Sparse deconvolution of seismic data with a regularized norm ratio

Audrey Repetti
Université Paris-Est Marne-la-Vallée, France

ICIAM 2015 – Beijing – August 11

In collaboration with


M. Q. Pham

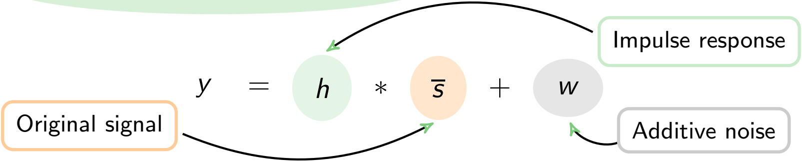

L. Duval


E. Chouzenoux


J.-C. Pesquet

Motivation: Inverse problems

Observation model



$$y = h * \overline{s} + w$$

Objective: Find an estimation $\hat{s} \in \mathbb{R}^N$ of $\overline{s}$ from y.

Motivation: Inverse problems

Observation model

$$y = h * \overline{s} + w$$

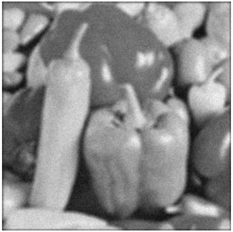
Impulse response (h)

Original signal ($\overline{s}$)

Additive noise (w)

Objective: Find an estimation $\hat{s} \in \mathbb{R}^N$ of $\overline{s}$ from y.

$\overline{s}$



y

$\hat{s}$

Variational formulation

Minimization problem

Define estimate $\hat{s}$ as a solution to $\underset{s \in \mathbb{R}^N}{\text{minimize}} \; F(s) + R(s)$.

- ⋆ **F is a data fidelity term** related to the observation model
- ⋆ **R is a regularization term** related to a priori assumptions on the target solution
  - e.g. a priori on the smoothness of an image,
  - e.g. a priori on the sparsity of a signal,
  - e.g. support constraint,
  - e.g. amplitude/energy bounds,
  - etc.

Variational formulation

Minimization problem

Define estimate $\hat{s}$ as a solution to $\underset{s \in \mathbb{R}^N}{\text{minimize}} \; F(s) + R(s)$.

- ⋆ F is a data fidelity term related to the observation model
- ⋆ R is a regularization term related to a priori assumptions on the target solution

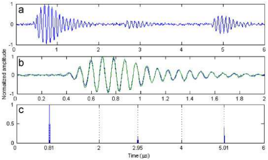
In the context of large scale problems, how to find an optimization algorithm able to deliver a reliable numerical solution in a reasonable time, with low memory requirement?

?

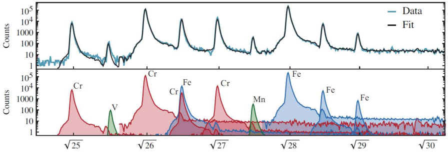
Blind deconvolution

Blind deconvolution problem : $y = \overline{h} * \overline{s} + w$, with

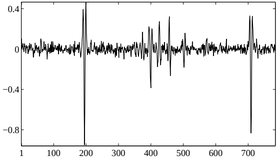
- ⋆ $\overline{s}$: unknown sparse latent signal



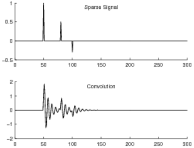
Ultrasonic NDT/NDE



Mass spectrometry/chromatography



Seismic deconvolution



Others (medical, comm., etc.)

Blind deconvolution

Blind deconvolution problem : $y = \overline{h} * \overline{s} + w$, with

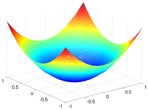
- $\overline{s}$: unknown sparse latent signal
- $\overline{h}$: unknown impulse response
  - blur, linear sensor response, point spread function, seismic wavelet, spectral broadening

OBJECTIVE: Find estimate $(\widehat{s}, \widehat{h}) \in \mathbb{R}^{N_1} \times \mathbb{R}^{N_2}$ from y.

MINIMIZATION PROBLEM

Define estimate $(\widehat{s}, \widehat{h})$ as a solution to $\underset{(s,h)\in\mathbb{R}^{N_1+N_2}}{\text{minimize}} \; F(s,h) + R_1(s) + R_2(h)$.

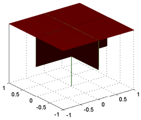
Choose a regularization term to promote sparsity in the signal



ℓ_2

- Smooth and convex
- Not efficient as a sparsity measure

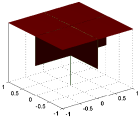
Choose a regularization term to promote sparsity in the signal



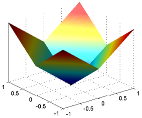
ℓ_0

- Nonsmooth and nonconvex
- Difficult to manage

Choose a regularization term to promote sparsity in the signal



ℓ_0

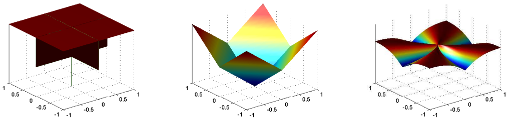


ℓ_1

- Convex relaxation of the ℓ_0-penalization function
- Nonsmooth and convex
- Do not lead to a good estimation of $\overline{s}$ in the context of blind deconvolution problems

  [Benichoux *et al.* – 2013]

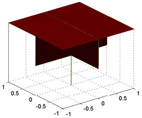
Choose a regularization term to promote sparsity in the signal



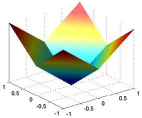
ℓ_0 ℓ_1 ℓ_1/ℓ_2

- Used in:
  - Non-negative Matrix Factorization (NMF) [Hoyer – 2004]
  - Sharpness constraint on wavelet coefficients in images
  - Non-destructive testing/evaluation (NDT/NDE)
  - Sparse recovery [Esser *et al.* – 2015]
  - Potential avoidance of pitfalls [Benichoux *et al.* – 2013]
  - Earlier mentions in geophysics [Gray – 1978]

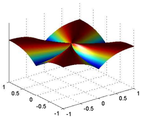
Choose a regularization term to promote sparsity in the signal



ℓ_0



ℓ_1

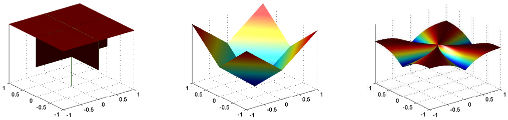


ℓ_1/ℓ_2

Comparison of different measures:

