

Convex nonsmooth optimization

Part III: Algorithms

Barbara Pascal

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

Collaboration

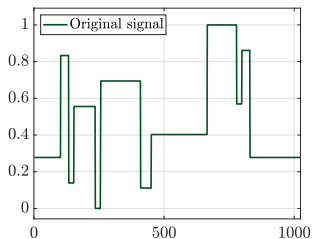
This course is a direct adaptation of the course built by Jean-Christophe Pesquet (CentraleSupélec) and Nelly Pustelnik (LPENSL)



Reconstruction of a piecewise noisy signal

Ground truth

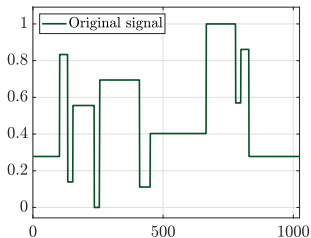
$$\bar{x} \in \mathbb{R}^N$$



Reconstruction of a piecewise noisy signal

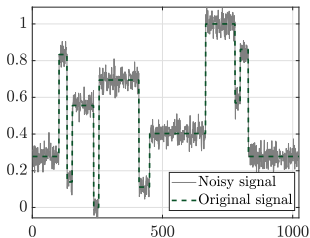
Ground truth

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



Gaussian noise with $\sigma = 0.04$

$$y = \bar{x} + \xi \in \mathbb{R}^N$$



Purpose: recover the true signal with *sharp* transitions

Denoising by functional minimization

Regularized scheme

\mathbf{D} : differential operator, $\|\cdot\|_p$: ℓ_p -norm

$$\hat{x}_\lambda \in \operatorname{Argmin}_{x \in \mathbb{R}^N} \frac{1}{2} \|x - y\|_2^2 + \lambda \|\mathbf{D}x\|_p^p$$

Denoising by functional minimization

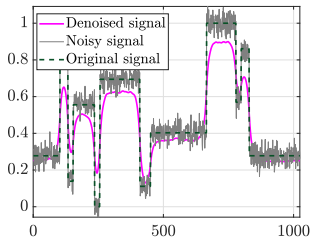
Regularized scheme

\mathbf{D} : differential operator, $\|\cdot\|_p$: ℓ_p -norm

$$\hat{x}_\lambda \in \operatorname{Argmin}_{x \in \mathbb{R}^N} \frac{1}{2} \|x - y\|_2^2 + \lambda \|\mathbf{D}x\|_p^p$$

Tikhonov regularizer $\|\mathbf{D}x\|_2^2$

Smooth: gradient descent



\times fuzzy transitions

Denoising by functional minimization

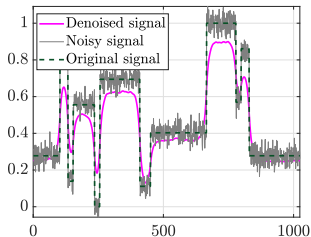
Regularized scheme

\mathbf{D} : differential operator, $\|\cdot\|_p$: ℓ_p -norm

$$\hat{x}_\lambda \in \operatorname{Argmin}_{x \in \mathbb{R}^N} \frac{1}{2} \|x - y\|_2^2 + \lambda \|\mathbf{D}x\|_p^p$$

Tikhonov regularizer $\|\mathbf{D}x\|_2^2$

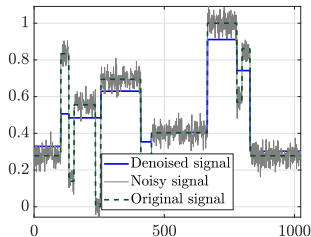
Smooth: gradient descent



✗ fuzzy transitions

Total Variation $\|\mathbf{D}x\|_1$

Nonsmooth: proximal algorithm



✓ sharp transitions

Formulation of the problem

Piecewise denoising

$$\hat{x}_\lambda \in \underset{x \in \mathbb{R}^N}{\operatorname{Argmin}} \frac{1}{2} \|x - y\|_2^2 + \lambda \|Dx\|_1$$

Formulation of the problem

Piecewise denoising

$$\hat{x}_\lambda \in \underset{x \in \mathbb{R}^N}{\operatorname{Argmin}} \frac{1}{2} \|x - y\|_2^2 + \lambda \|Dx\|_1$$

► *Smooth* data-fidelity $f(x) = \frac{1}{2} \|x - y\|_2^2$

Formulation of the problem

Piecewise denoising

$$\hat{x}_\lambda \in \underset{x \in \mathbb{R}^N}{\operatorname{Argmin}} \frac{1}{2} \|x - y\|_2^2 + \lambda \|\mathbf{D}x\|_1$$

- ▶ *Smooth* data-fidelity $f(x) = \frac{1}{2} \|x - y\|_2^2$
- ▶ *Non-smooth* regularizer $h(\mathbf{L}x) = \lambda \|\mathbf{D}x\|_1$, with $h(z) = \lambda \|z\|_1$

