012), pp. 557–580.

Proof

Fact 1: ${}^{\gamma}_{\omega}f(x) = \inf_{y\in\mathbb{R}^d}\left\{f(y) + \gamma\omega\left(\frac{x-y}{\gamma}\right)\right\} \le f(x) + \gamma\omega(0)$

(1)

(2)

Proof

Fact 1: ${}^{\gamma}_{\omega}f(x) = \inf_{y\in\mathbb{R}^d}\left\{f(y) + \gamma\omega\left(\frac{x-y}{\gamma}\right)\right\} \leq f(x) + \gamma\omega(0)$
Fact 2:

(1)

(2)

Proof

Fact 1: ${}^{\gamma}_{\omega}f(x) = \inf_{y\in\mathbb{R}^d}\left\{f(y) + \gamma\omega\left(\frac{x-y}{\gamma}\right)\right\} \leq f(x) + \gamma\omega(0)$
Fact 2:

$$ {}^{\gamma}_{\omega}f(x) - f(x) = \inf_{y\in\mathbb{R}^d}\left\{f(y) - f(x) + \gamma\omega\left(\frac{x-y}{\gamma}\right)\right\} $$

(1)

(2)

Proof

Fact 1: ${}^{\gamma}_{\omega}f(x) = \inf_{y\in\mathbb{R}^d} \left\{ f(y) + \gamma\omega\left(\frac{x-y}{\gamma}\right) \right\} \leq f(x) + \gamma\omega(0)$
Fact 2:

$$
\begin{aligned}
{}^{\gamma}_{\omega}f(x) - f(x) &= \inf_{y\in\mathbb{R}^d} \left\{ f(y) - f(x) + \gamma\omega\left(\frac{x-y}{\gamma}\right) \right\} \\
&\overset{(1)}{\geq} \inf_{y\in\mathbb{R}^d} \left\{ \langle\, g_x \,,\, y - x \,\rangle + \gamma\omega\left(\frac{x-y}{\gamma}\right) \right\}
\end{aligned}
$$

(1) sub-gradient definition at point x

(2)

Proof

Fact 1: ${}^{\gamma}_{\omega}f(x) = \inf_{y\in\mathbb{R}^d} \left\{ f(y) + \gamma\omega\left(\frac{x-y}{\gamma}\right)\right\} \leq f(x) + \gamma\omega(0)$
Fact 2:

$$
\begin{aligned}
{}^{\gamma}_{\omega}f(x) - f(x) &= \inf_{y\in\mathbb{R}^d} \left\{ f(y) - f(x) + \gamma\omega\left(\frac{x-y}{\gamma}\right)\right\} \\
&\overset{(1)}{\geq} \inf_{y\in\mathbb{R}^d} \left\{ \langle\, g_x \,,\, y - x \,\rangle + \gamma\omega\left(\frac{x-y}{\gamma}\right)\right\} \\
&\geq \inf_{z\in\mathbb{R}^d} \gamma \left\{ \langle\, g_x \,,\, -z \,\rangle + \omega\left(z\right)\right\}
\end{aligned}
$$

---

(1) sub-gradient definition at point x
(2)

Proof

Fact 1: ${}^{\gamma}_{\omega}f(x) = \inf_{y\in\mathbb{R}^d}\left\{f(y) + \gamma\omega\left(\frac{x-y}{\gamma}\right)\right\} \leq f(x) + \gamma\omega(0)$

Fact 2:

$$
\begin{aligned}
{}^{\gamma}_{\omega}f(x) - f(x) &= \inf_{y\in\mathbb{R}^d}\left\{f(y) - f(x) + \gamma\omega\left(\frac{x-y}{\gamma}\right)\right\} \\
&\overset{(1)}{\geq} \inf_{y\in\mathbb{R}^d}\left\{\langle\, g_x\,,\, y-x\,\rangle + \gamma\omega\left(\frac{x-y}{\gamma}\right)\right\} \\
&\geq \inf_{z\in\mathbb{R}^d}\gamma\left\{\langle\, g_x\,,\, -z\,\rangle + \omega\left(z\right)\right\} \\
&\geq -\gamma\omega^*(g_x)
\end{aligned}
$$

(1) sub-gradient definition at point x

(2) $\inf(-f) = -\sup f$

Approximation

Theorem

Let us denote $\omega_\gamma := \gamma\omega\left(\frac{\cdot}{\gamma}\right)$, so

$$ {}^{\gamma}_{\omega}f = \gamma\omega\left(\frac{\cdot}{\gamma}\right) \square f = (f^* + \gamma w^*)^* = (f^* + w^*_\gamma)^* $$

Let ω have gradient L_ω-Lipschitz. Then, for any $\gamma > 0$ and $x \in \mathbb{R}^d$, ${}^{\gamma}_{\omega}f$ has gradient $\frac{L_\omega}{\gamma}$-Lipschitz

Proof:

Fact 1: $\omega_\gamma := \gamma\omega\left(\frac{\cdot}{\gamma}\right)$ has gradient $\frac{L_\omega}{\gamma}$-Lipschitz

Fact 2: a function h has gradient L_h-Lipschitz iff its conjugate h^* is $\frac{1}{L_h}$-strongly convex; see Th. 4.2.1. Hiriart-Urruty and Lemarechal (1993b)

Relying on the two previous facts, the result holds true.