- Let $a = (a^{(n)})_{1 \leq n \leq N}$ such that $(\forall n \in \{1, \ldots, N\})\ a^{(n)} = 1/N$
- Let $b = (b^{(n)})_{1 \leq n \leq N}$ such that $b^{(1)} = 1$ and $(\forall n \in \{2, \ldots, N\})\ b^{(n)} = 0$
  - Same ℓ_1 norm: $\|a\|_1 = \|b\|_1 = 1$
  - $\|a\|_0 = N \geq \|b\|_0 = 1$
  - $\|a\|_1/\|a\|_2 = \sqrt{N} \geq \|b\|_1/\|b\|_2 = 1$

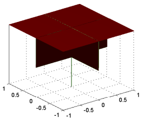
Choose a regularization term to promote sparsity in the signal



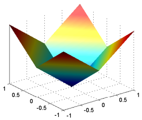
ℓ_0 ℓ_1 ℓ_1/ℓ_2

- Nonsmooth and nonconvex
- Efficient in the context of blind deconvolution problems [Benichoux *et al.* – 2013]
- Difficult to manage

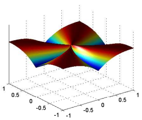
Choose a regularization term to promote sparsity in the signal



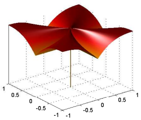
ℓ_0



ℓ_1



ℓ_1/ℓ_2



$\ell_{1,\alpha}/\ell_{2,\eta}$

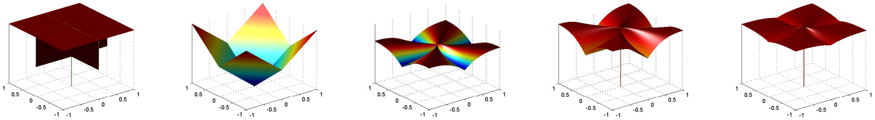
⋆ Use a smooth approximation of the ℓ_1/ℓ_2 penalization function.

- $(\forall s \in \mathbb{R}^N)\ \ell_{1,\alpha}(s) = \sum_{n=1}^{N_1} \left(\sqrt{(s^{(n)})^2 + \alpha^2} - \alpha\right)$, where $\alpha \in]0, +\infty[$

  ⇝ also known as the hybrid $\ell_1 - \ell_2$ or the hyperbolic norm

- $(\forall s \in \mathbb{R}^N)\ \ell_{2,\eta}(s) = \sqrt{\sum_{n=1}^{N_1} (s^{(n)})^2 + \eta^2}$, where $\eta \in]0, +\infty[$

Choose a regularization term to promote sparsity in the signal

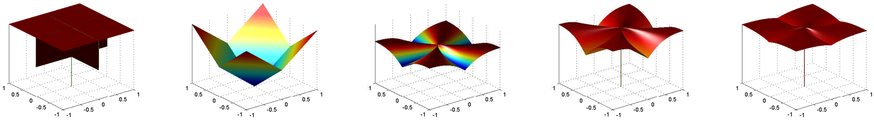


ℓ_0 | ℓ_1 | ℓ_1/ℓ_2 | $\ell_{1,\alpha}/\ell_{2,\eta}$ | $\log\left(\frac{\ell_{1,\alpha}+\beta}{\ell_{2,\eta}}\right)$

⋆ Use a smooth approximation of the ℓ_1/ℓ_2 penalization function.

- $(\forall s \in \mathbb{R}^N)\ \ell_{1,\alpha}(s) = \sum_{n=1}^{N_1} \left(\sqrt{(s^{(n)})^2 + \alpha^2} - \alpha\right)$, where $\alpha \in]0, +\infty[$
- $(\forall s \in \mathbb{R}^N)\ \ell_{2,\eta}(s) = \sqrt{\sum_{n=1}^{N_1} (s^{(n)})^2 + \eta^2}$, where $\eta \in]0, +\infty[$

Choose a regularization term to promote sparsity in the signal



ℓ_0 $\quad$ ℓ_1 $\quad$ ℓ_1/ℓ_2 $\quad$ $\ell_{1,\alpha}/\ell_{2,\eta}$ $\quad$ $\log\left(\frac{\ell_{1,\alpha}+\beta}{\ell_{2,\eta}}\right)$

⋆ Use a smooth approximation of the ℓ_1/ℓ_2 penalization function.

- The logarithm function strengthens the sparsity measure of the ℓ_1/ℓ_2 function.
- Differentiable nonconvex function.

Minimization problem

Optimization problem

$$\text{Find } \widehat{x} \in \underset{x \in \mathbb{R}^N}{\text{Argmin}} \left\{ G(x) = F(x) + R(x) \right\}$$

where

- $R\colon \mathbb{R}^N \to]-\infty, +\infty]$ is proper, lsc, bounded from below by an affine function, and the restriction to its domain is continuous,
- $F\colon \mathbb{R}^N \to]-\infty, +\infty[$ is β-Lipschitz differentiable,
- G is coercive.

Forward-Backward algorithm

Let $x_0 \in \operatorname{dom} R$.
Let, for every $k \in \mathbb{N}$, $\gamma_k \in]0, +\infty[$.
For $k = 0, 1, \ldots$
$\lfloor$ $x_{k+1} \in \operatorname{prox}_{\gamma_k R}\left(x_k - \gamma_k \nabla F(x_k)\right)$

Let $R\colon \mathbb{R}^N \to]-\infty, +\infty]$ be proper, lsc, and bounded from below by an affine function.
The proximity operator of R at $x \in \mathbb{R}^N$ is defined by

$$\operatorname{prox}_R(x) = \underset{y \in \mathbb{R}^N}{\operatorname{Argmin}}\, R(y) + \frac{1}{2}\|y - x\|^2.$$

Forward-Backward algorithm

Let $x_0 \in \operatorname{dom} R$.
Let, for every $k \in \mathbb{N}$, $\gamma_k \in]0, +\infty[$.
For $k = 0, 1, \ldots$
$\quad x_{k+1} \in \operatorname{prox}_{\gamma_k R}\left(x_k - \gamma_k \nabla F(x_k)\right)$

Let $R \colon \mathbb{R}^N \to]-\infty, +\infty]$ be proper, lsc, and bounded from below by an affine function.
The proximity operator of R at $x \in \mathbb{R}^N$ is defined by

$$\operatorname{prox}_R(x) = \underset{y \in \mathbb{R}^N}{\operatorname{Argmin}}\, R(y) + \frac{1}{2}\|y - x\|^2.$$

- When R is convex, then $\operatorname{prox}_R(x)$ is reduced to a singleton.

