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

Outline

General context

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

$$\boldsymbol{y} = \boldsymbol{H}\overline{\boldsymbol{x}} + \boldsymbol{w}, \qquad \boldsymbol{H} \in \mathbb{C}^{Q \times N}$$

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

Examples of complex-valued inverse problems:

⇝ Spectral analysis

⇝ Nuclear Magnetic Resonance

⇝ Mass Spectroscopy

⇝ Magnetic Resonance Imaging

General context

Penalized optimization problem

$$\underset{\boldsymbol{x}\in\mathbb{C}^N}{\text{minimize}} \;\; \big(F(\boldsymbol{x}) = \Phi(\boldsymbol{H}\boldsymbol{x}-\boldsymbol{y}) + \Psi(\boldsymbol{x})\big), \tag{1}$$

where

Φ : $\mathbb{C}^Q \to \mathbb{R} \rightsquigarrow$ Data fidelity term, related to noise model

Ψ : $\mathbb{C}^N \to \mathbb{R} \rightsquigarrow$ Regularization term, related to *a priori* assumptions

Considered penalization model:

$$\Psi(\boldsymbol{x}) = \sum_{s=1}^{S} \psi_s(|\boldsymbol{v}_s^{\mathrm{H}}\boldsymbol{x} - c_s|) + \frac{\varepsilon}{2}\|\boldsymbol{x}\|^2,$$

- For every $s \in \{1,\dots,S\}$, $\psi_s\colon \mathbb{R}\to\mathbb{R}$, $\boldsymbol{v}_s \in \mathbb{C}^N$, $c_s \in \mathbb{C}$,
- $\varepsilon \in [0,+\infty)$.

Examples of regularization functions

ℓ_2-ℓ_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 $\delta_s \to 0$, $\psi_s(t) = \lambda_s|t|$ (ℓ_1 penalty).

Examples of regularization functions

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

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

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

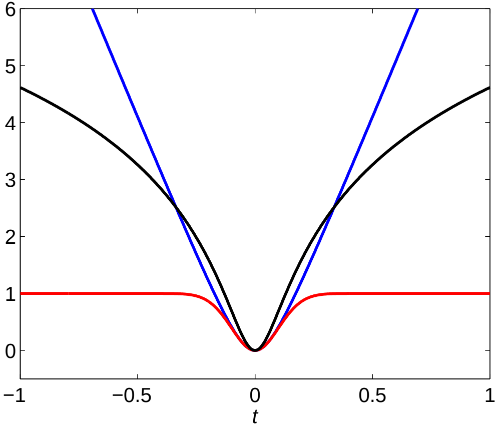
Limit case: When $\delta_s \to 0$, $\psi_s(t) \to 0$ if $t = 0$, λ_s otherwise (ℓ_0 penalty).

Examples of functions $(\psi_s)_{1\leq s\leq S}$

	$\lambda_s^{-1}\psi_s(t)$	Type	Name
Convex	$\lvert t\rvert - \delta_s \log(\lvert t\rvert/\delta_s + 1)$	$\ell_2 - \ell_1$	
Convex	$\begin{cases} t^2 & \text{if } \lvert t\rvert < \delta_s \\ 2\delta_s\lvert t\rvert - \delta_s^2 & \text{otherwise} \end{cases}$	$\ell_2 - \ell_1$	Huber
Convex	$\log(\cosh(t))$	$\ell_2 - \ell_1$	Green
Convex	$(1 + t^2/\delta_s^2)^{\kappa_s/2} - 1$	$\ell_2 - \ell_{\kappa_s}$	
Nonconvex	$1 - \exp(-t^2/(2\delta_s^2))$	$\ell_2 - \ell_0$	Welsch
Nonconvex	$t^2/(2\delta_s^2 + t^2)$	$\ell_2 - \ell_0$	Geman -McClure
Nonconvex	$\begin{cases} 1 - (1 - t^2/(6\delta_s^2))^3 & \text{if } \lvert t\rvert \leqslant \sqrt{6}\delta_s \\ 1 & \text{otherwise} \end{cases}$	$\ell_2 - \ell_0$	Tukey biweight
Nonconvex	$\tanh(t^2/(2\delta_s^2))$	$\ell_2 - \ell_0$	Hyberbolic tangent
Nonconvex	$\log(1 + t^2/\delta_s^2)$	$\ell_2 - \log$	Cauchy
Nonconvex	$1 - \exp(1 - (1 + t^2/(2\delta_s^2))^{\kappa_s/2})$	$\ell_2 - \ell_{\kappa_s} - \ell_0$	Chouzenoux

$(\lambda_s, \delta_s) \in]0, +\infty[^2$, $\kappa_s \in [1, 2]$

Examples of functions $(\psi_s)_{1 \leq s \leq S}$



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

Outline

Preliminaries

Notation:

- For every vector $\boldsymbol{x} \in \mathbb{C}^N$,
  - $\boldsymbol{x}_R \in \mathbb{R}^N$ (resp. $\boldsymbol{x}_I \in \mathbb{R}^N$) denotes the vector of real (resp. imaginary) parts of the components of $\boldsymbol{x}$.
  - $\widetilde{\boldsymbol{x}} \in \mathbb{R}^{2N}$ denotes the "concatenated" vector $\widetilde{\boldsymbol{x}} = [\boldsymbol{x}_R^\top \;\; \boldsymbol{x}_I^\top]^\top$ where $(\cdot)^\top$ is the transpose operation.
- If F is a function from $\mathbb{C}^N$ to $\mathbb{C}$, we define $\widetilde{F}$ the function of real variables associated with F, i.e. $(\forall \boldsymbol{x} \in \mathbb{C}^N)\ \widetilde{F}(\widetilde{\boldsymbol{x}}) = F(\boldsymbol{x})$.

