Majorization-Minimization methods for large scale inverse problems in signal and image processing.

Habilitation thesis defended by: Emilie CHOUZENOUX

University Paris East - LIGM UMR CNRS 8049, Champs-sur-Marne
CentraleSupélec - Center for Visual Computing, INRIA Saclay

December, 1st 2017

Short biography

Since Sep. 2016: Associate Researcher at Center for Visual Computing, INRIA Saclay, CentraleSupélec (délégation INRIA).

Since Sep. 2011: Assistant Professor at Université Paris-Est Marne-la-Vallée, Laboratoire d'Informatique Gaspard Monge.

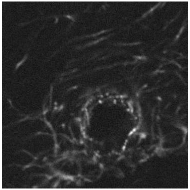
2010-2011: ATER at Université Paris-Est Marne-la-Vallée.

2007-2010: PhD Thesis at IRCCyN, Nantes, under the supervision of Jérôme Idier and Saïd Moussaoui, defended the 8th December 2010.

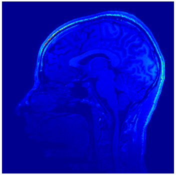
2006-2007: Master Degree in Automatic and Production Systems from Ecole Centrale Nantes. Graduated in September 2007 (with honors).

2004-2007: Engineer studies at Ecole Centrale de Nantes. Graduated in September 2007 (with honors).

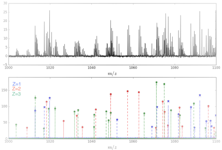
Context: inverse problems in signal/image processing



Microscopy

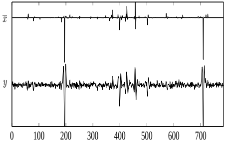


Parallel MRI

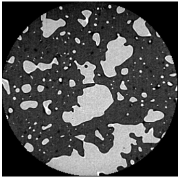


Mass spectrometry

Satellite imaging



Seismic data



Material science

Context: variational formulation

Observation model

Observation matrix

$$\mathbf{y} = \mathbf{H}\,\overline{\mathbf{x}} + \mathbf{w}$$

Original signal

Noise

Objective: Find an estimation $\hat{\mathbf{x}} \in \mathbb{R}^N$ of $\overline{\mathbf{x}}$ from $\mathbf{y}$.

$\Rightarrow$ Define estimate $\hat{\mathbf{x}}$ as a solution to $\underset{\mathbf{x}\in\mathbb{R}^N}{\text{minimize}} \; F(\mathbf{x})$.

✗ No closed form minimizer for F $\rightsquigarrow$ iterative method required.

✗ Large size of the problem (at least, $N = 10^6$ variables.)

Context: variational formulation

Observation model

Observation matrix

$$\mathbf{y} = \mathbf{H}\,\overline{\mathbf{x}} + \mathbf{w}$$

Original signal

Noise

Objective: Find an estimation $\hat{\mathbf{x}} \in \mathbb{R}^N$ of $\overline{\mathbf{x}}$ from $\mathbf{y}$.

$\Rightarrow$ Define estimate $\hat{\mathbf{x}}$ as a solution to $\underset{\mathbf{x}\in\mathbb{R}^N}{\text{minimize}} \quad F(\mathbf{x})$.

In the context of large scale problems, how to find an optimization algorithm able to deliver a reliable numerical solution in a reasonable time, with low memory requirement ?

A unified framework: Majorize-Minimize principle

PROBLEM: Find $\hat{\mathbf{x}} \in \operatorname{Argmin}_{\mathbf{x}\in\mathbb{R}^N} F(\mathbf{x})$

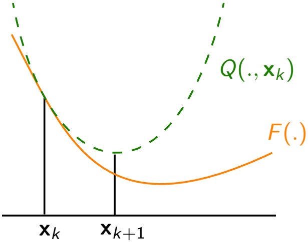
For all $\mathbf{x}' \in \mathbb{R}^N$, let $Q(.,\mathbf{x}')$ a **tangent majorant** of F at $\mathbf{x}'$ i.e.,

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad Q(\mathbf{x},\mathbf{x}') \geqslant F(\mathbf{x}) \quad \text{and} \quad Q(\mathbf{x}',\mathbf{x}') = F(\mathbf{x}')$$

MM algorithm:

$$(\forall k \in \mathbb{N})$$

$$\mathbf{x}_{k+1} \in \operatorname{Argmin}_{\mathbf{x}\in\mathbb{R}^N} Q(\mathbf{x},\mathbf{x}_k)$$



⋆ Quadratic majorants ⇝ **tractable inner minimization step**

Outline

MM FRAMEWORK: A simple and elegant methodology to build optimization algorithms for solving inverse problems of signal and image processing.