Forward-Backward algorithm

Let $x_0 \in \operatorname{dom} R$.
Let, for every $k \in \mathbb{N}$, $\gamma_k \in]0, +\infty[$.
For $k = 0, 1, \ldots$
$\quad x_{k+1} \in \operatorname{prox}_{\gamma_k R}(x_k - \gamma_k \nabla F(x_k))$

Let $R \colon \mathbb{R}^N \to]-\infty, +\infty]$ be proper, lsc, and bounded from below by an affine function.
The proximity operator of R at $x \in \mathbb{R}^N$ is defined by

$$\operatorname{prox}_R(x) = \underset{y \in \mathbb{R}^N}{\operatorname{Argmin}}\ R(y) + \frac{1}{2}\|y - x\|^2.$$

- When R is convex, then $\operatorname{prox}_R(x)$ is reduced to a singleton.
- When $R = \iota_{\mathcal{C}}$ is the indicator function of the non empty closed convex set $\mathcal{C} \subset \mathbb{R}^N$, then $\operatorname{prox}_{\iota_{\mathcal{C}}}(x) = \Pi_{\mathcal{C}}(x) = \underset{y \in \mathcal{C}}{\operatorname{argmin}}\ \|y - x\|^2$.

Forward-Backward algorithm

Let $x_0 \in \operatorname{dom} R$.
Let, for every $k \in \mathbb{N}$, $\gamma_k \in]0, +\infty[$.
For $k = 0, 1, \dots$
$\quad x_{k+1} \in \operatorname{prox}_{\gamma_k R}\left(x_k - \gamma_k \nabla F(x_k)\right)$

Let $R\colon \mathbb{R}^N \to]-\infty, +\infty]$ be proper, lsc, and bounded from below by an affine function. Let $U \in \mathbb{R}^{N\times N}$ be a symmetric positive definite (SPD) matrix.
The proximity operator of R at $x \in \mathbb{R}^N$ is defined by

$$\operatorname{prox}_{U,R}(x) = \underset{y\in\mathbb{R}^N}{\operatorname{Argmin}}\, R(y) + \frac{1}{2}\|y - x\|_U^2,$$

where $\|x\|_U^2 = \langle x \mid Ux\rangle$.

- ⋆ When R is convex, then $\operatorname{prox}_R(x)$ is reduced to a singleton.
- ⋆ When $R = \iota_{\mathcal{C}}$ is the indicator function of the non empty closed convex set $\mathcal{C} \subset \mathbb{R}^N$, then $\operatorname{prox}_{\iota_{\mathcal{C}}}(x) = \Pi_{\mathcal{C}}(x) = \underset{y\in\mathcal{C}}{\operatorname{argmin}}\, \|y - x\|^2$.

Forward-Backward algorithm

Let $x_0 \in \operatorname{dom} R$.
Let, for every $k \in \mathbb{N}$, $\gamma_k \in]0, +\infty[$.
For $k = 0, 1, \ldots$
$\lfloor$ $x_{k+1} \in \operatorname{prox}_{\gamma_k R}\left(x_k - \gamma_k \nabla F(x_k)\right)$

EXISTING CONVERGENCE RESULTS:

- Convergence of $(x_k)_{k\in\mathbb{N}}$ to a minimizer of G is ensured when F and R are convex, and $0 < \inf_{k\in\mathbb{N}} \gamma_k \leq \sup_{k\in\mathbb{N}} \gamma_k < 2\beta^{-1}$.

  [Combettes & Wajs – 2005]

Forward-Backward algorithm

Let $x_0 \in \operatorname{dom} R$.
Let, for every $k \in \mathbb{N}$, $\gamma_k \in]0, +\infty[$.
For $k = 0, 1, \ldots$
$\quad x_{k+1} \in \operatorname{prox}_{\gamma_k R}\left(x_k - \gamma_k \nabla F(x_k)\right)$

EXISTING CONVERGENCE RESULTS:

- ⋆ Convergence of $(x_k)_{k\in\mathbb{N}}$ to a minimizer of G is ensured when F and R are convex, and $0 < \inf_{k\in\mathbb{N}} \gamma_k \leq \sup_{k\in\mathbb{N}} \gamma_k < 2\beta^{-1}$.

  [Combettes & Wajs – 2005]

- ⋆ Convergence of $(x_k)_{k\in\mathbb{N}}$ to a critical point of G is ensured when F and/or R are nonconvex, and $0 < \inf_{k\in\mathbb{N}} \gamma_k \leq \sup_{k\in\mathbb{N}} \gamma_k < \beta^{-1}$.

  [Attouch, Bolte & Svaiter – 2011]

⇝ Proof based on **Kurdyka-Łojasiewicz inequality**

Kurdyka-Łojasiewicz inequality

Function G satisfies the Kurdyka-Łojasiewicz inequality i.e., for every $\xi \in \mathbb{R}$, and, for every bounded subset E of $\mathbb{R}^N$, there exist three constants $\kappa > 0$, $\zeta > 0$ and $\theta \in [0, 1)$ such that

$$\big(\forall t \in \partial G(x)\big) \qquad \|t\| \geq \kappa |G(x) - \xi|^{\theta},$$

for every $x \in E$ such that $|G(x) - \xi| \leq \zeta$.

Kurdyka-Łojasiewicz inequality

Function G satisfies the Kurdyka-Łojasiewicz inequality i.e., for every $\xi \in \mathbb{R}$, and, for every bounded subset E of $\mathbb{R}^N$, there exist three constants $\kappa > 0$, $\zeta > 0$ and $\theta \in [0, 1)$ such that

$$(\forall t \in \partial G(x)) \qquad \|t\| \geq \kappa |G(x) - \xi|^{\theta},$$

for every $x \in E$ such that $|G(x) - \xi| \leq \zeta$.

- ⋆ Note that other forms of the KL inequality can be found in the literature [Bolte *et al.* - 2007][Bolte *et al.* - 2010].
- ⋆ Technical assumption satisfied for a wide class of nonconvex functions :
  - real analytic functions
  - semi-algebraic functions
  - ...

Kurdyka-Łojasiewicz inequality

Function G satisfies the Kurdyka-Łojasiewicz inequality i.e., for every $\xi \in \mathbb{R}$, and, for every bounded subset E of $\mathbb{R}^N$, there exist three constants $\kappa > 0$, $\zeta > 0$ and $\theta \in [0,1)$ such that

$$(\forall t \in \partial G(x)) \qquad \|t\| \geq \kappa |G(x) - \xi|^{\theta},$$

for every $x \in E$ such that $|G(x) - \xi| \leq \zeta$.

