# A memory gradient algorithm for non-convex regularization with applications to image restoration

Emilie Chouzenoux, Jean-Christophe Pesquet and Audrey Repetti

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

*SIAM Conference on Imaging Science*
21 may 2012

# Outline

# Context

## Image restoration

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

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

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

$\boldsymbol{y}$

$\overline{\boldsymbol{x}}$

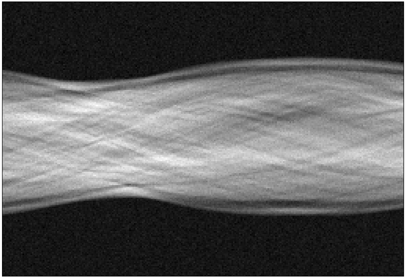
# Context

## Image restoration

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

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

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



$\boldsymbol{y}$

$\overline{\boldsymbol{x}}$

# Context

## Penalized optimization problem

Find

$$\min_{\boldsymbol{x}\in\mathbb{R}^N} \big(F(\boldsymbol{x}) = \Phi(\boldsymbol{H}\boldsymbol{x}-\boldsymbol{y}) + \Psi(\boldsymbol{x})\big),$$

$\Phi$ ⇝ Data fidelity term, related to noise

$\Psi$ ⇝ Regularization term, related to some *a priori* assumptions

**Considered model:**

$$G_\delta(\boldsymbol{x}) = \underbrace{\Phi(\boldsymbol{H}\boldsymbol{x}-\boldsymbol{y}) + \sum_{s=1}^{S} \psi_{s,\delta}(\|\boldsymbol{V}_s\boldsymbol{x}-\boldsymbol{c}_s\|)}_{F_\delta(\boldsymbol{x})} + \Psi_0(\boldsymbol{x}),$$

$\psi_{s,\delta}$ ⇝ (non necessarily convex) smooth function depending on $\delta > 0$

$\Psi_0$ ⇝ (non necessarily smooth) convex function

# Examples of regularization functions

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

*Example:* $(\forall t \in \mathbb{R})$, $\psi_\delta(t) = \lambda(\sqrt{1 + \frac{t^2}{\delta^2}} - 1)$, $\lambda > 0$

**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

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

*Example:* $(\forall t \in \mathbb{R})$, $\psi_\delta(t) = \lambda(\sqrt{1 + \frac{t^2}{\delta^2}} - 1)$, $\lambda > 0$

**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]*.

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

*Example:* $(\forall t \in \mathbb{R})$, $\psi_\delta(t) = \lambda \min(t^2/(2\delta^2), 1)$, $\lambda > 0$

**Limit case:** When $\delta \to 0$, $\psi_\delta(t) = \lambda$ if $t = 0$, $0$ otherwise ($\ell_0$ penalty).

**Non-convex functions** $\Rightarrow$ Which algorithms ?

# Outline

# Existence of minimizers

$$G_\delta(\boldsymbol{x}) = \Phi(\boldsymbol{H}\boldsymbol{x}-\boldsymbol{y})+\sum_{s=1}^{S}\psi_{s,\delta}(\|\boldsymbol{V}_s\boldsymbol{x}-\boldsymbol{c}_s\|)+\Psi_0(\boldsymbol{x})$$

**Difficulty:** $G_\delta$ is a **non convex** 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 $s\in\{1,\dots,S\}$, $\psi_{s,\delta}$ is continuous and takes nonnegative values.

(iii) $\Psi_0$ is a proper lower semi-continuous, lower bounded function.

(iv) $\operatorname{Ker}\boldsymbol{H}=\{\boldsymbol{0}\}$, or $\operatorname{dom}\Psi_0$ bounded.

Then, for every $\delta>0$, $G_\delta$ has a minimizer.

# Existence of minimizers

$$G_\delta(\boldsymbol{x}) = \Phi(\boldsymbol{H}\boldsymbol{x}-\boldsymbol{y}) + \sum_{s=1}^{S} \psi_{s,\delta}(\|\boldsymbol{V}_s\boldsymbol{x}-\boldsymbol{c}_s\|) + \Psi_0(\boldsymbol{x}) + \|\boldsymbol{V}_0\boldsymbol{x}\|^2$$

**Difficulty:** $G_\delta$ is a **non convex** 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) $\Psi_0$ is a proper lower semi-continuous, lower bounded function.

(iv) $\operatorname{Ker} \boldsymbol{H} \cap \operatorname{Ker} \boldsymbol{V}_0 = \{\boldsymbol{0}\}$, or $\operatorname{dom}\Psi_0$ bounded.