HOWEVER: A need for modernization !

Outline:

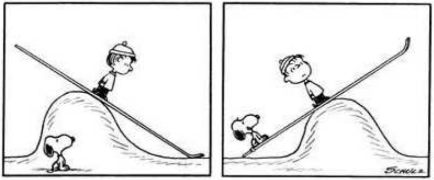
① Convergence analysis in the nonconvex case

② Block alternating and parallel strategies

③ Stochastic optimization at a large scale

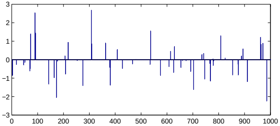
④ Data augmentation in the Bayesian framework

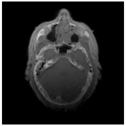
① - Convergence analysis in the nonconvex case



Motivations

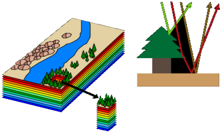
Nonconvexity in inverse problems:





Sparse signal recovery

Phase retrieval



Spectral unmixing

Blind deconvolution

How to design fast optimization algorithms with established convergence guarantees on their iterates in the nonconvex setting?

An essential tool: Kurdyka-Łojasiewicz inequality

Function F satisfies the Kurdyka-Łojasiewicz inequality i.e., for every $\xi \in \mathbb{R}$, and, for every bounded subset E of $\mathbb{R}^N$, there exist three constants $\kappa > 0$, $\zeta > 0$ and $\theta \in [0,1)$ such that

$$(\forall \mathbf{t} \in \partial F(\mathbf{x})) \qquad \|\mathbf{t}\| \geq \kappa |F(\mathbf{x}) - \xi|^{\theta},$$

for every $\mathbf{x} \in E$ such that $|F(\mathbf{x}) - \xi| \leq \zeta$.

- ⋆ Satisfied for a wide class of non necessarily convex functions :
  - real analytic functions
  - semi-algebraic functions
  - ...
- ⋆ Key ingredient to prove convergence of iterates under suitable descent properties

An essential tool: Kurdyka-Łojasiewicz inequality

Function F satisfies the Kurdyka-Łojasiewicz inequality i.e., for every $\xi \in \mathbb{R}$, and, for every bounded subset E of $\mathbb{R}^N$, there exist three constants $\kappa > 0$, $\zeta > 0$ and $\theta \in [0,1)$ such that

$$(\forall \mathbf{t} \in \partial F(\mathbf{x})) \qquad \|\mathbf{t}\| \geq \kappa |F(\mathbf{x}) - \xi|^{\theta},$$

for every $\mathbf{x} \in E$ such that $|F(\mathbf{x}) - \xi| \leq \zeta$.

- ⋆ Satisfied for a wide class of non necessarily convex functions :
  - real analytic functions
  - semi-algebraic functions
  - ...
- ⋆ Key ingredient to prove convergence of iterates under suitable descent properties ⬅ Provided by the MM framework.

Majorize-Minimize subspace algorithm [Chouzenoux *et al.*, 2013]

⋆ Minimize differentiable and nonconvex function F on $\mathbb{R}^N$.

At each iteration $k \in \mathbb{N}$:

❶ Build a quadratic majorant function $Q(\cdot, \mathbf{x}_k)$ of F at $\mathbf{x}_k$.

❷ Minimize it within the subspace spanned by the columns of a matrix $\mathbf{D}_k \in \mathbb{R}^{N \times M_k}$.

✘ MM algorithm: $\operatorname{rank}(\mathbf{D}_k) = N \rightsquigarrow$ Large computational cost.

Majorize-Minimize subspace algorithm [Chouzenoux *et al.*, 2013]

⋆ Minimize differentiable and nonconvex function F on $\mathbb{R}^N$.

At each iteration $k \in \mathbb{N}$:

1. Build a quadratic majorant function $Q(\cdot, \mathbf{x}_k)$ of F at $\mathbf{x}_k$.
2. Minimize it within the subspace spanned by the columns of a matrix $\mathbf{D}_k \in \mathbb{R}^{N \times M_k}$.

✘ MM algorithm: $\mathrm{rank}(\mathbf{D}_k) = N \rightsquigarrow$ Large computational cost.

☛ 3MG algorithm: $M_k = 2$ and $\mathbf{D}_k = [\nabla F(\mathbf{x}_k) \,|\, \mathbf{x}_k - \mathbf{x}_{k-1}]$.

