Optimization
– Proximity operator

Nelly Pustelnik

CNRS, Laboratoire de Physique de l'ENS de Lyon, France

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 $n = 0, 1 \ldots$
  $\quad x_{n+1} = x_n - \gamma \nabla f(x_n)$.

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:

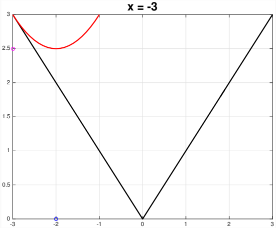
- How to determine x_{n+1} at each iteration n of this *implicit* scheme ?
- Which values of γ guarantee the convergence of $(x_n)_{n \in \mathbb{N}}$?
- What to do if f is nonsmooth ?

Proximity operator: definition

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

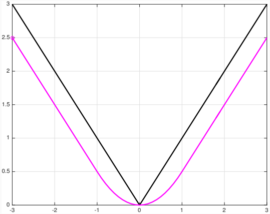
- The **Moreau envelope** of f of parameter $\gamma \in \,]0, +\infty[$ is
$$^{\gamma}f \colon \mathcal{H} \to \mathbb{R} \colon x \mapsto \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma}\|y - x\|^2.$$
- The **proximity operator** of f is
$$\mathrm{prox}_f \colon \mathcal{H} \to \mathcal{H} \colon x \mapsto \underset{y \in \mathcal{H}}{\mathrm{argmin}}\ f(y) + \frac{1}{2}\|y - x\|^2.$$

Proximity operator: definition



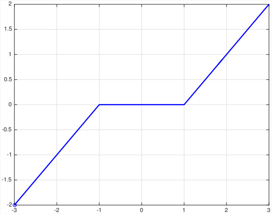
$f(y) = |y|$

$g(y;x) = |y| + \frac{1}{2}(y-x)^2$



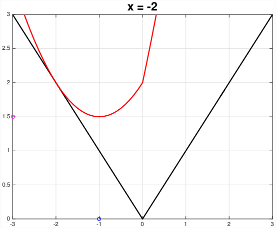
$f(x) = |x|$

$^{\gamma}f(x) = \inf_{y \in \mathcal{H}} g(y;x)$



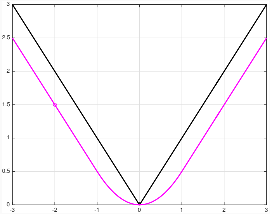
$\mathrm{prox}_f(x) = \operatorname{argmin}_{y \in \mathcal{H}} \; g(y;x)$

Proximity operator: definition



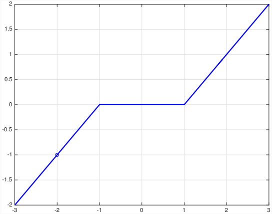
$f(y) = |y|$

$g(y;x) = |y| + \frac{1}{2}(y-x)^2$



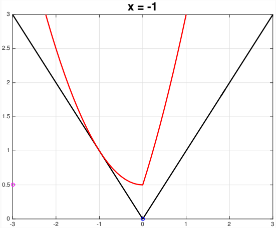
$f(x) = |x|$

${}^{\gamma}f(x) = \inf_{y\in\mathcal{H}} g(y;x)$



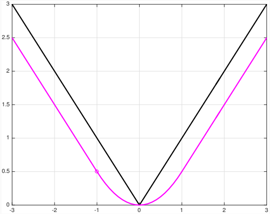
$\mathrm{prox}_f(x) = \underset{y\in\mathcal{H}}{\mathrm{argmin}}\ g(y;x)$

Proximity operator: definition



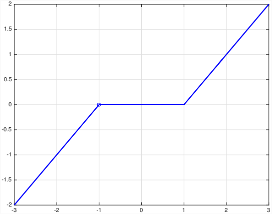
$$f(y) = |y|$$

$$g(y;x) = |y| + \frac{1}{2}(y-x)^2$$



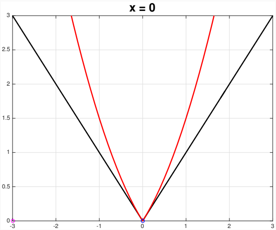
$$f(x) = |x|$$

$${}^{\gamma}f(x) = \inf_{y\in\mathcal{H}} g(y;x)$$



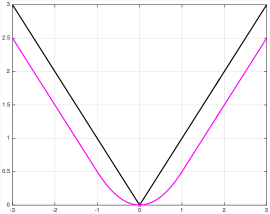
$$\mathrm{prox}_f(x) = \underset{y\in\mathcal{H}}{\mathrm{argmin}}\ g(y;x)$$

