A Majorize-Minimize subspace approach for $\ell_2$-$\ell_0$ regularization with applications to image processing

Emilie Chouzenoux
emilie.chouzenoux@univ-mlv.fr

Université Paris-Est
Lab. d’Informatique Gaspard Monge
UMR CNRS 8049

Séminaire Image, GREYC
2 février 2012
Outline

1 General context

2 $\ell_2$-$\ell_0$ regularization functions
   - Existence of minimizers
   - Epi-convergence property

3 Minimization of $F_\delta$
   - Proposed algorithm
   - Convergence results

4 Application to image processing
   - Image denoising
   - Image segmentation
   - Texture+Geometry decomposition
   - Image reconstruction

5 Conclusion
Image restoration

- We observe data $y \in \mathbb{R}^Q$, related to the original image $\overline{x} \in \mathbb{R}^N$ through:
  \[
  y = H\overline{x} + w, \quad H \in \mathbb{R}^{Q \times N}
  \]

- **Objective:** Restore the unknown original image $\overline{x}$ from $H$ and $y$. 

\[\begin{align*}
\text{y} & \quad \text{\overline{x}}
\end{align*}\]
We observe data \( y \in \mathbb{R}^Q \), related to the original image \( x \in \mathbb{R}^N \) through:

\[
y = Hx + w, \quad H \in \mathbb{R}^{Q \times N}
\]

**Objective:** Restore the unknown original image \( x \) from \( H \) and \( y \).
Penalized optimization problem

Find

\[
\min_{x \in \mathbb{R}^N} F(x) = \Phi(Hx - y) + \Psi(x),
\]

where

\[
\begin{align*}
\Phi & \leadsto \text{Data fidelity term, related to noise} \\
\Psi & \leadsto \text{Regularization term, related to some a priori assumptions}
\end{align*}
\]

**Assumption:** There exist \( V = [V_1^T | \ldots | V_S^T]^T \) and \( c = [c_1^T, \ldots, c_S^T]^T \), \( V_s \in \mathbb{R}^{P_s \times N} \) and \( c_s \in \mathbb{R}^{P_s} \), such that \( Vx - c \) is block-sparse.

\[
F_0(x) = \Phi(Hx - y) + \sum_{s=1}^{S} \lambda \chi_{\mathbb{R}\setminus\{0\}}(\|V_s x - c_s\|)
\]

where \( \chi_{\mathbb{R}\setminus\{0\}}(t) = 0 \) if \( t = 0 \), 1 otherwise \( \Rightarrow \) group \( \ell_0 \) penalty \([Eldar10]\)
Penalized optimization problem

Find

\[ \min_{x \in \mathbb{R}^N} (F(x) = \Phi(Hx - y) + \Psi(x)), \]

where

\[ \Phi \rightsquigarrow \text{Data fidelity term, related to noise} \]
\[ \Psi \rightsquigarrow \text{Regularization term, related to some } a \text{ priori } \text{assumptions} \]

**Assumption:** There exist \( V = [V_1^\top | \ldots | V_S^\top]^\top \) and \( c = [c_1^\top, \ldots, c_S^\top]^\top \), \( V_s \in \mathbb{R}^{P_s \times N} \) and \( c_s \in \mathbb{R}^{P_s} \), such that \( Vx - c \) is block-sparse.

\[ F_\delta(x) = \Phi(Hx - y) + \sum_{s=1}^{S} \psi_{s,\delta}(\|V_s x - c_s\|) \]

where \( \psi_{s,\delta} \) is an approximation of \( \lambda \chi_{\mathbb{R}\setminus\{0\}} \) depending on \( \delta > 0 \).
Examples of regularization functions

\(\ell_2-\ell_1\) functions: Asymptotically linear with a quadratic behavior near 0.

Example: \((\forall s \in \{1, \cdots, S\})(\forall t \in \mathbb{R}), \psi_{s,\delta}(t) = \lambda(\sqrt{1 + \frac{t^2}{\delta^2}} - 1)\)

Limit case: When \(\delta \to 0\), \(\psi_\delta(t) = \lambda|t|\) (\(\ell_1\) penalty).

Convex functions

\(\Rightarrow\) Majorize-Minimize algorithms [Allain06, Chouzenoux11]

\(\Rightarrow\) Proximal algorithms [Combettes11, Condat11, Raguet11].
Examples of regularization functions

$l_2$-$l_1$ functions: Asymptotically linear with a quadratic behavior near 0.

Example: $(\forall s \in \{1, \cdots, S\})(\forall t \in \mathbb{R}), \psi_{s,\delta}(t) = \lambda(\sqrt{1 + \frac{t^2}{\delta^2}} - 1)$

Limit case: When $\delta \to 0$, $\psi_{\delta}(t) = \lambda|t|$ ($l_1$ penalty).