Then, for every $\delta > 0$, $G_\delta$ has a minimizer.

# Epi-convergence property

**Assumptions:**

1. $(\forall s \in \{1, \ldots, S\})\ (\forall (\delta_1, \delta_2) \in (0, +\infty)^2)$
$$\delta_1 \leq \delta_2 \Rightarrow (\forall t \in \mathbb{R})\, \psi_{s,\delta_1}(t) \geq \psi_{s,\delta_2}(t)$$
2. $(\forall s \in \{1, \ldots, S\})(\forall t \in \mathbb{R})$, $\lim_{\substack{\delta \to 0 \\ \delta > 0}} \psi_{s,\delta}(t) = \lambda \psi_{s,0}(t)$
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,

$$\inf G_{\delta_n} \to \inf G_0 \qquad \text{as} \quad n \to +\infty$$

In addition, if for every $n \in \mathbb{N}$, $\widehat{\boldsymbol{x}}_n$ is a minimizer of $G_{\delta_n}$, then the sequence $(\widehat{\boldsymbol{x}}_n)_{n \in \mathbb{N}}$ is bounded and all its cluster points are minimizers of $G_0$.

# Outline

# Iterative minimization of $F_\delta(\boldsymbol{x})$

*From now, we assume $F_\delta$ differentiable*.

Descent algorithm

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k, \quad (\forall k \geq 0)$$

- $\boldsymbol{d}_k$ : search direction satisfying $\nabla F_\delta(\boldsymbol{x}_k)^\top \boldsymbol{d}_k < 0$
  Ex: Gradient, conjugate gradient, Newton, truncated Newton, ...
- stepsize $\alpha_k$: approximate minimizer of $f_{k,\delta}(\alpha)\colon \alpha \mapsto F_\delta(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k)$

# Iterative minimization of $F_\delta(\boldsymbol{x})$

*From now, we assume $F_\delta$ differentiable*.

Descent algorithm

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k, \quad (\forall k \geq 0)$$

- $\boldsymbol{d}_k$ : search direction satisfying $\nabla F_\delta(\boldsymbol{x}_k)^\top \boldsymbol{d}_k < 0$
  Ex: Gradient, conjugate gradient, Newton, truncated Newton, ...
- stepsize $\alpha_k$: approximate minimizer of $f_{k,\delta}(\alpha)\colon \alpha \mapsto F_\delta(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k)$

Generalization : subspace algorithm *[Zibulevsky10]*

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \sum_{m=1}^{M} u_{m,k} \boldsymbol{d}_k^m, \quad (\forall k \geq 0)$$

- $[\boldsymbol{d}_k^1, ..., \boldsymbol{d}_k^M] = \boldsymbol{D}_k$: Set of search directions
  Ex: Super-memory gradient $\boldsymbol{D}_k = [-\nabla F_\delta(\boldsymbol{x}_k), \boldsymbol{d}_{k-1}, .., \boldsymbol{d}_{k-l}]$
- stepsize $\boldsymbol{u}_k$: approximate minimizer of $f_{k,\delta}(\boldsymbol{u})\colon \boldsymbol{u} \mapsto F_\delta(\boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{u})$

# Majorize-Minimize principle *[Hunter04]*

**Objective:** Find $\hat{\boldsymbol{x}} \in \operatorname{Arg\,min}_{\boldsymbol{x}} F_\delta(\boldsymbol{x})$

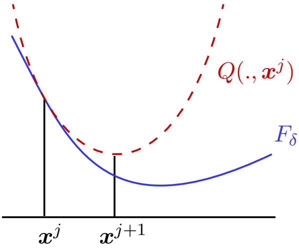
For all $\boldsymbol{x}'$, let $Q(., \boldsymbol{x}')$ a *tangent majorant* of $F_\delta$ at $\boldsymbol{x}'$ i.e.,

$$Q(\boldsymbol{x}, \boldsymbol{x}') \geqslant F_\delta(\boldsymbol{x}), \quad \forall \boldsymbol{x},$$
$$Q(\boldsymbol{x}', \boldsymbol{x}') = F_\delta(\boldsymbol{x}')$$

**MM algorithm:**

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

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



# Quadratic tangent majorant function

**Assumptions:**

(i) $\Phi$ is differentiable with an $L_\Phi$-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 $\overline{\omega_s} \in [0, +\infty)$ such that $(\forall t \in (0, +\infty))\ 0 \leq \dot{\psi}_{s,\delta}(t) \leq \overline{\omega_s} t$ where $\dot{\psi}_{s,\delta}$ is the derivative of $\psi_{s,\delta}$. In addition, $\lim_{\substack{t \to 0 \\ t \neq 0}} \omega_{s,\delta}(t) \in \mathbb{R}$ with $\omega_{s,\delta}(t) \triangleq \dot{\psi}_{s,\delta}(t)/t$.

