Majorize-Minimize stepsize strategies for image restoration

Jérôme Idier

jerome.idier@irccyn.ec-nantes.fr

http://www.irccyn.ec-nantes.fr/~idier

Team *Analyse et Décision en Traitement du Signal et de l'Image*

Institut de Recherche en Communication et Cybernétique de Nantes (FRANCE)

CIMI Workshop: Optimization and Statistics in Image Processing, Toulouse, 27 June 2013

General introduction

- Context: linear inverse problems, *e.g.*, image restoration

$$\text{Estimate } \boldsymbol{x} \text{ from } \boldsymbol{y} = \boldsymbol{K}\,\boldsymbol{x} + \boldsymbol{\varepsilon}$$

- Approach: Penalized least-square formulation

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \left(F(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|^2 + \lambda \Phi(\boldsymbol{x}) \right)$$

Φ: edge-preserving penalty function

Addressed issue

Design of provably convergent, maximally fast optimization algorithms inspired by the **majorize-minimize (MM) principle**

Warning

Until slide 28/29, Φ will be assumed **differentiable**

Bases of my talk...

- **Marc Allain**: *Penalized approach in helical tomography. Application to the conception of a personalized knee prosthesis*, PhD thesis, Universite de Paris-Sud / École Polytechnique de Montreal, Dec. 2002, co-advised by **Yves Goussard**
- **Christian Labat**: *Optimization algorithms for penalized criteria in image restoration. Application to spike train deconvolution in ultrasonic imaging*, PhD thesis, Ecole Centrale de Nantes, Dec. 2006
- **Émilie Chouzenoux**: *Majorize-Minimize algorithms for stepsize determination. Application to inverse problems*, PhD thesis, Ecole Centrale de Nantes, Dec. 2010, co-advised by **Saïd Moussaoui**

Plan

1. Introduction to MM algorithms
   - MM principle
   - Historical perspective
   - Two families of MM algorithms
   - Quadratic majorizing approximations
2. MM linesearch for image restoration
   - Application and limits of standard quadratic MM approach
   - Proposed scheme: classical descent direction + MM linesearch
   - Improved scheme based on a subspace approach
   - Criteria involving barrier functions
3. Epilogue: about the differentiability of Φ

Plan

Introduction: Majorize-Minimize (MM) principle

[Hunter and Lange 2004]

MM = *iterative majorization* = *optimization transfer*

Goal: find $\boldsymbol{x}$ that minimizes a function F over $\mathbb{R}^N$

Let H a Majorizing Approximation (MA) of F, *i.e.*,

- $H(\boldsymbol{x}, \boldsymbol{x}_0) \geq F(\boldsymbol{x}), \quad \forall \boldsymbol{x}_0, \boldsymbol{x} \in \mathbb{R}^N$,
- $H(\boldsymbol{x}_0, \boldsymbol{x}_0) = F(\boldsymbol{x}_0), \quad \forall \boldsymbol{x}_0 \in \mathbb{R}^N$

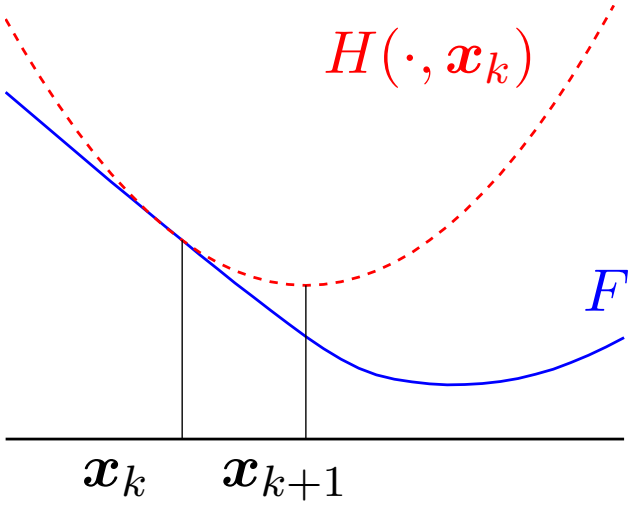
$$\Downarrow$$

$$F(\boldsymbol{x}_1) \leq H(\boldsymbol{x}_1, \boldsymbol{x}_0) \leq F(\boldsymbol{x}_0)$$

with $\boldsymbol{x}_1 = \arg\min_{\boldsymbol{x}} H(\boldsymbol{x}, \boldsymbol{x}_0)$

$$\rightsquigarrow$$

$$\boxed{\text{MM update: } \boldsymbol{x}_{k+1} = \arg\min_{\boldsymbol{x}} H(\boldsymbol{x}, \boldsymbol{x}_k)}$$



A brief historical perspective

- **Fermat-Weber problem** [Weiszfeld 1937]

$$\min_{\boldsymbol{x}} \sum_{m=1}^{M} \|\boldsymbol{x} - \boldsymbol{x}^m\|_2 \quad \rightsquigarrow \quad \text{Weiszfeld algorithm}$$

- **Rediscoveries of MM principle...**
  - EM algorithms [Dempster *et al.* 1977]
  - Robust regression: reweighted least-squares (IRLS, ...) [Beaton and Tukey 1974, Byrd and Payne 1979, Huber 1981]
  - Image restoration: half-quadratic (HQ) algorithms [Geman and Reynolds 1992, Geman and Yang 1995, Charbonnier *et al.* 1997]

Two families of MM algorithms

$\arg\min_{\boldsymbol{x}} H(\boldsymbol{x}, \boldsymbol{x}_0)$ must be easy to compute...