Convex functions

⇒ Majorize-Minimize algorithms [Allain06, Chouzenoux11]
⇒ Proximal algorithms [Combettes11, Condat11, Raguet11].

$l_2$-$l_0$ functions: Asymptotically constant with a quadratic behavior near 0.

Example: $(\forall s \in \{1, \cdots, S\})(\forall t \in \mathbb{R}), \psi_{s,\delta}(t) = \lambda \min(t^2/(2\delta^2), 1)$

Limit case: When $\delta \to 0$, $\psi_{\delta}(t) = \lambda \chi_{\mathbb{R}\setminus\{0\}}(t)$ ($l_0$ penalty).

Non-convex functions ⇒ Which algorithms?
1 General context

2 $\ell_2-\ell_0$ regularization functions
   - Existence of minimizers
   - Epi-convergence property

3 Minimization of $F_\delta$
   - Proposed algorithm
   - Convergence results

4 Application to image processing
   - Image denoising
   - Image segmentation
   - Texture+Geometry decomposition
   - Image reconstruction

5 Conclusion
Outline

1 General context

2 $\ell_2$-$\ell_0$ regularization functions
   - Existence of minimizers
   - Epi-convergence property

3 Minimization of $F_{\delta}$
   - Proposed algorithm
   - Convergence results

4 Application to image processing
   - Image denoising
   - Image segmentation
   - Texture+Geometry decomposition
   - Image reconstruction

5 Conclusion
\( \ell_2 - \ell_0 \) regularization functions

We consider the following class of potential functions:

1. \( (\forall s \in \{1, \ldots, S\}) (\forall \delta \in (0, +\infty)) \lim_{t \to \infty} \psi_{s,\delta}(t) = \lambda. \)

2. \( (\forall s \in \{1, \ldots, S\}) (\forall \delta \in (0, +\infty)) \psi_{s,\delta}(t) = O(t^2) \) for small \( t \).

Examples:

- \( \psi_\delta(t) = \min\left(\frac{t^2}{2\delta^2}, 1\right) \)
- \( \psi_\delta(t) = \frac{t^2}{2\delta^2 + t^2} \)
- \( \psi_\delta(t) = (1 - \exp(-\frac{t^2}{2\delta^2})) \)
- \( \psi_\delta(t) = \tanh\left(\frac{t^2}{2\delta^2}\right) \)
Existence of minimizers (I)

\[
F_\delta(x) = \Phi(Hx - y) + \sum_{s=1}^{S} \psi_{s,\delta}(\|V_s x - c_s\|)
\]

**Difficulty:** \( F_\delta \) is a **non convex, non coercive** function.

**Proposition 1**

Assume that

(i) \( \Phi \) is continuous and coercive, i.e. \( \lim_{\|x\| \to +\infty} \Phi(x) = +\infty \)

(ii) For every \( \delta > 0 \) and \( s \in \{1, \ldots, S\} \), \( \psi_{s,\delta} \) is continuous and takes nonnegative values

(iii) \( \text{Ker } H = \{0\} \)

Then, for every \( \delta > 0 \), \( F_\delta \) has a minimizer.
Existence of minimizers (II)

\[ F_\delta(x) = \Phi(Hx - y) + \sum_{s=1}^{S} \psi_{s,\delta}(\|V_s x - c_s\|) + \|V_0 x\|^2 \]

**Difficulty:** \( F_\delta \) is a non convex, non coercive function.

**Proposition 1**

Assume that

(i) \( \Phi \) is continuous and coercive, i.e. \( \lim_{\|x\| \to +\infty} \Phi(x) = +\infty \)
(ii) For every \( \delta > 0 \) and \( s \in \{1, \ldots, S\} \), \( \psi_{s,\delta} \) is continuous and takes nonnegative values
(iii) \( \ker H \cap \ker V_0 = \{0\} \)

Then, for every \( \delta > 0 \), \( F_\delta \) has a minimizer.
Existence of minimizers (III)

\[ F_\delta(x) = \Phi(Hx - y) + \sum_{s=1}^{S} \psi_{s,\delta}(\|V_s x - c_s\|) \]

**Difficulty:** \( F_\delta \) is a **non convex, non coercive** function.

**Proposition 1**

Assume that

(i) \( \Phi \) is continuous and coercive, i.e. \( \lim_{\|x\| \to +\infty} \Phi(x) = +\infty \)

(ii) For every \( \delta > 0 \) and \( s \in \{1, \ldots, S\} \), \( \psi_{s,\delta} \) is continuous and takes nonnegative values and \( \psi_{s,\delta}^{-1}(0) \) is a nonempty bounded set.

