Un taxi pour Euclide (et non Tobrouk) : déconvolution aveugle parcimonieuse, un algorithme préconditionné avec ratio de normes ℓ_1/ℓ_2

Laurent Duval
IFP Energies nouvelles

GdR ISIS — Problèmes inverses ; approches myopes et aveugles, semi- et non-supervisées — 6 novembre 2014

Taxi passengers

A. Repetti

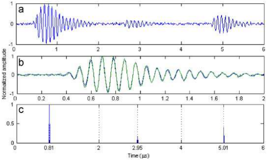
M. Q. Pham

E. Chouzenoux

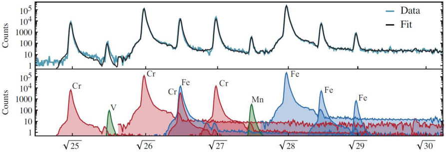
J.-C. Pesquet

Motivations on blind deconvolution

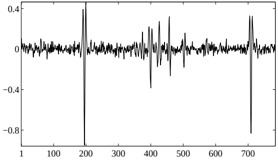
Blind deconvolution $y = \overline{h} * \overline{s} + w$, with sparse latent signals



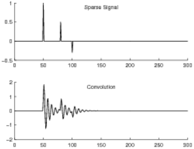
Ultrasonic NDT/NDE



Mass spectrometry/chromatography



Seismic deconvolution



Others (medical, comm., etc.)

Motivations on blind deconvolution

Blind deconvolution $y = \overline{h} * \overline{s} + w$, with sparse latent signals

- $\overline{h}$: (unknown) impulse response
  - blur, linear sensor response, point spread function, seismic wavelet, spectral broadening
- Objective: find estimates $(\widehat{s}, \widehat{h}) \in \mathbb{R}^{N_1} \times \mathbb{R}^{N_2}$ using an optimization approach
- Many works on Euclidean (ℓ_2) and Taxicab (ℓ_1) penalties

Scale-ambiguity $\rightsquigarrow$ focus on a scale-invariant contrast function

Motivations on ℓ_1/ℓ_2 (Taxicab-Euclidean ratio)

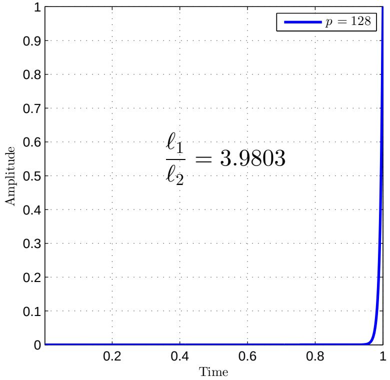
- Taxicab-Euclidean norm ratio
  - $\ell_2 \leq \ell_1 \leq \sqrt{N}\ell_2$
  - Scale-invariant "measure" of sparsity
- Used in the last decade in:
  - Non-negative Matrix Factorization (NMF, Hoyer, 2004)
  - Sharpness constraint on wavelet coefficients in images
  - Non-destructive testing/evaluation (NDT/NDE)
  - Sparse recovery
- Bonuses:
  - Potential avoidance of pitfalls (Benichoux *et al.*, 2013)
  - Earlier mentions in geophysics (Variable norm decon., 1978)

Motivations on ℓ_1/ℓ_2 (Taxicab-Euclidean ratio)

Comparison of different measures

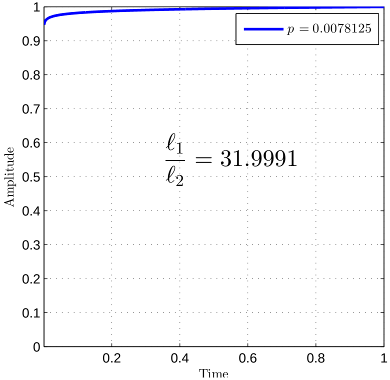
- $a_n = 1/N$ for $n \in \{0, \dots, N-1\}$
- $b_0 = 1$ and $b_n = 0$ for $n \in \{1, \dots, N-1\}$
  - Same ℓ_1 norm: $\|a\|_1 = \|b\|_1 = 1$
  - $\|a\|_0 = N \geq \|b\|_0 = 1$
  - $\|a\|_1/\|a\|_2 = \sqrt{N} \geq \|b\|_1/\|b\|_2 = 1$
- Evaluation of ℓ_1/ℓ_2 for power laws $x \to x^p$, $(p > 0)$