MM based on a *separable* MA function

$$H(\boldsymbol{x}, \boldsymbol{x}_0) = \sum_n h(x_n, \boldsymbol{x}_0) \rightsquigarrow \text{componentwise minimization}$$

- Many EM algorithms such as ISRA [De Pierro 1993]
- Iterative thresholding [Daubechies *et al.* 2004]

MM based on a *quadratic* MA function

Weiszfeld, IRLS, HQ, ...

The Quadratic MA (QMA) case

[Ortega and Rheinboldt 1970, Voss and Eckhardt 1980]

$$H_{\mathbf{B}}(\boldsymbol{x},\boldsymbol{x}_0) = F(\boldsymbol{x}_0) + (\boldsymbol{x}-\boldsymbol{x}_0)^{\mathrm{t}}\,\nabla F(\boldsymbol{x}_0) + (\boldsymbol{x}-\boldsymbol{x}_0)^{\mathrm{t}}\mathbf{B}(\boldsymbol{x}_0)(\boldsymbol{x}-\boldsymbol{x}_0)/2$$

Assume: $\exists\ \mathbf{B}(\cdot) > 0$ such that $H_{\mathbf{B}}(\boldsymbol{x},\boldsymbol{x}_0) \geq F(\boldsymbol{x}), \quad \forall \boldsymbol{x},\boldsymbol{x}_0 \in \mathbb{R}^N$

$$\Downarrow$$

- $H_{\mathbf{B}}$ is a QMA for F,
- $\arg\min_{\boldsymbol{x}} H_{\mathbf{B}}(\boldsymbol{x},\boldsymbol{x}_0) = \boldsymbol{x}_0 - \mathbf{B}^{-1}(\boldsymbol{x}_0)\nabla F(\boldsymbol{x}_0)$

$$\rightsquigarrow$$

MM algorithm: $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \mathbf{B}^{-1}(\boldsymbol{x}_k)\nabla F(\boldsymbol{x}_k)$

Plan

1. Introduction to MM algorithms

2. MM linesearch for image restoration
   - Application and limits of standard quadratic MM approach
   - Proposed scheme: classical descent direction + MM linesearch
   - Improved scheme based on a subspace approach
   - Criteria involving barrier functions

3. Epilogue: about the differentiability of Φ

Linear inverse problems

Estimate $\boldsymbol{x}$ from $\boldsymbol{y} = \boldsymbol{K}\,\boldsymbol{x} + \boldsymbol{\varepsilon}$

- **Example: image restoration (`pepper`)**

$\boldsymbol{y}$

$\boldsymbol{x}$

$N = 512 \times 512$
Gaussian PSF
($\sigma = 2.24$ pixels)
SNR 40dB

Penalized least-square formulation

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} F(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|^2 + \lambda\Phi(\boldsymbol{x})$$

- **Markov-type penalization**

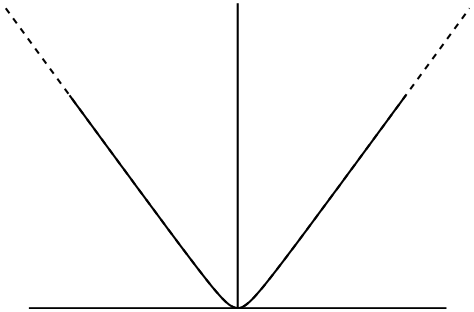
$$\Phi(\boldsymbol{x}) = \sum_{c\in\mathcal{C}} \phi(\boldsymbol{v}_c^{\mathrm{t}}\boldsymbol{x}), \quad \boldsymbol{v}_c^{\mathrm{t}}\boldsymbol{x} = x_r - x_s$$

$c = \{r, s\}$: pairs of horizontal or vertical neighboring pixels

Edge preserving $\rightsquigarrow$ nonquadratic ϕ

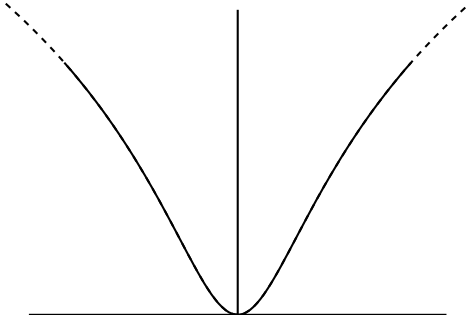
Typical examples of C^1 functions ϕ:

$$\phi(t) = \sqrt{\delta^2 + t^2}$$



[Charbonnier *et al.* 1997]

$$\phi(t) = \log(\delta^2 + t^2)$$



[Hebert and Leahy 1989]