(iii) \( \text{Ker } H \cap \bigcap_{s=1}^{S} \text{Ker } V_s = \{0\} \)

Then, for every \( \delta > 0 \), \( F_\delta \) has a minimizer.
Epi-convergence to the group $\ell_0$-penalized objective function

Assumptions:

1. $(\forall s \in \{1, \ldots, S\}) \ (\forall (\delta_1, \delta_2) \in (0, +\infty)^2) \quad \delta_1 \leq \delta_2 \implies (\forall t \in \mathbb{R}) \, \psi_{s,\delta_1}(t) \geq \psi_{s,\delta_2}(t)$

2. $(\forall s \in \{1, \ldots, S\})(\forall t \in \mathbb{R}), \lim_{\delta \to 0}^{\delta > 0} \psi_{s,\delta}(t) = \lambda \chi_{\mathbb{R}\{0\}}(t)$

3. Assumptions of Proposition 1

Proposition 2

Let $(\delta_n)_{n \in \mathbb{N}}$ be a decreasing sequence of positive real numbers converging to 0. Under the above assumptions,

$$\inf F_{\delta_n} \to \inf F_0 \quad \text{as} \quad n \to +\infty$$

In addition, if for every $n \in \mathbb{N}$, $\hat{x}_n$ is a minimizer of $F_{\delta_n}$, then the sequence $(\hat{x}_n)_{n \in \mathbb{N}}$ is bounded and all its cluster points are minimizers of $F_0$. 
Outline

1 General context

2 $\ell_2$-$\ell_0$ regularization functions
   - Existence of minimizers
   - Epi-convergence property

3 Minimization of $F_\delta$
   - Proposed algorithm
   - Convergence results

4 Application to image processing
   - Image denoising
   - Image segmentation
   - Texture+Geometry decomposition
   - Image reconstruction

5 Conclusion
Iterative minimization of $F_\delta(x)$

*In the sequel, we assume that $F_\delta$ is differentiable.*

Descent algorithm

$$x_{k+1} = x_k + \alpha_k d_k, \quad (\forall k \geq 0)$$

- $d_k$: search direction satisfying $g_k^T d_k < 0$ where $g_k \triangleq \nabla F_\delta(x_k)$
  - Ex: Gradient, conjugate gradient, Newton, truncated Newton, ...
- stepsize $\alpha_k$: approximate minimizer of $f_{k,\delta}(\alpha): \alpha \mapsto F_\delta(x_k + \alpha d_k)$
Iterative minimization of $F_\delta(x)$

In the sequel, we assume that $F_\delta$ is differentiable.

Descent algorithm

$$x_{k+1} = x_k + \alpha_k d_k, \quad (\forall k \geq 0)$$

- $d_k$: search direction satisfying $g_k^T d_k < 0$ where $g_k \triangleq \nabla F_\delta(x_k)$
  Ex: Gradient, conjugate gradient, Newton, truncated Newton, ...
- stepsize $\alpha_k$: approximate minimizer of $f_{k,\delta}(\alpha): \alpha \mapsto F_\delta(x_k + \alpha d_k)$

Generalization: subspace algorithm [Zibulevsky10]

$$x_{k+1} = x_k + \sum_{m=1}^{M} u_{m,k} d_k^m, \quad (\forall k \geq 0)$$

- $[d_k^1, \ldots, d_k^M] = D_k$: Set of search directions
  Ex: Super-memory gradient $D_k = [-g_k, d_{k-1}, \ldots, d_{k-l}]$
- stepsize $u_k$: approximate minimizer of $f_{k,\delta}(u): u \mapsto F_\delta(x_k + D_k u)$
Objective: Find $\hat{x} \in \text{Arg min}_x F_\delta(x)$

For all $x'$, let $Q(., x')$ a tangent majorant of $F_\delta$ at $x'$ i.e.,

$$Q(x, x') \geq F_\delta(x), \quad \forall x,$$
$$Q(x', x') = F_\delta(x')$$

**MM algorithm:**

$$\forall j \in \{0, \ldots, J\},$$

$$x^{j+1} \in \text{Arg min}_x Q(x, x^j)$$
Quadratic tangent majorant function

**Assumptions:**

(i) \( \Phi \) is differentiable with an \( L \)-Lipschitzian gradient

(ii) For every \( s \in \{1, \cdots, S\} \), \( \psi_{s,\delta} \) is a differentiable function.

(iii) For every \( s \in \{1, \cdots, S\} \), \( \psi_{s,\delta}(\sqrt{\cdot}) \) is concave on \([0, +\infty)\).

(iv) For every \( s \in \{1, \cdots, S\} \), there exists \( \omega_{s,\delta} \in [0, +\infty) \) such that

\[
(\forall t \in (0, +\infty)) \quad 0 \leq \dot{\psi}_{s,\delta}(t) \leq \omega_{s,\delta} t \text{ where } \dot{\psi}_{s,\delta} \text{ is the derivative of } \psi_{s,\delta}.
\]

In addition, \( \lim_{t \to 0} \omega_{s,\delta}(t) \in \mathbb{R} \) with \( \omega_{s,\delta}(t) \triangleq \dot{\psi}_{s,\delta}(t)/t \).