✓ Convergence of the sequence $(\mathbf{x}_k)_{k\in\mathbb{N}}$ to a critical point of F under KL assumption.

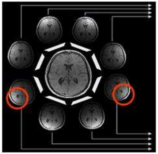
✓ 3MG algorithm outperforms state-of-the-art optimization algorithms in many image processing applications.

Application to parallel MRI [Florescu *et al.* - 2014]

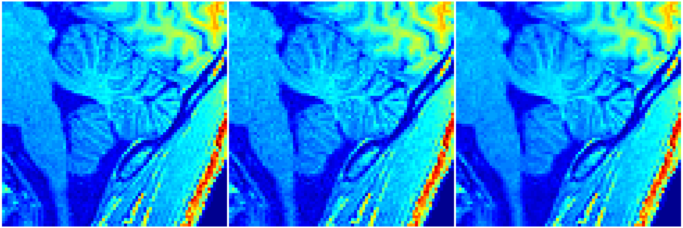
(Joint work with Ph. Ciuciu, CEA Neurospin)

CHALLENGES :

- Parallel acquisition and compressive sensing
- Complex-valued signals
- Nonconvex smoothed ℓ_0 prior on wavelet coefficients



RESULTS :

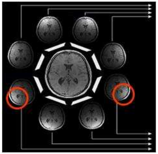


Original | 3MG result (convex) | 3MG result (nonconvex)

Application to parallel MRI [Florescu *et al.* - 2014]

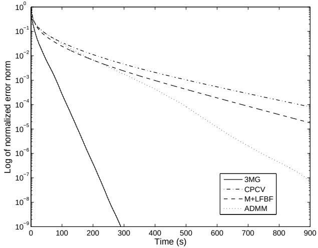
(Joint work with Ph. Ciuciu, CEA Neurospin)

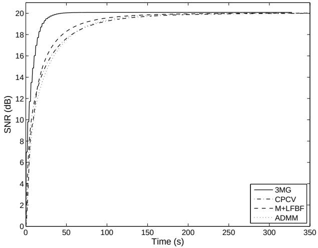
CHALLENGES :



- Parallel acquisition and compressive sensing
- Complex-valued signals
- Nonconvex smoothed ℓ_0 prior on wavelet coefficients

RESULTS :





Variable metric FB algorithm [Chouzenoux *et al.*, 2014]

(PhD Thesis of Audrey Repetti)

⋆ Minimize $F = f_1 + f_2$ with f_1 Lipschitz-differentiable and f_2 non smooth.

⇒ Forward-backward algorithm: gradient steps on f_1 and proximal steps on f_2:

$$(\forall k \in \mathbb{N}) \quad \mathbf{x}_{k+1} = \mathrm{prox}_{\theta_k f_2}\left(\mathbf{x}_k - \theta_k \nabla f_1(\mathbf{x}_k)\right).$$

✘ slow convergence in practice.

Variable metric FB algorithm [Chouzenoux *et al.*, 2014]

(PhD Thesis of Audrey Repetti)

⋆ Minimize $F = f_1 + f_2$ with f_1 Lipschitz-differentiable and f_2 non smooth.

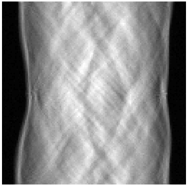
⇒ Forward-backward algorithm: gradient steps on f_1 and proximal steps on f_2:

☛ Use MM framework to propose an efficient variable metric strategy:

$$(\forall k \in \mathbb{N}) \quad \mathbf{x}_{k+1} = \operatorname{prox}_{\theta_k^{-1}\mathbf{A}_k, f_2}\left(\mathbf{x}_k - \theta_k \mathbf{A}_k^{-1} \nabla f_1(\mathbf{x}_k)\right).$$

- ✓ Convergence of the sequence $(\mathbf{x}_k)_{k\in\mathbb{N}}$ to a critical point of F under KL assumption.
- ✓ Robustness to errors in the computation of the proximity operator within the metric.
- ✓ Efficient construction of the preconditioning matrices thanks to the MM framework.

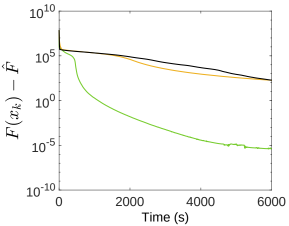
Application to image reconstruction

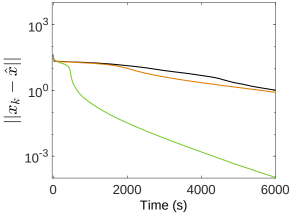


Observation model

$$\mathbf{y} = \mathbf{H}\overline{\mathbf{x}} + \mathbf{w}(\mathbf{H}\overline{\mathbf{x}})$$

with $\mathbf{H}$ Radon projector, and $\mathbf{w}(\mathbf{H}\overline{\mathbf{x}})$ non homogeneous Gaussian noise ($\approx$ Poisson-Gaussian model).
$\rightsquigarrow$ nonconvex data fidelity term.





