Un taxi pour Euclide (et non Tobrouk) : déconvolution aveugle parcimonieuse, un algorithme préconditionné avec ratio de normes $\ell_1/\ell_2$

Laurent Duval
IFP Energies nouvelles

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

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

Ultrasonic NDT/NDE

Mass spectrometry/chromatography

Seismic deconvolution

Others (medical, comm., etc.)
Motivations on blind deconvolution

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

- $\bar{h}$: (unknown) impulse response
  - blur, linear sensor response, point spread function, seismic wavelet, spectral broadening

- Objective: find estimates $(\hat{s}, \hat{h}) \in \mathbb{R}^{N_1} \times \mathbb{R}^{N_2}$ using an optimization approach

- Many works on Euclidean ($\ell_2$) and Taxicab ($\ell_1$) penalties

Scale-ambiguity $\sim$ focus on a scale-invariant contrast function
Motivations on $\ell_1/\ell_2$ (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 $\ell_1/\ell_2$ (Taxicab-Euclidean ratio)

Comparison of different measures

- $a_n = 1/N$ for $n \in \{0, \ldots, N - 1\}$
- $b_0 = 1$ and $b_n = 0$ for $n \in \{1, \ldots, N - 1\}$
  - Same $\ell_1$ 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 $\ell_1/\ell_2$ for power laws $x \rightarrow x^p$, ($p > 0$)
Motivations on $\ell_1/\ell_2$ (Taxicab-Euclidean ratio)

Power law $p = 128$
Motivations on $\ell_1/\ell_2$ (Taxicab-Euclidean ratio)

Power law $p = 1/128$
Motivations on $\ell_1/\ell_2$ (Taxicab-Euclidean ratio)
Motivations on $\ell_1/\ell_2$ (Taxicab-Euclidean ratio)

\[ \ell_1/\ell_2 \text{ ratios vs power law } p \]
Motivations on $\ell_1/\ell_2$ (Taxicab-Euclidean ratio)

$\ell_0$ quasi-norm

$\ell_1$ norm

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

SOOT $\ell_1/\ell_2$ norm ratio
Formulation

**Inverse problem**: Estimation of an object of interest $\bar{x} \in \mathbb{R}^N$ obtained by minimizing an objective function

\[ G = F + R \]

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 $\mathbf{x} \in \mathbb{R}^N$ obtained by minimizing an objective function

$$G = F + R$$

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

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

where:

- $F : \mathbb{R}^N \rightarrow \mathbb{R}$ is differentiable, and has an $L$-Lipschitz gradient on dom $R$, i.e.
  \[
  \left(\forall (x, y) \in (\text{dom} R)^2\right) \quad \|\nabla F(x) - \nabla F(y)\| \leq L\|x - y\|,
  \]

- $R : \mathbb{R}^N \rightarrow ]-\infty, +\infty]$ is proper, lower semicontinuous.

- $G$ is coercive, i.e. $\lim_{\|x\| \rightarrow +\infty} G(x) = +\infty$, and is non necessarily convex.
**Forward-Backward algorithm**

Let \( x_0 \in \mathbb{R}^N \)

For \( \ell = 0, 1, \ldots \)

\[
\begin{align*}
x_{\ell+1} & \in \text{prox}_{\gamma_\ell R} (x_\ell - \gamma_\ell \nabla F(x_\ell)), \quad \gamma_\ell \in ]0, +\infty[.
\end{align*}
\]

▶ Let \( x \in \mathbb{R}^N \). The **proximity operator** is defined by

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

⇝ 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 \( \ell = 0, 1, \ldots \)

\[
    x_{\ell+1} \in \text{prox}_{\gamma_\ell R} (x_\ell - \gamma_\ell \nabla F(x_\ell)), \quad \gamma_\ell \in ]0, +\infty[. 
\]

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

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

  \[\Rightarrow \text{When } R \text{ is nonconvex:} \]

  - 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 \( \ell = 0, 1, \ldots \)

\[
\begin{array}{c}
\text{Let } x_{\ell+1} \in \text{prox}_{\gamma_{\ell}^{-1} A_{\ell}(x_{\ell})}, R \left( x_{\ell} - \gamma_{\ell} A_{\ell}(x_{\ell})^{-1} \nabla F(x_{\ell}) \right), \\
\text{with } \gamma_{\ell} \in ]0, +\infty[, \text{ and } A_{\ell}(x_{\ell}) \text{ a SDP matrix.}
\end{array}
\]