Formulation using an overcomplete signal representation ("synthesis" approach)

$$\widehat{\boldsymbol{x}} = \boldsymbol{W}\widehat{\boldsymbol{z}} \text{ with } \boldsymbol{z} \in \mathbb{R}^L \text{ or } \mathbb{C}^L,\ L \gg N,$$

$$\widehat{\boldsymbol{z}} = \arg\min_{\boldsymbol{z}} \left(\|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{W}\boldsymbol{z}\|^2 + \mu \sum_{\ell=1}^{L} \phi(z_\ell) \right)$$

- $\boldsymbol{W}$ columns form an overcomplete basis of $\mathbb{R}^N$
- $\boldsymbol{z}$: vector of weights
- ϕ: sparsity inducing function, *e.g.*, L_p norm, $p < 2$

Two distinct forms of quadratic MA

Geman & Reynolds

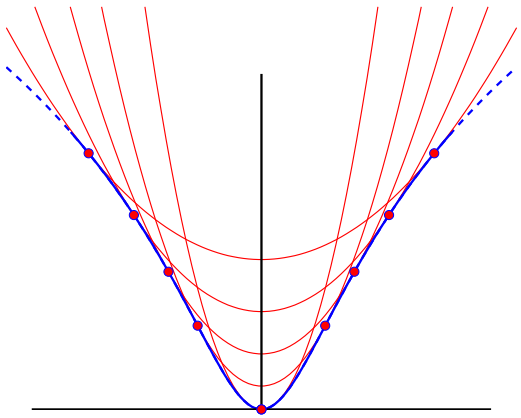
$$\phi\ \mathcal{C}^1,\ \text{even},\ \phi(\sqrt{\cdot})\ \text{concave on}\ \mathbb{R}^+$$

$$\Downarrow$$

$$\exists\psi\ \text{such that}\ \phi(u) = \inf_{b\in\mathbb{R}} \left(bu^2 + \psi(b)\right)$$

$$\Downarrow$$

$$\boldsymbol{B}_{\mathsf{GR}}(\boldsymbol{x}) = 2\boldsymbol{K}^{\mathrm{t}}\boldsymbol{K} + \mathbf{V}^{\mathrm{t}}\boldsymbol{L}(\boldsymbol{x})\mathbf{V},\ \boldsymbol{L}(\boldsymbol{x}) = \mathrm{Diag}\left\{\frac{\phi'(\boldsymbol{v}_c^{\mathrm{t}}\boldsymbol{x})}{\boldsymbol{v}_c^{\mathrm{t}}\boldsymbol{x}}\right\}$$



Geman & Yang

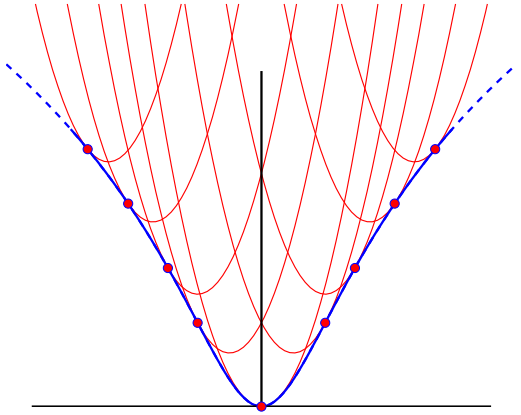
$$\tfrac{1}{2}u^2 - \phi\ \text{convex}$$

$$\Downarrow$$

$$\exists\zeta\ \text{such that}\ \phi(u) = \inf_{b\in\mathbb{R}} \left(\tfrac{1}{2}(u-b)^2 + \zeta(b)\right)$$

$$\Downarrow$$

$$\boldsymbol{B}_{\mathsf{GY}} = 2\boldsymbol{K}^{\mathrm{t}}\boldsymbol{K} + \mathbf{V}^{\mathrm{t}}\mathbf{V},\quad \mathbf{V} = [\boldsymbol{v}_1|...|\boldsymbol{v}_C]^{\mathrm{t}}$$



Convergence of resulting quadratic MM algorithms

[Charbonnier *et al.* 1997, Allain *et al.* 2006]

$$\boldsymbol{B}(\boldsymbol{x}) = \boldsymbol{B}_{\mathsf{GY}} = 2\boldsymbol{K}^{\mathrm{t}}\boldsymbol{K} + \boldsymbol{V}^{\mathrm{t}}\boldsymbol{V} \quad \text{or} \quad \boldsymbol{B}(\boldsymbol{x}) = \boldsymbol{B}_{\mathsf{GR}} = 2\boldsymbol{K}^{\mathrm{t}}\boldsymbol{K} + \boldsymbol{V}^{\mathrm{t}}\boldsymbol{L}(\boldsymbol{x})\boldsymbol{V}$$

$$F \text{ coercive}, \quad \boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \boldsymbol{B}^{-1}(\boldsymbol{x}_k)\nabla F(\boldsymbol{x}_k)$$

$$\Downarrow$$

$$\lim_{k\to\infty} \nabla F(\boldsymbol{x}_k) = \boldsymbol{0}$$

Drawback in image restoration

Large linear system to be solved at each iteration!