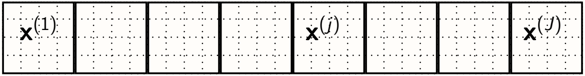
FB - FISTA - FBVM

② - Block alternating and parallel strategies

"Let's shrink Big Data into Small Data ...
and hope it magically becomes Great Data."

Block alternating strategy

The vector of unknowns **x** is partitioned into **block subsets**.
At each iteration, **one** or **several blocks** are updated.

$\mathbf{x} =$ 

Practical advantages:

- ✓ Control of memory for large scale image processing (eg, 3D, video).
- ✓ Flexibility of alternating scheme suitable to blind/unmixing problems.
- ✓ A first step towards parallel and distributed implementation.

How to find efficient and reliable block alternating schemes for nonconvex and/or non differentiable optimization problems ?

?

Block coordinate VMFB algorithm [Chouzenoux *et al.*, 2016]

(PhD Thesis of Audrey Repetti)

⋆ Minimize $F = f_1 + f_2$ with f_1 smooth and f_2 non differentiable.

At each iteration $k \in \mathbb{N}$:

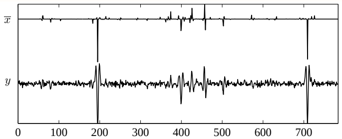
1. Choose a block index $j_k \in \{1, \dots, J\}$ according to a quasi-cyclic rule.
2. Perform a gradient step on the restriction of f_1 to block j_k, using a MM preconditioner.
3. Perform a proximal step on the restriction of f_2 to block j_k, within the MM metric.

✓ CONVERGENCE GUARANTEES on the sequence $(\mathbf{x}_k)_{k \in \mathbb{N}}$ under KL assumption.

✓ EXPERIMENTAL VALIDATION in numerous applications of image/signal processing (eg, phase retrieval, spectral unmixing, blind deconvolution).

Application to seismic data recovery [Repetti *et al.*, 2015]

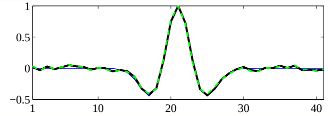
(Joint work with L. Duval, IFPEN)



$\mathbf{y} = \overline{\mathbf{h}} * \overline{\mathbf{x}} + \mathbf{w}$ with $\overline{\mathbf{x}}$ original sparse signal and $\overline{\mathbf{h}}$ unknown filter $\Rightarrow$ blind deconvolution problem.

Application to seismic data recovery [Repetti *et al.*, 2015]

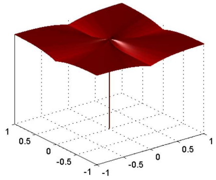
(Joint work with L. Duval, IFPEN)

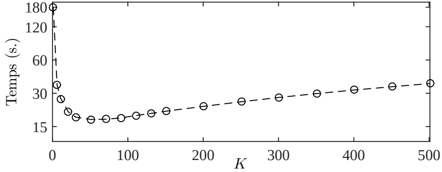


$\mathbf{y} = \overline{\mathbf{h}} * \overline{\mathbf{x}} + \mathbf{w}$ with $\overline{\mathbf{x}}$ original sparse signal and $\overline{\mathbf{h}}$ unknown filter $\Rightarrow$ blind deconvolution problem.

✓ Proposition of a novel nonconvex penalty for sparse signals: smoothed version of the ℓ_1/ℓ_2 prior.

✓ Application of the BC-VMFB algorithm alternating between $\mathbf{x}$ and $\mathbf{h}$.





Dual block alternating FB algorithm

(PhD Thesis of Feriel Abboud)

How to find a fast numerical solution for the computation of proximity operators of composite functions $F = \sum_{j=1}^{J} f_j \circ A_j$?

?

Dual block alternating FB algorithm

(PhD Thesis of Feriel Abboud)

How to find a fast numerical solution for the computation of proximity operators of composite functions $F = \sum_{j=1}^{J} f_j \circ A_j$?

?