- ⋆ Note that other forms of the KL inequality can be found in the literature [Bolte *et al.* - 2007][Bolte *et al.* - 2010].
- ⋆ Technical assumption satisfied for a wide class of nonconvex functions :
  - real analytic functions
  - semi-algebraic functions
  - ...

⇝ So far, almost every practically useful function imagined.

Variable metric forward-backward algorithm

⋆ Introduce preconditioning symmetric positive definite (SDP) matrices.

Let $x_0 \in \operatorname{dom} R$.
Let, for every $k \in \mathbb{N}$, $\gamma_k \in]0, +\infty[$ and $A_k(x_k) \in \mathbb{R}^{N \times N}$ an SPD matrix.
For $k = 0, 1, \ldots$

$$x_{k+1} \in \operatorname{prox}_{\gamma_k^{-1} A_k(x_k), R}\left(x_k - \gamma_k A_k(x_k)^{-1} \nabla F(x_k)\right)$$

Variable metric forward-backward algorithm

⋆ Introduce preconditioning symmetric positive definite (SDP) matrices.

Let $x_0 \in \operatorname{dom} R$.
Let, for every $k \in \mathbb{N}$, $\gamma_k \in]0,+\infty[$ and $A_k(x_k) \in \mathbb{R}^{N\times N}$ an SPD matrix.
For $k = 0, 1, \ldots$

$$x_{k+1} \in \operatorname{prox}_{\gamma_k^{-1} A_k(x_k), R}\left(x_k - \gamma_k A_k(x_k)^{-1} \nabla F(x_k)\right)$$

⋆ Existing convergence result:

- Convergence of $(x_k)_{k\in\mathbb{N}}$ to a minimizer of G when F and R are convex [Combettes & Vũ - 2012]

Variable metric forward-backward algorithm

⋆ Introduce preconditioning symmetric positive definite (SDP) matrices.

Let $x_0 \in \operatorname{dom} R$.
Let, for every $k \in \mathbb{N}$, $\gamma_k \in]0, +\infty[$ and $A_k(x_k) \in \mathbb{R}^{N\times N}$ an SPD matrix.
For $k = 0, 1, \ldots$

$$x_{k+1} \in \operatorname{prox}_{\gamma_k^{-1} A_k(x_k), R}\left(x_k - \gamma_k A_k(x_k)^{-1} \nabla F(x_k)\right)$$

⋆ Existing convergence result:

- Convergence of $(x_k)_{k\in\mathbb{N}}$ to a minimizer of G when F and R are convex [Combettes & Vũ - 2012]

⋆ Our contributions:

✓ Convergence in the nonconvex case

✓ Choice of the preconditioning matrices

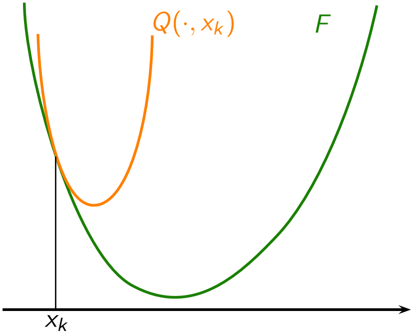
Majorize-Minimize strategy [Jacobson & Fessler – 2007]

For every $k \in \mathbb{N}$, there exists an SPD matrix $A_k(x_k) \in \mathbb{R}^{N \times N}$ such that

$$(\forall x \in \mathbb{R}^N) \quad Q(x, x_k) = F(x_k) + \langle x - x_k \mid \nabla F(x_k) \rangle + \frac{1}{2} \|x - x_k\|^2_{A_k(x_k)}$$

is a **majorant function** of F at x_k on $\operatorname{dom} R$, i.e.,

$$F(x_k) = Q(x_k, x_k) \quad \text{and} \quad (\forall x \in \operatorname{dom} R) \quad F(x) \leq Q(x, x_k).$$



Majorize-Minimize strategy [Jacobson & Fessler – 2007]

For every $k \in \mathbb{N}$, there exists an SPD matrix $A_k(x_k) \in \mathbb{R}^{N\times N}$ such that

$$(\forall x \in \mathbb{R}^N) \quad Q(x, x_k) = F(x_k) + \langle x - x_k \mid \nabla F(x_k)\rangle + \frac{1}{2}\|x - x_k\|^2_{A_k(x_k)}$$

is a **majorant function** of F at x_k on $\operatorname{dom} R$, i.e.,

$$F(x_k) = Q(x_k, x_k) \quad \text{and} \quad (\forall x \in \operatorname{dom} R) \quad F(x) \le Q(x, x_k).$$

F is differentiable with a β-Lipschitzian gradient on a convex subset of $\mathbb{R}^N$

→

$A_k(x_k) \equiv \beta\, \mathrm{I}_N$ satisfies the majorization condition

[Bertsekas - 1999]

VMFB algorithm: Convergence results

- G satisfies the KL inequality.
- $\exists(\underline{\nu}, \overline{\nu}) \in]0, +\infty[^2$ such that $(\forall k \in \mathbb{N})\ \underline{\nu} \mathrm{I}_N \preccurlyeq A_k(x_k) \preccurlyeq \overline{\nu} \mathrm{I}_N$.
- The step-size is chosen such that either:
  - $\exists(\underline{\gamma}, \overline{\gamma}) \in]0, +\infty[^2$ such that $(\forall k \in \mathbb{N})\ \underline{\gamma} \leq \gamma_k \leq 1 - \overline{\gamma}$.
  - R is convex and $\exists(\underline{\gamma}, \overline{\gamma}) \in]0, +\infty[^2$ such that $(\forall k \in \mathbb{N})\ \underline{\gamma} \leq \gamma_k \leq 2 - \overline{\gamma}$.

- Global convergence
  - $(x_k)_{k \in \mathbb{N}}$ converges to a critical point $\widehat{x}$ of G.
  - $(G(x_k))_{k \in \mathbb{N}}$ is a nonincreasing sequence converging to $G(\widehat{x})$.