---

**Lemma 1** [Allain06]

For every \( x \in \mathbb{R}^{N} \), let

\[
A(x) = \mu H^\top H + V^\top \text{Diag } \{b(x)\} V + 2V_0^\top V_0
\]

where \( \mu \in [L, +\infty) \) and \( b(x) \in \mathbb{R}^{SP} \) with \( b_{sp}(x) = \omega_{s,\delta}(\|V_s x - c_s\|) \).

Then, \( Q(x, x') = F_\delta(x') + \nabla F_\delta(x')^T(x - x') + \frac{1}{2}(x - x')^T A(x')(x - x') \) is a convex quadratic tangent majorant of \( F_\delta \) at \( x' \).
Majorize-Minimize multivariate stepsize \[\text{[Chouzenoux11]}\]

\[x_{k+1} = x_k + D_k u_k \quad (\forall k \geq 0)\]

- \(D_k\): set of directions
- \(u_k\): resulting from MM minimization of \(f_{k,\delta}(u) : u \mapsto F_\delta(x_k + D_k u)\)

\[q_k(u, u^j_k) : \text{Quadratic tangent majorant of } f_{k,\delta} \text{ at } u^j_k\]

with Hessian: \(B_{k,u^j_k} = D^T_k A(x_k + D_k u^j_k) D_k\)

MM minimization in the subspace:

\[
\begin{cases}
  u^0_k = 0, \\
  u^j_{k+1} \in \text{Arg min}_u q_k(u, u^j_k), \ (\forall j \in \{0, \ldots J - 1\}) \\
  u_k = u^J_k.
\end{cases}
\]
Proposed algorithm

Majorize-Minimize subspace algorithm

For all $k \geq 0$

1. Compute the set of directions $D_k = [d_k^1, \cdots, d_k^M]$

2. $u_k^0 = 0$

3. $\forall j \in \{0, \ldots, J - 1\},$
   
   - $B_{k,u_j} = D_k^\top A(x_k + D_k u_j^j) D_k$
   - $u_{j+1}^j = u_j^j - B_{u_j}^{-1} \nabla f_{k,\delta}(u_j^j)$

4. $u_k = u_{k}^J$

5. Update $x_{k+1} = x_k + D_k u_k$
Convergence results

Assumptions

1. Assumptions of Proposition 1

2. Assumptions of Lemma 1

3. For every $k \in \mathbb{N}$, the matrix of directions $D_k$ is of size $N \times M$ with $1 \leq M \leq N$ and the first subspace direction $d^1_k$ is gradient-related i.e., there exist $\gamma_0 > 0$ and $\gamma_1 > 0$ such that, for every $k \in \mathbb{N}$,

$$d^1_k \preceq -\gamma_0 \|g_k\|^2,$$

$$\|d^1_k\| \leq \gamma_1 \|g_k\|$$

4. $F_{\delta}$ satisfies the Łojasiewicz inequality [Attouch10a, Attouch10b]: For every $\tilde{x} \in \mathbb{R}^N$ and every bounded neighborhood of $E$ of $\tilde{x}$, there exist constants $\kappa > 0$, $\zeta > 0$ and $\theta \in [0, 1)$ such that

$$\|\nabla F_{\delta}(x)\| \geq \kappa |F_{\delta}(x) - F_{\delta}(\tilde{x})|^\theta,$$

for every $x \in E$ such that $|F_{\delta}(x) - F_{\delta}(\tilde{x})| < \zeta$. 
Convergence results

Theorem

Under Assumptions ①, ②, and ③, for all $J \geq 1$, the MM subspace algorithm is such that

$$\lim_{k \to \infty} \nabla F_\delta(x_k) = 0.$$ 

Furthermore, if Assumption ④ is fulfilled, then

- The MM subspace algorithm generates a sequence converging to a critical point $\tilde{x}$ of $F_\delta$.
- The sequence $(x_k)_{k \in \mathbb{N}}$ has a finite length in the sense that

$$\sum_{k=0}^{+\infty} \|x_{k+1} - x_k\| < +\infty.$$
Outline

1. General context

2. $\ell_2$-$\ell_0$ regularization functions
   - Existence of minimizers
   - Epi-convergence property

3. Minimization of $F_\delta$
   - Proposed algorithm
   - Convergence results

4. Application to image processing
   - Image denoising
   - Image segmentation
   - Texture+Geometry decomposition
   - Image reconstruction