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

\[
\text{prox}_{\gamma_{\ell}^{-1} A_{\ell}(x_{\ell})}, R(x) = \text{Argmin}_{y \in \mathbb{R}^N} R(y) + \frac{1}{2\gamma_{\ell}} \|y - x\|^2_{A_{\ell}(x_{\ell})}.
\]
Variable Metric Forward-Backward algorithm

**VMFB Algorithm**

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

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

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

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

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

**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.

\[
\begin{bmatrix}
x(1)
\end{bmatrix} \in \mathbb{R}^{N_1}
\]

\[
\begin{bmatrix}
x(2)
\end{bmatrix} \in \mathbb{R}^{N_2}
\]

\[
\begin{bmatrix}
\vdots
\end{bmatrix}
\]

\[
\begin{bmatrix}
x(J)
\end{bmatrix} \in \mathbb{R}^{N_J}
\]

\[
N = \sum_{j=1}^{J} N_j
\]
Block separable structure

- \( R \) is an additively block separable function.

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

\( (\forall j \in \{1, \ldots, J\}) \ R_j : \mathbb{R}^{N_j} \rightarrow (-\infty, +\infty] \) 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 \( \ell = 0, 1, \ldots \)

\[
\begin{align*}
\text{Let } & \; j_\ell \in \{1, \ldots, J\}, \\
\text{Let } & \; x_{\ell+1}^{(j_\ell)} \in \text{prox}_{\gamma_\ell R_{j_\ell}} \left( x_\ell^{(j_\ell)} - \gamma_\ell \nabla_{j_\ell} F(x_\ell) \right), \quad \gamma_\ell \in ]0, +\infty[, \\
\text{Let } & \; x_{\ell+1}^{(\overline{j}_\ell)} = x_\ell^{(\overline{j}_\ell)}. 
\end{align*}
\]

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

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

$x^{(j_\ell)}_{\ell+1} = x^{(j_\ell)}_\ell$,

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

Our contributions:

- How to choose the preconditioning matrices $(A_{j_\ell}(x_\ell))_{\ell \in \mathbb{N}}$?
  $\leadsto$ Majorize-Minimize principle.

- How to define a general update rule for $(j_\ell)_{\ell \in \mathbb{N}}$?
  $\leadsto$ Quasi-cyclic rule.
Euclid in a Taxicab: $\ell_1/\ell_2$ sparse blind deconvolution

**Majorize–Minimize assumption**

(Jacobson et al., 2007)

\[
F(x^{(1)}_\ell, \ldots, x^{(j_\ell-1)}_\ell, y, x^{(j_\ell+1)}_\ell, \ldots, x^{(J)}_\ell) \leq Q_{j_\ell}(y \mid x_\ell).
\]

**MM Assumption**

\[
(\forall \ell \in \mathbb{N}) \text{ there exists a lower and upper bounded SDP matrix } A_{j_\ell}(x_\ell) \in \mathbb{R}^{N_{j_\ell} \times N_{j_\ell}} \text{ such that } (\forall y \in \mathbb{R}^{N_{j_\ell}}) \\
Q_{j_\ell}(y \mid x_\ell) = F(x_\ell) + (y - x^{(j_\ell)}_\ell)^\top \nabla_{j_\ell} F(x_\ell) + \frac{1}{2} \|y - x^{(j_\ell)}_\ell\|_{A_{j_\ell}(x_\ell)}^2,
\]

is a majorant function on \( \text{dom } R_{j_\ell} \) of the restriction of \( F \) to its \( j_\ell \)-th block at \( x^{(j_\ell)}_\ell \), i.e.,

\[
(\forall y \in \text{dom } R_{j_\ell}) \\
F(x^{(1)}_\ell, \ldots, x^{(j_\ell-1)}_\ell, y, x^{(j_\ell+1)}_\ell, \ldots, x^{(J)}_\ell) \leq Q_{j_\ell}(y \mid x_\ell).
\]
**Majorize-Minimize assumption**

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

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

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

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

\( \text{dom } R \) is convex and \( F \) is \( L \)-Lipschitz differentiable \( \Rightarrow \)

The above assumption holds if

\[(\forall \ell \in \mathbb{N}) A_{j_\ell}(x_\ell) \equiv L I_{N_{j_\ell}} \]
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 $|G(x) - \xi| \leq \zeta$,
  
  $$\left( \forall r \in \partial R(x) \right) \quad \| \nabla F(x) + r \| \geq \kappa |G(x) - \xi|^\theta.$$  