Preliminaries

Notation:

- For every vector $\boldsymbol{x} \in \mathbb{C}^N$,
  - $\boldsymbol{x}_R \in \mathbb{R}^N$ (resp. $\boldsymbol{x}_I \in \mathbb{R}^N$) denotes the vector of real (resp. imaginary) parts of the components of $\boldsymbol{x}$.
  - $\widetilde{\boldsymbol{x}} \in \mathbb{R}^{2N}$ denotes the "concatenated" vector $\widetilde{\boldsymbol{x}} = [\boldsymbol{x}_R^\top \;\; \boldsymbol{x}_I^\top]^\top$ where $(\cdot)^\top$ is the transpose operation.
- If F is a function from $\mathbb{C}^N$ to $\mathbb{C}$, we define $\widetilde{F}$ the function of real variables associated with F, i.e. $(\forall \boldsymbol{x} \in \mathbb{C}^N)\ \widetilde{F}(\widetilde{\boldsymbol{x}}) = F(\boldsymbol{x})$.

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

$$(\forall \boldsymbol{x} \in \mathbb{C}^N) \qquad \nabla F(\boldsymbol{x}) = \frac{1}{2}\left(\frac{\partial \widetilde{F}(\widetilde{\boldsymbol{x}})}{\partial \boldsymbol{x}_R} + \imath \frac{\partial \widetilde{F}(\widetilde{\boldsymbol{x}})}{\partial \boldsymbol{x}_I}\right).$$

Assumptions

$$F(\boldsymbol{x}) = \Phi(\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y}) + \sum_{s=1}^{S} \psi_s(|\boldsymbol{v}_s^{\mathrm{H}}\boldsymbol{x} - c_s|) + \frac{\varepsilon}{2}\|\boldsymbol{x}\|^2$$

Assumption 1:

(i) $\widetilde{\Phi}$ is differentiable.

(ii) For every $s \in \{1, \ldots, S\}$, ψ_s is differentiable and $\lim_{\substack{t \to 0 \\ t \neq 0}} \dot{\psi}_s(t)/t \in \mathbb{R}$.

Assumption 2: One of the following conditions holds:

- Φ and $(\psi_s)_{1 \leqslant s \leqslant S}$ are lower bounded functions and $\varepsilon > 0$.
- (i) Φ is coercive (i.e. $\lim_{\|\boldsymbol{z}\| \to +\infty} \Phi(\boldsymbol{z}) = +\infty$).
  (ii) $(\psi_s)_{1 \leqslant s \leqslant S}$ are lower bounded functions.
  (iii) $\boldsymbol{H}$ is injective.
- (i) Φ is coercive.
  (ii) For every $s \in \{1, \ldots, S\}$, ψ_s is coercive.
  (iii) $\operatorname{Ker} \boldsymbol{H} \cap (\operatorname{span}\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_S\})^\perp = \{\boldsymbol{0}\}$

Properties

Complex-valued derivative of F

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

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

with

- $\boldsymbol{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_S] \in \mathbb{C}^{N \times S}$,
- $\boldsymbol{b}(\boldsymbol{x}) = (\omega_s(|\boldsymbol{v}_s^{\mathrm{H}} \boldsymbol{x} - c_s|))_{1 \leqslant s \leqslant S}$,
- $(\forall s \in \{1, \cdots, S\})(\forall a \in \mathbb{R}) \ \ \omega_s(a) = \begin{cases} \dfrac{\dot{\psi}_s(a)}{a} & \text{if } a \neq 0 \\ \lim_{\substack{t \to 0 \\ t \neq 0}} \dot{\psi}_s(t)/t & \text{otherwise.} \end{cases}$

Properties

Complex-valued derivative of F

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

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

with

- $\boldsymbol{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_S] \in \mathbb{C}^{N \times S}$,
- $\boldsymbol{b}(\boldsymbol{x}) = (\omega_s(|\boldsymbol{v}_s^{\mathrm{H}} \boldsymbol{x} - c_s|))_{1 \leqslant s \leqslant S}$,
- $(\forall s \in \{1, \cdots, S\})(\forall a \in \mathbb{R}) \;\; \omega_s(a) = \begin{cases} \dfrac{\dot{\psi}_s(a)}{a} & \text{if } a \neq 0 \\ \lim_{\substack{t \to 0 \\ t \neq 0}} \dot{\psi}_s(t)/t & \text{otherwise.} \end{cases}$

Existence of minimizers

Under **Assumptions 1 and 2**, Problem (1) has a solution.