Motivations on ℓ_1/ℓ_2 (Taxicab-Euclidean ratio)



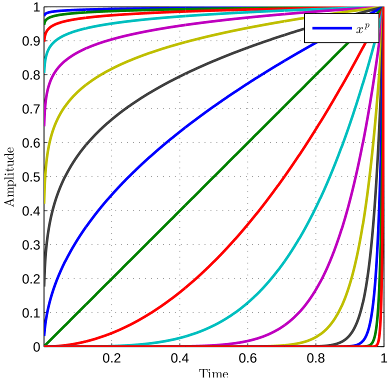
Power law $p = 128$

Motivations on ℓ_1/ℓ_2 (Taxicab-Euclidean ratio)



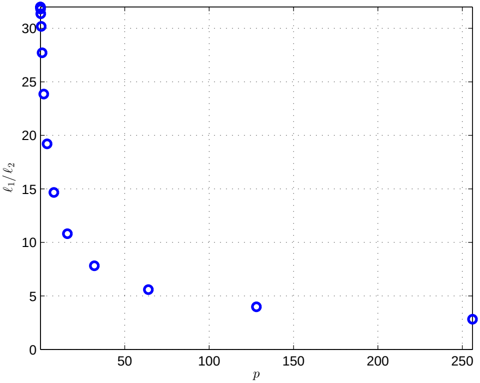
Power law $p = 1/128$

Motivations on ℓ_1/ℓ_2 (Taxicab-Euclidean ratio)



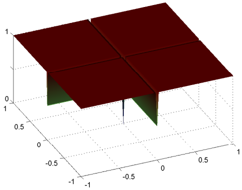
Power law series

Motivations on ℓ_1/ℓ_2 (Taxicab-Euclidean ratio)

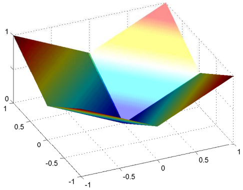


ℓ_1/ℓ_2 ratios vs power law p

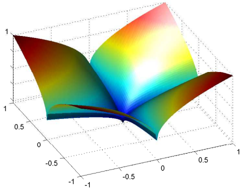
Motivations on ℓ_1/ℓ_2 (Taxicab-Euclidean ratio)



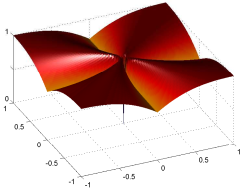
ℓ_0 quasi-norm



ℓ_1 norm



$\ell_{\frac{1}{2}}$ quasi-norm



SOOT ℓ_1/ℓ_2 norm ratio

Formulation

INVERSE PROBLEM: Estimation of an object of interest $\overline{x} \in \mathbb{R}^N$ obtained by minimizing an objective function

$$\boxed{G = F + R}$$

where

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

Formulation

INVERSE PROBLEM: Estimation of an object of interest $\overline{x} \in \mathbb{R}^N$ obtained by minimizing an objective function

$$\boxed{G = F + R}$$

where

- F is a data-fidelity term related to the observation model
- R is a regularization term related to some 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 ?

$\Rightarrow$ **Block alternating minimization.**

$\Rightarrow$ **Variable metric.**

Minimization problem

Problem

$$\text{Find} \quad \hat{x} \in \operatorname{Argmin}\{G = F + R\},$$

where:

- $F\colon \mathbb{R}^N \to \mathbb{R}$ is **differentiable**,
  and has an L-Lipschitz gradient on dom R, i.e.
  $$(\forall (x,y) \in (\operatorname{dom} R)^2) \quad \|\nabla F(x) - \nabla F(y)\| \leq L\|x-y\|,$$
- $R\colon \mathbb{R}^N \to]-\infty,+\infty]$ is proper, lower semicontinuous.
- G is coercive, i.e. $\lim_{\|x\|\to+\infty} G(x) = +\infty$,
  and is **non necessarily convex**.