Formulation of the problem

Piecewise denoising

$$\hat{x}_\lambda \in \underset{x \in \mathbb{R}^N}{\text{Argmin}} \frac{1}{2} \|x - y\|_2^2 + \lambda \|\mathbf{D}x\|_1$$

- ▶ *Smooth* data-fidelity $f(x) = \frac{1}{2} \|x - y\|_2^2$
- ▶ *Non-smooth* regularizer $h(\mathbf{L}x) = \lambda \|\mathbf{D}x\|_1, \quad \text{with } h(z) = \lambda \|z\|_1$

General form:

$$\hat{x} \in \underset{x \in \mathbb{R}^N}{\text{Argmin}} \{f(x) + h(\mathbf{L}x) = f(x) + g(x)\}$$

f smooth; h and $g = h(\mathbf{L}\cdot)$ nonsmooth.

Optimization algorithm: *Forward-Backward*

Let \mathcal{H} be a Hilbert space.

Let $f \in \Gamma_0(\mathcal{H})$ be differentiable with a ν -Lipschitzian gradient where $\nu \in]0, +\infty[$.

Let $g \in \Gamma_0(\mathcal{H})$.

Let $\gamma \in]0, 2/\nu[$ and $\delta = \min\{1, 1/(\nu\gamma)\} + 1/2$.

Let $(\lambda_n)_{n \in \mathbb{N}}$ be a sequence in $[0, \delta[$ such that $\sum_{n \in \mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$.

Assume that $\text{Argmin}(f + g) \neq \emptyset$. Let $x_0 \in \mathcal{H}$ and

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

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

Example: bounded least-squares

Observation model:

$$y = \mathbf{A}\bar{x} + \xi \in \mathbb{R}^P,$$

linear operator $\mathbf{A} \in \mathbb{R}^{P \times N}$, ξ Gaussian noise, ground truth $\bar{x} \in \mathbb{R}^N$, s.t.

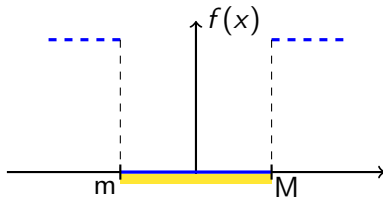
$$\forall i \in \{1, \dots, N\}, \quad m \leq \bar{x}_i \leq M$$

Bounded least-squares

$$C = \{x \in \mathbb{R}^N \mid \forall i, x_i \in [m, M]\}$$

$$\hat{x} \in \underset{x \in C}{\operatorname{Argmin}} \frac{1}{2} \|y - \mathbf{A}x\|_2^2$$

$$\iff \hat{x} \in \underset{x \in \mathbb{R}^N}{\operatorname{Argmin}} \frac{1}{2} \|y - \mathbf{A}x\|_2^2 + \iota_C(x)$$



Optimization algorithm: projected gradient

Let \mathcal{H} be a Hilbert space.

Let $f \in \Gamma_0(\mathcal{H})$ be differentiable with a ν -Lipschitzian gradient where $\nu \in]0, +\infty[$.

Let C a nonempty closed convex subset of \mathcal{H} and P_C the projection on C .

Let $\gamma \in]0, 2/\nu[$ and $\delta = \min\{1, 1/(\nu\gamma)\} + 1/2$.

Let $(\lambda_n)_{n \in \mathbb{N}}$ be a sequence in $[0, \delta[$ such that $\sum_{n \in \mathbb{N}} \lambda_n(\delta - \lambda_n) = +\infty$.

Assume that $\text{Argmin}_{x \in C} g(x) \neq \emptyset$. Let $x_0 \in \mathcal{H}$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = x_n - \gamma \nabla f(x_n) \\ x_{n+1} = x_n + \lambda_n (P_C y_n - x_n). \end{cases}$$

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

Optimization algorithm: gradient descent

Let \mathcal{H} be a Hilbert space.

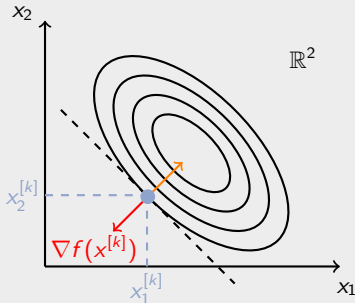
Let $f \in \Gamma_0(\mathcal{H})$ differentiable with a ν -Lipschitz gradient, $\nu \in]0, +\infty[$.

Let $\gamma \in]0, 2/\nu[$.

Assume that $\text{Argmin } f \neq \emptyset$. Let $x_0 \in \mathcal{H}$ and

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n - \gamma \nabla f(x_n)$$

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



Optimization algorithm: Douglas-Rachford

Let \mathcal{H} be a Hilbert space.

Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$.

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

Optimization algorithm: Douglas-Rachford

Let \mathcal{H} be a Hilbert space.

Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{H})$.