Majorize-Minimize principle [Hunter04]

Objective: Find $\hat{\boldsymbol{x}} \in \operatorname{Arg\,min} F$

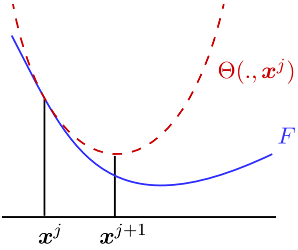
For all $\boldsymbol{x}'$, let $\Theta(.,\boldsymbol{x}')$ a *tangent majorant* of F at $\boldsymbol{x}'$ i.e.,

$$\begin{aligned} \Theta(\boldsymbol{x},\boldsymbol{x}') &\geqslant F(\boldsymbol{x}) \quad (\forall \boldsymbol{x}), \\ \Theta(\boldsymbol{x}',\boldsymbol{x}') &= F(\boldsymbol{x}') \end{aligned}$$

MM algorithm:

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

$\boldsymbol{x}^{j+1} \in \operatorname{Arg\,min}_{\boldsymbol{x}} \Theta(\boldsymbol{x},\boldsymbol{x}^j)$



Quadratic majorization

Assumption 3:

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

$$(\forall \boldsymbol{z} \in \mathbb{C}^Q)(\forall \boldsymbol{z}' \in \mathbb{C}^Q) \qquad \|\nabla\Phi(\boldsymbol{z}) - \nabla\Phi(\boldsymbol{z}')\| \leqslant \beta\|\boldsymbol{z} - \boldsymbol{z}'\|.$$

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

(iii) There exists $\overline{\omega} \in [0, +\infty)$ such that $(\forall s \in \{1, \ldots, S\})$ $(\forall t \in (0, +\infty))$ $0 \leqslant \omega_s(t) \leqslant \overline{\omega}$.

Proposition

If, for every $\boldsymbol{x}' \in \mathbb{C}^N$, $\boldsymbol{A}(\boldsymbol{x}') = \mu \boldsymbol{H}^{\mathrm{H}}\boldsymbol{H} + \boldsymbol{V}\ \mathrm{Diag}\left(\boldsymbol{b}(\boldsymbol{x}')\right) \boldsymbol{V}^{\mathrm{H}} + \varepsilon \boldsymbol{I}_N$ with $\mu \in [2\beta, +\infty)$, then

$$\Theta(\boldsymbol{x}, \boldsymbol{x}') = F(\boldsymbol{x}') + 2\,\mathrm{Re}\left\{\nabla F(\boldsymbol{x}')^{\mathrm{H}}(\boldsymbol{x} - \boldsymbol{x}')\right\} + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}')^{\mathrm{H}}\boldsymbol{A}(\boldsymbol{x}')(\boldsymbol{x} - \boldsymbol{x}')$$

is a quadratic tangent majorant of F at $\boldsymbol{x}'$.

Proposed algorithm

MM Subspace algorithm:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{u}_k \qquad (\forall k \geqslant 0)$$

Proposed algorithm

MM Subspace algorithm:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{u}_k \qquad (\forall k \geqslant 0)$$

- $\boldsymbol{D}_k \in \mathbb{C}^{N \times M}$: matrix of M directions

Example: Memory gradient $\boldsymbol{D}_k = [-\nabla F(\boldsymbol{x}_k), \boldsymbol{x}_k - \boldsymbol{x}_{k-1}]$

Proposed algorithm

MM Subspace algorithm:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{u}_k \qquad (\forall k \geqslant 0)$$

- $\boldsymbol{D}_k \in \mathbb{C}^{N \times M}$: matrix of M directions
- $\boldsymbol{u}_k \in \mathbb{C}^{M}$: multivariate stepsize resulting from MM minimization of $f_k(\boldsymbol{u})\colon \boldsymbol{u} \mapsto F(\boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{u})$

Proposed algorithm

MM Subspace algorithm:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{u}_k \qquad (\forall k \geqslant 0)$$

- $\boldsymbol{D}_k \in \mathbb{C}^{N \times M}$: matrix of M directions
- $\boldsymbol{u}_k \in \mathbb{C}^M$: multivariate stepsize resulting from MM minimization of $f_k(\boldsymbol{u})$: $\boldsymbol{u} \mapsto F(\boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{u})$

MM minimization in the subspace:

$$\begin{cases} \boldsymbol{u}_k^0 & = & \boldsymbol{0}, \\ \boldsymbol{u}_k^j & \in & \operatorname{Arg\,min}_{\boldsymbol{u}} \vartheta_k(\boldsymbol{u}, \boldsymbol{u}_k^{j-1}) \ (\forall j \in \{1, \dots J\}) \end{cases}$$

with, for all $\boldsymbol{u} \in \mathbb{C}^M$, $\vartheta_k(\boldsymbol{u}, \boldsymbol{u}_k^j) = \Theta(\boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{u}, \boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{u}_k^j)$, quadratic tangent majorant of f_k at $\boldsymbol{u}_k^j$ with Hessian:

$$\boldsymbol{B}_k^j = \boldsymbol{D}_k^{\mathrm{H}} \boldsymbol{A}(\boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{u}_k^j) \boldsymbol{D}_k$$