$\star$ Apply the block coordinate VMFB to the dual problem $\Leftrightarrow$ dual ascent technique [Abboud *et al.*, 2016]:

- ✓ Acceleration thanks to MM preconditioning strategy.
- ✓ Convergence guarantees on the primal and dual iterates.

Dual block alternating FB algorithm

(PhD Thesis of Feriel Abboud)

How to find a fast numerical solution for the computation of proximity operators of composite functions $F = \sum_{j=1}^{J} f_j \circ A_j$?

?

⋆ Apply the block coordinate VMFB to the dual problem ⇔ dual ascent technique [Abboud *et al.*, 2016]:

- ✓ Acceleration thanks to MM preconditioning strategy.
- ✓ Convergence guarantees on the primal and dual iterates.

⋆ Introduction of a consensus constraint decomposed into hyperedges of a connected hypergraph ⇔ distributed implementation [Abboud *et al.*, 2015]:

- ✓ Suitable to multicore computing architectures.
- ✓ Convergence guarantees on the primal and dual iterates.

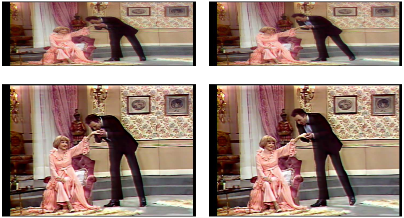
Application to video restoration

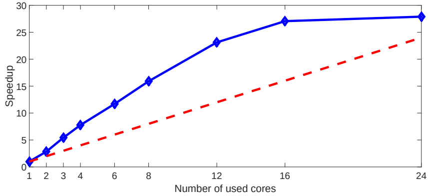
(Joint work with J.-H. Chenot and L. Laborelli, INA)

Observation model

At each frame $t \in \{1, \ldots, T\}$: $\mathbf{y}_t = \mathrm{S}_t(\mathbf{h} * \mathbf{x}_t) + \mathbf{w}_t$

with S_t decimation operator and $\mathbf{h}$ horizontal blur.





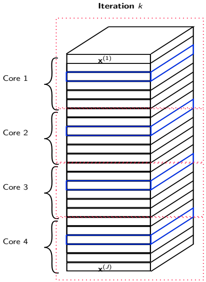
Acceleration using distributed implementation
(joint work with H. Talbot, LIGM)

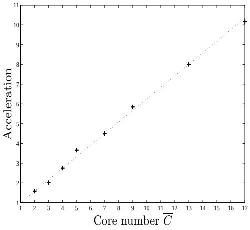
Parallel 3MG algorithm [Cadoni *et al.*, 2016]

How to make 3MG algorithm efficient for parallel implementation ?

At each iteration $k \in \mathbb{N}$:

1. Choose a subset of block indexes $\mathcal{S}_k \subset \{1, \ldots, J\}$.
2. Update the selected blocks using a 3MG step performed in parallel thanks to a block-diagonal MM metric.





- Application to 3D image deblurring with space-variant PSF *(CNRS OPTIMISM project)*.
- SPMD implementation on Matlab Parallel Toolbox.
- Great potential for parallelization.

③ - Stochastic optimization at a large scale

"Why Gramma, what big data you have!"

Problem statement

STOCHASTIC PROBLEM

$$\underset{\mathbf{x}\in\mathbb{R}^N}{\text{minimize}} \quad \left(F(\mathbf{x}) = \frac{1}{2}\mathbb{E}(\|\mathbf{y}_j - \mathbf{h}_j^\top \mathbf{x}\|^2) + \Psi(\mathbf{x}) \right)$$

⋆ The second-order statistics of $(\mathbf{h}_j, \mathbf{y}_j)_{j\geq 1}$ are estimated online in an adaptive manner.

NUMEROUS APPLICATIONS:

- supervised classification
- inverse problems
- system identification
- linear prediction/interpolation
- echo cancellation
- channel equalization

How to find a fast and flexible stochastic optimization algorithm with theoretical convergence guarantees ?

Stochastic 3MG algorithm [Chouzenoux and Pesquet, 2017]

At each iteration $j \in \mathbb{N}^*$:

1. Build an estimate of the objective function:

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad F_j(\mathbf{x}) = \frac{1}{2j} \sum_{k=1}^{j} \|\mathbf{y}_k - \mathbf{h}_k^\top \mathbf{x}\|^2 + \Psi(\mathbf{x})$$

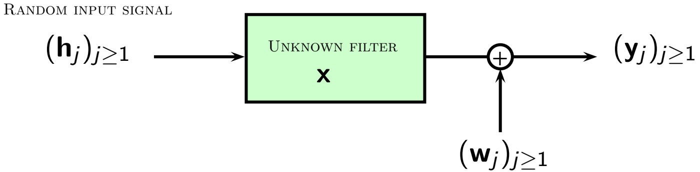
2. Construct a quadratic majorant for F_j.
3. Minimize in a memory gradient subspace.
4. Perform recursive updates of the second-order statistics.

