Majorize-Minimize Memory Gradient Methods for Data Processing

Jean-Christophe Pesquet

Laboratoire d’Informatique Gaspard Monge - CNRS
Univ. Paris-Est Marne-la-Vallée, France

3 Dec. 2015
In collaboration with

A. Florescu    E. Chouzenoux    P. Ciuciu    S. Ciochina
Outline

1. Introduction

2. Proposed optimization method
   - Preliminaries
   - Proposed algorithm
   - Convergence result

3. Application to CS-PMRI
   - Model
   - Simulation results

4. Online algorithm

5. Conclusion
Outline

1 Introduction

2 Proposed optimization method
   - Preliminaries
   - Proposed algorithm
   - Convergence result

3 Application to CS-PMRI
   - Model
   - Simulation results

4 Online algorithm

5 Conclusion
General context

We observe data $y \in \mathbb{C}^Q$, related to the original image $\bar{x} \in \mathbb{C}^N$ through:

$$y = H\bar{x} + w, \quad H \in \mathbb{C}^{Q \times N}$$

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

Examples of complex-valued inverse problems:

- Spectral analysis
- Nuclear Magnetic Resonance
- Mass Spectroscopy
- Magnetic Resonance Imaging
General context

Penalized optimization problem

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

(1)

where

\[ \Phi : \mathbb{C}^Q \to \mathbb{R} \sim \text{Data fidelity term, related to noise model} \]
\[ \Psi : \mathbb{C}^N \to \mathbb{R} \sim \text{Regularization term, related to a priori assumptions} \]

Considered penalization model:

\[
\Psi(x) = \sum_{s=1}^{S} \psi_s(|v_s^Hx - c_s|) + \frac{\varepsilon}{2} \|x\|^2,
\]

- For every \( s \in \{1, \ldots, S\} \), \( \psi_s : \mathbb{R} \to \mathbb{R} \), \( v_s \in \mathbb{C}^N \), \( c_s \in \mathbb{C} \),
- \( \varepsilon \in [0, +\infty) \).
Examples of regularization functions

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

Example: \((\forall t \in \mathbb{R}), \psi_s(t) = \lambda_s(\sqrt{\delta_s^2 + t^2} - \delta_s), \lambda_s > 0, \delta_s > 0\)

Limit case: When \(\delta_s \to 0\), \(\psi_s(t) = \lambda_s |t| \) (\(\ell_1\) penalty).
Examples of regularization functions

**$\ell_2$-$\ell_1$ functions:** Asymptotically linear with a quadratic behavior near 0.

**$\ell_2$-$\ell_0$ functions:** Asymptotically constant with a quadratic behavior near 0.

*Example:* $(\forall t \in \mathbb{R}), \psi_s(t) = \lambda_s (2\delta_s^2 + t^2)^{-1}t^2, \lambda_s > 0, \delta_s > 0$

*Limit case:* When $\delta_s \to 0$, $\psi_s(t) \to 0$ if $t = 0$, $\lambda_s$ otherwise ($\ell_0$ penalty).
Examples of functions \( (\psi_s)_1 \leq s \leq S \)

<table>
<thead>
<tr>
<th>( \lambda_s^{-1} \psi_s(t) )</th>
<th>Type</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>(</td>
<td>t</td>
<td>- \delta_s \log(</td>
</tr>
<tr>
<td>(t^2) if (</td>
<td>t</td>
<td>&lt; \delta_s) (2\delta_s</td>
</tr>
<tr>
<td>(\log(\cosh(t)))</td>
<td>(\ell_2 - \ell_1)</td>
<td>Green</td>
</tr>
<tr>
<td>((1 + t^2/\delta_s^2)^{\kappa_s/2} - 1)</td>
<td>(\ell_2 - \ell_{\kappa_s})</td>
<td></td>
</tr>
</tbody>
</table>