$$\rightsquigarrow$$

efficiency $\searrow\searrow$ when $N \nearrow$

Proposed alternate scheme: "classical" descent direction + MM linesearch

Preconditioned conjugate gradient (PCG)

$$\begin{aligned} \boldsymbol{d}_k &= -\mathbf{M}^{-1}\nabla F(\boldsymbol{x}_k) + \beta_k \boldsymbol{d}_{k-1} \\ \boldsymbol{x}_{k+1} &= \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k \end{aligned}$$

$\mathbf{M} > 0$: preconditioning matrix

Stepsize strategy

Minimization of $f(\alpha) = F(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k)$

by I_k sub-iterations of a 1D quadratic MM algorithm

Theoretical convergence [Labat and Idier 2008]

$$\forall I_k \geq 1, \quad \liminf_{k\to\infty} \nabla F(\boldsymbol{x}_k) = \mathbf{0} \text{ (Polak-Ribiere, ...)}$$

Practical behavior

- **faster convergence** than classical stepsize schemes using Wolfe stopping conditions (*e.g.*, Moré-Thuente) About Wolfe
- Best choice is $I_k = 1$!
  $\rightsquigarrow$ no subiteration, α_k is given by a **closed-form formula**

More convergence results

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k, \quad k = 1, \ldots$$

Theorem [Chouzenoux *et al.* 2011]

If $\boldsymbol{d}_k$ is a *gradient related* direction [Bertsekas 1999], then $(\boldsymbol{x}_k)$ converges in the sense $\lim_k \|\nabla F(\boldsymbol{x}_k)\| = 0$, $\forall I_k \geq 1$.

Applicable to:

- Unconstrained case: gradient, Newton, truncated Newton,...
- Constrained case: projected gradient, gradient splitting,...

Improved scheme based on a subspace approach

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \sum_{r=1}^{R} s_{k,r} \boldsymbol{d}_k^r$$

- $[\boldsymbol{d}_k^1, ..., \boldsymbol{d}_k^R] = \boldsymbol{D}_k$: set of directions
  e.g., super-memory gradient (SMG): $\boldsymbol{D}_k = [-\boldsymbol{g}_k, \boldsymbol{d}_{k-1}, .., \boldsymbol{d}_{k-R+1}]$
- $\boldsymbol{s}_k \in \mathbb{R}^R$: multidimensional step
- Proposed stepsize strategy:
  MM algorithm in dimension R to minimize $f(\boldsymbol{s}) = F(\boldsymbol{x}_k + \boldsymbol{D}_k \boldsymbol{s})$

Convergence theorem [Chouzenoux *et al.* 2011]

If $\boldsymbol{D}_k$ contains a gradient related direction, then $(\boldsymbol{x}_k)$ converges in the sense $\lim_k \|\nabla F(\boldsymbol{x}_k)\| = 0$, whatever the number of MM subiterations I_k.

$$\Downarrow$$

Convergence of many subspace algorithms:

SMG, SESOP-TN, PCD-SESOP [Zibulevsky and Elad 2010], ...

Image restoration tests

pepper

noisy blurred image

reconstructed image

- $F(\boldsymbol{x}) = \|\boldsymbol{K}\boldsymbol{x} - \boldsymbol{y}\|^2 + \lambda \sum_c \phi([\boldsymbol{V}\boldsymbol{x}]_c)$, $\phi(u) = \sqrt{\delta^2 + u^2}$
- SMG / non linear conjugate gradient
- Stopping criterion: $\|\boldsymbol{g}_k\|/\sqrt{N} \leqslant 10^{-4}$

Results for SMG

$$\boldsymbol{D}_k = [-\boldsymbol{g}_k, \boldsymbol{d}_{k-1}, ..., \boldsymbol{d}_{k-m}]$$

Iteration number K / Time T before convergence (s.)

SMG(m)		1	2	5	10
MM (J)	1	67/119	68/125	67/140	67/163
	2	66/141	66/147	67/172	67/206
	5	74/211	72/225	71/255	72/323
	10	76/297	74/319	73/394	74/508

- Influence of number of MM subiterations: $J \nearrow \Rightarrow T \nearrow$ and $K \nearrow$
- Influence memory size: $m \nearrow \Rightarrow T \nearrow$ et $K \rightarrow$
- Comparison with SESOP-TN : 76 iterations / 578 s.