Proposed algorithm

Complex-valued 3MG algorithm

$\boldsymbol{x}_0 \in \mathbb{C}^N$, $\boldsymbol{x}_{-1} = \boldsymbol{0}$

For all $k = 0, \ldots$

$$\boldsymbol{D}_k = [-\nabla F(\boldsymbol{x}_k), \boldsymbol{x}_k - \boldsymbol{x}_{k-1}]$$

$$\boldsymbol{u}_k^0 = \boldsymbol{0},$$

For all $j = 1, \ldots, J$

$$\boldsymbol{B}_k^{j-1} = \boldsymbol{D}_k^{\mathrm{H}} \boldsymbol{A}(\boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{u}_k^{j-1}) \boldsymbol{D}_k,$$

$$\boldsymbol{u}_k^j = \boldsymbol{u}_k^{j-1} - 2(\boldsymbol{B}_k^{j-1})^{\dagger} \boldsymbol{D}_k^{\mathrm{H}} \nabla F(\boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{u}_k^{j-1}),$$

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{u}_k^J.$$

Proposed algorithm

Complex-valued 3MG algorithm

$$
\begin{array}{l}
\boldsymbol{x}_0 \in \mathbb{C}^N,\ \boldsymbol{x}_{-1} = \boldsymbol{0} \\
\text{For all } k = 0, \ldots \\
\left\lfloor
\begin{array}{l}
\boldsymbol{D}_k = [-\nabla F(\boldsymbol{x}_k), \boldsymbol{x}_k - \boldsymbol{x}_{k-1}] \\
\boldsymbol{u}_k^0 = \boldsymbol{0}, \\
\text{For all } j = 1, \ldots, J \\
\left\lfloor
\begin{array}{l}
\boldsymbol{B}_k^{j-1} = \boldsymbol{D}_k^{\mathrm{H}} \boldsymbol{A}(\boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{u}_k^{j-1}) \boldsymbol{D}_k, \\
\boldsymbol{u}_k^j = \boldsymbol{u}_k^{j-1} - 2(\boldsymbol{B}_k^{j-1})^{\dagger} \boldsymbol{D}_k^{\mathrm{H}} \nabla F(\boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{u}_k^{j-1}),
\end{array}
\right. \\
\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{u}_k^J .
\end{array}
\right.
\end{array}
$$

⇝ Equivalent to 3MG algorithm for minimizing real-valued function $\widetilde{F}$, taking

$$\widetilde{\boldsymbol{D}}_k = \begin{bmatrix} \boldsymbol{D}_{k,R} & -\boldsymbol{D}_{k,I} \\ \boldsymbol{D}_{k,I} & \boldsymbol{D}_{k,R} \end{bmatrix} \in \mathbb{R}^{2N \times 2M}$$

Convergence result

Assumptions

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

$$\|\nabla F(\boldsymbol{x})\| \geqslant \kappa |F(\boldsymbol{x}) - F(\hat{\boldsymbol{x}})|^{\theta},$$

for every $\boldsymbol{x} \in \mathbb{B}$ such that $|F(\boldsymbol{x}) - F(\hat{\boldsymbol{x}})| \leqslant \zeta$.

Convergence result

Proposition

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

$$\sum_{k=0}^{+\infty} \|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\| < +\infty.$$

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

$$F(\boldsymbol{x}_0) \leqslant \eta + \inf F,$$

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

Outline

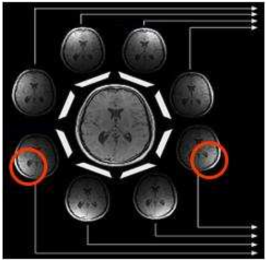
Parallel Magnetic Resonance Imaging

Objective:

- Reduce the acquisition time
- Maintain good image quality

Principle:

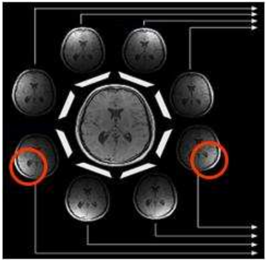
- k-space subsampling
- Multiple receiver coils



Parallel Magnetic Resonance Imaging

Objective:

- Reduce the acquisition time
- Maintain good image quality



Principle:

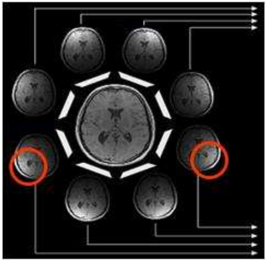
- k-space subsampling
- Multiple receiver coils

Acquisition model: $(\forall \ell \in \{1, \dots, L\}) \qquad \boldsymbol{d}_\ell = \boldsymbol{\Sigma F S}_\ell \overline{\boldsymbol{\rho}} + \boldsymbol{w}_\ell$

Parallel Magnetic Resonance Imaging

Objective:

- Reduce the acquisition time
- Maintain good image quality



Principle:

- k-space subsampling
- Multiple receiver coils

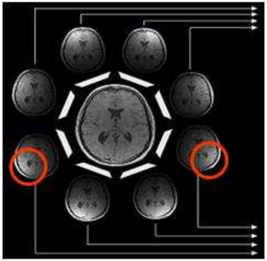
Acquisition model: $(\forall \ell \in \{1,\dots,L\}) \qquad \boldsymbol{d}_\ell = \boldsymbol{\Sigma F S}_\ell \overline{\boldsymbol{\rho}} + \boldsymbol{w}_\ell$

- $\forall \ell \in \{1,\dots,L\}$, $\boldsymbol{S}_\ell \in \mathbb{C}^{K\times K}$: diagonal sensitivity matrix,

Parallel Magnetic Resonance Imaging

Objective:

- Reduce the acquisition time
- Maintain good image quality



Principle:

- k-space subsampling
- Multiple receiver coils

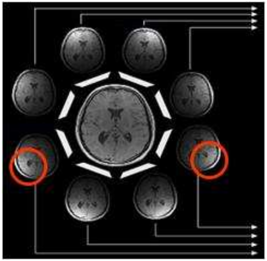
Acquisition model: $(\forall \ell \in \{1, \dots, L\})$ $\quad \boldsymbol{d}_\ell = \boldsymbol{\Sigma} \boldsymbol{F} \boldsymbol{S}_\ell \overline{\boldsymbol{\rho}} + \boldsymbol{w}_\ell$

- $\forall \ell \in \{1, \dots, L\}$, $\boldsymbol{S}_\ell \in \mathbb{C}^{K \times K}$: diagonal sensitivity matrix,
- $\boldsymbol{F} \in \mathbb{C}^{K \times K}$: 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: $(\forall \ell \in \{1, \ldots, L\}) \qquad \boldsymbol{d}_\ell = \boldsymbol{\Sigma} \boldsymbol{F} \boldsymbol{S}_\ell \overline{\boldsymbol{\rho}} + \boldsymbol{w}_\ell$

- $\forall \ell \in \{1, \ldots, L\}$, $\boldsymbol{S}_\ell \in \mathbb{C}^{K \times K}$: diagonal sensitivity matrix,
- $\boldsymbol{F} \in \mathbb{C}^{K \times K}$: 2D discrete Fourier transform,
- $\boldsymbol{\Sigma} \in \{0, 1\}^{\lfloor \frac{K}{R} \rfloor \times K}$: subsampling matrix

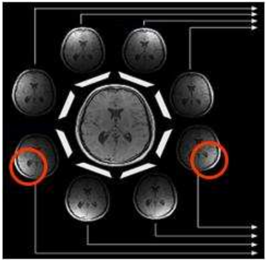
Parallel Magnetic Resonance Imaging

Objective:

- Reduce the acquisition time
- Maintain good image quality

Principle:

- k-space subsampling
- Multiple receiver coils



Acquisition model: $(\forall \ell \in \{1, \ldots, L\}) \qquad \boldsymbol{d}_\ell = \boldsymbol{\Sigma F S}_\ell \overline{\boldsymbol{\rho}} + \boldsymbol{w}_\ell$

- $\forall \ell \in \{1, \ldots, L\}$, $\boldsymbol{S}_\ell \in \mathbb{C}^{K \times K}$: diagonal sensitivity matrix,
- $\boldsymbol{F} \in \mathbb{C}^{K \times K}$: 2D discrete Fourier transform,
- $\boldsymbol{\Sigma} \in \{0,1\}^{\lfloor \frac{K}{R} \rfloor \times K}$: subsampling matrix
- $\forall \ell \in \{1, \ldots, L\}$, $\boldsymbol{w}_\ell \in \mathbb{C}^{\lfloor \frac{K}{R} \rfloor}$: realization of circular complex Gaussian noise with zero-mean and covariance matrix $\boldsymbol{\Lambda}_\ell$.

Variational formulation

$$\underset{\boldsymbol{\rho}\in\mathbb{E}}{\text{minimize}} \quad \left(\sum_{\ell=1}^{L} \|\boldsymbol{\Sigma}\boldsymbol{F}\boldsymbol{S}_{\ell}\boldsymbol{\rho} - \boldsymbol{d}_{\ell}\|^2_{\boldsymbol{\Lambda}_{\ell}^{-1}} + \sum_{s=1}^{S} \psi_s(|\boldsymbol{f}_s^{\mathrm{H}}\boldsymbol{\rho}|) \right)$$

Variational formulation

$$\underset{\boldsymbol{\rho}\in\mathbb{E}}{\text{minimize}} \quad \left(\sum_{\ell=1}^{L} \|\boldsymbol{\Sigma F S}_\ell \boldsymbol{\rho} - \boldsymbol{d}_\ell\|^2_{\boldsymbol{\Lambda}_\ell^{-1}} + \sum_{s=1}^{S} \psi_s(|\boldsymbol{f}_s^{\mathrm{H}} \boldsymbol{\rho}|) \right)$$

$$\rightsquigarrow \quad \underset{\boldsymbol{x}\in\mathbb{C}^N}{\text{minimize}} \quad \left(\sum_{\ell=1}^{L} \|\boldsymbol{\Sigma F S}_\ell \boldsymbol{E x} - \boldsymbol{d}_\ell\|^2_{\boldsymbol{\Lambda}_\ell^{-1}} + \sum_{s=1}^{S} \psi_s(|\boldsymbol{f}_s^{\mathrm{H}} \boldsymbol{E x}|) + \frac{\varepsilon}{2}\|\boldsymbol{x}\|^2 \right)$$

where $\boldsymbol{E} \in \mathbb{C}^{K\times N}$ allows us to set the background pixels to zero.