Convex

| \(1 - \exp(-t^2/(2\delta_s^2))\) | \(\ell_2 - \ell_0\) | Welsch |
| \(t^2/(2\delta_s^2 + t^2)\) | \(\ell_2 - \ell_0\) | Geman-\(-McClure\) |
| \(1 - (1 - t^2/(6\delta_s^2))^3\) if \(|t| \leq \sqrt{6}\delta_s\) \(1\) otherwise | \(\ell_2 - \ell_0\) | Tukey biweight |
| \(\tanh(t^2/(2\delta_s^2))\) | \(\ell_2 - \ell_0\) | Hyperbolic tangent |
| \(\log(1 + t^2/\delta_s^2)\) | \(\ell_2 - \ell_{\log}\) | Cauchy |
| \(1 - \exp(1 - (1 + t^2/(2\delta_s^2))^{\kappa_s/2})\) | \(\ell_2 - \ell_{\kappa_s} - \ell_0\) | Chouzenoux |

Nonconvex

\((\lambda_s, \delta_s) \in [0, +\infty]^2, \kappa_s \in [1, 2]\)
Examples of functions \((\psi_s)_{1 \leq s \leq S}\)

\[
\psi_s(t) = (1 + \frac{t^2}{\delta^2})^{1/2} - 1, \quad \psi_s(t) = \log \left(1 + \frac{t^2}{\delta^2}\right), \quad \psi_s(t) = 1 - \exp(-\frac{t^2}{2\delta^2}).
\]
Outline

1. Introduction

2. Proposed optimization method
   - Preliminaries
   - Proposed algorithm
   - Convergence result

3. Application to CS-PMRI
   - Model
   - Simulation results

4. Online algorithm

5. Conclusion
Preliminaries

Notation:

- For every vector \( x \in \mathbb{C}^N \),
  - \( x_R \in \mathbb{R}^N \) (resp. \( x_I \in \mathbb{R}^N \)) denotes the vector of real (resp. imaginary) parts of the components of \( x \).
  - \( \tilde{x} \in \mathbb{R}^{2N} \) denotes the “concatenated” vector \( \tilde{x} = [x_R^\top \ x_I^\top]^\top \) where \((\cdot)^\top\) is the transpose operation.

- If \( F \) is a function from \( \mathbb{C}^N \) to \( \mathbb{C} \), we define \( \tilde{F} \) the function of real variables associated with \( F \), i.e. \( \forall x \in \mathbb{C}^N \) \( \tilde{F}(\tilde{x}) = F(x) \).
Preliminaries

Notation:

- For every vector $\mathbf{x} \in \mathbb{C}^N$,
  - $\mathbf{x}_R \in \mathbb{R}^N$ (resp. $\mathbf{x}_I \in \mathbb{R}^N$) denotes the vector of real (resp. imaginary) parts of the components of $\mathbf{x}$.
  - $\tilde{\mathbf{x}} \in \mathbb{R}^{2N}$ denotes the “concatenated” vector $\tilde{\mathbf{x}} = [\mathbf{x}_R^\top \mathbf{x}_I^\top]^\top$ where $(\cdot)^\top$ is the transpose operation.

- If $F$ is a function from $\mathbb{C}^N$ to $\mathbb{C}$, we define $\tilde{F}$ the function of real variables associated with $F$, i.e. $(\forall \mathbf{x} \in \mathbb{C}^N) \tilde{F}(\tilde{\mathbf{x}}) = F(\mathbf{x})$.

Complex-valued differential calculus:
According to Wirtinger’s calculus, the derivative of $F$ with respect to the conjugate of its variable is formally defined as

$$(\forall \mathbf{x} \in \mathbb{C}^N) \quad \nabla F(\mathbf{x}) = \frac{1}{2} \left( \frac{\partial \tilde{F}(\tilde{\mathbf{x}})}{\partial \mathbf{x}_R} + i \frac{\partial \tilde{F}(\tilde{\mathbf{x}})}{\partial \mathbf{x}_I} \right).$$
Assumptions

\[ F(x) = \Phi(Hx - y) + \sum_{s=1}^{S} \psi_s(|v_s^H x - c_s|) + \frac{\varepsilon}{2} ||x||^2 \]

Assumption 1:
(i) \( \tilde{\Phi} \) is differentiable.
(ii) For every \( s \in \{1, \ldots, S\} \), \( \psi_s \) is differentiable and \( \lim_{t \to 0} \frac{\dot{\psi}_s(t)}{t} \in \mathbb{R} \).