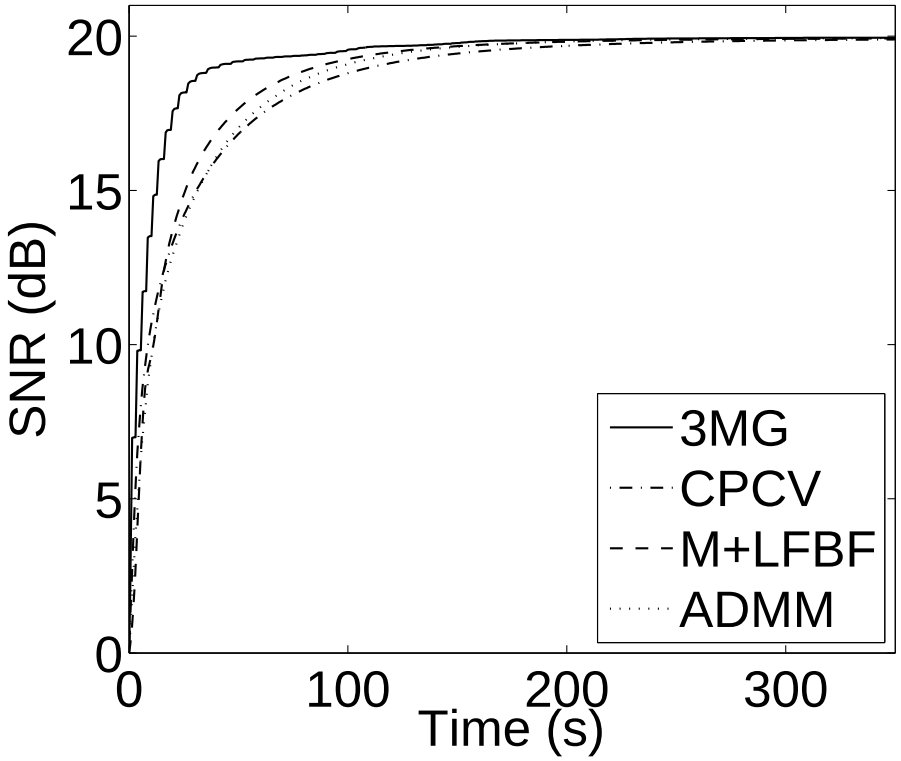
Comparisons between SMG and CG

Iteration number K / Time T before convergence (s.)

	boat	lena	peppers
CG-FR	$77/141$	$98/179$	$145/270$
CG-DY	$86/161$	$127/240$	$234/447$
CG-PRP	$40/74$	$55/99$	$77/137$
CG-HS	$39/71$	$50/93$	$68/122$
CG-LS	$42/81$	$57/103$	$82/149$
SMG(1)	$37/67$	$47/85$	$67/119$

Comparisons with state-of-the-art approaches

Context: parallel magnetic resonance imaging [Florescu *et al.* 2013]



3MG: proposed MM-SMG
CPCV: Chambolle-Pock algorithm [Chambolle and Pock 2011]
M+LFBF: Primal-dual splitting [Combettes and Pesquet 2012]
ADMM: Alternating-Direction Method of Multipliers [Afonso *et al.* 2011]

Criteria involving barrier functions

Definition

A strictly convex function B is a **barrier** corresponding to $\boldsymbol{x} \in \mathcal{C}$ if $\|\nabla B\|$ is unbounded at the frontier of $\mathcal{C}$

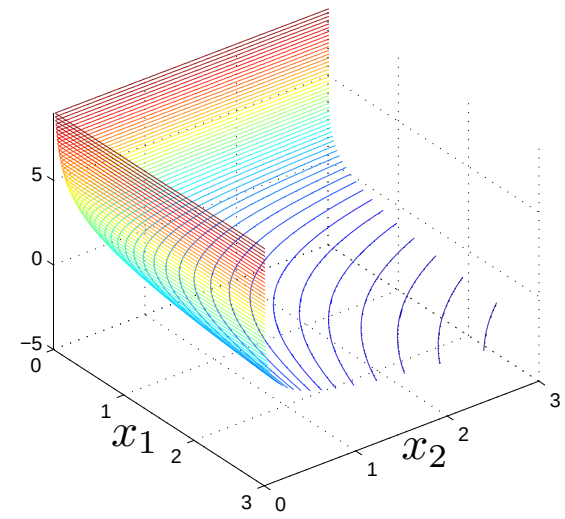
Examples

1. *Poisson noise (e.g., emission tomography)*

$$F(\boldsymbol{x}) = \sum_i [\boldsymbol{K}\boldsymbol{x}]_i - y_i \log[\boldsymbol{K}\boldsymbol{x}]_i$$

2. *Maximum entropy penalty*

$$F(\boldsymbol{x}) = \|\boldsymbol{K}\boldsymbol{x} - \boldsymbol{y}\|^2 + \lambda \sum_n x_n \log x_n$$