- Local convergence

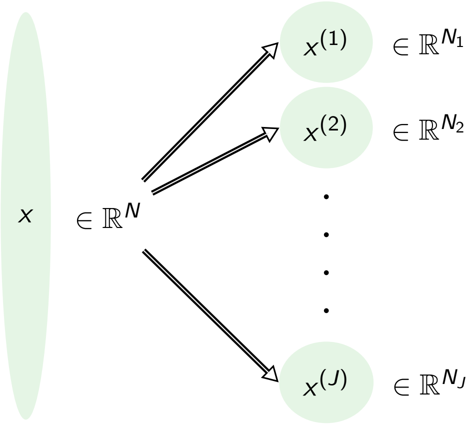
  If $(\exists \upsilon > 0)$ such that $G(x_0) \leq \inf_{x \in \mathbb{R}^N} G(x) + \upsilon$,
  then $(x_k)_{k \in \mathbb{N}}$ converges to a solution $\widehat{x}$ to the minimization problem.

Block separable structure

- R is an additively block separable function.

Block separable structure

- R is an additively block separable function.



$$N = \sum_{j=1}^{J} N_j$$

Block separable structure

- R is an **additively block separable** function.

$$R\left(x\right) = R\begin{pmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(J)} \end{pmatrix} = \sum_{j=1}^{J} R(x^{(j)})$$

$(\forall j \in \{1,\ldots,J\})$ $R_j \colon \mathbb{R}^{N_j} \to]-\infty,+\infty]$ is a proper, lsc function, continuous on its domain and bounded from below by an affine function.

Block coordinate approach

Optimization problem

$$\text{Find } \widehat{x} \in \underset{x \in \mathbb{R}^N}{\text{Argmin}} \left\{ G(x) = F(x) + \sum_{j=1}^{J} R_j(x^{(j)}) \right\}$$

⋆ **Principle**

At each iteration $k \in \mathbb{N}$, update only a subset of components ($\sim$ Gauss-Seidel methods)

⋆ **Advantages**

- more flexibility,
- reduce computational cost at each iteration,
- reduce memory requirement.

Block coordinate VMFB algorithm

Let $x_0 \in \operatorname{dom} R$.
For $k = 0, 1, \ldots$

$$
\left\lfloor
\begin{array}{l}
\text{Let } j_k \in \{1, \ldots, J\}, A_{j_k}(x_k) \in \mathbb{R}^{N_{j_k} \times N_{j_k}} \text{ and } \gamma_k \in]0, +\infty[. \\
x_{k+1}^{(j_k)} \in \operatorname{prox}_{\gamma_k^{-1} A_{j_k}(x_k), R_{j_k}} \left(x_k^{(j_k)} - \gamma_k A_{j_k}(x_k)^{-1} \nabla_{j_k} F(x_k) \right) \\
x_{k+1}^{(\bar{\jmath}_k)} = x_k^{(\bar{\jmath}_k)}
\end{array}
\right.
$$

where $(\forall k \in \mathbb{N})\ x_k^{(\bar{\jmath}_k)} = \left(x^{(1)}, \ldots, x^{(\bar{\jmath}_k - 1)}, x^{(\bar{\jmath}_k + 1)}, \ldots, x^{(J)} \right)$.

Block coordinate VMFB algorithm

Let $x_0 \in \operatorname{dom} R$.
For $k = 0, 1, \dots$
- Let $j_k \in \{1, \dots, J\}$, $A_{j_k}(x_k) \in \mathbb{R}^{N_{j_k} \times N_{j_k}}$ and $\gamma_k \in]0, +\infty[$.
- $x_{k+1}^{(j_k)} \in \operatorname{prox}_{\gamma_k^{-1} A_{j_k}(x_k), R_{j_k}} \left(x_k^{(j_k)} - \gamma_k A_{j_k}(x_k)^{-1} \nabla_{j_k} F(x_k) \right)$
- $x_{k+1}^{(\bar{\jmath}_k)} = x_k^{(\bar{\jmath}_k)}$

Existing convergence results:

⋆ [Bolte, Sabach & Teboulle – 2013]
When $A_{j_k}(x_k) \equiv \mathrm{I}_{N_{j_k}}$ and a cyclic updating rule is adopted.

⋆ [Frankel, Garrigos & Peypouquet – 2014]
When $A_{j_k}(x_k)$ is a general SPD matrix and a cyclic updating rule is adopted.

⋆ [Combettes & Pesquet – 2014]
In the convex case, when $A_{j_k}(x_k) \equiv \mathrm{I}_{N_{j_k}}$ and a random updating rule is adopted.

Block coordinate VMFB algorithm

Let $x_0 \in \operatorname{dom} R$.
For $k = 0, 1, \ldots$

$$
\left\lfloor
\begin{array}{l}
\text{Let } j_k \in \{1, \ldots, J\}, A_{j_k}(x_k) \in \mathbb{R}^{N_{j_k} \times N_{j_k}} \text{ and } \gamma_k \in]0, +\infty[. \\
x_{k+1}^{(j_k)} \in \operatorname{prox}_{\gamma_k^{-1} A_{j_k}(x_k), R_{j_k}} \left(x_k^{(j_k)} - \gamma_k A_{j_k}(x_k)^{-1} \nabla_{j_k} F(x_k) \right) \\
x_{k+1}^{(\bar{\jmath}_k)} = x_k^{(\bar{\jmath}_k)}
\end{array}
\right.
$$

⋆ Our contributions:

- ✓ Convergence in the nonconvex case.
- ✓ Choice of preconditioning matrices $(A_{j_k}(x_k))_{k \in \mathbb{N}}$.
- ✓ General updating rule for $(j_k)_{k \in \mathbb{N}}$.

BC-VMFB algorithm: Convergence results

- Choice of preconditioning matrices $(A_{j_k}(x_k))_{k\in\mathbb{N}}$
  For every $k \in \mathbb{N}$, for every $j_k \in \{1, \ldots, J\}$, $A_{j_k}(x_k)$ satisfies the MM assumption at $x_k^{(j_k)}$ for the restriction of F to the block j_k:

$$y \in \mathbb{R}^{N_{j_k}} \mapsto F\left(x_k^{(1)}, \ldots, x_k^{(j_k-1)}, y, x_k^{(j_k+1)}, \ldots, x_k^{(J)}\right)$$

BC-VMFB algorithm: Convergence results

- Choice of preconditioning matrices $(A_{j_k}(x_k))_{k\in\mathbb{N}}$
  For every $k \in \mathbb{N}$, for every $j_k \in \{1,\dots,J\}$, $A_{j_k}(x_k)$ satisfies the MM assumption at $x_k^{(j_k)}$ for the restriction of F to the block j_k:

$$y \in \mathbb{R}^{N_{j_k}} \mapsto F\left(x_k^{(1)},\dots,x_k^{(j_k-1)},y,x_k^{(j_k+1)},\dots,x_k^{(J)}\right)$$

- Updating rule for $(j_k)_{k\in\mathbb{N}}$
  Blocks $(j_k)_{k\in\mathbb{N}}$ updated according to a quasi-cyclic rule, i.e., there exists $K \geq J$ such that, for every $\ell \in \mathbb{N}$, $\{1,\dots,J\} \subset \{j_k,\dots,j_{k+K-1}\}$.

