A Memory Gradient Algorithm for ℓ_2-ℓ_0 Regularization with Applications to Image Restoration

Emilie Chouzenoux, Jean-Christophe Pesquet, Hugues Talbot and Anna Jezierska

Université Paris-Est, Institut Gaspard Monge, France, {first.last}@univ-paris-est.fr

INTRODUCTION

State of the art:

✓ Non-convex priors have good ability to promote sparsity

✗ However, they lead to a difficult optimization problem

Proposal: Majorize-Minimize Memory Gradient algorithm

✓ Proof of convergence of the iterates of the algorithm

✓ Good numerical performance on image restoration problems

PROBLEM

Objective: Restore the unknown original image $\overline{\boldsymbol{x}} \in \mathbb{R}^N$ from $\boldsymbol{y} \in \mathbb{R}^Q$, related to $\overline{\boldsymbol{x}}$ through:

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

Goal of the algorithm:

$$\boxed{\underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{minimize}}\; F_\delta(\boldsymbol{x}) = \Phi(\boldsymbol{H}\boldsymbol{x}-\boldsymbol{y}) + \Psi_\delta(\boldsymbol{x})}$$

Φ - some measure of data fidelity

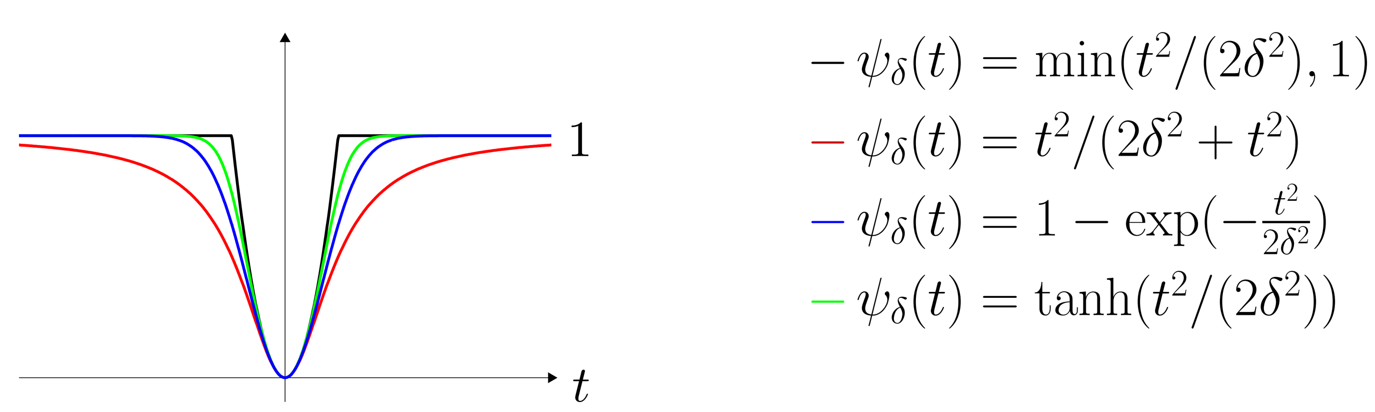
Ψ_δ - regularization term defined as:

$$\boxed{\Psi_\delta(\boldsymbol{x}) = \lambda \sum_{c=1}^{C} \psi_\delta(\boldsymbol{V}_c^\top \boldsymbol{x}) + \|\boldsymbol{\Pi}\boldsymbol{x}\|^2}$$

with $\boldsymbol{V}_c$ a dictionary of analysis vectors, λ a positive real value, $\boldsymbol{\Pi}$ a matrix in $\mathbb{R}^{P\times N}$ and ψ_δ a **differentiable**, **non-convex** approximation of the ℓ_0 norm.

Property: Epi-convergence to the ℓ_0 solution.