Forward-Backward algorithm

FB Algorithm

Let $x_0 \in \mathbb{R}^N$
For $\ell = 0, 1, \ldots$
$\lfloor\; x_{\ell+1} \in \text{prox}_{\gamma_\ell R}(x_\ell - \gamma_\ell \nabla F(x_\ell)), \quad \gamma_\ell \in]0, +\infty[.$

- Let $x \in \mathbb{R}^N$. The **proximity operator** is defined by

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

$\rightsquigarrow$ When R is nonconvex:

- Non necessarily uniquely defined.
- Existence guaranteed if R is bounded from below by an affine function.

Forward-Backward algorithm

FB Algorithm

Let $x_0 \in \mathbb{R}^N$
For $\ell = 0, 1, \dots$
$\lfloor\ x_{\ell+1} \in \operatorname{prox}_{\gamma_\ell R}\left(x_\ell - \gamma_\ell \nabla F(x_\ell)\right), \quad \gamma_\ell \in]0, +\infty[.$

- Let $x \in \mathbb{R}^N$. The **proximity operator** is defined by

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

  $\rightsquigarrow$ When R is nonconvex:
  - Non necessarily uniquely defined.
  - Existence guaranteed if R is bounded from below by an affine function.

- Slow convergence.

Variable Metric Forward-Backward algorithm

VMFB Algorithm

Let $x_0 \in \mathbb{R}^N$

For $\ell = 0, 1, \dots$

$$x_{\ell+1} \in \operatorname{prox}_{\gamma_\ell^{-1} A_\ell(x_\ell), R}\left(x_\ell - \gamma_\ell A_\ell(x_\ell)^{-1} \nabla F(x_\ell)\right),$$

with $\gamma_\ell \in]0, +\infty[$, and $A_\ell(x_\ell)$ a SDP matrix.

- Let $x \in \mathbb{R}^N$. The proximity operator relative to the metric induced by $A_\ell(x_\ell)$ is defined by

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

Variable Metric Forward-Backward algorithm

VMFB Algorithm

Let $x_0 \in \mathbb{R}^N$

For $\ell = 0, 1, \dots$

$$x_{\ell+1} \in \mathrm{prox}_{\gamma_\ell^{-1} A_\ell(x_\ell), R}\left(x_\ell - \gamma_\ell A_\ell(x_\ell)^{-1} \nabla F(x_\ell)\right),$$

with $\gamma_\ell \in]0, +\infty[$, and $A_\ell(x_\ell)$ a SDP matrix.

- Let $x \in \mathbb{R}^N$. The proximity operator relative to the metric induced by $A_\ell(x_\ell)$ is defined by

$$\mathrm{prox}_{\gamma_\ell^{-1} A_\ell(x_\ell), R}(x) = \underset{y \in \mathbb{R}^N}{\mathrm{Argmin}}\, R(y) + \frac{1}{2\gamma_\ell}\|y - x\|^2_{A_\ell(x_\ell)}.$$

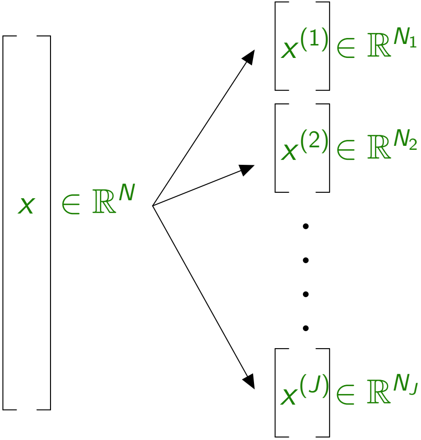
- Convergence is established for a wide class of nonconvex functions G and $(A_\ell(x_\ell))_{\ell \in \mathbb{N}}$ are general SDP matrices in [Chouzenoux *et al.*, 2013]

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.

