Variable Metric Forward Backward algorithm

Image reconstruction corrupted by a dependent Gaussian noise

A. Repetti
in collaboration with E. Chouzenoux and J.-C. Pesquet

Laboratoire d'Informatique Gaspard Monge - UMR CNRS 8049, Université Paris-Est, France.

GRETSI 2013 - Brest

Outline

1. Introduction
2. Theoretical background
3. Variable Metric Forward-Backward Algorithm
   - State of the art
   - Proposed algorithm
   - Convergence results
4. Application to image reconstruction
   - Signal-dependent Gaussian noise
   - Results

Context

Image Reconstruction

- We observe data $z \in \mathbb{R}^M$, related to the original image $\bar{x} \in \mathbb{R}^N$ through $z = H\bar{x} + w$, where
  - $H \in \mathbb{R}^{M \times N}$ observation matrix,
  - $w \in \mathbb{R}^M$ additive noise.
- Objective: Produce an estimate $\hat{x} \in \mathbb{R}^N$ of the target image $\bar{x}$ from the observed data z.

$\bar{x}$

z

$\hat{x}$

Minimization problem

Penalized optimization problem

$$\text{Find} \quad \hat{x} \in \operatorname{Argmin}\{G = F + R\}, \tag{1}$$

$F \rightsquigarrow$ Data fidelity term, related to noise

$R \rightsquigarrow$ Regularization term, related to some *a priori* informations

- F has an L-Lipschitz gradient on dom R, i.e.
$$\big(\forall (x,y) \in (\text{dom } R)^2\big) \quad \|\nabla F(x) - \nabla F(y)\| \leq L\|x-y\|,$$
- $R\colon \mathbb{R}^N \to (-\infty, +\infty]$ proper, lsc, convex and continuous on its domain,
- G is coercive, i.e. $\lim_{\|x\|\to+\infty} G(x) = +\infty$.

$$\Rightarrow \operatorname{Argmin} G \neq \varnothing$$

Proximity operator

Proximity operator

Let $\psi\colon \mathbb{R}^N \to (-\infty, +\infty]$ proper, lsc, convex. Let $x \in \mathbb{R}^N$.

$$\operatorname{prox}_\psi(x) = \underset{y\in\mathbb{R}^N}{\operatorname{argmin}}\ \psi(y) + \frac{1}{2}\|y - x\|^2.$$

$\rightsquigarrow$ Characterization: $p = \operatorname{prox}_\psi(x) \Leftrightarrow x - p \in \partial\psi(p)$.

Particular case:

Let $\mathcal{C} \subset \mathbb{R}^N$ be a convex set, $x \in \mathbb{R}^N$, $\iota_\mathcal{C}(x) = \begin{cases} 0 & \text{if } x \in \mathcal{C} \\ +\infty & \text{otherwise} \end{cases}$.

$$\psi(x) = \iota_\mathcal{C}(x) \Rightarrow \operatorname{prox}_\psi = \mathrm{P}_\mathcal{C}.$$

Proximity operator relative to a metric

Let $U \in \mathbb{R}^{N\times N}$ be a symmetric positive definite matrix. Let $x \in \mathbb{R}^N$.

- Weighted norm: $\|x\|_U = \left(x^\top U x\right)^{1/2}$.
- Loewner partial ordering on $\mathbb{R}^{N\times N}$ $(\forall (U_1, U_2) \in (\mathbb{R}^{N\times N})^2)$

$$U_1 \succeq U_2 \quad \Leftrightarrow \quad x^\top U_1 x \geq x^\top U_2 x.$$

Proximity operator relative to the metric induced by U

Let $\psi \colon \mathbb{R}^N \to (-\infty, +\infty]$ proper, lsc, convex.

$$\mathrm{prox}_{U,\psi}(x) = \underset{y\in\mathbb{R}^N}{\mathrm{argmin}}\ \psi(y) + \frac{1}{2}\|y - x\|_U^2.$$

$\rightsquigarrow$ $\mathrm{prox}_{\mathrm{I_N},\psi} = \mathrm{prox}_\psi$.

$\rightsquigarrow$ Characterization: $p = \mathrm{prox}_{U,\psi}(x) \Leftrightarrow U(x - p) \in \partial\psi(p)$.