Assumption 2: One of the following conditions holds:
- \( \Phi \) and \( (\psi_s)_{1 \leq s \leq S} \) are lower bounded functions and \( \varepsilon > 0 \).
- (i) \( \Phi \) is coercive (i.e. \( \lim_{||z|| \to +\infty} \Phi(z) = +\infty \)).
  (ii) \( (\psi_s)_{1 \leq s \leq S} \) are lower bounded functions.
  (iii) \( H \) is injective.
- (i) \( \Phi \) is coercive.
  (ii) For every \( s \in \{1, \ldots, S\} \), \( \psi_s \) is coercive.
  (iii) \( \text{Ker} \, H \cap (\text{span}\{v_1, \ldots, v_S\})^\perp = \{0\} \)
Properties

Complex-valued derivative of $F$

Under **Assumption 1**, for all $x \in \mathbb{C}^N$,

$$\nabla F(x) = H^H \nabla \Phi(Hx - y) + \frac{1}{2} V \ \text{Diag} \ (b(x)) \ (V^H x - c) + \frac{\varepsilon}{2} x,$$

with

- $V = [v_1, \ldots, v_S] \in \mathbb{C}^{N \times S}$,
- $b(x) = (\omega_s(|v_s^H x - c_s|))_{1 \leq s \leq S}$,
- $(\forall s \in \{1, \cdots, S\})(\forall a \in \mathbb{R}) \ \omega_s(a) = \begin{cases} \dfrac{\dot{\psi}_s(a)}{a} & \text{if } a \neq 0 \\ \lim_{t \to 0} \frac{\dot{\psi}_s(t)}{t} & \text{otherwise.} \end{cases}$
Properties

Complex-valued derivative of $F$

Under Assumption 1, for all $x \in \mathbb{C}^N$, \n\[ \nabla F(x) = H^H \nabla \Phi(Hx - y) + \frac{1}{2} V \text{Diag}(b(x)) (V^H x - c) + \frac{\varepsilon}{2} x, \]

with

- $V = [v_1, \ldots, v_S] \in \mathbb{C}^{N \times S}$,
- $b(x) = (\omega_s(|v_s^H x - c_s|))_{1 \leq s \leq S}$,
- $(\forall s \in \{1, \ldots, S\})(\forall a \in \mathbb{R}) \omega_s(a) = \begin{cases} \frac{\dot{\psi}_s(a)}{a} & \text{if } a \neq 0 \\ \lim_{t \to 0} \frac{\dot{\psi}_s(t)}{t} & \text{otherwise}. \end{cases}$

Existence of minimizers

Under Assumptions 1 and 2, Problem (1) has a solution.
Majorize-Minimize principle [Hunter04]

**Objective:** Find $\hat{x} \in \text{Arg min } F$

For all $x'$, let $\Theta(., x')$ a *tangent majorant* of $F$ at $x'$ i.e.,

$$\Theta(x, x') \geq F(x) \quad (\forall x),$$

$$\Theta(x', x') = F(x')$$

**MM algorithm:**

$$(\forall j \in \{1, \ldots, J\})$$

$$x^{j+1} \in \text{Arg min}_x \Theta(x, x^j)$$
Quadratic majorization

Assumption 3:

(i) \( \Phi \) has a \( \beta \)-Lipschitzian derivative with \( \beta \in (0, +\infty) \), i.e.

\[
(\forall z \in \mathbb{C}^Q) (\forall z' \in \mathbb{C}^Q) \quad \|\nabla \Phi(z) - \nabla \Phi(z')\| \leq \beta \|z - z'\|.
\]

(ii) For every \( s \in \{1, \ldots, S\} \), \( \psi_s(\sqrt{\cdot}) \) is concave on \([0, +\infty)\).