  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 $|G(x) - \xi| \leq \zeta$,

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

  Technical assumption satisfied for a wide class of nonconvex functions

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

  So far, almost every practically useful function imagined
Convergence results

Additional assumptions

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

\[
\text{For every } \xi \in \mathbb{R}, \text{ for every bounded } E \subset \mathbb{R}^N, \text{ there exist } \kappa, \zeta > 0 \text{ and } \theta \in [0, 1) \text{ such that, for every } x \in E \text{ such that } |G(x) - \xi| \leq \zeta, \\
(\forall r \in \partial R(x)) \quad \|\nabla F(x) + r\| \geq \kappa |G(x) - \xi|^{\theta}.
\]

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 \( |G(x) - \xi| \leq \zeta \),

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

  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 $|G(x) - \xi| \leq \zeta$,
  
  $$\left( \forall r \in \partial R(x) \right) \quad \|\nabla F(x) + r\| \geq \kappa |G(x) - \xi|^{\theta}.$$ 

  Technical assumption satisfied for a wide class of nonconvex functions

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

  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
    - $\{1, 3, 2, 2, 2, 1, 3, \ldots\}$
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 $|G(x) - \xi| \leq \zeta$,

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

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

Convergence theorem

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

- **Global convergence:**
  \[ (x_\ell)_{\ell \in \mathbb{N}} \] converges to a critical point \(\hat{x}\) of \(G\).
  
  \[ (G(x_\ell))_{\ell \in \mathbb{N}} \] is a nonincreasing sequence converging to \(G(\hat{x})\).

- **Local convergence:**
  If \((\exists \nu > 0)\) such that \(G(x_0) \leq \inf_{x \in \mathbb{R}^N} G(x) + \nu\), then \((x_\ell)_{\ell \in \mathbb{N}}\) converges to a solution \(\hat{x}\) 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:**
  \(\Rightarrow (x_\ell)_{\ell \in \mathbb{N}}\) converges to a critical point \(\hat{x}\) of \(G\).

- **Local convergence:**
  If \((\exists \nu > 0)\) such that \(G(x_0) \leq \inf_{x \in \mathbb{R}^N} G(x) + \nu\), then \((x_\ell)_{\ell \in \mathbb{N}}\) converges to a solution \(\hat{x}\) to the minimization problem.

\(\Rightarrow\) Similar results in [Frankel et al., 2014] restricted to a cyclic updating rule for \((j_\ell)_{\ell \in \mathbb{N}}\).
Seismic blind deconvolution problem

\[
y = \bar{h} \ast \bar{s} + w
\]

where

- \( y \in \mathbb{R}^{N_1} \) observed signal (\( N_1 = 784 \))
- \( \bar{s} \in \mathbb{R}^{N_1} \) unknown sparse original seismic signal
- \( \bar{h} \in \mathbb{R}^{N_2} \) 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 \( \sigma^2 \)
Proposed criterion

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

\[
\begin{align*}
\text{minimize}_{s \in \mathbb{R}^{N_1}, h \in \mathbb{R}^{N_2}} & \quad (G(s, h) = F(s, h) + R_1(s) + R_2(h)) \\
\quad & \quad F(s, h) = \frac{1}{2} \| h \ast s - y \|^2 + \lambda \log \left( \frac{\ell_1,\alpha(s) + \beta}{\ell_2,\eta(s)} \right) \\
\quad & \quad \text{data fidelity term} \quad \text{smooth regularization term}
\end{align*}
\]

with $\ell_1,\alpha$ (resp. $\ell_2,\eta$) smooth approximation of $\ell_1$-norm (resp. $\ell_2$-norm),

for $(\alpha, \beta, \eta, \lambda) \in ]0, +\infty[^4$. 

Euclid in a Taxicab: $\ell_1/\ell_2$ sparse blind deconvolution
Proposed criterion

**Observation model:** \( y = h \ast s + w \)

\[
\min_{s \in \mathbb{R}^{N_1}, h \in \mathbb{R}^{N_2}} (G(s, h) = F(s, h) + R_1(s) + R_2(h))
\]

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

  with \( \ell_{1,\alpha} \) (resp. \( \ell_{2,\eta} \)) smooth approximation of \( \ell_1 \)-norm (resp. \( \ell_2 \)-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:** \( y = \overline{h} \ast \overline{s} + w \)

\[
\begin{align*}
\text{minimize } \quad & (G(s, h) = F(s, h) + R_1(s) + R_2(h)) \\
\text{subject to } & s \in \mathbb{R}^{N_1}, h \in \mathbb{R}^{N_2}
\end{align*}
\]