Variational formulation

$$\underset{\boldsymbol{\rho}\in\mathbb{E}}{\text{minimize}} \quad \left(\sum_{\ell=1}^{L} \|\boldsymbol{\Sigma F S}_\ell \boldsymbol{\rho} - \boldsymbol{d}_\ell\|^2_{\boldsymbol{\Lambda}_\ell^{-1}} + \sum_{s=1}^{S} \psi_s(|\boldsymbol{f}_s^{\mathrm{H}} \boldsymbol{\rho}|) \right)$$

$$\rightsquigarrow \quad \underset{\boldsymbol{x}\in\mathbb{C}^N}{\text{minimize}} \quad \left(\sum_{\ell=1}^{L} \|\boldsymbol{\Sigma F S}_\ell \boldsymbol{E x} - \boldsymbol{d}_\ell\|^2_{\boldsymbol{\Lambda}_\ell^{-1}} + \sum_{s=1}^{S} \psi_s(|\boldsymbol{f}_s^{\mathrm{H}} \boldsymbol{E x}|) + \frac{\varepsilon}{2}\|\boldsymbol{x}\|^2 \right)$$

$$\Leftrightarrow \quad \underset{\boldsymbol{x}\in\mathbb{C}^N}{\text{minimize}} \quad \left(\Phi(\boldsymbol{H x} - \boldsymbol{y}) + \sum_{s=1}^{S} \psi_s(|\boldsymbol{v}_s^{\mathrm{H}} \boldsymbol{x} - c_s|) + \frac{\varepsilon}{2}\|\boldsymbol{x}\|^2 \right)$$

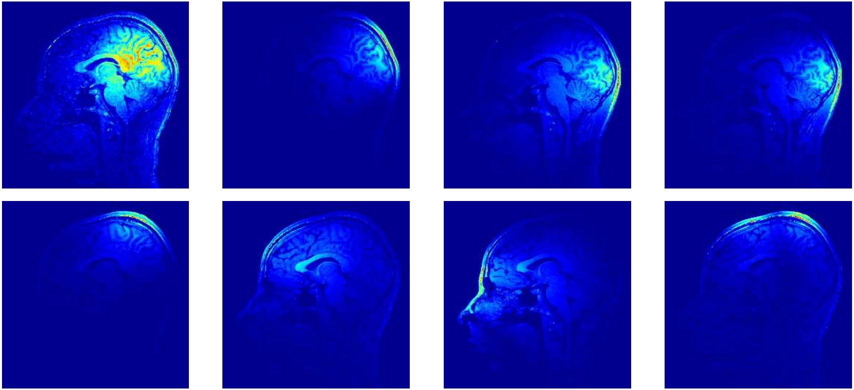
with Φ squared Hermitian norm of $\mathbb{C}^Q$ with $Q = L\lfloor K/R \rfloor$ and

- $\boldsymbol{H} = [\boldsymbol{H}_1^\top, \ldots, \boldsymbol{H}_L^\top]^\top, \quad (\forall \ell \in \{1, \ldots, L\}) \;\; \boldsymbol{H}_\ell = \boldsymbol{\Lambda}_\ell^{-1/2} \boldsymbol{\Sigma F S}_\ell \boldsymbol{E}$
- $\boldsymbol{y} = [\boldsymbol{y}_1^\top, \ldots, \boldsymbol{y}_L^\top]^\top, \quad (\forall \ell \in \{1, \ldots, L\}) \;\; \boldsymbol{y}_\ell = \boldsymbol{\Lambda}_\ell^{-1/2} \boldsymbol{d}_\ell$
- $(\boldsymbol{v}_s)_{1\leqslant s\leqslant S} = (\boldsymbol{E}^{\mathrm{H}} \boldsymbol{f}_s)_{1\leqslant s\leqslant S}$, $(c_s)_{1\leqslant s\leqslant S} = \boldsymbol{0}$

Simulation settings

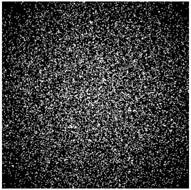
- 3 Tesla Siemens Trio magnet with $L = 32$ channel receiver coil
- Reconstruction of sagittal views of a 3D anatomical image with 256×256 pixels
- Reference image $\bar{\boldsymbol{\rho}}$ defined as the reconstruction result from a non-accelerated acquisition $(R = 1)$
- Different sampling patterns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

Effect of the sensitivity matrices

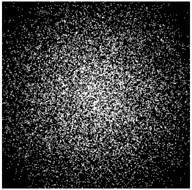


Moduli of the images corresponding to $(\boldsymbol{S}_\ell \bar{\boldsymbol{\rho}})_{1 \leqslant \ell \leqslant L}$ *for 8 channels out of 32.*