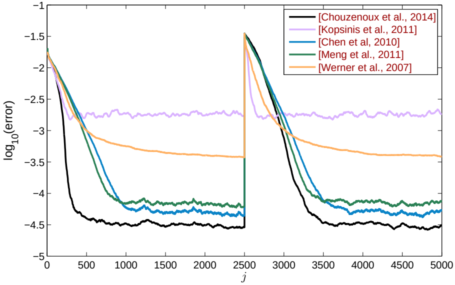
✓ CONVERGENCE GUARANTEES on the sequence $(\mathbf{x}_j)_{j\geq 1}$.

✓ REDUCED COMPLEXITY thanks to recursive update scheme.

✓ CONVERGENCE RATE ANALYSIS in stochastic and batch case ([Chouzenoux and Pesquet, 2016]).

Application to sparse adaptive filtering





- $\mathbf{x}$: sparse linear filter with abrupt change at $j = 2500$.
- S3MG algorithm with forgetting factor and smoothed ℓ_0 penalty.
- Minimal estimation error, and good tracking properties.

④ - Data augmentation in the Bayesian framework

Motivation: Bayesian formulation

Bayes framework

We observe $\mathbf{y} \in \mathbb{R}^N$ according to the model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$. We seek for an entire distribution describing $\boldsymbol{\Theta} = \{\mathbf{x}, \mathbf{H}, \ldots\}$:

$$\underbrace{\mathrm{p}(\boldsymbol{\Theta}|\mathbf{y})}_{\text{posterior}} = \frac{\overbrace{\mathrm{p}(\mathbf{y}|\boldsymbol{\Theta})}^{\text{likelihood}} \; \overbrace{\mathrm{p}(\boldsymbol{\Theta})}^{\text{prior}}}{\underbrace{\int \mathrm{p}(\mathbf{y}|\boldsymbol{\Theta}')\mathrm{p}(\boldsymbol{\Theta}')\, d\boldsymbol{\Theta}'}_{\mathrm{p}(\mathbf{z}):\ \text{marginal density}}}$$

How to find fast and flexible Bayesian algorithms for approximating $\mathrm{p}(\boldsymbol{\Theta}|\mathbf{y})$ in the context of large scale inverse problems ?

?

☛ Take advantages from optimization tools developed in the deterministic framework.

Optimization tool: Half-quadratic strategies

(PhD Thesis of Yosra Marnissi)

Half-quadratic scheme

⋆ For a wide class of cost functions F in inverse problems:

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad F(\mathbf{x}) = \underset{\mathbf{b}\in\mathbb{R}^P}{\operatorname{argmin}}\ \Phi(\mathbf{x}, \mathbf{b})$$

with $\mathbf{x} \mapsto \Phi(\mathbf{x}, \mathbf{b})$ quadratic and $\mathbf{b} \mapsto \Phi(\mathbf{x}, \mathbf{b})$ separable.

☛ Minimize F using an alternating minimization scheme on Φ

➡ Half-quadratic algorithm ⇔ MM quadratic algorithm.

Optimization tool: Half-quadratic strategies

(PhD Thesis of Yosra Marnissi)

Half-quadratic scheme

⋆ For a wide class of cost functions F in inverse problems:

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad F(\mathbf{x}) = \underset{\mathbf{b} \in \mathbb{R}^P}{\text{argmin}} \; \Phi(\mathbf{x}, \mathbf{b})$$

with $\mathbf{x} \mapsto \Phi(\mathbf{x}, \mathbf{b})$ quadratic and $\mathbf{b} \mapsto \Phi(\mathbf{x}, \mathbf{b})$ separable.

☛ Minimize F using an alternating minimization scheme on Φ

➨ Half-quadratic algorithm $\Leftrightarrow$ MM quadratic algorithm.