- \( F(s, h) = \frac{1}{2} \| h \ast s - y \|^2 + \lambda \log \left( \frac{\ell_1,\alpha(s) + \beta}{\ell_2,\eta(s)} \right) \)
  - data fidelity term
  - smooth regularization term
  - with \( \ell_1,\alpha \) (resp. \( \ell_2,\eta \)) smooth approximation of \( \ell_1 \)-norm (resp. \( \ell_2 \)-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 \((\nu, \overline{\nu}) \in ]0, +\infty[^2\) such that, for all \(k \in \mathbb{N}\),
   \[
   (\forall j \in \{0, \ldots, J_k - 1\}) \quad \nu I_N \preceq A_1(s^k_j, h^k) \preceq \overline{\nu} I_N, \\
   (\forall i \in \{0, \ldots, I_k - 1\}) \quad \nu I_S \preceq A_2(s^{k+1}, h^k, i) \preceq \overline{\nu} I_S .
   \]

2. Step-sizes \(\gamma_\ell\) for \(s\) and \(h\) are chosen in the interval \([\gamma, 2 - \gamma]\).

3. \(G\) is a semi-algebraic function.

Then \((s^k, h^k)_{k\in\mathbb{N}}\) converges to a critical point \((\hat{s}, \hat{h})\) of \(G(s, h)\).
\((G(s^k, h^k))_{k\in\mathbb{N}}\) is a nonincreasing sequence converging to \(G(\hat{s}, \hat{h})\).
SOOT algorithm: propositions

Construction of the quadratic majorants

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

\[
A_1(s, h) = \left( L_1(h) + \frac{9\lambda}{8\eta^2} \right) I_{N_1} + \frac{\lambda}{\ell_{1,\alpha}(s) + \beta} A_{\ell_{1,\alpha}}(s),
\]

\[
A_2(s, h) = L_2(s) I_{N_2},
\]

where

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

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 \(h\)).
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

<table>
<thead>
<tr>
<th></th>
<th>Noise level ($\sigma$)</th>
<th>0.01</th>
<th>0.02</th>
<th>0.03</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observation error</strong></td>
<td>$\ell_2 \times 10^{-2}$</td>
<td>7.14</td>
<td>7.35</td>
<td>7.68</td>
</tr>
<tr>
<td></td>
<td>$\ell_1 \times 10^{-2}$</td>
<td>2.85</td>
<td>3.44</td>
<td>4.09</td>
</tr>
<tr>
<td><strong>Signal error</strong></td>
<td>Krishnan <em>et al.</em>, 2011</td>
<td>$\ell_2 \times 10^{-2}$</td>
<td>1.23</td>
<td>1.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\ell_1 \times 10^{-3}$</td>
<td>3.79</td>
<td>4.69</td>
</tr>
<tr>
<td></td>
<td>SOOT</td>
<td>$\ell_2 \times 10^{-2}$</td>
<td>1.09</td>
<td>1.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\ell_1 \times 10^{-3}$</td>
<td>3.42</td>
<td>4.30</td>
</tr>
<tr>
<td><strong>Kernel error</strong></td>
<td>Krishnan <em>et al.</em>, 2011</td>
<td>$\ell_2 \times 10^{-2}$</td>
<td>1.88</td>
<td>2.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\ell_1 \times 10^{-2}$</td>
<td>1.44</td>
<td>1.96</td>
</tr>
<tr>
<td></td>
<td>SOOT</td>
<td>$\ell_2 \times 10^{-2}$</td>
<td>1.62</td>
<td>2.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\ell_1 \times 10^{-2}$</td>
<td>1.22</td>
<td>1.77</td>
</tr>
<tr>
<td><strong>Time (s.)</strong></td>
<td>Krishnan <em>et al.</em>, 2011</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SOOT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>106</td>
<td>61</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td></td>
<td>56</td>
<td>22</td>
<td>18</td>
</tr>
</tbody>
</table>
Numerical results

Sparse seismic reflectivity signal recovery

- Continuous red line: $\bar{s}$
- Dashed black line: $\hat{s}$
Numerical results

Band-pass seismic “wavelet” recovery

- Continuous red line: $\bar{h}$
- Dashed black line: $\hat{h}$
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 $\ell_1/\ell_2$ 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.*  

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.*  

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 $\ell_1/\ell_2$ regularization.*  
So, why Tobrouk (or Tobruk)?

A bunker named Tobruk

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