A memory gradient algorithm for ℓ_2-ℓ_0 regularization with applications to image restoration

E. Chouzenoux, A. Jezierska, J.-C. Pesquet and H. Talbot

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

Journées GDR MOA-MSPC
Optimisation et traitement d'images

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:

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

- **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

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

where

- Φ $\rightsquigarrow$ Fidelity to data term, related to noise
- R $\rightsquigarrow$ Regularization term, related to some *a priori* assumptions
- λ $\rightsquigarrow$ Regularization weight

Assumption: $\boldsymbol{x}$ is **sparse** in a dictionary $\mathcal{V}$ of analysis vectors in $\mathbb{R}^N$

$$F_0(\boldsymbol{x}) = \Phi(\boldsymbol{H}\boldsymbol{x}-\boldsymbol{y}) + \lambda\,\ell_0(\boldsymbol{V}\boldsymbol{x})$$

Context

Penalized optimization problem

Find

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

where

- $\Phi \rightsquigarrow$ Fidelity to data term, related to noise
- $R \rightsquigarrow$ Regularization term, related to some *a priori* assumptions
- $\lambda \rightsquigarrow$ Regularization weight

Assumption: $\boldsymbol{x}$ is **sparse** in a dictionary $\mathcal{V}$ of analysis vectors in $\mathbb{R}^N$

$$F_\delta(\boldsymbol{x}) = \Phi(\boldsymbol{H}\boldsymbol{x}-\boldsymbol{y}) + \lambda \sum_{c=1}^{C} \psi_\delta(\boldsymbol{V}_c^\top \boldsymbol{x})$$

where ψ_δ is a differentiable, non-convex approximation of the ℓ_0 norm.

Outline

Outline

ℓ_2-ℓ_0 regularization functions

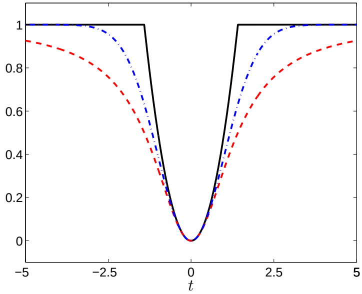
We consider the following class of potential functions:

1. $(\forall \delta \in (0,+\infty))$ ψ_δ is differentiable.
2. $(\forall \delta \in (0,+\infty))$ $\lim_{t\to\infty} \psi_\delta(t) = 1$.
3. $(\forall \delta \in (0,+\infty))$ $\psi_\delta(t) = \mathcal{O}(t^2)$ for small t.

Examples:

— — — $\psi_\delta(t) = \frac{t^2}{2\delta^2+t^2}$

— · —· $\psi_\delta(t) = 1 - \exp(-\frac{t^2}{2\delta^2})$



Existence of minimizers

$$F_\delta(\boldsymbol{x}) = \Phi(\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y}) + \lambda \sum_{c=1}^{C} \psi_\delta(\boldsymbol{V}_c^\top \boldsymbol{x})$$

Difficulty: F_δ is a **non convex**, **non coercive** function.

Proposition

Assume that

$$\lim_{\|x\| \to +\infty} \Phi(x) = +\infty$$

and that

$$\mathrm{Ker}\, \boldsymbol{H} = \{\boldsymbol{0}\}$$

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

Existence of minimizers

$$F_\delta(\boldsymbol{x}) = \Phi(\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y}) + \lambda \sum_{c=1}^{C} \psi_\delta(\boldsymbol{V}_c^\top \boldsymbol{x}) + \|\boldsymbol{\Pi}\boldsymbol{x}\|^2$$

Difficulty: F_δ is a **non convex**, **non coercive** function.