In the Bayesian setting:

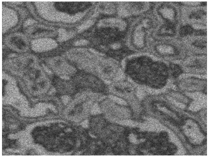
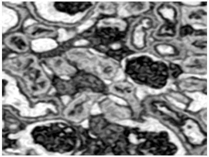
$$\left.\begin{array}{l} \text{Quadratic} \Rightarrow \text{Gaussian statistics} \\ \text{Separable} \Rightarrow \text{Independent statistics} \end{array}\right\} \text{Efficient strategies available}$$

Fast variational Bayesian approach [Marnissi *et al.*, *2017*]

TARGET PARAMETERS: $\boldsymbol{\Theta} = \{\mathbf{x}, \gamma\}$

BAYES VARIATIONAL STRATEGY: Approximate $\mathrm{p}(\boldsymbol{\Theta}|\mathbf{y})$ by a separable density $q(\boldsymbol{\Theta}) = q_X(\mathbf{x})q_\Gamma(\gamma) = \underset{q}{\operatorname{argmin}}\ \mathcal{KL}(q(\boldsymbol{\Theta})\|\mathrm{p}(\boldsymbol{\Theta}|\mathbf{y}))$.

1. Replace $\mathrm{p}(\boldsymbol{\Theta}|\mathbf{y})$ by an augmented function $L(\boldsymbol{\Theta}|\mathbf{y};\boldsymbol{\omega},\boldsymbol{\lambda})$ resulting from half-quadratic construction strategies.
2. Minimize the distance $\mathcal{KL}(q_X(\mathbf{x})q_\Gamma(\gamma)\|L(\boldsymbol{\Theta}|\mathbf{y};\boldsymbol{\omega},\boldsymbol{\lambda}))$ using an alternating scheme on $(q_X(\mathbf{x}), q_\Gamma(\gamma), \boldsymbol{\omega}, \boldsymbol{\lambda})$.

Application to image deblurring with Poisson-Gaussian noise *(ANR GRAPHSIP)*

✓ Flexibility of the half-quadratic construction.

✓ Reduced computational cost.

✓ Automatic determination of the regularization parameter.

Accelerated MH algorithm [Marnissi *et al.*, 2016a]

(Joint work with A. Benazza, SUPCOM Tunis)

⋆ Metropolis-Hastings sampling method to explore $\mathrm{p}(\boldsymbol{\Theta} \mid \mathbf{y})$:

For every iteration $k \in \mathbb{N}$:

1. Generate $\widetilde{\boldsymbol{\Theta}}_k$ from a proposal distribution of density $\mathrm{g}(\cdot|\boldsymbol{\Theta}_k)$.
2. Accept $\boldsymbol{\Theta}_{k+1} = \widetilde{\boldsymbol{\Theta}}_k$ with probability $\min\left(1, \frac{\mathrm{p}(\widetilde{\boldsymbol{\Theta}}_k|\mathbf{y})\mathrm{g}(\widetilde{\boldsymbol{\Theta}}_k|\boldsymbol{\Theta}_k)}{\mathrm{p}(\boldsymbol{\Theta}_k|\mathbf{y})\mathrm{g}(\boldsymbol{\Theta}_k|\widetilde{\boldsymbol{\Theta}}_k)}\right)$.

✗ Slow convergence in the context of large scale problems.

Accelerated MH algorithm [Marnissi *et al.*, 2016a]

(Joint work with A. Benazza, SUPCOM Tunis)

⋆ Metropolis-Hastings sampling method to explore $\mathrm{p}(\boldsymbol{\Theta} \mid \mathbf{y})$:

For every iteration $k \in \mathbb{N}$:

1. Generate $\widetilde{\boldsymbol{\Theta}}_k$ from a proposal distribution of density $\mathrm{g}(\cdot|\boldsymbol{\Theta}_k)$.
2. Accept $\boldsymbol{\Theta}_{k+1} = \widetilde{\boldsymbol{\Theta}}_k$ with probability $\min\left(1, \frac{\mathrm{p}(\widetilde{\boldsymbol{\Theta}}_k|\mathbf{y})\mathrm{g}(\widetilde{\boldsymbol{\Theta}}_k|\boldsymbol{\Theta}_k)}{\mathrm{p}(\boldsymbol{\Theta}_k|\mathbf{y})\mathrm{g}(\boldsymbol{\Theta}_k|\widetilde{\boldsymbol{\Theta}}_k)}\right)$.

☛ **3MH:** Langevin proposal with MM preconditioning

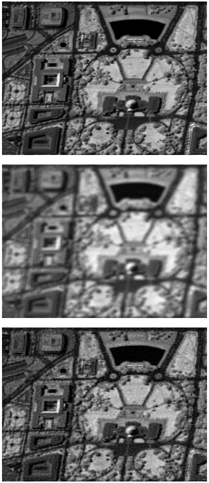
$$\widetilde{\boldsymbol{\Theta}}_k \sim \mathcal{N}\left(\boldsymbol{\Theta}_k + \frac{\epsilon^2}{2}\mathbf{A}(\boldsymbol{\Theta}_k)^{-1}\nabla \log \mathrm{p}(\boldsymbol{\Theta}_k \mid \mathbf{y}), \epsilon^2 \mathbf{A}(\boldsymbol{\Theta}_k)^{-1}\right)$$

☛ Data augmentation to facilitate preconditioning.