Examples of ℓ_2-ℓ_0 penalty functions



ALGORITHM

Majorize-Minimize Memory-Gradient algorithm:

⇝ Subspace algorithm

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{s}_k.$$

⇝ $\boldsymbol{D}_k$: set of Memory-Gradient directions

⇝ $\boldsymbol{s}_k$ resulting from MM minimization of $f_{k,\delta}(\boldsymbol{s})\colon \boldsymbol{s} \mapsto F_\delta(\boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{s})$

OPTIMIZATION

Construction of the directions: Memory-Gradient subspace [Cantrell 1969]

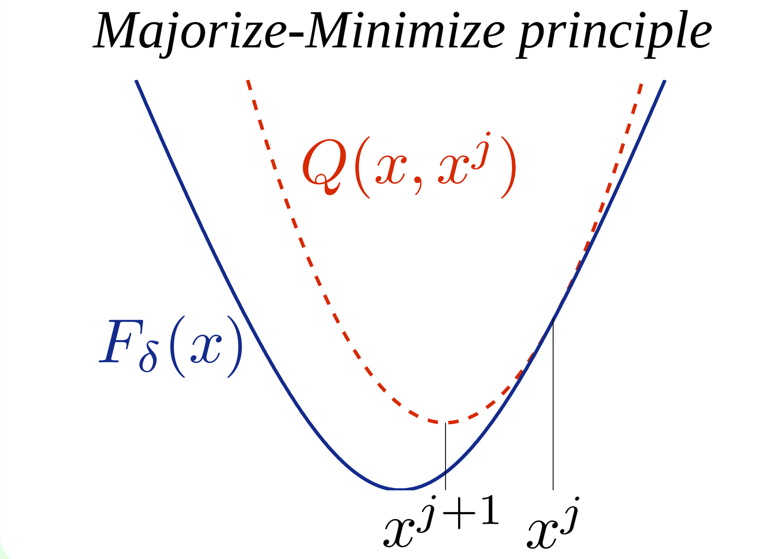
$$\boldsymbol{D}_k = [-\nabla F_\delta(\boldsymbol{x}_k), \boldsymbol{x}_k - \boldsymbol{x}_{k-1}] \in \mathbb{R}^{N\times 2}$$

Computation of the stepsize: [Chouzenoux *et al.* 2011] We assume that for all $\boldsymbol{x}'$, there exists $\boldsymbol{A}(\boldsymbol{x}')$, definite positive, such that

$$Q(\boldsymbol{x}, \boldsymbol{x}') = F_\delta(\boldsymbol{x}') + \nabla F_\delta(\boldsymbol{x}')^\top(\boldsymbol{x}-\boldsymbol{x}') + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}')^\top \boldsymbol{A}(\boldsymbol{x}')(\boldsymbol{x}-\boldsymbol{x}')$$

is a quadratic tangent majorant of F_δ at $\boldsymbol{x}'$ i.e., $Q(\boldsymbol{x}, \boldsymbol{x}') \geqslant F_\delta(\boldsymbol{x})$, $\forall \boldsymbol{x}$.

Majorize-Minimize principle



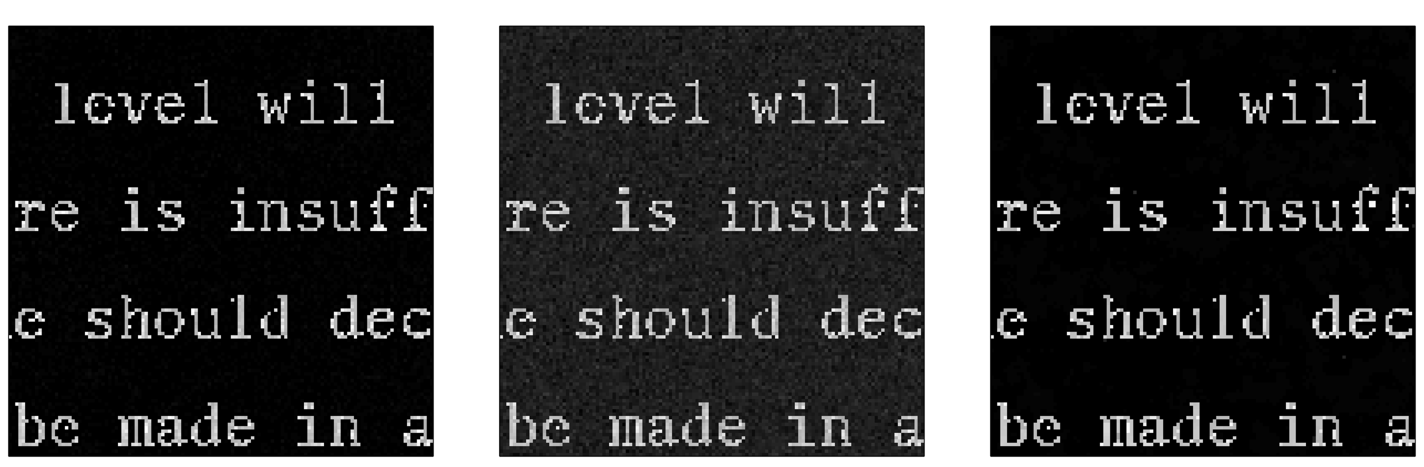
MM minimization in the subspace:

$$\begin{cases} \boldsymbol{s}_k^0 = \boldsymbol{0}, \\ (\forall j \in \{0,\dots J-1\}) \\ \quad \boldsymbol{s}_k^{j+1} \in \text{Argmin}_{\boldsymbol{s}} q_k(\boldsymbol{s}, \boldsymbol{s}_k^j), \\ \boldsymbol{s}_k = \boldsymbol{s}_k^J. \end{cases}$$

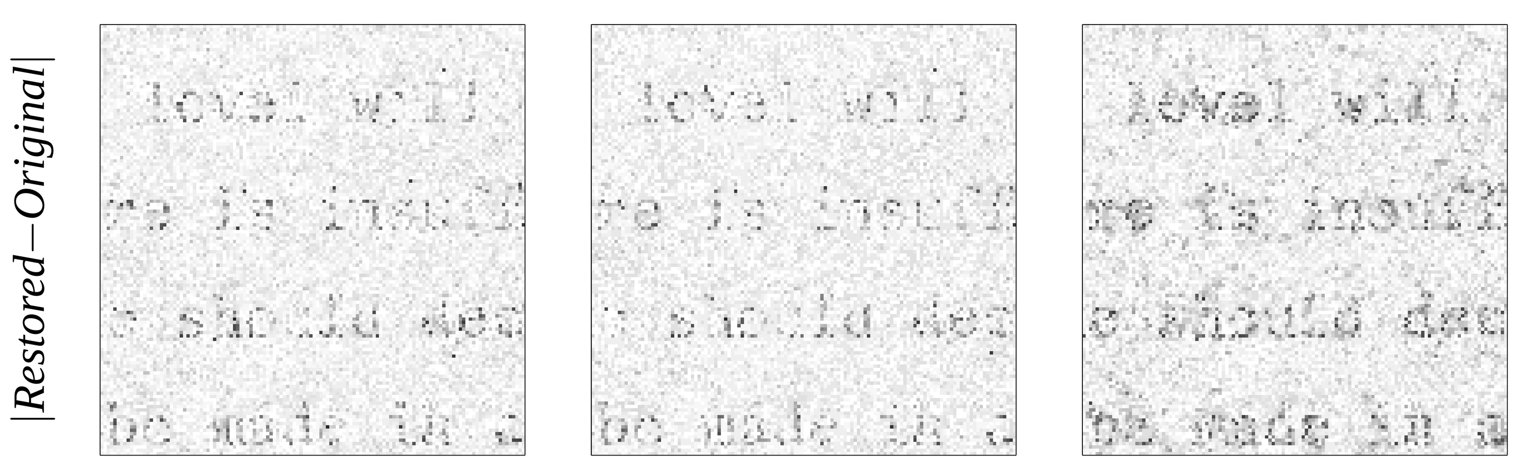
where $q_k(\boldsymbol{s}, \boldsymbol{s}_k^j)$ is a quadratic tangent majorant of $f_{k,\delta}$ at $\boldsymbol{s}_k^j$ with Hessian: $\boldsymbol{B}_{\boldsymbol{s}_k^j} = \boldsymbol{D}_k^\top \boldsymbol{A}(\boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{s}_k^j)\boldsymbol{D}_k$.

Convergence result

- If Φ is coercive, $\operatorname{Ker} \boldsymbol{H} \cap \operatorname{Ker} \boldsymbol{\Pi} = \boldsymbol{0}$, and the gradient of Φ is L-Lipschitzian, then, for all $J \geqslant 1$, $\lim_{k\to\infty} \nabla F_\delta(\boldsymbol{x}_k) = \boldsymbol{0}$.
- Morever, if F_δ satisfies the Łojasiewicz inequality [Attouch *et al.* 2010], the sequence $(\boldsymbol{x}_k)_{k\in\mathbb{N}}$ converges to a critical point $\widetilde{\boldsymbol{x}}$ of F_δ.