$B(x) = -\log(x_1) - \log(x_2)$

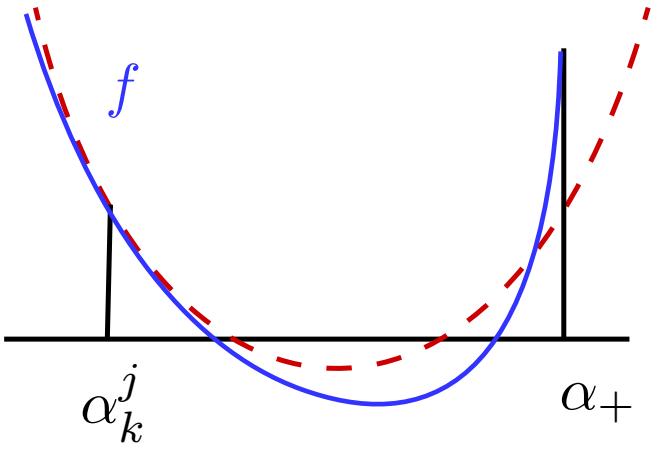
Proposed approach

Optimization problems of the form:

$$\min_{\boldsymbol{x}} \left(F(\boldsymbol{x}) = P(\boldsymbol{x}) + \mu B(\boldsymbol{x})\right), \quad \mu > 0$$

- $P(\boldsymbol{x})$: Possibility to build a QMA
- $B(\boldsymbol{x}) = \sum_{i=1}^{I} b_i(\boldsymbol{c}_i^T \boldsymbol{x} + \rho_i)$: barrier function,
  e.g., $b_i(u) = -\log u$ or $u \log u$

There is no QMA for
$f(\alpha) = F(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k)$



1D nonquadratic MM stepsize

Majorizing approximations of the form

$$\boxed{h(\alpha, \alpha_k^j) = h_0 + h_1\alpha + h_2\alpha^2 - h_3 \log(h_4 - \alpha)}$$

- $\arg\min_\alpha h(\alpha, \alpha_k^j)$ is a root of a degree 2 polynomial
- Available for barriers $-\log u$ and $u \log u$

Proposed optimization scheme [Chouzenoux *et al.* 2012; 2010]

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k, \quad \text{for} \quad k = 1, \ldots$$

- $\boldsymbol{d}_k$ "standard" descent direction (conjugate gradient, ...)
- stepsize strategy: **1D nonquadratic MM** algorithm

Convergence results + practical efficiency: see [Chouzenoux *et al.* 2012]

Conclusion

MM strategy in small dimension gives very efficient stepsize schemes:

- simplicity (no nested subiterations)
- mathematical convergence
- practical efficiency

- Adaptation of **quadratic MM** algorithms to large scale problems
- Proposition of a **nonquadratic MM** stepsize rule to deal with barrier optimization

Plan

Φ differentiable or not? (for better solutions to linear inverse problems)

- **Reconstruction using total variation**

Nonsmoothness of Φ $\rightsquigarrow$ *staircasing effect* [Nikolova 2002]

- **Reconstruction in an overcomplete wavelet dictionnary**

Restoration of $\boldsymbol{x}$ from $\boldsymbol{y} = \boldsymbol{K}\boldsymbol{x} + \boldsymbol{\epsilon}$

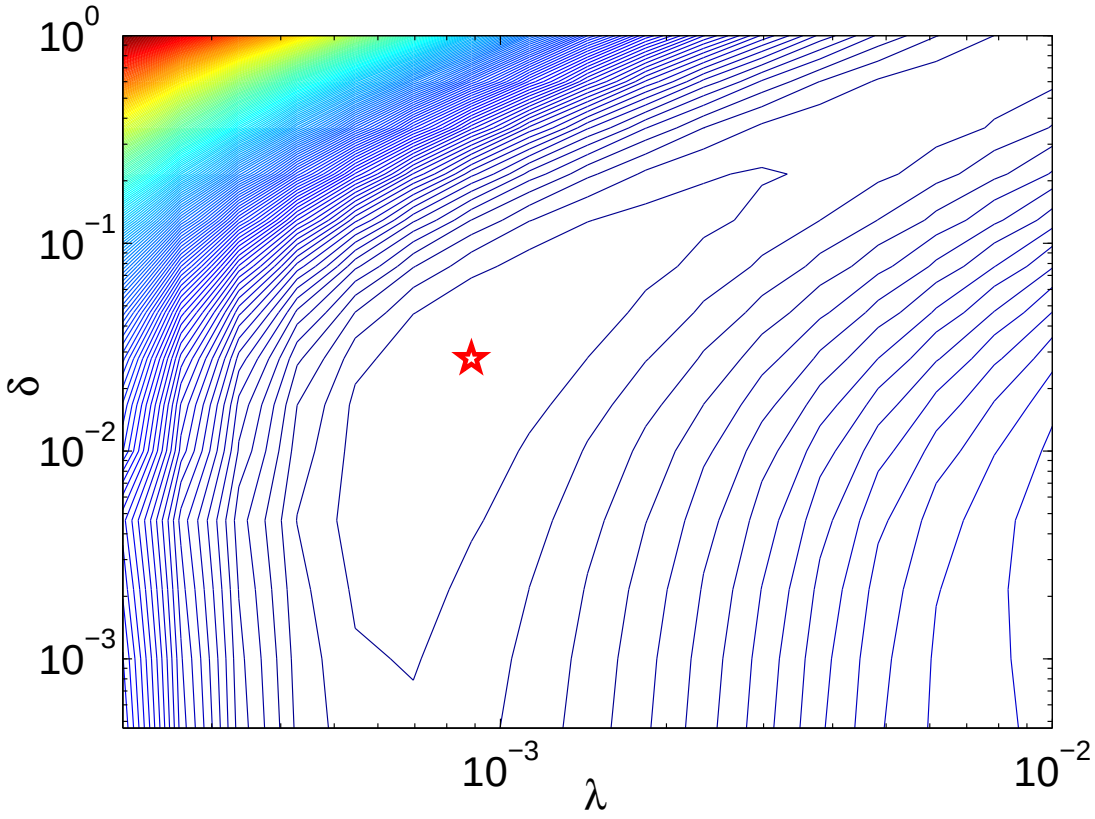
$$\widehat{\boldsymbol{z}} = \arg\min_{\boldsymbol{z}} \left(F(\boldsymbol{z}) = \|\boldsymbol{K}\boldsymbol{W}\boldsymbol{z} - \boldsymbol{y}\|_2^2 + \lambda \sum_{\ell} \phi(z_\ell) \right)$$