5. Conclusion
Simulation settings

Considered penalization functions:

\[
\begin{align*}
\psi_{s,\delta}(t) &= \lambda \sqrt{1 + \frac{t^2}{\delta^2}} - 1 \quad \text{SC} \\
\psi_{s,\delta}(t) &= \lambda \frac{\lambda t^2}{\lambda^2 + t^2} \quad \text{SNC-(i)} \\
\psi_{s,\delta}(t) &= \lambda (1 - \exp(-\frac{t^2}{2\delta^2})) \quad \text{SNC-(ii)} \\
\psi_{s,\delta}(t) &= \lambda \tanh\left(\frac{t^2}{2\delta^2}\right) \quad \text{SNC-(iii)} \\
\psi_{s,\delta}(t) &= \lambda \min\left(\frac{t^2}{\delta^2}, 1\right) \quad \text{NSNC}
\end{align*}
\]

Optimization algorithms:

\[\mapsto\] MM subspace algorithm with \(D_k = [-g_k \mid x_k - x_{k-1}]\) (MM-MG)

\[\mapsto\] NLCG [Hager06], L-BFGS [Liu89] and HQ [Allain06] algorithms

\[\mapsto\] NSNC: Four state-of-the-art combinatorial optimization algorithms:
\(\alpha\)-EXP [Boykov01], QCSM [Jezierska11], TRW [Kolmogorov06] and BP [Felzenszwalb10]
Image denoising

Original image $\bar{x}$ with $128 \times 128$ pixels (left) and noisy image $y$, degraded by i.i.d. Gaussian noise, $SNR = 15 \, dB$ (right).

$$F_\delta(x) = \frac{1}{2}\|x - y\|^2 + \sum_{s=1}^{S} \psi_{s,\delta}(\| V_s x \|) + \beta d_B^2(x).$$

$\Rightarrow$ $d_B$ is the quadratic distance to the closed convex interval $B = [0, 255]$

$\Rightarrow$ Anisotropic penalization on the gradients of $x$
Results

Denoising result (left) and absolute reconstruction error (right) with SC penalty using MM-MG, SNR = 20.41 dB, MSSIM = 0.89.
Denoising result (left) and absolute reconstruction error (right) with SNC-(i) penalty using MM-MG, SNR = 22.74 dB, MSSIM = 0.92.
Denoising result (left) and absolute reconstruction error (right) with NSNC penalty using TRW, SNR = 22.8 dB, MSSIM = 0.93.
### Results

<table>
<thead>
<tr>
<th>Penalty</th>
<th>Algorithm</th>
<th>Iteration</th>
<th>Time</th>
<th>$F_\delta$</th>
<th>SNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC</td>
<td>MM-MG</td>
<td>122</td>
<td>0.22</td>
<td>$2.7 \cdot 10^6$</td>
<td>20.41</td>
</tr>
<tr>
<td></td>
<td>NLCG</td>
<td>138</td>
<td>0.35</td>
<td>$2.7 \cdot 10^6$</td>
<td>20.41</td>
</tr>
<tr>
<td></td>
<td>L-BFGS</td>
<td>209</td>
<td>0.73</td>
<td>$2.7 \cdot 10^6$</td>
<td>20.41</td>
</tr>
<tr>
<td></td>
<td>HQ</td>
<td>670</td>
<td>3.03</td>
<td>$2.7 \cdot 10^6$</td>
<td>20.41</td>
</tr>
<tr>
<td>SNC-(i)</td>
<td>MM-MG</td>
<td>270</td>
<td>0.35</td>
<td>$1.54 \cdot 10^6$</td>
<td>22.74</td>
</tr>
<tr>
<td></td>
<td>NLCG</td>
<td>1250</td>
<td>2.34</td>
<td>$1.54 \cdot 10^6$</td>
<td>22.74</td>
</tr>
<tr>
<td></td>
<td>L-BFGS</td>
<td>332</td>
<td>0.96</td>
<td>$1.54 \cdot 10^6$</td>
<td>22.73</td>
</tr>
<tr>
<td></td>
<td>HQ</td>
<td>1025</td>
<td>3.84</td>
<td>$1.54 \cdot 10^6$</td>
<td>22.74</td>
</tr>
<tr>
<td>NSNC</td>
<td>$\alpha$-EXP</td>
<td>4</td>
<td>4.67</td>
<td>$1.31 \cdot 10^6$</td>
<td>22.69</td>
</tr>
<tr>
<td></td>
<td>QCSM</td>
<td>2</td>
<td>1.25</td>
<td>$1.31 \cdot 10^6$</td>
<td>22.60</td>
</tr>
<tr>
<td></td>
<td>TRW</td>
<td>5</td>
<td>1.65</td>
<td>$1.31 \cdot 10^6$</td>
<td>22.80</td>
</tr>
<tr>
<td></td>
<td>BP</td>
<td>18</td>
<td>5.33</td>
<td>$1.31 \cdot 10^6$</td>
<td>22.73</td>
</tr>
</tbody>
</table>
Image segmentation