BC-VMFB algorithm: Convergence results

- Choice of preconditioning matrices $(A_{j_k}(x_k))_{k\in\mathbb{N}}$

  For every $k \in \mathbb{N}$, for every $j_k \in \{1,\dots,J\}$, $A_{j_k}(x_k)$ satisfies the MM assumption at $x_k^{(j_k)}$ for the restriction of F to the block j_k:

$$y \in \mathbb{R}^{N_{j_k}} \mapsto F\left(x_k^{(1)},\dots,x_k^{(j_k-1)},y,x_k^{(j_k+1)},\dots,x_k^{(J)}\right)$$

- Updating rule for $(j_k)_{k\in\mathbb{N}}$

  Blocks $(j_k)_{k\in\mathbb{N}}$ updated according to a quasi-cyclic rule, i.e., there exists $K \geq J$ such that, for every $\ell \in \mathbb{N}$, $\{1,\dots,J\} \subset \{j_k,\dots,j_{k+K-1}\}$.

  Example: $J = 3$ blocks denoted $\{1,2,3\}$

BC-VMFB algorithm: Convergence results

- Choice of preconditioning matrices $(A_{j_k}(x_k))_{k\in\mathbb{N}}$
  For every $k \in \mathbb{N}$, for every $j_k \in \{1,\dots,J\}$, $A_{j_k}(x_k)$ satisfies the MM assumption at $x_k^{(j_k)}$ for the restriction of F to the block j_k:

$$y \in \mathbb{R}^{N_{j_k}} \mapsto F\left(x_k^{(1)},\dots,x_k^{(j_k-1)},y,x_k^{(j_k+1)},\dots,x_k^{(J)}\right)$$

- Updating rule for $(j_k)_{k\in\mathbb{N}}$
  Blocks $(j_k)_{k\in\mathbb{N}}$ updated according to a quasi-cyclic rule, i.e., there exists $K \geq J$ such that, for every $\ell \in \mathbb{N}$, $\{1,\dots,J\} \subset \{j_k,\dots,j_{k+K-1}\}$.
  Example: $J = 3$ blocks denoted $\{1,2,3\}$
  - $K = 3$:
    - cyclic updating order: $\{1,2,3,\ 1,2,3,\ \dots\}$
    - example of quasi-cyclic updating order: $\{1,3,2,\ 2,1,3,\ \dots\}$

BC-VMFB algorithm: Convergence results

- Choice of preconditioning matrices $(A_{j_k}(x_k))_{k\in\mathbb{N}}$

  For every $k \in \mathbb{N}$, for every $j_k \in \{1,\dots,J\}$, $A_{j_k}(x_k)$ satisfies the MM assumption at $x_k^{(j_k)}$ for the restriction of F to the block j_k:

$$y \in \mathbb{R}^{N_{j_k}} \mapsto F\left(x_k^{(1)},\dots,x_k^{(j_k-1)},y,x_k^{(j_k+1)},\dots,x_k^{(J)}\right)$$

- Updating rule for $(j_k)_{k\in\mathbb{N}}$

  Blocks $(j_k)_{k\in\mathbb{N}}$ updated according to a quasi-cyclic rule, i.e., there exists $K \geq J$ such that, for every $\ell \in \mathbb{N}$, $\{1,\dots,J\} \subset \{j_k,\dots,j_{k+K-1}\}$.

  Example: $J = 3$ blocks denoted $\{1,2,3\}$

  - $K = 3$:
    - cyclic updating order: $\{1,2,3,\ 1,2,3,\ \dots\}$
    - example of quasi-cyclic updating order: $\{1,3,2,\ 2,1,3,\ \dots\}$
  - $K = 4$: possibility to update some blocks more than once every K iteration
    - $\{1,3,2,2,\ 2,2,1,3,\ \dots\}$

BC-VMFB algorithm: Convergence results

- Choice of preconditioning matrices $(A_{j_k}(x_k))_{k\in\mathbb{N}}$
  For every $k \in \mathbb{N}$, for every $j_k \in \{1, \dots, J\}$, $A_{j_k}(x_k)$ satisfies the MM assumption at $x_k^{(j_k)}$ for the restriction of F to the block j_k:

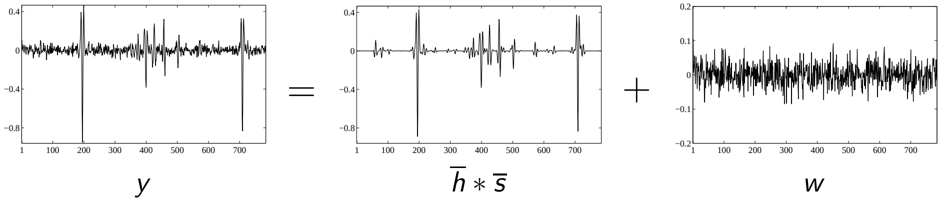
$$y \in \mathbb{R}^{N_{j_k}} \mapsto F\left(x_k^{(1)}, \dots, x_k^{(j_k-1)}, y, x_k^{(j_k+1)}, \dots, x_k^{(J)}\right)$$

- Updating rule for $(j_k)_{k\in\mathbb{N}}$
  Blocks $(j_k)_{k\in\mathbb{N}}$ updated according to a quasi-cyclic rule, i.e., there exists $K \geq J$ such that, for every $\ell \in \mathbb{N}$, $\{1, \dots, J\} \subset \{j_k, \dots, j_{k+K-1}\}$.

Same convergence results as for the VMFB algorithm:

- Global convergence to a critical point of G.
- Local convergence to a minimizer of G.