**Lemma 1** *[Allain06]*

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

$$\boldsymbol{A}(\boldsymbol{x}) = \mu \boldsymbol{H}^\top \boldsymbol{H} + \boldsymbol{V}^\top \mathrm{Diag}\left\{\boldsymbol{b}(\boldsymbol{x})\right\} \boldsymbol{V} + 2\boldsymbol{V}_0^\top \boldsymbol{V}_0$$

$\mu \in [L_\Phi, +\infty)$, $\boldsymbol{b}(\boldsymbol{x}) \in \mathbb{R}^{SP}$ with $b_{P_1+\cdots+P_{s-1}+p}(\boldsymbol{x}) = \omega_{s,\delta}(\|\boldsymbol{V}_s \boldsymbol{x} - \boldsymbol{c}_s\|)$.
Then, $Q(\boldsymbol{x}, \boldsymbol{x}') = F_\delta(\boldsymbol{x}') + \nabla F_\delta(\boldsymbol{x}')^\top (\boldsymbol{x} - \boldsymbol{x}') + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}')^\top \boldsymbol{A}(\boldsymbol{x}')(\boldsymbol{x} - \boldsymbol{x}')$ is a convex quadratic tangent majorant of $F_\delta$ at $\boldsymbol{x}'$.

# Majorize-Minimize multivariate stepsize *[Chouzenoux11]*

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

- $\boldsymbol{D}_k$: set of directions
- $\boldsymbol{u}_k$ resulting from MM minimization of $f_{k,\delta}(\boldsymbol{u})$: $\boldsymbol{u} \mapsto F_\delta(\boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{u})$

$q_k(\boldsymbol{u}, \boldsymbol{u}_k^j)$ : Quadratic tangent majorant of $f_{k,\delta}$ at $\boldsymbol{u}_k^j$

with Hessian: $\boldsymbol{B}_{k,\boldsymbol{u}_k^j} = \boldsymbol{D}_k^\top \boldsymbol{A}(\boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{u}_k^j) \boldsymbol{D}_k$

MM minimization in the subspace:

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

# Proposed algorithm

## Majorize-Minimize Memory Gradient algorithm

For all $k \geq 0$

1. Compute $\boldsymbol{D}_k = \begin{cases} -\nabla F_\delta(\boldsymbol{x}_0) & \text{if} \quad k = 0 \\ [-\nabla F_\delta(\boldsymbol{x}_k) \quad \boldsymbol{x}_k - \boldsymbol{x}_{k-1}] & \text{if} \quad k > 0 \end{cases}$
2. $\boldsymbol{u}_k^0 = \boldsymbol{0}$
3. $\forall j \in \{0, \dots J-1\}$,
   - $\boldsymbol{B}_{k,\boldsymbol{u}_k^j} = \boldsymbol{D}_k^\top \boldsymbol{A}(\boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{u}_k^j)\boldsymbol{D}_k$
   - $\boldsymbol{u}_k^{j+1} = \boldsymbol{u}_k^j - \boldsymbol{B}_{k,\boldsymbol{u}_k^j}^{-1} \nabla f_{k,\delta}(\boldsymbol{u}_k^j)$
4. Update $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{u}_k^J$

$\rightsquigarrow$ Converges to a critical point of $F_\delta$ *[Chouzenoux11]*

# Outline

# Minimization of $F_\delta + \Psi_0$

$$G_\delta(\boldsymbol{x}) = \underbrace{\Phi(\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y}) + \sum_{s=1}^{S} \psi_{s,\delta}(\|\boldsymbol{V}_s\boldsymbol{x} - \boldsymbol{c}_s\|) + \|\boldsymbol{V}_0\boldsymbol{x}\|^2}_{F_\delta(\boldsymbol{x})} + \Psi_0(\boldsymbol{x})$$

where $\Psi_0$ is a proper semicontinuous **convex** function.