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

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

BC Forward-Backward algorithm

BC-FB Algorithm [Bolte *et al.*, 2013]

Let $x_0 \in \mathbb{R}^N$

For $\ell = 0, 1, \ldots$

$$
\left\lfloor
\begin{array}{l}
\text{Let } j_\ell \in \{1, \ldots, J\}, \\
x_{\ell+1}^{(j_\ell)} \in \text{prox}_{\gamma_\ell R_{j_\ell}} \left(x_\ell^{(j_\ell)} - \gamma_\ell \nabla_{j_\ell} F(x_\ell) \right), \quad \gamma_\ell \in]0, +\infty[, \\
x_{\ell+1}^{(\bar{\jmath}_\ell)} = x_\ell^{(\bar{\jmath}_\ell)}.
\end{array}
\right.
$$

- Advantages of a block coordinate strategy:
  - more flexibility,
  - reduce computational cost at each iteration,
  - reduce memory requirement.

BC Variable Metric Forward-Backward algorithm

BC-VMFB Algorithm

Let $x_0 \in \mathbb{R}^N$

For $\ell = 0, 1, \ldots$

Let $j_\ell \in \{1, \ldots, J\}$,

$$x_{\ell+1}^{(j_\ell)} \in \operatorname{prox}_{\gamma_\ell^{-1} A_{j_\ell}(x_\ell), R_{j_\ell}} \left(x_\ell^{(j_\ell)} - \gamma_\ell A_{j_\ell}(x_\ell)^{-1} \nabla_{j_\ell} F(x_\ell) \right),$$

$$x_{\ell+1}^{(\bar{\jmath}_\ell)} = x_\ell^{(\bar{\jmath}_\ell)},$$

with $\gamma_\ell \in]0, +\infty[$, and $A_{j_\ell}(x_\ell)$ a SDP matrix.

Our contributions:

- How to choose the preconditioning matrices $(A_{j_\ell}(x_\ell))_{\ell \in \mathbb{N}}$?
  ⇝ Majorize-Minimize principle.
- How to define a general update rule for $(j_\ell)_{\ell \in \mathbb{N}}$?
  ⇝ Quasi-cyclic rule.

Majorize-Minimize assumption [Jacobson *et al.*, 2007]

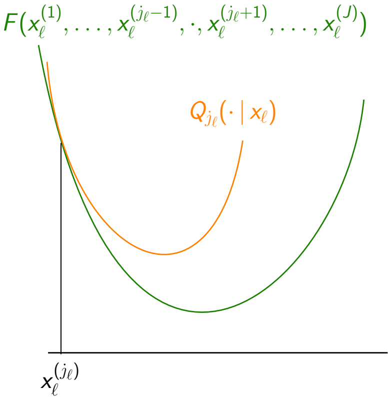
MM Assumption

$(\forall \ell \in \mathbb{N})$ there exists a lower and upper bounded SDP matrix $A_{j_\ell}(x_\ell) \in \mathbb{R}^{N_{j_\ell} \times N_{j_\ell}}$ such that $(\forall y \in \mathbb{R}^{N_{j_\ell}})$

$$Q_{j_\ell}(y \,|\, x_\ell) = F(x_\ell) + (y - x_\ell^{(j_\ell)})^\top \nabla_{j_\ell} F(x_\ell) + \frac{1}{2}\|y - x_\ell^{(j_\ell)}\|^2_{A_{j_\ell}(x_\ell)},$$

is a *majorant function* on dom R_{j_ℓ} of the restriction of F to its j_ℓ-th block at $x_\ell^{(j_\ell)}$, i.e., $(\forall y \in \operatorname{dom} R_{j_\ell})$

$$F\left(x_\ell^{(1)}, \ldots, x_\ell^{(j_\ell - 1)}, y, x_\ell^{(j_\ell+1)}, \ldots, x_\ell^{(J)}\right) \leq Q_{j_\ell}(y \,|\, x_\ell).$$



Majorize-Minimize assumption [Jacobson *et al.*, 2007]