Seismic blind deconvolution problem



where

- $y \in \mathbb{R}^{N_1}$ observed signal ($N_1 = 784$)
- $\overline{s} \in \mathbb{R}^{N_1}$ unknown sparse original seismic signal
- $\overline{h} \in \mathbb{R}^{N_2}$ unknown original blur kernel ($N_2 = 41$)
- $w \in \mathbb{R}^{N_1}$ additive noise: realization of a zero-mean white Gaussian noise with variance σ^2

Proposed criterion

OBSERVATION MODEL: $y = \overline{h} * \overline{s} + w$

$$\underset{s\in\mathbb{R}^{N_1},h\in\mathbb{R}^{N_2}}{\text{minimize}} \quad (G(s,h) = F(s,h) + R_1(s) + R_2(h))$$

- ⋆ $F(s,h) = \rho(s,h) + \varphi(s)$, where
  - $\rho(s,h) = \frac{1}{2}\|h * s - y\|^2$ ⇝ data fidelity term,
  - $\varphi(s) = \lambda \log\left(\frac{\ell_{1,\alpha}(s) + \beta}{\ell_{2,\eta}(s)}\right)$ ⇝ smooth regularization term,

    with $\ell_{1,\alpha}$ (resp. $\ell_{2,\eta}$) smooth approximation of ℓ_1-norm (resp. ℓ_2-norm), for $(\alpha,\beta,\eta,\lambda) \in]0,+\infty[^4$.

Proposed criterion

OBSERVATION MODEL: $y = \overline{h} * \overline{s} + w$

$$\underset{s\in\mathbb{R}^{N_1}, h\in\mathbb{R}^{N_2}}{\text{minimize}} \quad (G(s,h) = F(s,h) + R_1(s) + R_2(h))$$

- ⋆ $F(s,h) = \rho(s,h) + \varphi(s)$, where
  - $\rho(s,h) = \frac{1}{2}\|h * s - y\|^2 \rightsquigarrow$ data fidelity term,
  - $\varphi(s) = \lambda \log\left(\dfrac{\ell_{1,\alpha}(s) + \beta}{\ell_{2,\eta}(s)}\right) \rightsquigarrow$ smooth regularization term,

  with $\ell_{1,\alpha}$ (resp. $\ell_{2,\eta}$) smooth approximation of ℓ_1-norm (resp. ℓ_2-norm), for $(\alpha, \beta, \eta, \lambda) \in]0, +\infty[^4$.

- ⋆ $\ell_{1,\alpha}(s) = \sum_{n=1}^{N}\left(\sqrt{(s^{(n)})^2 + \alpha^2} - \alpha\right)$.
- ⋆ $\ell_{2,\eta}(s) = \sqrt{\sum_{n=1}^{N}(s^{(n)})^2 + \eta^2}$.

Proposed criterion

OBSERVATION MODEL: $y = \overline{h} * \overline{s} + w$

$$\underset{s\in\mathbb{R}^{N_1}, h\in\mathbb{R}^{N_2}}{\text{minimize}} \quad (G(s,h) = F(s,h) + R_1(s) + R_2(h))$$

- ⋆ $F(s,h) = \rho(s,h) + \varphi(s)$, where
  - $\rho(s,h) = \frac{1}{2}\|h * s - y\|^2 \rightsquigarrow$ data fidelity term,
  - $\varphi(s) = \lambda \log\left(\frac{\ell_{1,\alpha}(s)+\beta}{\ell_{2,\eta}(s)}\right) \rightsquigarrow$ smooth regularization term,

    with $\ell_{1,\alpha}$ (resp. $\ell_{2,\eta}$) smooth approximation of ℓ_1-norm (resp. ℓ_2-norm), for $(\alpha,\beta,\eta,\lambda) \in]0,+\infty[^4$.

- $R_1(s) = \iota_{[s_{\min},s_{\max}]^{N_1}}(s)$, with $(s_{\min}, s_{\max}) \in]0,+\infty[^2$.

- $R_2(h) = \iota_{\mathcal{C}}(h)$, with $\mathcal{C} = \{h \in [h_{\min}, h_{\max}]^{N_2} \mid \|h\| \leq \delta\}$, for $(h_{\min}, h_{\max}, \delta) \in]0,+\infty[^3$.

SOOT algorithm

Let $s_0 \in \operatorname{dom} R_1$ and $h_0 \in \operatorname{dom} R_2$.
For $k = 0, 1, \ldots$
Let $(K_s, K_h) \in (\mathbb{N}^*)^2, A_1(s_k, h_k) \in \mathbb{R}^{N_1 \times N_1}, A_2(s_k, h_k) \in \mathbb{R}^{N_2 \times N_2}$,
and $\gamma_k \in]0, +\infty[$. Let $s_{k,0} = s_k$, and $h_{k,0} = h_k$.
For $j = 1, \ldots, K_s$
$$s_{k+1,j} \in \operatorname{prox}_{\gamma_k^{-1} A_1(s_{k,j}, h_k), R_1} \left(s_{k,j} - \gamma_k A_1(s_{k,j}, h_k)^{-1} \nabla_1 F(s_{k,j}, h_k)\right)$$
$s_{k+1} = s_{k,K_s}$.
For $i = 1, \ldots, K_h$
$$h_{k+1,i} \in \operatorname{prox}_{\gamma_k^{-1} A_2(s_{k+1}, h_{k,i}), R_1} \left(s_{k,j} - \gamma_k A_2(s_{k+1}, h_{k,i})^{-1} \nabla_1 F(s_{k+1}, h_{k,i})\right)$$
$h_{k+1} = h_{k,K_h}$.

Assume that there exists $(\underline{\nu}, \overline{\nu}) \in]0, +\infty[^2$ such that, for all $k \in \mathbb{N}$,
$$(\forall j \in \{0, \ldots, K_s - 1\}) \quad \underline{\nu}\, \mathrm{I}_{N_1} \preceq A_1(s_{k,j}, h_k) \preceq \overline{\nu}\, \mathrm{I}_{N_1},$$
$$(\forall i \in \{0, \ldots, K_h - 1\}) \quad \underline{\nu}\, \mathrm{I}_{N_2} \preceq A_2(s_{k+1}, h_{k,i}) \preceq \overline{\nu}\, \mathrm{I}_{N_2}.$$
Thus $(s_k, h_k)_{k \in \mathbb{N}}$ converges to a critical point $(\widehat{s}, \widehat{h})$ of G and $\big(G(s_k, h_k)\big)_{k \in \mathbb{N}}$ is a nonincreasing sequence converging to $G(\widehat{s}, \widehat{h})$.