Original image — *Noisy image (15 dB)* — *Restored image*

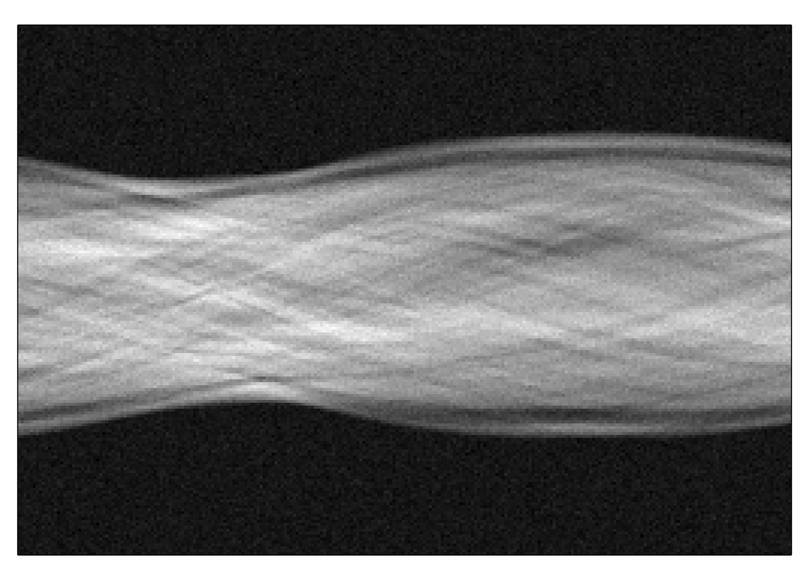


Proposed non-convex — *Truncated quadratic* — *Smooth convex*

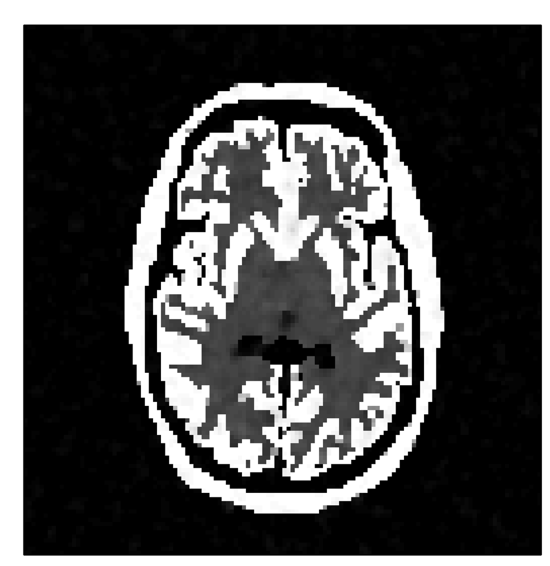
22.4 dB, MSSIM =0.92 — *22.8 dB, MSSIM =0.93* — *20.2 dB, MSSIM =0.89*

Proposed MM-MG algorithm	0.74 s
Conjugate Gradient algorithm [Hager 2006]	1.11 s
Quasi-Newton algorithm [Liu & Nocedal 1989]	1.05 s
Half-Quadratic algorithm [Allain *et al.* 2006]	4.16 s
TRW algorithm [Kolmogorov 2006]	6.96 s
BP algorithm [Felzenszwalb & Huttenlocher 2010]	20.95 s

Convergence speed of several optimization algorithms for the considered denoising problem, with ℓ_2-ℓ_0 penalties.



Noisy sinogram (25 dB)



Reconstructed image

Original image (Detail) — *Proposed non-convex* — *Smooth convex*

20.4 dB, MSSIM =0.79 — *18.4 dB, MSSIM =0.78*

Proposed MM-MG algorithm	36 s
Conjugate Gradient algorithm [Hager 2006]	48 s
Quasi-Newton algorithm [Liu & Nocedal 1989]	42 s
Half-Quadratic algorithm [Allain *et al.* 2006]	779 s

Convergence speed of several optimization algorithms for the considered reconstruction problem, with ℓ_2-ℓ_0 penalties.