Original image $\bar{x}$ with $256 \times 256$ pixels.

$$F_\delta(x) = \frac{1}{2} \| x - \bar{x} \|^2 + \sum_{s=1}^{S} \psi_{s,\delta}(\| V_s x \|)$$

$\sim$ Anisotropic penalization on the gradients of $x$
Results

Segmented image (left) and its gradient (right) with $SC$ penalty using MM-MG.
Results

Segmented image (left) and its gradient (right) with \textit{SNC-(2)} penalty using MM-MG.
Results

Segmented image (left) and its gradient (right) with NSNC penalty using TRW.
Results

Detail of segmented image (left) and its gradient (right) with \textit{SC} penalty using MM-MG.
Results

Detail of segmented image (left) and its gradient (right) with *SNC-(2)* penalty using MM-MG.
Results

Detail of segmented image (left) and its gradient (right) with \textit{NSNC} penalty using TRW.
Comparison of 50th line of segmented images using \textit{NSNC} (\times), \textit{SNC}-\textit{(ii)} (\textdagger) or \textit{SC} (\textdaggerdbl) potential functions. The 50th line of the original image is indicated in dotted plot.
## Results

<table>
<thead>
<tr>
<th>Penalty</th>
<th>Algorithm</th>
<th>Iteration</th>
<th>Time</th>
<th>$F_\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC</td>
<td>MM-MG</td>
<td>132</td>
<td>0.99</td>
<td>$6.69 \cdot 10^6$</td>
</tr>
<tr>
<td></td>
<td>NLCG</td>
<td>144</td>
<td>1.49</td>
<td>$6.69 \cdot 10^6$</td>
</tr>
<tr>
<td></td>
<td>L-BFGS</td>
<td>215</td>
<td>3.44</td>
<td>$6.69 \cdot 10^6$</td>
</tr>
<tr>
<td></td>
<td>HQ</td>
<td>898</td>
<td>18.19</td>
<td>$6.69 \cdot 10^6$</td>
</tr>
<tr>
<td>SNC-(ii)</td>
<td>MM-MG</td>
<td>622</td>
<td>4.35</td>
<td>$1.59 \cdot 10^7$</td>
</tr>
<tr>
<td></td>
<td>NLCG</td>
<td>1578</td>
<td>14.93</td>
<td>$1.59 \cdot 10^7$</td>
</tr>
<tr>
<td></td>
<td>L-BFGS</td>
<td>632</td>
<td>9.57</td>
<td>$1.59 \cdot 10^7$</td>
</tr>
<tr>
<td></td>
<td>HQ</td>
<td>3553</td>
<td>65.2</td>
<td>$1.59 \cdot 10^7$</td>
</tr>
<tr>
<td>NSNC</td>
<td>$\alpha$-EXP</td>
<td>9</td>
<td>57.97</td>
<td>$5.58 \cdot 10^6$</td>
</tr>
<tr>
<td></td>
<td>QCSM</td>
<td>1</td>
<td>7.05</td>
<td>$5.52 \cdot 10^6$</td>
</tr>
<tr>
<td></td>
<td>TRW</td>
<td>5</td>
<td>6.71</td>
<td>$5.52 \cdot 10^6$</td>
</tr>
<tr>
<td></td>
<td>BP</td>
<td>50</td>
<td>61.83</td>
<td>$5.52 \cdot 10^6$</td>
</tr>
</tbody>
</table>
Texture + Geometry decomposition

Original image $\bar{x}$ with $256 \times 256$ pixels (left) and noisy image $y$, degraded with i.i.d. Gaussian noise, SNR=15 dB (right)

$y = \hat{x} + \check{x}$ with

\[
\begin{align*}
\hat{x} & \quad \text{geometry} \\
\check{x} & \quad \text{texture + noise}
\end{align*}
\]

where $\hat{x}$ minimizes ([Osher03])

\[
F_\delta(x) = \frac{1}{2} \| \nabla \Delta^{-1}(x - y) \|^2 + \lambda \sum_{s=1}^{S} \psi_{s,\delta}(\| V_s x \|)
\]

\[\leadsto\] Isotropic penalization on the gradients of $x$
Results

Recovered geometry part $\hat{x}$ (left) and texture+noise part $\check{x}$ (right) with \textit{SC} penalty using MM-MG.
Recovered geometry part $\hat{x}$ (left) and texture+noise part $\tilde{x}$ (right) with $SNC-(iii)$ penalty using MM-MG.
## Results