Proposition

Assume that

$$\lim_{\|x\|\to+\infty} \Phi(x) = +\infty$$

and that

$$\mathrm{Ker}\, \boldsymbol{H} \cap \mathrm{Ker}\, \boldsymbol{\Pi} = \{\boldsymbol{0}\}$$

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

Epi-convergence to the ℓ_0-penalized objective function

Assumptions:

1. $(\forall(\delta_1,\delta_2)\in(0,+\infty)^2)\ \delta_1\le\delta_2 \Rightarrow (\forall t\in\mathbb{R})\ \psi_{\delta_1}(t)\ge\psi_{\delta_2}(t)\ge 0.$
2. $(\forall t\in\mathbb{R})\ \ \lim_{\delta\to 0}\psi_\delta(t)=\begin{cases}0 & \text{if } t=0\\ 1 & \text{otherwise.}\end{cases}$
3. Φ is coercive and $\operatorname{Ker}\boldsymbol{H}\cap\operatorname{Ker}\boldsymbol{\Pi}=\boldsymbol{0}$

Proposition

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

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

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

Outline

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

Descent algorithm

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

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

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

Descent algorithm

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

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

Generalization : subspace algorithm *[Zibulevsky10]*

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \sum_{i=1}^{r} s_{i,k} \boldsymbol{d}_k^i, \quad (\forall k \in \{0, \dots, K\})$$

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

Majorize-Minimize principle *[Hunter04]*

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

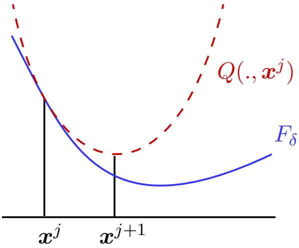
For all $\boldsymbol{x}'$, let $Q(., \boldsymbol{x}')$ a *tangent majorant* of F_δ at $\boldsymbol{x}'$ i.e.,

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

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

Assumption: For all $\boldsymbol{x}'$, there exists $\boldsymbol{A}(\boldsymbol{x}')$, definite positive, such that

$$Q(\boldsymbol{x}, \boldsymbol{x}') = F_\delta(\boldsymbol{x}') + \nabla F_\delta(\boldsymbol{x}')^T(\boldsymbol{x} - \boldsymbol{x}') + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}')^T \boldsymbol{A}(\boldsymbol{x}')(\boldsymbol{x} - \boldsymbol{x}')$$

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

Construction of $\boldsymbol{A}(.)$

$$G(\boldsymbol{x}) = \sum_c \phi([\boldsymbol{V}\boldsymbol{x} - \boldsymbol{w}]_c)$$

$$\begin{cases} \phi \text{ is } \mathcal{C}^1 \text{ on } \mathbb{R} \\ \phi \text{ has } L-\text{Lipschitz gradient on } \mathbb{R} \end{cases} \Rightarrow \boldsymbol{A}(\boldsymbol{x}') = a\boldsymbol{V}^T\boldsymbol{V}, \quad a \geqslant L$$

Quadratic tangent majorant function

Assumption: For all $\boldsymbol{x}'$, there exists $\boldsymbol{A}(\boldsymbol{x}')$, definite positive, such that

$$Q(\boldsymbol{x},\boldsymbol{x}') = F_\delta(\boldsymbol{x}') + \nabla F_\delta(\boldsymbol{x}')^T(\boldsymbol{x}-\boldsymbol{x}') + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}')^T\boldsymbol{A}(\boldsymbol{x}')(\boldsymbol{x}-\boldsymbol{x}')$$

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

Construction of $\boldsymbol{A}(.)$

$$G(\boldsymbol{x}) = \sum_c \phi([\boldsymbol{V}\boldsymbol{x}-\boldsymbol{w}]_c)$$

$$\begin{cases} \phi \text{ is } \mathcal{C}^1 \text{ on } \mathbb{R} \\ \phi \text{ has } L-\text{Lipschitz gradient on } \mathbb{R} \end{cases} \Rightarrow \boldsymbol{A}(\boldsymbol{x}') = a\boldsymbol{V}^T\boldsymbol{V}, \quad a \geqslant L$$