(iii) There exists \( \omega \in [0, +\infty) \) such that

\[
(\forall s \in \{1, \ldots, S\} \quad (\forall t \in (0, +\infty)) \quad 0 \leq \omega_s(t) \leq \omega.
\]

Proposition

If, for every \( x' \in \mathbb{C}^N \), \( A(x') = \mu H^H H + V \text{ Diag} (b(x')) V^H + \epsilon I_N \) with \( \mu \in [2\beta, +\infty) \), then

\[
\Theta(x, x') = F(x') + 2 \text{Re} \left\{ \nabla F(x')^H (x - x') \right\} + \frac{1}{2} (x - x')^H A(x')(x - x')
\]

is a quadratic tangent majorant of \( F \) at \( x' \).
Proposed algorithm

MM Subspace algorithm:

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

MM Subspace algorithm:

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

- \( D_k \in \mathbb{C}^{N \times M} \): matrix of \( M \) directions

Example: Memory gradient \( D_k = [-\nabla F(x_k), x_k - x_{k-1}] \)
Proposed algorithm

MM Subspace algorithm:

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

- \( D_k \in \mathbb{C}^{N \times M} \): matrix of \( M \) directions
- \( u_k \in \mathbb{C}^M \): multivariate stepsize resulting from MM minimization of \( f_k(u) : u \mapsto F(x_k + D_k u) \)
Proposed algorithm

**MM Subspace algorithm:**

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

- \( D_k \in \mathbb{C}^{N \times M} \): matrix of \( M \) directions
- \( u_k \in \mathbb{C}^M \): multivariate stepsize resulting from MM minimization of \( f_k(u) : u \mapsto F(x_k + D_k u) \)

**MM minimization in the subspace:**

\[
\begin{aligned}
\begin{cases}
    u_k^0 &= 0, \\
    u_k^j &\in \text{Arg min}_u \vartheta_k(u, u_k^{j-1}) (\forall j \in \{1, \ldots J\})
\end{cases}
\end{aligned}
\]

with, for all \( u \in \mathbb{C}^M \), \( \vartheta_k(u, u_k^j) = \Theta(x_k + D_k u, x_k + D_k u_k^j) \), quadratic tangent majorant of \( f_k \) at \( u_k^j \) with Hessian:

\[ B_k^j = D_k^H A(x_k + D_k u_k^j) D_k \]
Proposed algorithm

Complex-valued 3MG algorithm

\( x_0 \in \mathbb{C}^N, \; x_{-1} = 0 \)

For all \( k = 0, \ldots \)

\[
D_k = [-\nabla F(x_k), x_k - x_{k-1}]
\]

\( u^0_k = 0, \)

For all \( j = 1, \ldots, J \)

\[
B_k^{j-1} = D_k^H A(x_k + D_k u_k^{j-1})D_k,
\]

\[
u_k^j = u_k^{j-1} - 2(B_k^{j-1})^* D_k^H \nabla F(x_k + D_k u_k^{j-1}),
\]

\( x_{k+1} = x_k + D_k u_k^J. \)
Proposed algorithm

Complex-valued 3MG algorithm

\[ x_0 \in \mathbb{C}^N, \ x_{-1} = 0 \]

For all \( k = 0, \ldots \)

\[
D_k = \begin{bmatrix}
-\nabla F(x_k), x_k - x_{k-1}
\end{bmatrix}
\]

\( u_0^k = 0, \)