<table>
<thead>
<tr>
<th>Penalty</th>
<th>Algorithm</th>
<th>Iteration</th>
<th>Time</th>
<th>$F_\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC</td>
<td>MM-MG</td>
<td>633</td>
<td>20.8</td>
<td>$2.99 \cdot 10^{12}$</td>
</tr>
<tr>
<td></td>
<td>NLCG</td>
<td>591</td>
<td>17.5</td>
<td>$2.99 \cdot 10^{12}$</td>
</tr>
<tr>
<td></td>
<td>L-BFGS</td>
<td>674</td>
<td>22.9</td>
<td>$2.99 \cdot 10^{12}$</td>
</tr>
<tr>
<td></td>
<td>HQ</td>
<td>6067</td>
<td>306</td>
<td>$2.99 \cdot 10^{12}$</td>
</tr>
<tr>
<td>SC</td>
<td>MM-MG</td>
<td>448</td>
<td>11.84</td>
<td>$2.14 \cdot 10^{12}$</td>
</tr>
<tr>
<td>SNC-(iii)</td>
<td>NLCG</td>
<td>1058</td>
<td>25.1</td>
<td>$2.14 \cdot 10^{12}$</td>
</tr>
<tr>
<td></td>
<td>L-BFGS</td>
<td>457</td>
<td>13.2</td>
<td>$2.14 \cdot 10^{12}$</td>
</tr>
<tr>
<td></td>
<td>HQ</td>
<td>3882</td>
<td>168</td>
<td>$2.17 \cdot 10^{12}$</td>
</tr>
</tbody>
</table>
Image reconstruction

Original image $\bar{x}$ with $128 \times 128$ pixels (left) and noisy sinogram $y$ with $181 \times 256$ measurements, degraded by i.i.d. Laplacian noise, $SNR=23.5$ dB (right).

$$F_\delta(x) = \sum_{q=1}^{Q} \phi_{q,\rho}((Rx)_q - y_q) + \sum_{s=1}^{S} \psi_{s,\delta}(\|V_s x\|) + \beta d_B^2(x) + \tau \|x\|^2$$

$\Rightarrow R$ is the Radon projection matrix, $\phi_{q,\rho}$ is the SC function

$\Rightarrow$ Isotropic penalization on the gradients of $x$
Results

Reconstructed image (left) and detail (right) with SC penalty using MM-MG, SNR = 18.05 dB, MSSIM = 0.81.
Reconstructed image (left) and detail (right) with \textit{SNC-(i)} penalty using MM-MG, SNR $= 21.13$ dB, MSSIM $= 0.92$. 
## Results

<table>
<thead>
<tr>
<th>Penalty</th>
<th>Algorithm</th>
<th>Iteration</th>
<th>Time</th>
<th>$F_\delta$</th>
<th>SNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SC</td>
<td>MM-MG</td>
<td>253</td>
<td>59.3</td>
<td>$1.1 \cdot 10^6$</td>
<td>18.05</td>
</tr>
<tr>
<td></td>
<td>NLCG</td>
<td>358</td>
<td>84.1</td>
<td>$1.1 \cdot 10^6$</td>
<td>18.05</td>
</tr>
<tr>
<td></td>
<td>L-BFGS</td>
<td>349</td>
<td>82.3</td>
<td>$1.1 \cdot 10^6$</td>
<td>18.05</td>
</tr>
<tr>
<td></td>
<td>HQ</td>
<td>728</td>
<td>337</td>
<td>$1.1 \cdot 10^6$</td>
<td>18.05</td>
</tr>
<tr>
<td>SNC-(i)</td>
<td>MM-MG</td>
<td>516</td>
<td>119.8</td>
<td>$8.6214 \cdot 10^6$</td>
<td>21.13</td>
</tr>
<tr>
<td></td>
<td>NLCG</td>
<td>618</td>
<td>143</td>
<td>$8.6228 \cdot 10^6$</td>
<td>20.89</td>
</tr>
<tr>
<td></td>
<td>L-BFGS</td>
<td>870</td>
<td>203</td>
<td>$8.6225 \cdot 10^6$</td>
<td>21.17</td>
</tr>
<tr>
<td></td>
<td>HQ</td>
<td>1152</td>
<td>530</td>
<td>$8.6236 \cdot 10^6$</td>
<td>20.85</td>
</tr>
</tbody>
</table>
Outline

1. General context
2. $\ell_2$-$\ell_0$ regularization functions
   - Existence of minimizers
   - Epi-convergence property
3. Minimization of $F_\delta$
   - Proposed algorithm
   - Convergence results
4. Application to image processing
   - Image denoising
   - Image segmentation
   - Texture+Geometry decomposition
   - Image reconstruction
5. Conclusion
Conclusion

- Majorize-Minimize subspace algorithm for $\ell_2$-$\ell_0$ minimization
  - Faster methods w.r.t. combinatorial optimization techniques
  - Simplicity of implementation

- Future work
  - Constrained case
  - Non differentiable case
  - Application to a wider class of problems (Ex: IRM)
M. Allain, J. Idier and Y. Goussard
On global and local convergence of half-quadratic algorithms