Proximity operator: definition



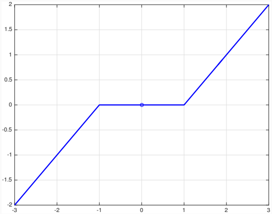
$$f(y) = |y|$$

$$g(y;x) = |y| + \frac{1}{2}(y-x)^2$$



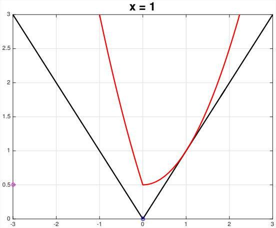
$$f(x) = |x|$$

$$^{\gamma}f(x) = \inf_{y\in\mathcal{H}} g(y;x)$$



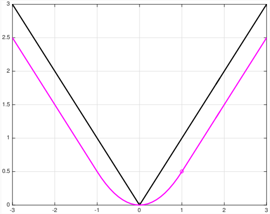
$$\mathrm{prox}_f(x) = \operatorname*{argmin}_{y\in\mathcal{H}} \; g(y;x)$$

Proximity operator: definition



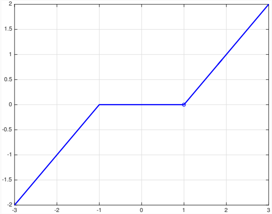
$$f(y) = |y|$$

$$g(y;x) = |y| + \frac{1}{2}(y-x)^2$$



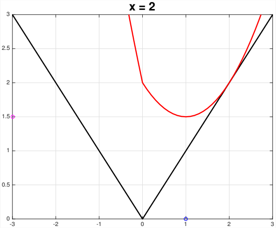
$$f(x) = |x|$$

$${}^{\gamma}f(x) = \inf_{y\in\mathcal{H}} g(y;x)$$



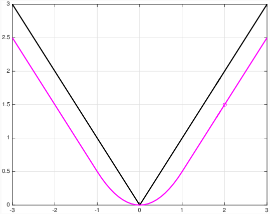
$$\mathrm{prox}_f(x) = \operatorname*{argmin}_{y\in\mathcal{H}} \; g(y;x)$$

Proximity operator: definition



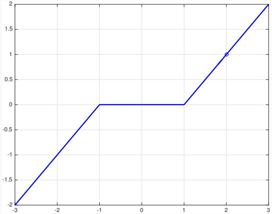
$$f(y) = |y|$$

$$g(y; x) = |y| + \frac{1}{2}(y - x)^2$$



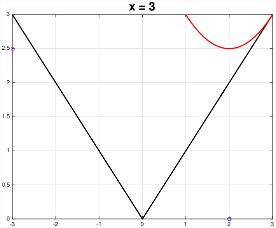
$$f(x) = |x|$$

$$^{\gamma} f(x) = \inf_{y \in \mathcal{H}} g(y; x)$$



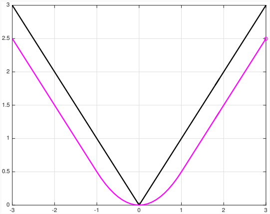
$$\mathrm{prox}_f(x) = \operatorname*{argmin}_{y \in \mathcal{H}} \; g(y; x)$$

Proximity operator: definition



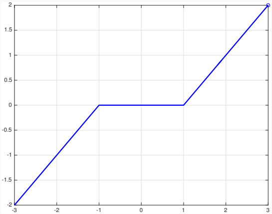
$$f(y) = |y|$$

$$g(y;x) = |y| + \tfrac{1}{2}(y-x)^2$$



$$f(x) = |x|$$

$${}^{\gamma}f(x) = \inf_{y\in\mathcal{H}} g(y;x)$$



$$\mathrm{prox}_f(x) = \operatorname*{argmin}_{y\in\mathcal{H}} \; g(y;x)$$

Proximity operator: characterization

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

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

Proximity operator: characterization

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

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

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

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

Proximity operator: characterization

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

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

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

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

- **Proximal step** :

$$x_{k+1} = \mathrm{prox}_{\gamma f}(x_k) \quad \Leftrightarrow \quad x_{k+1} = x_k - u_k \text{ where } u_k \in \gamma \partial f(x_{k+1})$$

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

Power q function with $q \geq 1$:

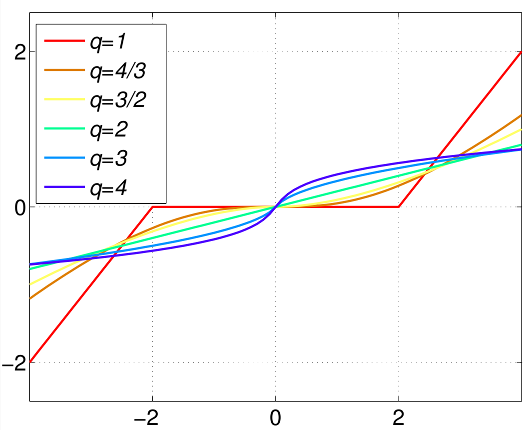
Let $\chi > 0$, $q \in [1, +\infty[$ and $\varphi \colon \mathbb{R} \to\ ]-\infty, +\infty] : \eta \mapsto \chi|\xi|^q$.