## Definition

For every $\boldsymbol{\xi} \in \mathbb{R}^N$, the minimization problem

$$\text{minimize}_{\boldsymbol{\pi} \in \mathbb{R}^N} \Psi_0(\boldsymbol{\pi}) + \frac{1}{2}\|\boldsymbol{\xi} - \boldsymbol{\pi}\|^2,$$

admits a unique solution, which is denoted by $\mathrm{prox}_{\Psi_0}(\boldsymbol{\xi})$. The so-defined function $\mathrm{prox}_{\Psi_0} : \mathbb{R}^N \to \mathbb{R}^N$ is the *proximity operator* of $\Psi_0$.

# Proposed algorithm

## Proximal Majorize-Minimize Memory Gradient algorithm

For all $k \geq 0$

1. $\bar{\boldsymbol{x}}_k = \mathrm{prox}_{\gamma_k \Psi_0}(\boldsymbol{x}_k - \gamma_k \nabla F_\delta(\boldsymbol{x}_k))$
2. Compute $\boldsymbol{D}_k = \begin{cases} \bar{\boldsymbol{x}}_0 - \boldsymbol{x}_0 & \text{if} \quad k = 0 \\ [\bar{\boldsymbol{x}}_k - \boldsymbol{x}_k \quad \boldsymbol{x}_k - \boldsymbol{x}_{k-1}] & \text{if} \quad k > 0 \end{cases}$
3. $\bar{\boldsymbol{u}}_k^0 = \boldsymbol{0}$
4. $\forall j \in \{0, \ldots J-1\}$,
   - $\bar{\boldsymbol{u}}_k^{j+1} \in \mathrm{Arg\,min}_{\boldsymbol{u} \in \mathbb{R}^2} \left( \bar{q}_k(\boldsymbol{u}, \bar{\boldsymbol{u}}_k^j) + \Psi_0(\bar{\boldsymbol{x}}_k + \boldsymbol{D}_k \boldsymbol{u}) \right)$

   where $\bar{q}_k(\boldsymbol{u}, \boldsymbol{u}') \triangleq Q(\bar{\boldsymbol{x}}_k + \boldsymbol{D}_k \boldsymbol{u}, \bar{\boldsymbol{x}}_k + \boldsymbol{D}_k \boldsymbol{u}')$
5. Update $\boldsymbol{x}_{k+1} = \bar{\boldsymbol{x}}_k + \boldsymbol{D}_k \bar{\boldsymbol{u}}_k^J$

# Proposed algorithm

## Inexact Proximal Majorize-Minimize Memory Gradient algorithm

For all $k \geq 0$

1. $\bar{\boldsymbol{x}}_k = \mathrm{prox}_{\gamma_k \Psi_0}(\boldsymbol{x}_k - \gamma_k \nabla F_\delta(\boldsymbol{x}_k))$
2. Compute $\boldsymbol{D}_k = \begin{cases} \bar{\boldsymbol{x}}_0 - \boldsymbol{x}_0 & \text{if} \quad k = 0 \\ [\bar{\boldsymbol{x}}_k - \boldsymbol{x}_k \quad \boldsymbol{x}_k - \boldsymbol{x}_{k-1}] & \text{if} \quad k > 0 \end{cases}$
3. $\bar{\boldsymbol{u}}_k^0 = \boldsymbol{0}$
4. $\forall j \in \{0, \dots J-1\}$,
   - $\boldsymbol{B}_{k,\bar{\boldsymbol{u}}_k^j} = \boldsymbol{D}_k^\top \boldsymbol{A}(\bar{\boldsymbol{x}}_k + \boldsymbol{D}_k \bar{\boldsymbol{u}}_k^j)\boldsymbol{D}_k$
   - $\bar{\boldsymbol{u}}_k^{j+1}$ is such that

   $$\bar{q}_k(\bar{\boldsymbol{u}}_k^{j+1}, \bar{\boldsymbol{u}}_k^j) + \Psi_0(\bar{\boldsymbol{x}}_k + \boldsymbol{D}_k \bar{\boldsymbol{u}}_k^{j+1}) \leqslant G_\delta(\bar{\boldsymbol{x}}_k + \boldsymbol{D}_k \bar{\boldsymbol{u}}_k^j) - \frac{\alpha}{2} \|\bar{\boldsymbol{u}}_k^{j+1}\|^2_{\boldsymbol{B}_{k,\bar{\boldsymbol{u}}_k^j}}$$
5. Update $\boldsymbol{x}_{k+1} = \bar{\boldsymbol{x}}_k + \boldsymbol{D}_k \bar{\boldsymbol{u}}_k^J$

# Assumptions

① Assumptions of Proposition 1.

② Assumptions of Lemma 1.