Forward-Backward algorithm

FB Algorithm

$$
\begin{array}{l}
x_0 \in \mathbb{R}^N \\
\text{For } \; k = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\bar{y}_k = x_k - \gamma_k \nabla F(x_k), \\
y_k = \mathrm{prox}_{\gamma_k R}(\bar{y}_k), \\
x_{k+1} = x_k + \lambda_k (y_k - x_k),
\end{array}
\right.
\end{array}
$$

Convergence is established if:

- [Combettes and Pesquet - 2007]
  - $\rightsquigarrow$ F convex with L-Lipschitzian gradient, R convex lsc proper,
  - $\rightsquigarrow$ $(\gamma_k)_{k \in \mathbb{N}}$ and $(\lambda_k)_{k \in \mathbb{N}}$ bounded.
- [Attouch, Bolte and Svaiter - 2011]
  - $\rightsquigarrow$ F and R non convex, F with L-Lipschitzian gradient,
  - $\rightsquigarrow$ $\lambda_k \equiv 1$ and $(\gamma_k)_{k \in \mathbb{N}}$ bounded.

Variable Metric Forward-Backward algorithm

VMFB Algorithm

$$
\begin{array}{l}
x_0 \in \mathbb{R}^N \\
\text{For} \quad k = 0, 1, \ldots \\
\left\lfloor \begin{array}{l}
\bar{y}_k = x_k - \gamma_k A_k^{-1} \nabla F(x_k), \\
y_k = \operatorname{prox}_{\gamma_k^{-1} A_k, R}(\bar{y}_k), \\
x_{k+1} = x_k + \lambda_k (y_k - x_k),
\end{array} \right.
\end{array}
$$

Convergence is established if:

- [Combettes and Vũ - 2013]
  - ⇝ F convex with L-Lipschitzian gradient, R convex lsc proper,
  - ⇝ $(\gamma_k)_{k\in\mathbb{N}}$ and $(\lambda_k)_{k\in\mathbb{N}}$ bounded,
  - ⇝ $\exists (\eta_k)_{k\in\mathbb{N}} \in \ell_1^+(\mathbb{N})$, such that $(\forall k \in \mathbb{N})$ $(1+\eta_k)A_{k+1} \succeq A_k$,
  - ⇝ $\exists (\underline{\nu}, \overline{\nu}) \in (0, +\infty)^2$ such that $(\forall k \in \mathbb{N})$ $\underline{\nu}\, \mathrm{I}_N \preceq A_k \preceq \overline{\nu}\, \mathrm{I}_N$.
- Non convex case ?

Our contribution [Chouzenoux *et al.* - 2013]

- Convergence of the VMFB algorithm for F non convex ?
  $\rightsquigarrow$ Kurdyka-Łojasiewicz Inequality.
- Choice of variable metric $(A_k)_{k\in\mathbb{N}}$?
  $\rightsquigarrow$ Majorize-Minimize principle.
- Calculation of the proximity operator ?
  $\rightsquigarrow$ Inexact VMFB algorithm.

Majorize-Minimize assumption

MM Assumption

- For every $k \in \mathbb{N}$, there exists a symmetric positive definite matrix $A_k \in \mathbb{R}^{N \times N}$ such that for every $x \in \mathbb{R}^N$

$$Q(x, x_k) = F(x_k) + (x - x_k)^\top \nabla F(x_k) + \frac{1}{2}(x - x_k)^\top A_k (x - x_k),$$

  is a *majorant function* of F at x_k on $\operatorname{dom} R$, i.e.,

$$F(x_k) = Q(x_k, x_k) \quad \text{and} \quad (\forall x \in \operatorname{dom} R) \quad F(x) \leq Q(x, x_k).$$

- There exists $(\underline{\nu}, \overline{\nu}) \in (0, +\infty)^2$ such that $(\forall k \in \mathbb{N})$ $\underline{\nu}\, \mathrm{I}_N \preceq A_k \preceq \overline{\nu}\, \mathrm{I}_N$.