More references

- Material mostly inspired by the lecture notes by Pontus Giselsson: http://www.control.lth.se/ls-convex-2015/
- Remark on naming : inf-convolution see Hiriart-Urruty and Lemarechal (1993), Remark 2.1.4.
- https://statweb.stanford.edu/~candes/math301/Lectures/Moreau-Yosida.pdf
- full details and proof in Bauschke and Combettes (2011)

Examples for (geometric) median computation

Definition

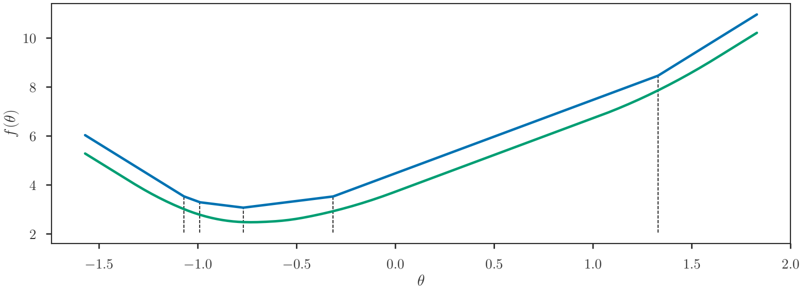
(Geometric) Median : $$\text{Med}_n(\mathbf{x}) \in \underset{\theta\in\mathbb{R}^d}{\arg\min} \sum_{i=1}^{n} \|x - x_i\| = f(x)$$

A possible approach to solve this problem is to perform smoothing with $\omega = \frac{1}{2}\|\cdot\|^2$ and γ, which leads to solve the Huber-mean :

$$\underset{\theta\in\mathbb{R}^d}{\arg\min} \sum_{i=1}^{n} H_\gamma(x - x_i)$$

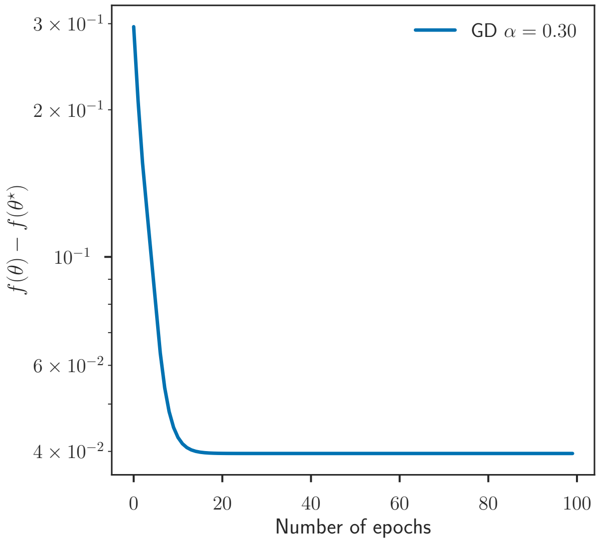
where $H_\gamma(z) = \begin{cases} \frac{\|z\|^2}{2\gamma} & \|y\| \leq \gamma \\ \|y\| - \frac{\gamma}{2} & \|y\| > \gamma \end{cases}$

Visualisation



TO DO: add legend:Green is the smooth approximation

Optimization



References I

- Bauschke, H. H. and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. New York: Springer, 2011, pp. xvi+468.
- Beck, A. and M. Teboulle. "Smoothing and first order methods: A unified framework". In: *SIAM J. Optim.* 22.2 (2012), pp. 557–580.
- Hiriart-Urruty, J.-B. and C. Lemaréchal. *Convex analysis and minimization algorithms. I*. Vol. 305. Berlin: Springer-Verlag, 1993.
- – .*Convex analysis and minimization algorithms. II*. Vol. 306. Berlin: Springer-Verlag, 1993.
- Moreau, J.-J. "Fonctions convexes duales et points proximaux dans un espace hilbertien". In: *C. R. Acad. Sci. Paris* 255 (1962), pp. 2897–2899.
- Nesterov, Y. "Smooth minimization of non-smooth functions". In: *Math. Program.* 103.1 (2005), pp. 127–152.