③ $\nabla F_\delta$ is Lipschitzian continuous.

④ $\Psi_0$ is convex, and continuous on its domain.

⑤ $\alpha \in (0,1]$ and there exist constants $\underline{\gamma}$ and $\overline{\gamma}$ such that $(\forall k \in \mathbb{N})\ 0 < \underline{\gamma} \leqslant \gamma_k \leqslant \overline{\gamma} < \nu^{-1}$ where $\nu$ is an upper bound of the spectrum of $\boldsymbol{A}(\boldsymbol{z}), \boldsymbol{z} \in \mathbb{R}^N$.

⑥ $G_\delta$ satisfies the Łojasiewicz inequality *[Attouch10a,Attouch10b]*:
For every $\widetilde{\boldsymbol{x}} \in \mathbb{R}^N$ and every bounded neighborhood of $E$ of $\widetilde{\boldsymbol{x}}$, there exist constants $\kappa > 0$, $\zeta > 0$ and $\theta \in [0,1)$ such that for all $h(\boldsymbol{x}) \in \partial G_\delta(\boldsymbol{x})$,

$$\|h(\boldsymbol{x})\| \geqslant \kappa |G_\delta(\boldsymbol{x}) - G_\delta(\widetilde{\boldsymbol{x}})|^\theta,$$

for every $\boldsymbol{x} \in E$ such that $|G_\delta(\boldsymbol{x}) - G_\delta(\widetilde{\boldsymbol{x}})| < \zeta$.

# Convergence results

## Lemma 2

Under Assumptions ② and ⑤, there exists $\mu > 0$ such that

$$(\forall k \in \mathbb{N})(\forall j \in \{0, \ldots, J\}) \quad G_\delta(\boldsymbol{x}_k) - G_\delta(\bar{\boldsymbol{x}}_k^j) \geqslant \frac{\mu}{2} \|\bar{\boldsymbol{x}}_k - \boldsymbol{x}_k\|^2,$$

where $\bar{\boldsymbol{x}}_k^j \triangleq \bar{\boldsymbol{x}}_k + \boldsymbol{D}_k \bar{\boldsymbol{u}}_k^j$.

## Lemma 3

Under Assumptions ② and ⑤, there exists $\eta > 0$ such that

$$(\forall k \in \mathbb{N})(\forall j \in \{0, \ldots, J-1\}) \quad G_\delta(\bar{\boldsymbol{x}}_k^j) - G_\delta(\bar{\boldsymbol{x}}_k^{j+1}) \geqslant \frac{\alpha\eta}{2} \|\bar{\boldsymbol{x}}_k^{j+1} - \bar{\boldsymbol{x}}_k^j\|^2.$$

# Convergence results

## Theorem 2

Under Assumptions ①, ②, ③, ④, ⑤ and ⑥

- The Proximal MM Memory Gradient algorithm generates a sequence converging to a critical point $\widetilde{\boldsymbol{x}}$ of $G_\delta = F_\delta + \Psi_0$.
- The sequences $(\boldsymbol{x}_k)_{k\in\mathbb{N}}$ and $(\bar{\boldsymbol{x}}_k)_{k\in\mathbb{N}}$ have a finite length in the sense that

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

# Outline

# Simulation settings

Considered penalization functions:

$$\begin{cases} \psi_{s,\delta}(t) = & \lambda(\sqrt{1+\frac{t^2}{\delta^2}}-1) \quad \text{SC} \\ \psi_{s,\delta}(t) = & \lambda\frac{t^2}{2\delta^2+t^2} \quad \text{SNC} \\ \psi_{s,\delta}(t) = & \lambda \min\left(\frac{t^2}{2\delta^2}, 1\right) \quad \text{NSNC} \end{cases}$$

Optimization algorithms for minimizing $F_\delta$:

- ⇝ MM Memory Gradient algorithm (MM-MG) *[Chouzenoux11]*
- ⇝ NLCG *[Hager06]*
- ⇝ L-BFGS *[Liu89]*