SOOT algorithm: preconditioning matrices

Construction of the quadratic majorants

For every $(s,h) \in \mathbb{R}^{N_1} \times \mathbb{R}^{N_2}$, let

$$A_1(s,h) = \left(L_1(h) + \frac{9\lambda}{8\eta^2}\right) \mathrm{I}_{N_1} + \frac{\lambda}{\ell_{1,\alpha}(s)+\beta} A_{\ell_{1,\alpha}}(s),$$
$$A_2(s,h) = L_2(s)\, \mathrm{I}_{N_2},$$

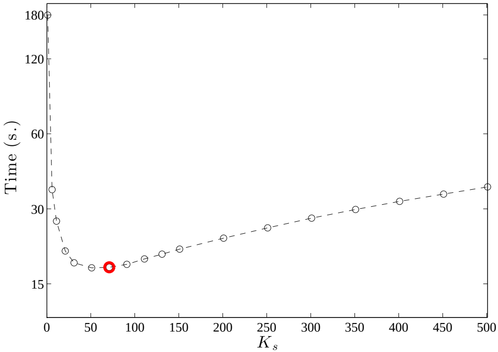
where

$$A_{\ell_{1,\alpha}}(s) = \mathrm{Diag}\left(\left(((s^{(n)})^2+\alpha^2)^{-1/2}\right)_{1\leq n\leq N_1}\right),$$

and $L_1(h)$ (resp. $L_2(s)$) is a Lipschitz constant for $\nabla_1\rho(\cdot,h)$ (resp. $\nabla_2\rho(s,\cdot)$). Then, $A_1(s,h)$ (resp. $A_2(s,h)$) satisfies the majoration condition for $F(\cdot,h)$ at s (resp. $F(s,\cdot)$ at h).

Algorithm behavior

Effect of the quasi-cyclic rule on convergence speed



K_s: number of iterations on s for one iteration on h

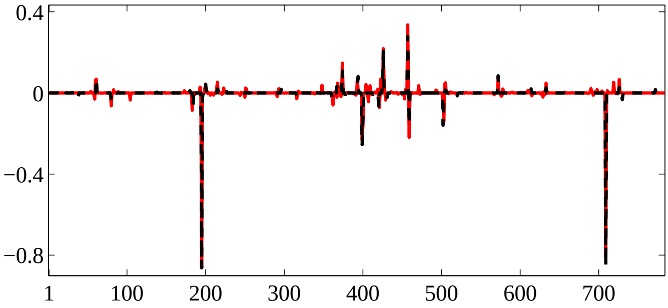
Numerical results

Noise level (σ)			0.01	0.02	0.03
Observation error		ℓ_2 ($\times 10^{-2}$)	7.14	7.35	7.68
		ℓ_1 ($\times 10^{-2}$)	2.85	3.44	4.09
Signal error	Krishnan *et al.*, 2011	ℓ_2 ($\times 10^{-2}$)	1.23	1.66	1.84
		ℓ_1 ($\times 10^{-3}$)	3.79	4.69	5.30
	SOOT	ℓ_2 ($\times 10^{-2}$)	1.09	1.63	1.83
		ℓ_1 ($\times 10^{-3}$)	3.42	4.30	4.85
Kernel error	Krishnan *et al.*, 2011	ℓ_2 ($\times 10^{-2}$)	1.88	2.51	3.21
		ℓ_1 ($\times 10^{-2}$)	1.44	1.96	2.53
	SOOT	ℓ_2 ($\times 10^{-2}$)	1.62	2.26	2.93
		ℓ_1 ($\times 10^{-2}$)	1.22	1.77	2.31
Time (s.)	Krishnan *et al.*, 2011		106	61	56
	SOOT		56	22	18

Numerical results

Sparse seismic reflectivity signal recovery

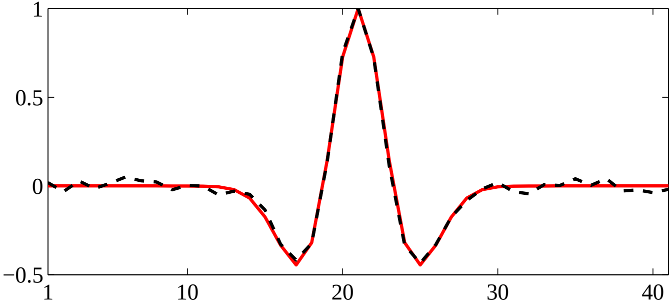
- Continuous red line: $\overline{s}$
- Dashed black line: $\widehat{s}$



Numerical results

Band-pass seismic "wavelet" recovery

- Continuous red line: $\overline{h}$
- Dashed black line: $\widehat{h}$



Conclusion

⇝ Smooth parametric approximations to the ℓ_1/ℓ_2 norm ratio.

Conclusion

- Smooth parametric approximations to the ℓ_1/ℓ_2 norm ratio.
- Proposition of the SOOT algorithm based on a new BC-VMFB algorithm for minimizing the sum of
  - a nonconvex smooth function F,
  - a nonconvex non necessarily smooth function R.
- Convergence results both on iterates and function values.
- Blocks updated according to a flexible quasi-cyclic rule.
- Acceleration of the convergence thanks to the choice of preconditioning matrices based on MM principle.

Conclusion

- Smooth parametric approximations to the ℓ_1/ℓ_2 norm ratio.
- Proposition of the SOOT algorithm based on a new BC-VMFB algorithm for minimizing the sum of
  - a nonconvex smooth function F,
  - a nonconvex non necessarily smooth function R.
- Convergence results both on iterates and function values.
- Blocks updated according to a flexible quasi-cyclic rule.
- Acceleration of the convergence thanks to the choice of preconditioning matrices based on MM principle.
- Application to sparse blind deconvolution .
- Results demonstrated on sparse seismic reflectivity series.

Some references

E. Chouzenoux, J.-C. Pesquet and A. Repetti.
A block coordinate variable metric Forward-Backward algorithm.
Tech. Rep., 2013. Available on
http://www.optimization-online.org/DB_HTML/2013/12/4178.html.

E. Chouzenoux, J.-C. Pesquet and A. Repetti.
Variable metric Forward-Backward algorithm for minimizing the sum of a differentiable function and a convex function.
J. Optim. Theory and Appl., vol.162, no. 1, pp. 107-132, Jul. 2014.

E. Chouzenoux, J.-C. Pesquet and A. Repetti.
A preconditioned Forward-Backward approach with application to large-scale nonconvex spectral unmixing problems.
ICASSP 2014, Florence, Italie, 4-9 May 2014.

A. Repetti, M.Q. Pham, L. Duval, E. Chouzenoux and J.-C. Pesquet.
Euclid in a taxicab: sparse blind deconvolution with smoothed ℓ_1/ℓ_2 regularization.
IEEE Signal Processing Letters, May 2015.