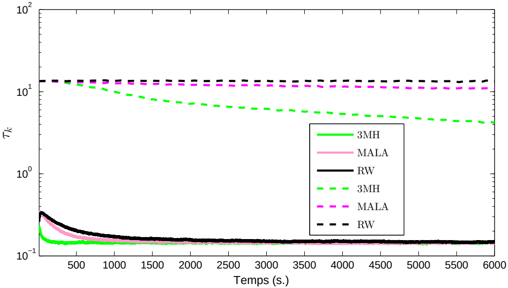
✓ Geometric ergodicity of the generated Markov chain.

✓ Good practical performance on image/signal restoration problems.

Application to image deblurring [Marnissi *et al.*, 2016b]



- Restoration of a multispectral cube degraded by blur and noise.
- GMEP prior on wavelet coefficients to account for cross-component similarities [Marnissi *et al.*, 2013].
- Auxiliary variables to split Fourier / Wavelet transformed domains.



Conclusions and future work

Conclusion

SIGNAL/IMAGE APPLICATIONS

✗ Nonconvex and non smooth cost functions.

✗ Large number of variables.

✗ Limited time or limited accessibility to dataset.

Conclusion

SIGNAL/IMAGE APPLICATIONS

✓ Flexible and robust algorithms which take into account the characteristics of the problems.

✓ Convergence guarantees on the iterates.

✓ Online/parallel/distributed processing.

✗ Nonconvex and non smooth cost functions.

✗ Large number of variables.

✗ Limited time or limited accessibility to dataset.

OPTIMIZATION THEORY

Future works

⋆ **Nonconvex optimization:**

⇝ Efficient resolution of nonlinear inverse problems ?

⇝ Interior points in the nonconvex setting ?

⋆ **Huge scale problems:**

⇝ Efficient online schemes for non quadratic losses ?

⇝ Practical implementation on multicore computers ?

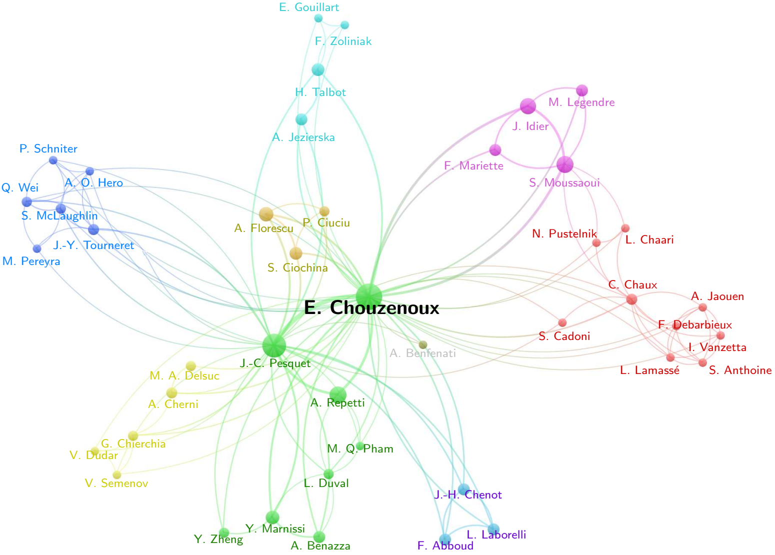
⋆ **Neural networks models :**

⇝ Resolution of complex inverse problems with CNN ?

⇝ Acceleration of back-propagation algorithm ?

☞ ANR JCJC MajIC starting in 2018.
☞ CNRS-Cefipra project (collab. IIIT Delhi).

⋆ Collaboration graph:



PhD students


A. Repetti


Y. Marnissi


F. Abboud


A. Cherni


M.-C. Corbineau


M. Sghaier

Supervision	6 PhD students (3 defended)
	2 Post Docs students
	6 Master students
Dissemination	18 journal papers (14 since PhD)
	40 conference papers (14 invited)
	30 invited seminars
	6 open-source software + 2 web platforms
Grants	ANR JCJC
	Univ. Paris Saclay
	CNRS-Cefipra
	CNRS Mastodons
	GDR ISIS JCJC

THANK YOU !