# Simulation settings

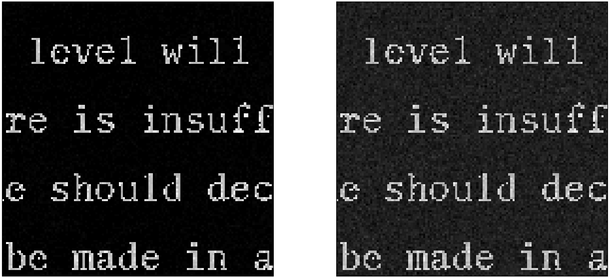
Considered penalization functions:

$$\begin{cases} \psi_{s,\delta}(t) = & \lambda(\sqrt{1+\frac{t^2}{\delta^2}}-1) \quad \text{SC} \\ \psi_{s,\delta}(t) = & \lambda\frac{t^2}{2\delta^2+t^2} \quad \text{SNC} \\ \psi_{s,\delta}(t) = & \lambda \min\left(\frac{t^2}{2\delta^2},1\right) \quad \text{NSNC} \end{cases}$$

Optimization algorithms for minimizing $G_\delta$:

- ⇝ Proximal Memory Gradient algorithm (P-MM-MG), inner constrained minimization with ADMM *[Glow89]*
- ⇝ Forward-Backward algorithm (FB) *[Chen97]*
- ⇝ NSNC: Three state-of-the-art combinatorial optimization algorithms : $\alpha$-EXP *[Boykov01]*, TRW *[Kolmogorov06]* and BP *[Felzenszwalb10]*

# Image denoising



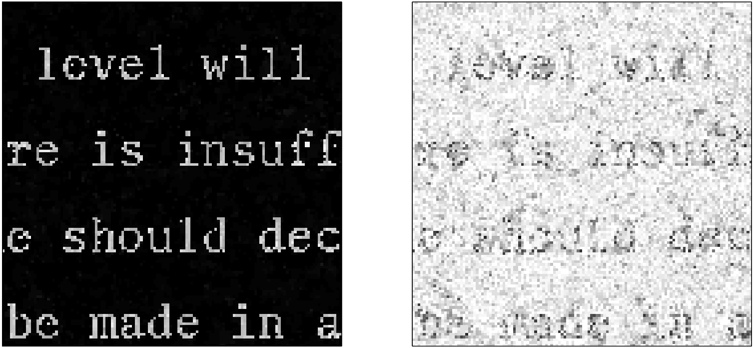

*Original image* $\bar{x}$ *with* $128 \times 128$ *pixels (left) and noisy image* $\boldsymbol{y}$, *degraded by i.i.d. Gaussian noise, SNR*$= 15$ *dB (right).*

$$G_\delta(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{y}\|^2 + \sum_{s=1}^{S} \psi_{s,\delta}(\|\boldsymbol{V}_s \boldsymbol{x}\|) + \iota_B(\boldsymbol{x}).$$

$\iota_B$ $\rightsquigarrow$ Indicator function of the closed convex interval $B = [0, 255]^N$

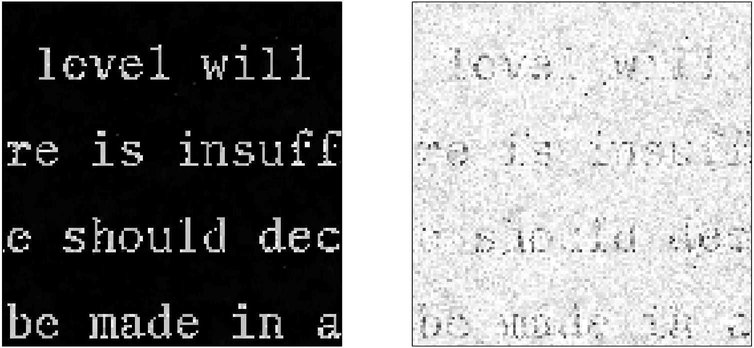
$\boldsymbol{V}_s$ $\rightsquigarrow$ Anisotropic penalization on the gradients of $\boldsymbol{x}$

# Results



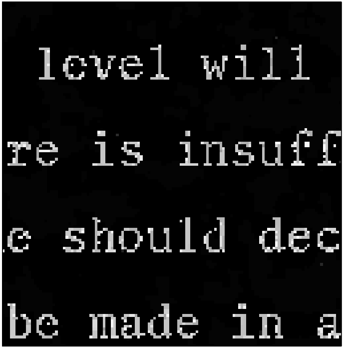

*Denoising result (left) and absolute reconstruction error (right) with SC penalty using P-MM-MG, SNR* $= 20.5$ *dB.*

# Results




*Denoising result (left) and absolute reconstruction error (right) with SNC penalty using P-MM-MG, SNR* $= 22.8$ *dB.*

# Results




*Denoising result (left) and absolute reconstruction error (right) with NSNC penalty using TRW, SNR $= 22.8$ dB.*