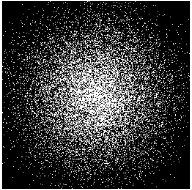
Different types of subsampling



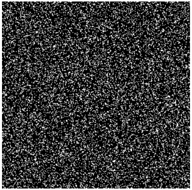
Poly1



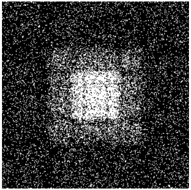
Poly2



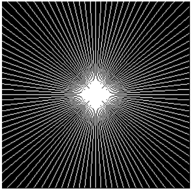
Poly3



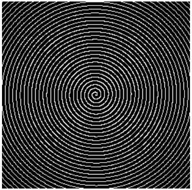
Uniform



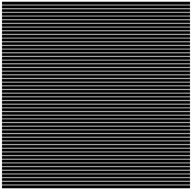
π



Radial



Spiral



Regular

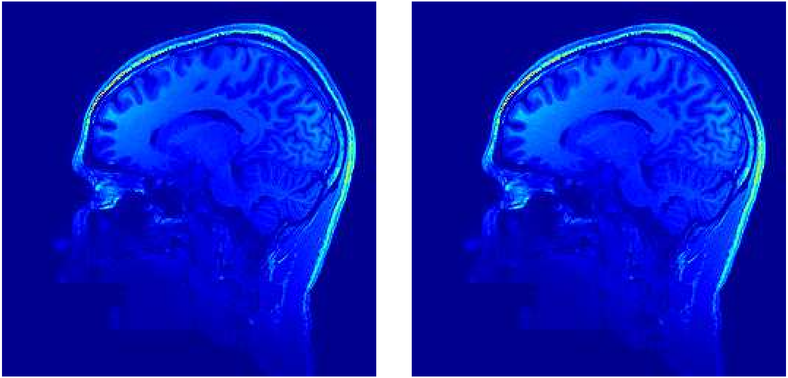
Simulation results

	SNR (dB)		
Sampling pattern	Slide No 70	Slice No 82	Slice No 121
Poly1	**21.15**	**19.96**	**20.89**
Poly2	20.32	19.34	20.07
Poly3	19.43	18.53	19.18
Poly4	18.47	17.50	18.35
Poly5	17.67	16.95	17.52
Uniform	21.02	19.71	20.68
π	20.46	19.31	20.08
Radial	20.27	19.20	20.01
Spiral	20.35	19.17	20.03
Regular	19.18	18.13	18.66

SNR values for various subsampling strategies using 3MG and ℓ_2 - ℓ_1 regularization

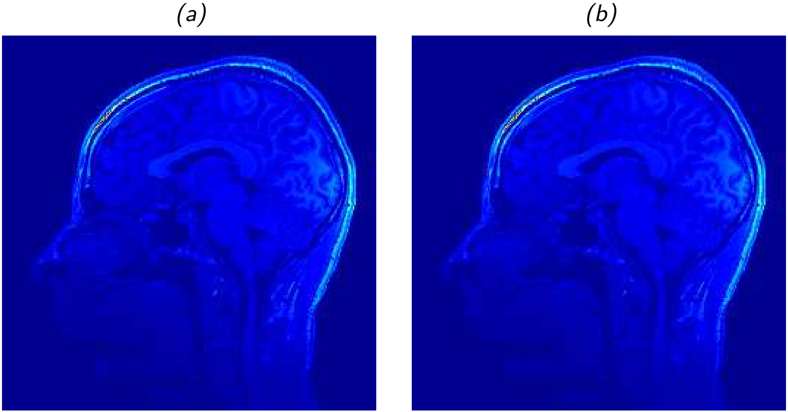
Simulation results

(a) (b)



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.*

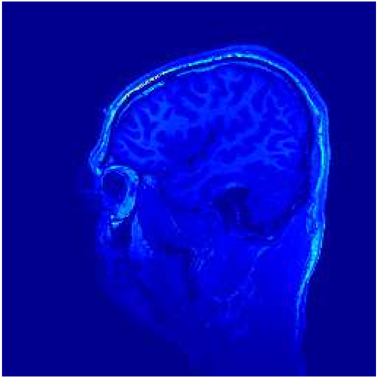
Simulation results

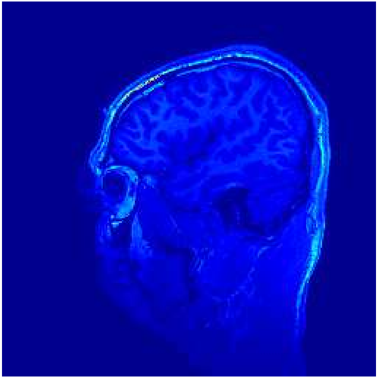


Slice No 82: *Moduli of the original image* $\bar{\rho}$ *(a) and the reconstructed one (b) with SNR* $=$ *19.95 dB using Poly1 sampling, 3MG algorithm and* $\ell_2 - \ell_1$ *regularization.*