with $\phi(u) = |u| - \delta \log(1 + |u|/\delta)$

Question: What is the best value for δ ?

$\delta = 0^+ \rightsquigarrow \phi \equiv L_1 \rightsquigarrow \widehat{\boldsymbol{z}}$ **sparse**?

Deconvolution of `peppers` (128×128)
Minimization of F using *iterative thresholding*



Quadratic restoration error

Sparsity is not the best option!

References I

[Afonso *et al.* 2011] M. Afonso, J. Bioucas-Dias and M. A. T. Figueiredo. An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Trans. Image Process.*, 20 (3): 681–695, 2011.

[Allain *et al.* 2006] M. Allain, J. Idier and Y. Goussard. On global and local convergence of half-quadratic algorithms. *IEEE Trans. Image Process.*, 15 (5): 1130–1142, May 2006.

[Beaton and Tukey 1974] A. E. Beaton and J. W. Tukey. The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*, 16: 147–185, 1974.

[Bertsekas 1999] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, MA, 2nd edition, 1999.

[Byrd and Payne 1979] R. H. Byrd and D. A. Payne. Convergence of the iteratively reweighted least squares algorithm for robust regression. Rapport Interne 313, The Johns Hopkins Univ., Baltimore, MD, June 1979.

[Chambolle and Pock 2011] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imag. Vision*, 40 (1): 120–145, 2011.

[Charbonnier *et al.* 1997] P. Charbonnier, L. Blanc-Féraud, G. Aubert and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Trans. Image Process.*, 6 (2): 298–311, Feb. 1997.

[Chouzenoux *et al.* 2011] E. Chouzenoux, J. Idier and S. Moussaoui. A majorize-minimize strategy for subspace optimization applied to image restoration. *IEEE Trans. Image Process.*, 20 (6): 1517–1528, July 2011.

References II

[Chouzenoux *et al.* 2012] E. Chouzenoux, S. Moussaoui and J. Idier. Majorize-minimize linesearch for inversion methods involving barrier function optimization. *Inverse Probl.*, 28: 065011 (24 pages), Oct. 2012.

[Chouzenoux *et al.* 2010] E. Chouzenoux, S. Moussaoui, J. Idier and F. Mariette. Efficient maximum entropy reconstruction of nuclear magnetic resonance T1-T2 spectra. *IEEE Trans. Signal Process.*, 58 (12): 6040–6051, Dec. 2010.

[Combettes and Pesquet 2012] P. L. Combettes and J.-C. Pesquet. Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators. *Set-Valued and Variational Analysis*, 20 (2): 307–330, 2012.

[Daubechies *et al.* 2004] I. Daubechies, M. Defrise and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.*, 57 (11): 1413–1457, 2004.

[De Pierro 1993] A. De Pierro. On the relation between the ISRA and the EM algorithm for positron emission tomography. *IEEE Trans. Med. Imag.*, 12 (2): 328–333, 1993.

[Dempster *et al.* 1977] A. P. Dempster, N. M. Laird and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. B*, 39: 1–38, 1977.

[Florescu *et al.* 2013] A. Florescu, E. Chouzenoux, J.-C. Pesquet, P. Ciuciu and S. Ciochina. A complex-valued majorize-minimize memory gradient method with application to parallel MRI. In *EUSIPCO*, submitted, 2013.

References III

[Fuchs 2007] J.-J. Fuchs. Convergence of a sparse representations algorithm applicable to real or complex data. *IEEE J. Sel. Topics Signal Process.*, 1 (4): 598–605, Dec. 2007. Issue: Convex Optimization Methods for Signal Processing.

[Geman and Reynolds 1992] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14 (3): 367–383, Mar. 1992.

[Geman and Yang 1995] D. Geman and C. Yang. Nonlinear image recovery with half-quadratic regularization. *IEEE Trans. Image Process.*, 4 (7): 932–946, July 1995.