# Results

| Penalty | Algorithm | Iterations | Time (s) | SNR (dB) |
|---|---|---|---|---|
| SC | MM-MG | 134 | 2 | 20.2 |
| | NLCG | 144 | 2.7 | 20.2 |
| | L-BFGS | 172 | 3.4 | 20.2 |
| SC $+\iota_B$ | P-MM-MG | 76 | 0.9 | 20.5 |
| | FB | 2478 | 5.2 | 20.5 |
| SNC | MM-MG | 119 | 1 | 22.6 |
| | NLCG | 149 | 1.8 | 22.6 |
| | L-BFGS | 314 | 4.4 | 22.6 |
| SNC $+\iota_B$ | P-MM-MG | 103 | 1 | 22.8 |
| | FB | 861 | 1.2 | 22.8 |
| NSNC $+\iota_B$ | $\alpha$-EXP | 4.9 | 4.67 | 22.7 |
| | TRW | 5 | 3.1 | 22.8 |
| | BP | 18 | 11.7 | 22.7 |

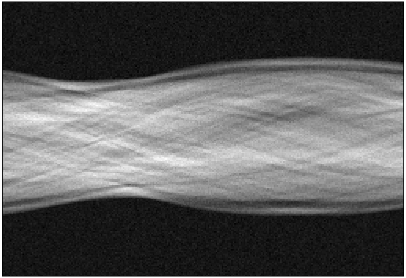
# Results

| Penalty | Algorithm | Iterations | Time (s) | SNR (dB) |
|---|---|---|---|---|
| SC | MM-MG | 134 | 2 | 20.2 |
| | NLCG | 144 | 2.7 | 20.2 |
| | L-BFGS | 172 | 3.4 | 20.2 |
| SC $+\iota_B$ | P-MM-MG | 76 | 0.9 | **20.5** |
| | FB | 2478 | 5.2 | **20.5** |
| SNC | MM-MG | 119 | 1 | 22.6 |
| | NLCG | 149 | 1.8 | 22.6 |
| | L-BFGS | 314 | 4.4 | 22.6 |
| SNC $+\iota_B$ | P-MM-MG | 103 | 1 | **22.8** |
| | FB | 861 | 1.2 | **22.8** |
| NSNC $+\iota_B$ | $\alpha$-EXP | 4.9 | 4.67 | **22.7** |
| | TRW | 5 | 3.1 | **22.8** |
| | BP | 18 | 11.7 | **22.7** |

# Results

| Penalty | Algorithm | Iterations | Time (s) | SNR (dB) |
|---|---|---|---|---|
| SC | **MM-MG** | 134 | **2** | 20.2 |
| | NLCG | 144 | 2.7 | 20.2 |
| | L-BFGS | 172 | 3.4 | 20.2 |
| SC $+\iota_B$ | **P-MM-MG** | 76 | **0.9** | **20.5** |
| | FB | 2478 | 5.2 | **20.5** |
| SNC | **MM-MG** | 119 | **1** | 22.6 |
| | NLCG | 149 | 1.8 | 22.6 |
| | L-BFGS | 314 | 4.4 | 22.6 |
| SNC $+\iota_B$ | **P-MM-MG** | 103 | **1** | **22.8** |
| | FB | 861 | 1.2 | **22.8** |
| NSNC $+\iota_B$ | $\alpha$-EXP | 4.9 | 4.67 | **22.7** |
| | TRW | 5 | 3.1 | **22.8** |
| | BP | 18 | 11.7 | **22.7** |

# Image reconstruction



*Original image* $\bar{\boldsymbol{x}}$ *with* $128 \times 128$ *pixels (left) and noisy sinogram* $\boldsymbol{y}$ *with* $181 \times 256$ *measurements, degraded by i.i.d. Gaussian noise, SNR=25 dB (right).*

$$G_\delta(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{R}\boldsymbol{x} - \boldsymbol{y}\|^2 + \sum_{s=1}^{S} \psi_{s,\delta}(\|\boldsymbol{V}_s\boldsymbol{x}\|) + \iota_B(\boldsymbol{x})$$

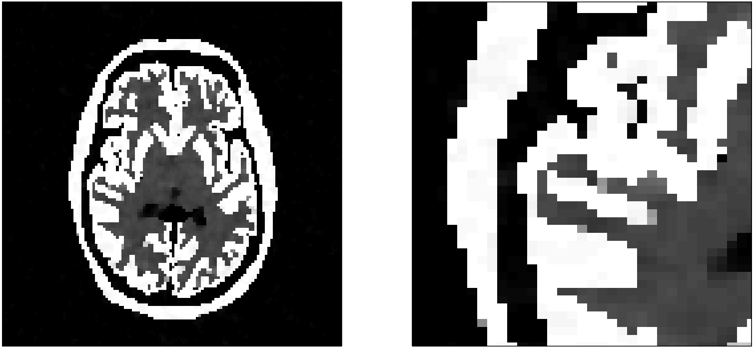
$\boldsymbol{R}$ $\rightsquigarrow$ Radon projection matrix