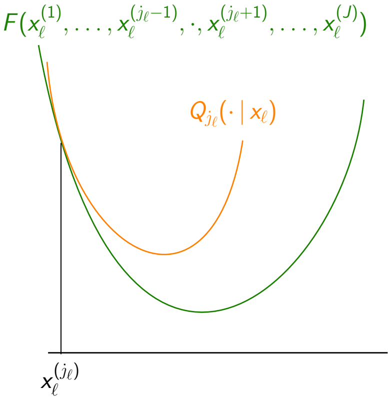
MM Assumption

$(\forall \ell \in \mathbb{N})$ there exists a lower and upper bounded SDP matrix $A_{j_\ell}(x_\ell) \in \mathbb{R}^{N_{j_\ell} \times N_{j_\ell}}$ such that $(\forall y \in \mathbb{R}^{N_{j_\ell}})$

$$Q_{j_\ell}(y \mid x_\ell) = F(x_\ell) + (y - x_\ell^{(j_\ell)})^\top \nabla_{j_\ell} F(x_\ell) + \frac{1}{2}\|y - x_\ell^{(j_\ell)}\|^2_{A_{j_\ell}(x_\ell)},$$

is a *majorant function* on dom R_{j_ℓ} of the restriction of F to its j_ℓ-th block at $x_\ell^{(j_\ell)}$, i.e., $(\forall y \in \text{dom } R_{j_\ell})$

$$F\left(x_\ell^{(1)}, \ldots, x_\ell^{(j_\ell-1)}, y, x_\ell^{(j_\ell+1)}, \ldots, x_\ell^{(J)}\right) \leq Q_{j_\ell}(y \mid x_\ell).$$



dom R is convex and F is L-Lipschitz differentiable $\Rightarrow$ The above assumption holds if $(\forall \ell \in \mathbb{N})$ $A_{j_\ell}(x_\ell) \equiv L\,\mathrm{I}_{N_{j_\ell}}$

Convergence results

Additional assumptions

- G satisfies the Kurdyka-Łojasiewicz inequality [Attouch *et al.*, 2011]:

  For every $\xi \in \mathbb{R}$, for every bounded $E \subset \mathbb{R}^N$, there exist $\kappa, \zeta > 0$ and $\theta \in [0,1)$ such that, for every $x \in E$ such that $|G(x) - \xi| \leq \zeta$,

$$(\forall r \in \partial R(x)) \qquad \|\nabla F(x) + r\| \geq \kappa |G(x) - \xi|^{\theta}.$$

  Technical assumption satisfied for a wide class of nonconvex functions

  - semi-algebraic functions
  - real analytic functions
  - ...

Convergence results

Additional assumptions

- G satisfies the Kurdyka-Łojasiewicz inequality [Attouch *et al.*, 2011]:

  For every $\xi \in \mathbb{R}$, for every bounded $E \subset \mathbb{R}^N$, there exist $\kappa, \zeta > 0$ and $\theta \in [0,1)$ such that, for every $x \in E$ such that $|G(x) - \xi| \leq \zeta$,

$$(\forall r \in \partial R(x)) \qquad \|\nabla F(x) + r\| \geq \kappa |G(x) - \xi|^{\theta}.$$

  Technical assumption satisfied for a wide class of nonconvex functions
  - semi-algebraic functions
  - real analytic functions
  - ...

  $\rightsquigarrow$ So far, almost every practically useful function imagined

Convergence results

Additional assumptions

- G satisfies the Kurdyka-Łojasiewicz inequality [Attouch *et al.*, 2011]:

  For every $\xi \in \mathbb{R}$, for every bounded $E \subset \mathbb{R}^N$, there exist $\kappa, \zeta > 0$ and $\theta \in [0, 1)$ such that, for every $x \in E$ such that $|G(x) - \xi| \leq \zeta$,

  $$(\forall r \in \partial R(x)) \qquad \|\nabla F(x) + r\| \geq \kappa |G(x) - \xi|^{\theta}.$$

  Technical assumption satisfied for a wide class of nonconvex functions

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

Convergence results

Additional assumptions

- G satisfies the Kurdyka-Łojasiewicz inequality [Attouch *et al.*, 2011]:

  For every $\xi \in \mathbb{R}$, for every bounded $E \subset \mathbb{R}^N$, there exist $\kappa, \zeta > 0$ and $\theta \in [0,1)$ such that, for every $x \in E$ such that $|G(x) - \xi| \leq \zeta$,

  $$(\forall r \in \partial R(x)) \qquad \|\nabla F(x) + r\| \geq \kappa |G(x) - \xi|^{\theta}.$$