For all \( j = 1, \ldots, J \)

\[
B_{k}^{j-1} = D_k^H A (x_k + D_k u_{k}^{j-1}) D_k,
\]

\[
u_{k}^{j} = u_{k}^{j-1} - 2(B_{k}^{j-1})^\dagger D_k^H \nabla F(x_k + D_k u_{k}^{j-1}),
\]

\[
x_{k+1} = x_k + D_k u_{k}^{J}.
\]

\[\mapsto\text{ Equivalent to 3MG algorithm for minimizing real-valued function } \tilde{F}, \]

taking

\[
\tilde{D}_k = \begin{bmatrix}
D_{k,R} & -D_{k,I} \\
D_{k,I} & D_{k,R}
\end{bmatrix} \in \mathbb{R}^{2N \times 2M}
\]
Convergence result

Assumptions

- Assumption 1
- Assumption 2
- Assumption 3
- **Assumption 4:** $F$ satisfies the Kurdyka-Łojasiewicz inequality, i.e. for every $\hat{x} \in \mathbb{C}^N$ and every bounded neighborhood $B$ of $\hat{x}$, there exist constants $\kappa > 0$, $\zeta > 0$ and $\theta \in [0, 1)$ such that

  \[
  \|\nabla F(x)\| \geq \kappa |F(x) - F(\hat{x})|^\theta, \\
  \text{for every } x \in B \text{ such that } |F(x) - F(\hat{x})| \leq \zeta.
  \]
Proposition

Assume that there exists $\alpha \in (0, +\infty)$ such that $(\forall x \in \mathbb{C}^N) \ A(x) - \alpha I_N$ is a positive semi-definite matrix. Then, under Assumptions 1-4, the 3MG algorithm generates a sequence $(x_k)_{k \in \mathbb{N}}$ converging to a critical point of $F$. Moreover, $(F(x_k))_{k \in \mathbb{N}}$ is a nonincreasing sequence and $(x_k)_{k \in \mathbb{N}}$ has a finite length in the sense that

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

Finally, there exists $\eta \in (0, +\infty)$ such that, if

$$F(x_0) \leq \eta + \inf F,$$

then $(x_k)_{k \in \mathbb{N}}$ converges to a global solution to Problem (1).
Outline

1. Introduction

2. Proposed optimization method
   - Preliminaries
   - Proposed algorithm
   - Convergence result

3. Application to CS-PMRI
   - Model
   - Simulation results

4. Online algorithm

5. Conclusion
Parallel Magnetic Resonance Imaging

Objective:
- Reduce the acquisition time
- Maintain good image quality

Principle:
- k-space subsampling
- Multiple receiver coils
Parallel Magnetic Resonance Imaging

Objective:
- Reduce the acquisition time
- Maintain good image quality

Principle:
- k-space subsampling
- Multiple receiver coils

Acquisition model: \((\forall \ell \in \{1, \ldots, L\}) \quad d_\ell = \Sigma F S_\ell \rho + w_\ell\)
Parallel Magnetic Resonance Imaging

Objective:
- Reduce the acquisition time
- Maintain good image quality

Principle:
- k-space subsampling
- Multiple receiver coils

Acquisition model: \((\forall \ell \in \{1, \ldots, L\})\) \quad d_\ell = \Sigma F S_\ell \rho + w_\ell

- \(\forall \ell \in \{1, \ldots, L\}, S_\ell \in \mathbb{C}^{K \times K}\): diagonal sensitivity matrix,
Parallel Magnetic Resonance Imaging

Objective:
- Reduce the acquisition time
- Maintain good image quality

Principle:
- k-space subsampling
- Multiple receiver coils

Acquisition model: \((\forall \ell \in \{1, \ldots, L\})\)
\[ d_\ell = \Sigma F S_\ell \rho + w_\ell \]
- \(\forall \ell \in \{1, \ldots, L\}, S_\ell \in \mathbb{C}^{K \times K} : \) diagonal sensitivity matrix,
- \(F \in \mathbb{C}^{K \times K} : \) 2D discrete Fourier transform,
Parallel Magnetic Resonance Imaging