[Allain06]

$$\begin{cases} \phi \text{ is } \mathcal{C}^1 \text{ and even on } \mathbb{R} \\ \phi(\sqrt{.}) \text{ is concave on } \mathbb{R}^+ \\ \omega(u) = \dot{\phi}(u)/u \in (0,\infty) \end{cases} \Rightarrow \boldsymbol{A}(\boldsymbol{x}') = \boldsymbol{V}^T \mathrm{Diag}\left\{\omega([\boldsymbol{V}\boldsymbol{x}'-\boldsymbol{w}])\right\}\boldsymbol{V}$$

Majorize-Minimize multivariate stepsize *[Chouzenoux11]*

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{s}_k \qquad (\forall k \in \{0, \dots K\})$$

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

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

with Hessian: $\boldsymbol{B}_{\boldsymbol{s}_k^j} = \boldsymbol{D}_k^T \boldsymbol{A}(\boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{s}_k^j)\boldsymbol{D}_k$

MM minimization in the subspace:

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

Proposed algorithm

Majorize-Minimize Memory Gradient (MM-MG) algorithm

For $k = 1, \ldots, K$

1. Compute the set of directions, for example $\boldsymbol{D}_k = [-\boldsymbol{g}_k, \boldsymbol{x}_k - \boldsymbol{x}_{k-1}]$
2. $\boldsymbol{s}_k^0 = \boldsymbol{0}$
3. $\forall j \in \{0, \ldots J-1\}$,
   - $\boldsymbol{B}_{\boldsymbol{s}_k^j} = \boldsymbol{D}_k^\top \boldsymbol{A}(\boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{s}_k^j)\boldsymbol{D}_k$
   - $\boldsymbol{s}_k^{j+1} = \boldsymbol{s}_k^j - \boldsymbol{B}_{\boldsymbol{s}_k^j}^{-1} \nabla f_{k,\delta}(\boldsymbol{s}_k^j)$
4. $\boldsymbol{s}_k = \boldsymbol{s}_k^J$
5. Update $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{s}_k$

Convergence results

Assumptions

① Φ is coercive and $\operatorname{Ker} \boldsymbol{H} \cap \operatorname{Ker} \boldsymbol{\Pi} = \boldsymbol{0}$

② The gradient of Φ is L-Lipschitzian

③ ψ_δ is even and $\psi_\delta(\sqrt{.})$ is concave on $\mathbb{R}^+$. Moreover, there exists $\overline{\omega} \in [0, +\infty)$ such that $(\forall t \in (0, +\infty)) \quad 0 \leq \dot{\psi}_\delta(t) \leq \overline{\omega} t$.
In addition, $\lim_{\substack{t \to 0 \\ t \neq 0}} \dot{\psi}_\delta(t)/t \in \mathbb{R}$.

④ F_δ 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

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

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

Convergence results

Theorem

Under Assumptions ①, ②, and ③, for all $J \geqslant 1$, the MM-MG algorithm is such that

$$\lim_{k \to \infty} \nabla F_\delta(\boldsymbol{x}_k) = \mathbf{0}.$$

Furthermore, if Assumption ④ is fulfilled, then

- The MM-MG algorithm generates a sequence converging to a critical point $\widetilde{\boldsymbol{x}}$ of F_δ.
- The sequence $(\boldsymbol{x}_k)_{k \in \mathbb{N}}$ has a finite length in the sense that

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

Outline

Image denoising

Original image $\overline{\boldsymbol{x}}$
512×512

Noisy image $\boldsymbol{y}$
SNR $= 15$ dB

- $F_\delta(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{y}\|^2 + \lambda \sum_c \psi_\delta(\boldsymbol{V}_c^\top \boldsymbol{x})$, $\psi_\delta(t) = 1 - \exp(-t^2/2\delta^2)$
- MM-MG / Beck Teboulle / Half quadratic algorithms
- Comparison with discrete methods for $\psi_\delta(t) = \delta^{-2}\min(t^2, \delta^2)$