H. Attouch, J. Bolte and B. F. Svaiter
Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods
http://www.optimization-online.org, 2010

H. Attouch, J. Bolte, P. Redont and A. Soubeyran
Proximal alternating minimization and projection methods for nonconvex problems. An approach based on the Kurdyka-Łojasiewicz inequality

E. Chouzenoux, J. Idier and S. Moussaoui
A Majorize-Minimize strategy for subspace optimization applied to image restoration

E. Chouzenoux, A. Jezierska, J.-C. Pesquet and H. Talbot
A Majorize-Minimize subspace approach for $\ell_2$-$\ell_0$ image regularization
Thanks for your attention!
Existence of minimizers (I)

\[ F_\delta(x) = \Phi(Hx - y) + \sum_{s=1}^{S} \psi_{s,\delta}(\|V_s x - c_s\|) \]

**Difficulty:** \( F_\delta \) is a non convex, non coercive function.

**Proposition 1**

Assume that

(i) \( \Phi \) is continuous and coercive, i.e. \( \lim_{\|x\| \to +\infty} \Phi(x) = +\infty \)

(ii) For every \( \delta > 0 \) and \( s \in \{1, \ldots, S\} \), \( \psi_{s,\delta} \) is continuous and takes nonnegative values

(iii) \( \text{Ker} \ H = \{0\} \)

Then, for every \( \delta > 0 \), \( F_\delta \) has a minimizer.
Existence of minimizers (II)

$$F_\delta(x) = \Phi(Hx - y) + \sum_{s=1}^{S} \psi_{s,\delta}(\|V_s x - c_s\|) + \|V_0 x\|^2$$

**Difficulty:** $F_\delta$ is a non convex, non coercive function.

**Proposition 1**

Assume that

(i) $\Phi$ is continuous and coercive, i.e. $\lim_{\|x\| \to +\infty} \Phi(x) = +\infty$

(ii) For every $\delta > 0$ and $s \in \{1, \ldots, S\}$, $\psi_{s,\delta}$ is continuous and takes nonnegative values

(iii) $\text{Ker } H \cap \text{Ker } V_0 = \{0\}$

Then, for every $\delta > 0$, $F_\delta$ has a minimizer.
Existence of minimizers (III)

\[
F_\delta(x) = \Phi(Hx - y) + \sum_{s=1}^{S} \psi_{s,\delta}(\|V_s x - c_s\|)
\]

**Difficulty:** \(F_\delta\) is a non convex, non coercive function.

**Proposition 1**

Assume that

(i) \(\Phi\) is continuous and coercive, i.e. \(\lim_{\|x\| \to +\infty} \Phi(x) = +\infty\)

(ii) For every \(\delta > 0\) and \(s \in \{1, \ldots, S\}\), \(\psi_{s,\delta}\) is continuous and takes nonnegative values and \(\psi_{s,\delta}^{-1}(0)\) is a nonempty bounded set.

(iii) \(\text{Ker } H \cap \bigcap_{s=1}^{S} \text{Ker } V_s = \{0\}\)

Then, for every \(\delta > 0\), \(F_\delta\) has a minimizer.
Quadratic tangent majorant function

Assumptions:
(i) Φ is differentiable with an $L$-Lipschitzian gradient
(ii) For every $s \in \{1, \cdots, S\}$, $\psi_{s,\delta}$ is a differentiable function.
(iii) For every $s \in \{1, \cdots, S\}$, $\psi_{s,\delta}(\sqrt{\cdot})$ is concave on $[0, +\infty)$.
(iv) For every $s \in \{1, \cdots, S\}$, there exists $\omega_{s} \in [0, +\infty)$ such that 
    $$(\forall t \in (0, +\infty)) \ 0 \leq \psi_{s,\delta}(t) \leq \omega_{s} t$$
In addition, $\lim_{t \to 0} \omega_{s,\delta}(t) \in \mathbb{R}$ with $\omega_{s,\delta}(t) \triangleq \psi_{s,\delta}(t)/t$.

Lemma 1 [Allain06]

For every $x \in \mathbb{R}^{N}$, let

$$A(x) = \mu H^{\top}H + 2V_{0}^{\top}V_{0} + V^{\top}\text{Diag}\{b(x)\}V$$

where $\mu \in [L, +\infty)$ and $b(x) \in \mathbb{R}^{SP}$ with $b_{sp}(x) = \omega_{s,\delta}(\|V_{s}x - c_{s}\|)$.
Then, $Q(x, x') = F_{\delta}(x') + \nabla F_{\delta}(x')^{T}(x - x') + \frac{1}{2}(x - x')^{T}A(x')(x - x')$ is a convex quadratic tangent majorant of $F_{\delta}$ at $x'$. 

Emilie Chouzenoux
Séminaire Image, GREYC