  Technical assumption satisfied for a wide class of nonconvex functions

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

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

  - $K = 3$:
    - cyclic updating order: $\{1, 2, 3, \;\; 1, 2, 3, \;\; \ldots\}$
    - example of quasi-cyclic updating order: $\{1, 3, 2, \;\; 2, 1, 3, \;\; \ldots\}$

Convergence results

Additional assumptions

- G satisfies the Kurdyka-Łojasiewicz inequality [Attouch *et al.*, 2011]:

  For every $\xi \in \mathbb{R}$, for every bounded $E \subset \mathbb{R}^N$, there exist $\kappa, \zeta > 0$ and $\theta \in [0, 1)$ such that, for every $x \in E$ such that $|G(x) - \xi| \leq \zeta$,

  $$(\forall r \in \partial R(x)) \qquad \|\nabla F(x) + r\| \geq \kappa |G(x) - \xi|^\theta.$$

  Technical assumption satisfied for a wide class of nonconvex functions

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

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

Convergence results

Additional assumptions

- G satisfies the Kurdyka-Łojasiewicz inequality [Attouch *et al.*, 2011]:

  For every $\xi \in \mathbb{R}$, for every bounded $E \subset \mathbb{R}^N$, there exist $\kappa, \zeta > 0$ and $\theta \in [0,1)$ such that, for every $x \in E$ such that $|G(x) - \xi| \leq \zeta$,

  $$(\forall r \in \partial R(x)) \qquad \|\nabla F(x) + r\| \geq \kappa |G(x) - \xi|^{\theta}.$$

  Technical assumption satisfied for a wide class of nonconvex functions

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

- The step-size is chosen such that:
  - $\exists (\underline{\gamma}, \overline{\gamma}) \in (0, +\infty)^2$ such that $(\forall \ell \in \mathbb{N})$ $\underline{\gamma} \leq \gamma_\ell \leq 1 - \overline{\gamma}$.
  - For every $j \in \{1, \ldots, J\}$, R_j is a convex function and $\exists (\underline{\gamma}, \overline{\gamma}) \in (0, +\infty)^2$ such that $(\forall \ell \in \mathbb{N})$ $\underline{\gamma} \leq \gamma_\ell \leq 2 - \overline{\gamma}$.

Convergence results

Convergence theorem

Let $(x_\ell)_{\ell\in\mathbb{N}}$ be a sequence generated by the BC-VMFB algorithm.

- Global convergence:
  $\rightsquigarrow$ $(x_\ell)_{\ell\in\mathbb{N}}$ converges to a critical point $\widehat{x}$ of G.

  $\rightsquigarrow$ $(G(x_\ell))_{\ell\in\mathbb{N}}$ 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 $(x_\ell)_{\ell\in\mathbb{N}}$ converges to a solution $\widehat{x}$ to the minimization problem.

Convergence results

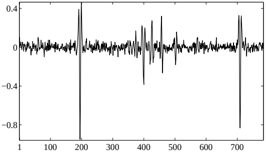
Convergence theorem

Let $(x_\ell)_{\ell\in\mathbb{N}}$ be a sequence generated by the BC-VMFB algorithm.

- Global convergence:
  $\rightsquigarrow$ $(x_\ell)_{\ell\in\mathbb{N}}$ converges to a critical point $\widehat{x}$ of G.
  $\rightsquigarrow$ $(G(x_\ell))_{\ell\in\mathbb{N}}$ 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 $(x_\ell)_{\ell\in\mathbb{N}}$ converges to a solution $\widehat{x}$ to the minimization problem.

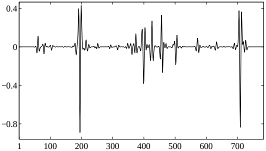
$\rightsquigarrow$ Similar results in [Frankel *et al.*, 2014] restricted to a cyclic updating rule for $(j_\ell)_{\ell\in\mathbb{N}}$.