Objective:
- Reduce the acquisition time
- Maintain good image quality

Principle:
- k-space subsampling
- Multiple receiver coils

Acquisition model: \((\forall \ell \in \{1, \ldots, L\}) \quad d_\ell = \sum F S_\ell \rho + w_\ell\)

- \(\forall \ell \in \{1, \ldots, L\}\), \(S_\ell \in \mathbb{C}^{K \times K}\): diagonal sensitivity matrix,
- \(F \in \mathbb{C}^{K \times K}\): 2D discrete Fourier transform,
- \(\Sigma \in \{0, 1\}^{\lfloor K/R \rfloor \times K}\): subsampling matrix
Parallel Magnetic Resonance Imaging

Objective:
- Reduce the acquisition time
- Maintain good image quality

Principle:
- k-space subsampling
- Multiple receiver coils

Acquisition model:  \( \forall \ell \in \{1, \ldots, L\} \) \( d_{\ell} = \Sigma F S_{\ell} \rho + w_{\ell} \)

- \( \forall \ell \in \{1, \ldots, L\} \), \( S_{\ell} \in \mathbb{C}^{K \times K} \): diagonal sensitivity matrix,
- \( F \in \mathbb{C}^{K \times K} \): 2D discrete Fourier transform,
- \( \Sigma \in \{0, 1\}^{\lfloor \frac{K}{R} \rfloor \times K} \): subsampling matrix
- \( \forall \ell \in \{1, \ldots, L\} \), \( w_{\ell} \in \mathbb{C}^{\lfloor \frac{K}{R} \rfloor} \): realization of circular complex Gaussian noise with zero-mean and covariance matrix \( \Lambda_{\ell} \).
Variational formulation

\[
\min_{\rho \in \mathbb{E}} \left( \sum_{\ell=1}^{L} \| \Sigma F S_{\ell} \rho - d_{\ell} \|_{\Lambda_{\ell}^{-1}}^{2} + \sum_{s=1}^{S} \psi_{s}(|f_{s}^{H} \rho|) \right)
\]
Variational formulation

\[
\begin{align*}
\min_{\rho \in \mathbb{E}} \quad & \left( \sum_{\ell=1}^{L} \| \Sigma F S_{\ell} \rho - d_{\ell} \|^{2}_{\Lambda_{\ell}^{-1}} + \sum_{s=1}^{S} \psi_{s}(|f_{s}^{H} \rho|) \right) \\
\Downarrow \quad & \min_{x \in \mathbb{C}^{N}} \left( \sum_{\ell=1}^{L} \| \Sigma F S_{\ell} E x - d_{\ell} \|^{2}_{\Lambda_{\ell}^{-1}} + \sum_{s=1}^{S} \psi_{s}(|f_{s}^{H} E x|) + \frac{\varepsilon}{2} \| x \|^{2} \right)
\end{align*}
\]

where \( E \in \mathbb{C}^{K \times N} \) allows us to set the background pixels to zero.
Variational formulation

\[
\begin{align*}
\text{minimize } \rho \in \mathbb{B}X & \quad \left( \sum_{\ell=1}^{L} \| \Sigma F S_{\ell} \rho - d_{\ell} \|_{\Lambda_{\ell}^{-1}}^{2} + \sum_{s=1}^{S} \psi_s(|f_{s}^{H} \rho|) \right) \\
\approx \text{minimize } x \in \mathbb{C}^N & \quad \left( \sum_{\ell=1}^{L} \| \Sigma F S_{\ell} E x - d_{\ell} \|_{\Lambda_{\ell}^{-1}}^{2} + \sum_{s=1}^{S} \psi_s(|f_{s}^{H} E x|) + \frac{\varepsilon}{2} \| x \|^{2} \right) \\
\Leftrightarrow \text{minimize } x \in \mathbb{C}^N & \quad \left( \Phi(H x - y) + \sum_{s=1}^{S} \psi_s(|v_{s}^{H} x - c_{s}|) + \frac{\varepsilon}{2} \| x \|^{2} \right)
\end{align*}
\]

with \( \Phi \) squared Hermitian norm of \( \mathbb{C}^Q \) with \( Q = L\lfloor K/R \rfloor \) and