Simulation results

(a) *(b)*





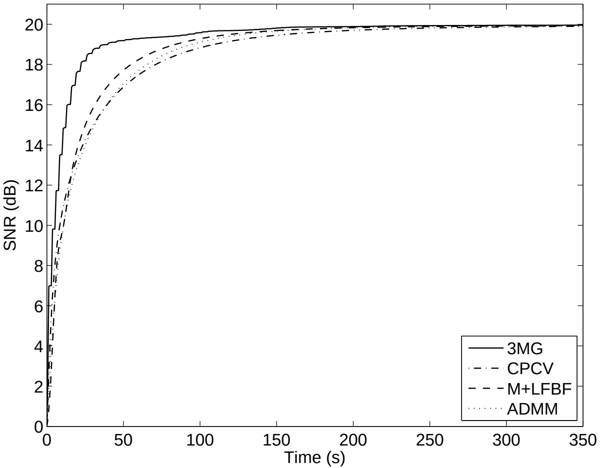
Slice No 121: *Moduli of the original image* $\bar{\rho}$ *(a) and the reconstructed one (b) with SNR* $=$ *20.89 dB using Poly1 sampling, 3MG algorithm and* $\ell_2 - \ell_1$ *regularization.*

Simulation results

Decomp.	Algorithm	Penalization	SNR (dB)		
			Slice No 70	Slice No 82	Slice No 121
Wav. basis	M+LFBF	ℓ_1	21.15	19.96	20.89
	CPCV	ℓ_1	21.15	19.96	20.89
	ADMM	ℓ_1	21.15	19.96	20.89
	3MG	ℓ_2 - ℓ_1	21.15	19.96	20.89
	3MG	ℓ_2 - ℓ_0 (H)	21.09	20.05	20.97
	3MG	ℓ_2 - ℓ_0 (W)	21.21	20.17	21.10
	3MG	ℓ_2 - ℓ_0 (G)	21.33	20.27	21.20
Redundant wav. frame	3MG	ℓ_2 - ℓ_1	21.67	20.46	21.39
	3MG	ℓ_2 - ℓ_0 (G)	**22.10**	**20.94**	**21.84**

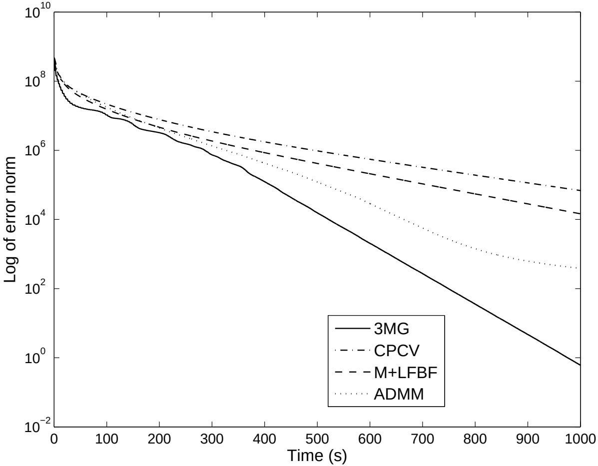
Reconstruction results for several optimization and regularization strategies using two different decompositions (Poly1 subsampling pattern)

Simulation results



SNR evolution as a function of computation time

Simulation results

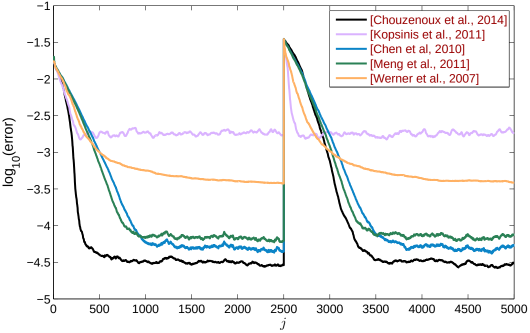


Error $\|x_k - \hat{x}\|$ *as a function of computation time*

Outline

S3MG

- Stochastic version for solving online/adaptive problems



Estimation error along time, for various sparse adaptive filtering strategies

- 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

Conclusion

- Majorize-Minimize Memory Gradient algorithm for optimization of smooth nonconvex complex-valued functions.
- Application to Parallel Magnetic Resonance Imaging
  - $\rightsquigarrow$ Faster than standard proximal techniques
- Future work
  - $\rightsquigarrow$ Application to other inverse problems (CEA-LETI: microscopy imaging)
  - $\rightsquigarrow$ 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
IEEE Transactions on Image Processing, vol. 20, no 6, pp. 1517-1528, June 2011

E. Chouzenoux, A. Jezierska, J.-C. Pesquet and H. Talbot
A Majorize-Minimize subspace approach for ℓ_2-ℓ_0 image regularization
SIAM Journal on Imaging Science, vol. 6, no 1, pp. 563–591, Jan. 2013.

E. Chouzenoux, J.-C. Pesquet and A. Florescu.
A Stochastic 3MG Algorithm with Application to 2D Filter Identification
European Signal Processing Conference (EUSIPCO 2014), pp. 1587-1591, Lisboa, Portugal, 1-5 Sept. 2014.

A. Florescu, E. Chouzenoux, J.-C. Pesquet, P. Ciuciu and S. Ciochina
A Majorize-Minimize memory gradient method for complex-valued inverse problems
Signal Processing, Special issue on Image Restoration and Enhancement: Recent Advances and Applications, vol. 103, pp. 285-295, Oct. 2014.