Results

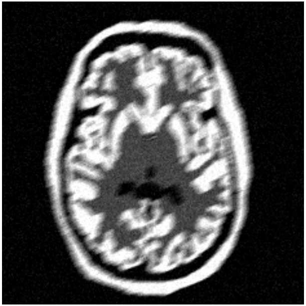
Denoised image
SNR = 24.4 dB

Results

Algorithm	Time	SNR (dB)
MM-MG	28 s	24.4
Beck-Teboulle *[Beck09]*	45 s	24.45
Half-Quadratic *[Allain06]*	97 s	24.42
Graph Cut Swap *[Boykov02]*	365 s	23.98
Tree-Reweighted *[Felzenszwalb10]*	181 s	24.28
Belief Propagation *[Kolmogorov06]*	1958 s	24.28

Image deblurring

Original image $\overline{\boldsymbol{x}}$
256×256

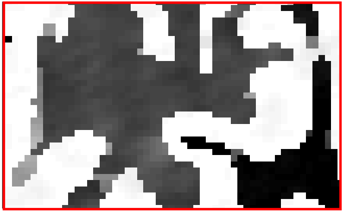


Noisy blurred image $\boldsymbol{y}$
Motion blur (9 pixels)
SNR $= 10$ dB

- $F_\delta(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{H}\boldsymbol{x} - \boldsymbol{y}\|^2 + \lambda \sum_c \psi_\delta(\boldsymbol{V}_c^\top \boldsymbol{x}) + \tau\|\boldsymbol{x}\|^2$, $\tau = 10^{-10}$
- Non convex penalty: $\psi_\delta(t) = 1 - \exp(-t^2/2\delta^2)$
- Convex penalty: $\psi_\delta(t) = \sqrt{1 + t^2/\delta^2} - 1$

Results: Non convex penalty

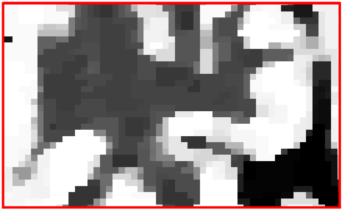
Restored image
SNR $= 18.82$ dB



Algorithm	Time	SNR (dB)
MM-MG	29 s	18.82
B-T	32 s	18.67
H-Q	141 s	18.52

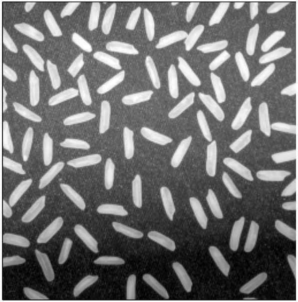
Results: Convex penalty

Denoised image
SNR = 17.91 dB



Algorithm	Time	SNR (dB)
MM-MG	194 s	17.91
B-T	900 s	17.91
H-Q	1824 s	17.91

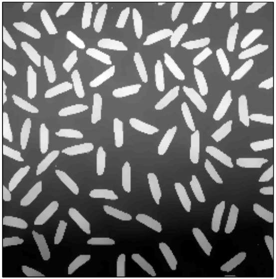
Image segmentation



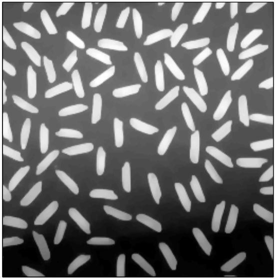
Original image $\overline{\boldsymbol{x}}$
256×256

- $F_\delta(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{x} - \overline{\boldsymbol{x}}\|^2 + \lambda \sum_c \psi_\delta(\boldsymbol{V}_c^\top \boldsymbol{x})$ with large λ and small δ
- Non convex penalty: $\psi_\delta(t) = 1 - \exp(-t^2/2\delta^2)$
- Convex penalty: $\psi_\delta(t) = \sqrt{1 + t^2/\delta^2} - 1$