- \( H = [H_{1}^{T}, \ldots, H_{L}^{T}]^{T}, \quad (\forall \ell \in \{1, \ldots, L\}) \quad H_{\ell} = \Lambda_{\ell}^{-1/2} \Sigma F S_{\ell} E \)
- \( y = [y_{1}^{T}, \ldots, y_{L}^{T}]^{T}, \quad (\forall \ell \in \{1, \ldots, L\}) \quad y_{\ell} = \Lambda_{\ell}^{-1/2} d_{\ell} \)
- \( (v_{s})_{1 \leq s \leq S} = (E^{H} f_{s})_{1 \leq s \leq S}, \quad (c_{s})_{1 \leq s \leq S} = 0 \)
Simulation settings

- 3 Tesla Siemens Trio magnet with $L = 32$ channel receiver coil
- Reconstruction of sagittal views of a 3D anatomical image with $256 \times 256$ pixels
- Reference image $\tilde{\rho}$ defined as the reconstruction result from a non-accelerated acquisition ($R = 1$)
- Different sampling patterns with $R = 5$ acceleration factor
- Circular complex Gaussian noise with zero-mean and covariance matrices $\Lambda_\ell = \sigma^2 I_{\lfloor K/R \rfloor}$, $\ell \in \{1, \ldots, L\}$, $\sigma^2 = 6 \times 10^9$
- $(f_s)_{1 \leq s \leq S}$ ($S = K$) corresponds to an orthonormal wavelet basis using Symmlet filters of length 10 and 3 resolution levels
Effect of the sensitivity matrices

Moduli of the images corresponding to \((S_{\ell \bar{\rho}})_{1\leq \ell \leq L}\) for 8 channels out of 32.
Different types of subsampling

Poly1

Poly2

Poly3

Uniform

π

Radial

Spiral

Regular
Simulation results

<table>
<thead>
<tr>
<th>Sampling pattern</th>
<th>Slide No 70</th>
<th>Slice No 82</th>
<th>Slice No 121</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poly1</td>
<td>21.15</td>
<td>19.96</td>
<td>20.89</td>
</tr>
<tr>
<td>Poly2</td>
<td>20.32</td>
<td>19.34</td>
<td>20.07</td>
</tr>
<tr>
<td>Poly3</td>
<td>19.43</td>
<td>18.53</td>
<td>19.18</td>
</tr>
<tr>
<td>Poly4</td>
<td>18.47</td>
<td>17.50</td>
<td>18.35</td>
</tr>
<tr>
<td>Poly5</td>
<td>17.67</td>
<td>16.95</td>
<td>17.52</td>
</tr>
<tr>
<td>Uniform</td>
<td>21.02</td>
<td>19.71</td>
<td>20.68</td>
</tr>
<tr>
<td>π</td>
<td>20.46</td>
<td>19.31</td>
<td>20.08</td>
</tr>
<tr>
<td>Radial</td>
<td>20.27</td>
<td>19.20</td>
<td>20.01</td>
</tr>
<tr>
<td>Spiral</td>
<td>20.35</td>
<td>19.17</td>
<td>20.03</td>
</tr>
<tr>
<td>Regular</td>
<td>19.18</td>
<td>18.13</td>
<td>18.66</td>
</tr>
</tbody>
</table>

SNR values for various subsampling strategies using 3MG and $\ell_2 - \ell_1$ regularization
Simulation results

Slice No 70: Moduli of the original image $\bar{\rho}$ (a) and the reconstructed one (b) with SNR = 21.15 dB using Poly1 sampling, 3MG algorithm and $\ell_2 - \ell_1$ regularization.
Simulation results