[Hebert and Leahy 1989] T. Hebert and R. Leahy. A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors. *IEEE Trans. Med. Imag.*, 8 (2): 194–202, June 1989.

[Huber 1981] P. J. Huber. *Robust Statistics*. John Wiley, New York, NY, 1981.

[Hunter and Lange 2004] D. R. Hunter and K. Lange. A tutorial on MM algorithms. *Amer. Statist.*, 58 (1): 30–37, Feb. 2004.

[Kuhn 1973] H. W. Kuhn. A note on Fermat's problem. *Math. Program.*, 4: 98–107, 1973.

[Labat and Idier 2008] C. Labat and J. Idier. Convergence of conjugate gradient methods with a closed-form stepsize formula. *J. Optim. Theory Appl.*, 136 (1): 43–60, Jan. 2008.

[Nikolova 2002] M. Nikolova. Minimizers of cost-functions involving non-smooth data-fidelity terms. Application to the processing of outliers. *SIAM J. Num. Anal.*, 40 (3): 965–994, 2002.

References IV

[Ortega and Rheinboldt 1970] J. Ortega and W. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York, NY, 1970.

[Voss and Eckhardt 1980] H. Voss and U. Eckhardt. Linear Convergence of Generalized Weiszfeld's Method. *Computing*, 25: 243–251, 1980.

[Weiszfeld 1937] E. Weiszfeld. Sur le point pour lequel la somme des distances de n points donnés est minimum. *Tôhoku Mathematical Journal*, 43: 355–386, 1937.

[Zibulevsky and Elad 2010] M. Zibulevsky and M. Elad. $\ell_2 - \ell_1$ optimization in signal and image processing. *IEEE Sig. Proc. Mag.*, 27 (3): 76–88, May 2010.

Usual stepsize strategies

Scalar case

Minimization of f by dichotomy/interpolation until sufficient conditions are met, *e.g.*, Wolfe conditions of parametres c_1, $c_2 \in]0,1[$:

$$F(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k) \leq F(\boldsymbol{x}_k) + c_1 \alpha_k \boldsymbol{g}_k^T \boldsymbol{d}_k$$
$$\nabla F(\boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k)^T \boldsymbol{d}_k \geq c_2 \boldsymbol{g}_k^T \boldsymbol{d}_k$$

Multidimensional case

No natural extension of Wolfe conditions to several dimensions

In pratice: iterative minimization of $f(\boldsymbol{s})$ (*e.g.*, Newton [Zibulevsky and Elad 2010])

Back

Quadratic MM linesearch

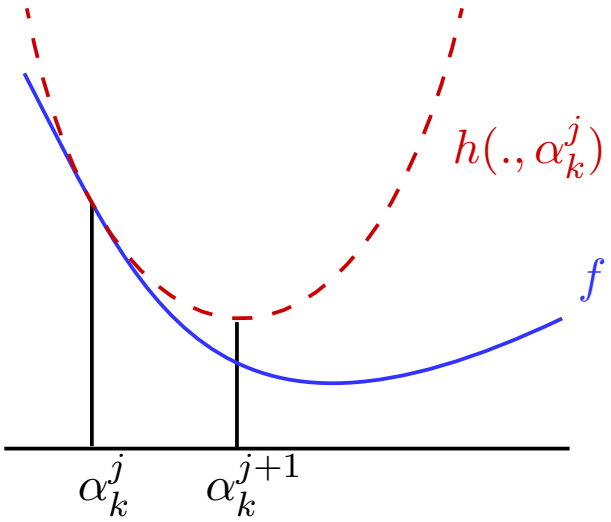
Goal: build a minimizing sequence (α_k) for $f(\alpha) = F(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k)$

For all β, let $h(., \beta)$ a QMA of f:

$$h(\alpha, \beta) = f(\beta) + (\alpha - \beta)\dot{f}(\beta) + \frac{1}{2} m_\beta (\alpha - \beta)^2$$

MM linesearch:

$$\begin{aligned} \alpha_k^{j+1} &= \arg\min_\alpha h(\alpha, \alpha_k^j),\ j < J \\ &= \alpha_k^j - \dot{f}(\alpha_k^j)/m_k^j,\ j < J \\ \alpha_k &= \alpha_k^J \end{aligned}$$



Quadratic MM applied to L_1 cases

- **Quadratic majorizing approximation (GR type)**

$$\phi(x) = |x| \leq \gamma(x, x_0) = \frac{x^2}{2\,|x_0|} + \frac{|x_0|}{2}$$

- **Proved convergence in specific cases**
  - Fermat-Weber problem (Weiszfeld algorithm) [Kuhn 1973]
  - $\|\boldsymbol{y} - \boldsymbol{K}\boldsymbol{x}\|^2 + \lambda\,\|\boldsymbol{x}\|_1$ [Fuchs 2007]

- **Conjectured convergence in a wider case including total variation (ongoing work)**

- **Iteration cost is prohibitive for large N!**

Perspective

iterative thresholding as a descent direction + **1D MM stepsize**