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

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

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 $z \in \mathcal{G}$.

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

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	${}^{\gamma}f(x) = \inf_{y \in \mathcal{H}} f(y) + \frac{1}{2\gamma}\|x-y\|^2$ $\gamma > 0$	$\frac{1}{1+\gamma}\left(\gamma x + \mathrm{prox}_{(1+\gamma)f}(x)\right)$

Proximity operator: properties

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

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

then

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

Proximity operator: properties

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

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

then

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

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

Proximity operator: properties

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

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

then

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

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

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

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 $LL^* = \mu \mathrm{Id}$ where $\mu \in \,]0, +\infty[$. Then

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

Proximity operator: properties

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

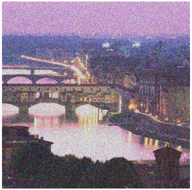
$$\mathrm{prox}_{f \circ L} = \mathrm{Id} - \mu^{-1} L^* \circ (\mathrm{Id} - \mathrm{prox}_{\mu f}) \circ L.$$

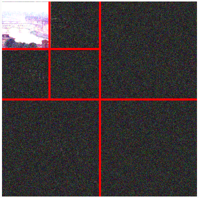
Proof: $LL^* = \mu \mathrm{Id} \Rightarrow \mathrm{ran}\, L = \mathcal{H}$ is closed, hence $V = \mathrm{ran}\,(L^*) = (\ker L)^\perp$ is closed. The orthogonal projection onto V is $P_V = L^*(LL^*)^{-1}L = \mu^{-1}L^*L$.
For every $x \in \mathcal{H}$, $p = \mathrm{prox}_{f \circ L} x \Leftrightarrow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

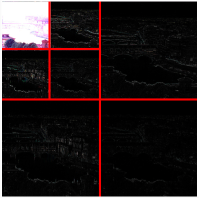
Proximity operator: properties

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

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







Proximity operator: closed form expression

- $\mathrm{prox}_{\lambda\|\cdot\|_1}$: soft-thresholding with a fixed threshold $\lambda > 0$.
- $\mathrm{prox}_{\|\cdot\|_{1,2}}$[Peyré,Fadili,2011].
- $\mathrm{prox}_{\|\|_p^p}$ with $p = \{\frac{4}{3}, \frac{3}{2}, 2, 3, 4\}$[Chaux et al.,2005].
- $\mathrm{prox}_{D_{KL}}$[Combettes,Pesquet,2007].
- $\mathrm{prox}_{\iota_C} = P_C$ projection onto the convex set C.

  $\rightarrow$ range constraint hypercube projection,

  $\rightarrow$ $\ell_{1,p}$-ball constraint [Quattoni,2007] [VanDenBerg,2008]
- $\mathrm{prox}_{\sum_{g\in\mathcal{G}}\|\cdot\|_q}$ with overlapping groups [Jenatton et al., 2011]
- Composition with a linear operator: $\mathrm{prox}_{\varphi\circ L}$ closed form if $LL^* = \nu\mathrm{Id}$ [Pustelnik et al., 2016]

Proximity operator: closed form expression

$$\text{prox}_{\varphi_1+\varphi_2} = \text{prox}_{\varphi_2} \circ \text{prox}_{\varphi_1}$$

- [Combettes-Pesquet, 2007] $N = 1$, $\varphi_2 = \iota_C$ of a non-empty closed convex subset of C and φ_1 is differentiable at 0 with $h'(0) = 0$.
- [Chaux-Pesquet-Pustelnik,2009] C and φ_2 are separable in the same basis.
- [Yu, 2013][Shi et al., 2017] $\partial\varphi_2(x) \subset \partial\varphi_2(\text{prox}\varphi_1(x))$.
- Other recent results [Pustelnik, Condat, 2017][Yukawa, Kagami, 2017][del Aguila Pla, Jaldén, 2017]

Useful websites

- Exhaustive list of proximity operators, Matlab and Python codes:
  `http://proximity-operator.net/`
  authors: Chierchia, Chouzenoux, Combettes, Pesquet

- Table by A. Beck: [Link]

- 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