Slice No 82: Moduli of the original image $\tilde{p}$ (a) and the reconstructed one (b) with $\text{SNR} = 19.95 \text{ dB}$ using Poly1 sampling, 3MG algorithm and $\ell_2 - \ell_1$ regularization.
Simulation results

Slice No 121: Moduli of the original image $\bar{p}$ (a) and the reconstructed one (b) with $SNR = 20.89$ dB using Poly1 sampling, 3MG algorithm and $\ell_2 - \ell_1$ regularization.
Simulation results

<table>
<thead>
<tr>
<th>Decomp.</th>
<th>Algorithm</th>
<th>Penalization</th>
<th>SNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Slice No 70</td>
</tr>
<tr>
<td>Wav. basis</td>
<td>M+LFBF</td>
<td>(\ell_1)</td>
<td>21.15</td>
</tr>
<tr>
<td></td>
<td>CPCV</td>
<td>(\ell_1)</td>
<td>21.15</td>
</tr>
<tr>
<td></td>
<td>ADMM</td>
<td>(\ell_1)</td>
<td>21.15</td>
</tr>
<tr>
<td></td>
<td>3MG</td>
<td>(\ell_2 - \ell_1)</td>
<td>21.15</td>
</tr>
<tr>
<td></td>
<td>3MG</td>
<td>(\ell_2 - \ell_0) (H)</td>
<td>21.09</td>
</tr>
<tr>
<td></td>
<td>3MG</td>
<td>(\ell_2 - \ell_0) (W)</td>
<td>21.21</td>
</tr>
<tr>
<td></td>
<td>3MG</td>
<td>(\ell_2 - \ell_0) (G)</td>
<td>21.33</td>
</tr>
<tr>
<td>Redundant wav. frame</td>
<td>3MG</td>
<td>(\ell_2 - \ell_1)</td>
<td>21.67</td>
</tr>
<tr>
<td></td>
<td>3MG</td>
<td>(\ell_2 - \ell_0) (G)</td>
<td><strong>22.10</strong></td>
</tr>
</tbody>
</table>

Reconstruction results for several optimization and regularization strategies using two different decompositions (Poly1 subsampling pattern)
Simulation results

SNR evolution as a function of computation time
Simulation results

Error $\|x_k - \hat{x}\|$ as a function of computation time
Outline

1 Introduction

2 Proposed optimization method
   - Preliminaries
   - Proposed algorithm
   - Convergence result

3 Application to CS-PMRI
   - Model
   - Simulation results

4 Online algorithm

5 Conclusion
S3MG

- Stochastic version for solving online/adaptive problems

The parameters of each tested method are optimized manually.

The Stochastic Majorize-Minimize Memory gradient (S3MG) algorithm leads to a minimal estimation error, while benefiting from good tracking properties.
Outline

1. Introduction

2. Proposed optimization method
   - Preliminaries
   - Proposed algorithm
   - Convergence result

3. Application to CS-PMRI
   - Model
   - Simulation results

4. Online algorithm

5. Conclusion
Conclusion

- Majorize-Minimize Memory Gradient algorithm for optimization of smooth nonconvex complex-valued functions.

- Application to Parallel Magnetic Resonance Imaging
  - Faster than standard proximal techniques

- Future work
  - Application to other inverse problems (CEA-LETI: microscopy imaging)
  - Non-smooth case
Some references …

L. Chaâri, J.-C. Pesquet, A. Benazza-Benyahia and Ph. Ciuciu,
A wavelet-based regularized reconstruction algorithm for SENSE parallel MRI with applications to neuroimaging
*Medical Image Analysis*, vol. 15, pp. 185-201, Apr. 2011.

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

E. Chouzenoux, J.-C. Pesquet and A. Florescu.
A Stochastic 3MG Algorithm with Application to 2D Filter Identification

A. Florescu, E. Chouzenoux, J.-C. Pesquet, P. Ciuciu and S. Ciochina
A Majorize-Minimize memory gradient method for complex-valued inverse problems