Results



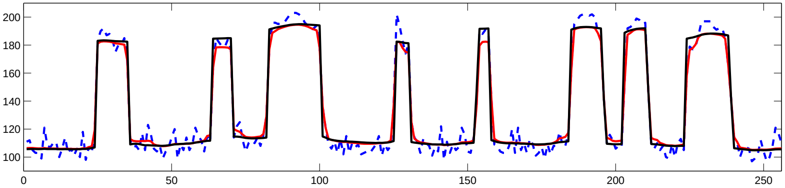
Segmentation with non convex penalty



Segmentation with convex penalty

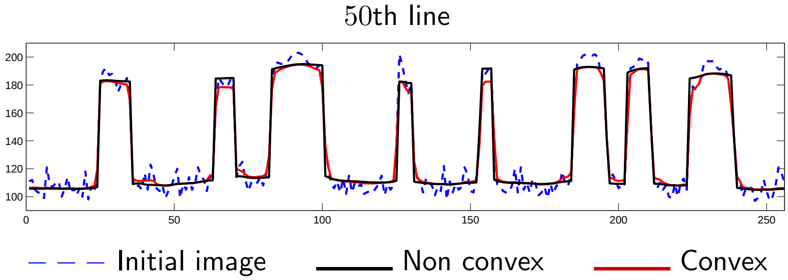
Results

50th line



Initial image Non convex Convex

Results



Algorithm	Time	
	Non convex	Convex
MM-MG	15 s	5 s
Beck-Teboulle	21 s	50 s
Half-Quadratic	39 s	31 s

Texture+Geometry decomposition

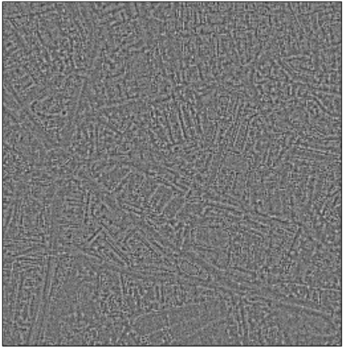
Original image $\overline{\boldsymbol{x}}$
256×256

Noisy image $\boldsymbol{y}$
SNR=15 dB

- $\boldsymbol{y} = \hat{\boldsymbol{x}} + \check{\boldsymbol{x}}$ with $\begin{cases} \hat{\boldsymbol{x}} & \text{geometry} \\ \check{\boldsymbol{x}} & \text{texture + noise} \end{cases}$
- $\hat{x} \in \operatorname{Arg\,min}_{\boldsymbol{x}} \left(\frac{1}{2}\|\nabla\Delta^{-1}(\boldsymbol{x}-\boldsymbol{y})\|_2 + \lambda \sum_c (\psi_\delta(\boldsymbol{V}_c^T \boldsymbol{x}))\right)$ *[Osher03]*
- Non convex penalty / convex penalty

Results: Non convex penalty

Geometry $\hat{x}$

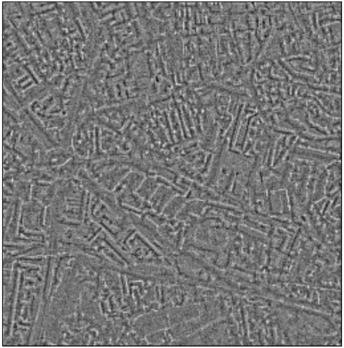


Texture + noise $\check{x}$

MM-MG algorithm: Convergence in 91 s

Results: Convex penalty

Geometry $\hat{x}$



Texture + noise $\check{x}$

MM-MG algorithm: Convergence in 134 s

Outline

Conclusion

- Majorize-Minimize memory gradient algorithm for ℓ_2-ℓ_0 minimization
  - $\rightsquigarrow$ Faster methods w.r.t. combinatorial optimization techniques
  - $\rightsquigarrow$ Simplicity of implementation

- Future work
  - $\rightsquigarrow$ Constrained case
  - $\rightsquigarrow$ Non differentiable case
  - $\rightsquigarrow$ How to ensure the global convergence?

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

A. Beck and M. Teboulle
Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems
IEEE Transactions on Image Processing, 18(11):2419–2434, 2009

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

Thanks for you attention !