$\boldsymbol{V}_s$ $\rightsquigarrow$ Anisotropic penalization on the gradients of $\boldsymbol{x}$

# Results

*Reconstructed image (left) and detail (right) with*
*SC* $+\iota_B$ *penalty using P-MM-MG, SNR* $= 19.05$ *dB.*

# Results



*Reconstructed image (left) and detail (right) with*
*SNC* $+\iota_B$ *penalty using P-MM-MG, SNR* $= 22.85$ *dB.*

# Results

| Penalty | Algorithm | Iterations | Time (s) | SNR (dB) |
|---|---|---|---|---|
| SC | MM-MG | 111 | 36 | 18.4 |
| | NLCG | 129 | 40 | 18.4 |
| | L-BFGS | 113 | 38 | 18.4 |
| SC $+\iota_B$ | P-MM-MG | 141 | 118 | 19.05 |
| | FB | 3276 | 625 | 19.05 |
| SNC | MM-MG | 236 | 92 | 21.12 |
| | NLCG | 242 | 94 | 21.22 |
| | L-BFGS | 244 | 96 | 21.14 |
| SNC $+\iota_B$ | P-MM-MG | 245 | 190 | 22.85 |
| | FB | 5459 | 1011 | 22.85 |

# Results

| Penalty | Algorithm | Iterations | Time (s) | SNR (dB) |
|---|---|---|---|---|
| SC | MM-MG | 111 | 36 | 18.4 |
| | NLCG | 129 | 40 | 18.4 |
| | L-BFGS | 113 | 38 | 18.4 |
| SC $+\iota_B$ | P-MM-MG | 141 | 118 | **19.05** |
| | FB | 3276 | 625 | **19.05** |
| SNC | MM-MG | 236 | 92 | 21.12 |
| | NLCG | 242 | 94 | 21.22 |
| | L-BFGS | 244 | 96 | 21.14 |
| SNC $+\iota_B$ | P-MM-MG | 245 | 190 | **22.85** |
| | FB | 5459 | 1011 | **22.85** |

# Results

| Penalty | Algorithm | Iterations | Time (s) | SNR (dB) |
|---|---|---|---|---|
| SC | **MM-MG** | 111 | **36** | 18.4 |
| | NLCG | 129 | 40 | 18.4 |
| | L-BFGS | 113 | 38 | 18.4 |
| SC $+\iota_B$ | **P-MM-MG** | 141 | **118** | **19.05** |
| | FB | 3276 | 625 | **19.05** |
| SNC | **MM-MG** | 236 | **92** | 21.12 |
| | NLCG | 242 | 94 | 21.22 |
| | L-BFGS | 244 | 96 | 21.14 |
| SNC $+\iota_B$ | **P-MM-MG** | 245 | **190** | **22.85** |
| | FB | 5459 | 1011 | **22.85** |

# Outline

# Conclusion

- New Proximal Majorize-Minimize Memory Gradient algorithm for non-smooth nonconvex optimization.

- When $\Psi_0 = \iota_B$:
  - ⇝ Faster methods w.r.t. combinatorial optimization techniques
  - ⇝ Faster methods w.r.t. first order proximal techniques

- Future work
  - ⇝ Numerical performance for other choices of $\Psi_0$
  - ⇝ Splitting in more than two terms

# Bibliography


M. Allain, J. Idier and Y. Goussard
On global and local convergence of half-quadratic algorithms
*IEEE Transactions on Image Processing*, 15(5):1130–1142, 2006

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
*Mathematics of Operations Research*, 35(2):438–457, 2010

E. Chouzenoux, J. Idier and S. Moussaoui
A Majorize-Minimize strategy for subspace optimization applied to image restoration
*IEEE Transactions on Image Processing*, 20(6):1517–1528, 2011

E. Chouzenoux, A. Jezierska, J.-C. Pesquet and H. Talbot
A Majorize-Minimize subspace approach for $\ell_2$-$\ell_0$ image regularization
http://arxiv.org/abs/1112.6272, 2011

Thanks for your attention !