Let $\gamma \in]0, +\infty[$ and let $(\lambda_n)_{n \in \mathbb{N}}$ a sequence in $[0, 2]$ s.t. $\sum_{n \in \mathbb{N}} \lambda_n(2 - \lambda_n) = +\infty$.

Assume that $\text{Argmin}(f + g) \neq \emptyset$. Let $x_0 \in \mathcal{H}$ and

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

The following properties are satisfied:

- ▶ $x_n \rightharpoonup \hat{x}$
- ▶ $z_n - y_n \rightarrow 0$, $y_n \rightharpoonup \hat{y}$, $z_n \rightharpoonup \hat{y}$ where $\hat{y} = \text{prox}_{\gamma g} \hat{x} \in \text{Argmin}(f + g)$.

Optimization algorithm: Douglas-Rachford

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

Let $g \in \Gamma_0(\mathcal{H})$ and $L \in \mathcal{B}(\mathcal{G}, \mathcal{H})$ s.t. 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 $\sum_{n \in \mathbb{N}} \lambda_n(2 - \lambda_n) = +\infty$.

Assume that $\text{Argmin}(g \circ L) \neq \emptyset$. Let $x_0 \in \mathcal{H}$, $v_0 = (L^*L)^{-1}L^*x_0$ and

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

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

Optimization algorithm: Douglas-Rachford

Sketch of proof:

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

where $E = \text{ran } L$.

We apply Douglas-Rachford algorithm with

$f = \iota_E \Rightarrow \text{prox}_{\gamma f} = P_E$ by setting

$$(\forall n \in \mathbb{N}) \quad P_E y_n = Lc_n \quad \text{and} \quad P_E x_n = Lv_n$$

where $c_n = \underset{c \in \mathcal{H}}{\text{argmin}} \quad \|y_n - Lc\|^2 = (L^*L)^{-1}L^*y_n$.

Optimization algorithm: Douglas-Rachford

Particular case of Douglas-Rachford algorithm:

$\mathcal{H} = \mathcal{H}_1 \times \cdots \times \mathcal{H}_m$ where $\mathcal{H}_1, \dots, \mathcal{H}_m$ Hilbert spaces

$(\forall x = (x_1, \dots, x_m) \in \mathcal{H}) \quad g(x) = \sum_{i=1}^m g_i(x_i)$

where $(\forall i \in \{1, \dots, m\}) \quad g_i \in \Gamma_0(\mathcal{H}_i)$

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

PPXA+ algorithm

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

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

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

Optimization algorithm: Douglas-Rachford

Particular case of Douglas-Rachford algorithm:

$\mathcal{H} = \mathcal{H}_1 \times \cdots \times \mathcal{H}_m$ where $\mathcal{H}_1 = \cdots = \mathcal{H}_m$ Hilbert spaces

$(\forall x = (x_1, \dots, x_m) \in \mathcal{H}) \quad g(x) = \sum_{i=1}^m g_i(x_i)$

where $(\forall i \in \{1, \dots, m\}) \quad g_i \in \Gamma_0(\mathcal{H}_i)$

$L: v \mapsto (L_1 v, \dots, L_m v)$ where $L_1 = \cdots = L_m = \text{Id}$.

PPXA algorithm

Let $(x_{0,i})_{1 \leq i \leq m} \in \mathcal{H}$, $v_0 = \frac{1}{m} \sum_{i=1}^m x_{0,i}$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_{n,i} = \text{prox}_{\gamma g_i} x_{n,i}, & i \in \{1, \dots, m\} \\ c_n = \frac{1}{m} \sum_{i=1}^m y_{n,i} \\ x_{n+1,i} = x_{n,i} + \lambda_n (2c_n - v_n - y_{n,i}), & i \in \{1, \dots, m\} \\ v_{n+1} = v_n + \lambda_n (c_n - v_n). \end{cases}$$

We have then $v_n \rightharpoonup \hat{v} \in \text{Argmin} \sum_{i=1}^m g_i$.

Optimization algorithms

Forward-Backward	$f_1 + f_2$	f_1 gradient Lipschitz prox_{f_2}	[Combettes,Wajs,2005]
ISTA	$f_1 + f_2$	f_1 gradient Lipschitz $f_2 = \lambda \ \cdot\ _1$	[Daubechies et al, 2003]
Projected gradient	$f_1 + f_2$	f_1 gradient Lipschitz $f_2 = \iota_C$	
Gradient descent	$f_1 + f_2$	f_1 gradient Lipschitz $f_2 = 0$	
Douglas-Rachford	$f_1 + f_2$	prox_{f_1} prox_{f_2}	[Combettes,Pesquet, 2007]
PPXA	$\sum_i f_i$	prox_{f_i}	[Combettes,Pesquet, 2008]
PPXA+	$\sum_i f_i \circ L_i$	prox_{f_i} $(\sum_{i=1}^m L_i^* L_i)^{-1}$	[Pesquet, Pustelnik, 2012]