Seismic blind deconvolution problem



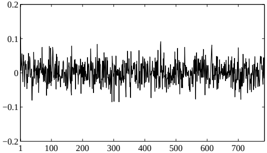
y

$=$



$\overline{h} * \overline{s}$

$+$



w

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

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

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

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

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

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

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

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

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

  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: propositions

Convergence

Let $(s^k)_{k\in\mathbb{N}}$ and $(h^k)_{k\in\mathbb{N}}$ be sequences generated by SOOT. If:

1. There exists $(\underline{\nu}, \overline{\nu}) \in]0, +\infty[^2$ such that, for all $k \in \mathbb{N}$,

$$(\forall j \in \{0, \dots, J_k - 1\}) \quad \underline{\nu}\, \mathrm{I}_N \preceq A_1(s^{k,j}, h^k) \preceq \overline{\nu}\, \mathrm{I}_N,$$
$$(\forall i \in \{0, \dots, I_k - 1\}) \quad \underline{\nu}\, \mathrm{I}_S \preceq A_2(s^{k+1}, h^{k,i}) \preceq \overline{\nu}\, \mathrm{I}_S\,.$$

2. Step-sizes γ_ℓ for s and h are chosen in the interval $[\underline{\gamma}, 2 - \overline{\gamma}]$.
3. G is a semi-algebraic function.

Then $(s^k, h^k)_{k\in\mathbb{N}}$ converges to a critical point $(\widehat{s}, \widehat{h})$ of $G(s, h)$. $\big(G(s^k, h^k)\big)_{k\in\mathbb{N}}$ is a nonincreasing sequence converging to $G(\widehat{s}, \widehat{h})$.

SOOT algorithm: propositions

Construction of the quadratic majorants

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

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

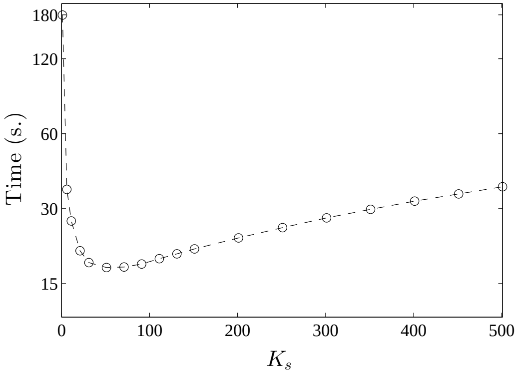
where

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

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).

Numerical results

Effect of the quasi-cyclic rule on convergence speed



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

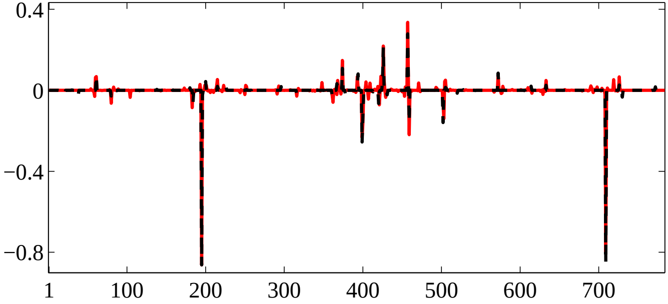
Numerical results

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

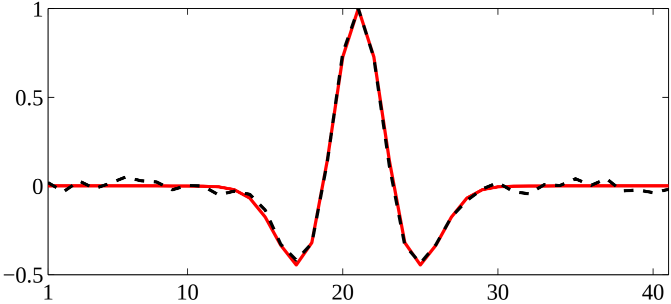
- Continuous red line: $\overline{s}$
- Dashed black line: $\widehat{s}$



Numerical results

Band-pass seismic "wavelet" recovery

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



Conclusion

- 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.
- Smooth parametric approximations to the ℓ_1/ℓ_2 norm ratio
- 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 matrices $(A_{j_\ell}(x_\ell))_{\ell\in\mathbb{N}}$ 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.

So, why Tobrouk (or Tobruk)?

A bunker named Tobruk

or a concrete $\ell_1 \subset \ell_2$ embedding