F is differentiable with an L-Lipschitzian gradient on $\operatorname{dom} R$	$\Rightarrow$	$A_k \equiv L\, \mathrm{I}_\mathrm{N}$ satisfies the above assumption [Bertsekas - 1999]

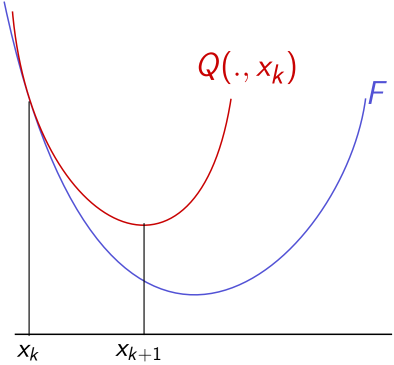
Majorize-Minimize algorithm [Jacobson and Fessler - 2007]

MM Algorithm

$$x_{k+1} \in \underset{x}{\mathrm{Argmin}}\, Q(x, x_k)$$

$\Leftrightarrow$ VMFB Algorithm with

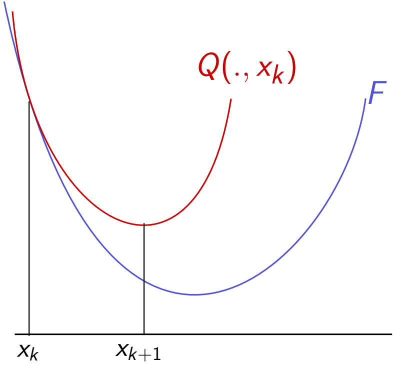
- $R \equiv 0$
- $\lambda_k \equiv 1$
- $\gamma_k \equiv 1$



Majorize-Minimize algorithm [Jacobson and Fessler - 2007]

MM Algorithm

$$x_{k+1} \in \underset{x}{\operatorname{Argmin}}\, Q(x, x_k) + R(x)$$



$\Leftrightarrow$ VMFB Algorithm with

- $\lambda_k \equiv 1$
- $\gamma_k \equiv 1$

Proposed algorithm

VMFB Algorithm

$x_0 \in \operatorname{dom} R$

For $k = 0, 1, \ldots$

$$
\left\lfloor
\begin{array}{l}
\bar{y}_k = x_k - \gamma_k A_k^{-1} \nabla F(x_k), \\
y_k = \operatorname{prox}_{\gamma_k^{-1} A_k, R}(\bar{y}_k), \\
x_{k+1} = (1 - \lambda_k) x_k + \lambda_k y_k,
\end{array}
\right.
$$

where

- $\exists (\underline{\eta}, \overline{\eta}) \in (0, +\infty)^2$ such that $(\forall k \in \mathbb{N})$ $\underline{\eta} \leq \gamma_k \lambda_k \leq 2 - \overline{\eta}$.
- $\exists \underline{\lambda} \in (0, +\infty)$ such that $(\forall k \in \mathbb{N})$ $\underline{\lambda} \leq \lambda_k \leq 1$.

Proposed algorithm

Inexact VMFB Algorithm

$x_0 \in \operatorname{dom} R, \tau \in (0, +\infty)$
For $k = 0, 1, \ldots$
Find $y_k \in \mathbb{R}^N$ and $r(y_k) \in \partial R(y_k)$ such that
$$R(y_k) + (y_k - x_k)^\top \nabla F(x_k) + \gamma_k^{-1} \|y_k - x_k\|_{A_k}^2 \le R(x_k),$$
$$\|\nabla F(x_k) + r(y_k)\| \le \tau \|y_k - x_k\|_{A_k},$$
$x_{k+1} = (1 - \lambda_k) x_k + \lambda_k y_k,$

where

- $\exists (\underline{\eta}, \overline{\eta}) \in (0, +\infty)^2$ such that $(\forall k \in \mathbb{N})$ $\underline{\eta} \le \gamma_k \lambda_k \le 2 - \overline{\eta}$.
- $\exists \underline{\lambda} \in (0, +\infty)$ such that $(\forall k \in \mathbb{N})$ $\underline{\lambda} \le \lambda_k \le 1$.

Inexact proximal step

$$\begin{cases} y_k = \operatorname{prox}_{\gamma_k^{-1} A_k, R}(x_k - \gamma_k A_k^{-1} \nabla F(x_k)) \\ \text{Convexity of } R \end{cases}$$

$$\Leftrightarrow (\exists r(y_k) \in \partial R(y_k)) \quad \begin{cases} r(y_k) = -\nabla F(x_k) + \gamma_k^{-1} A_k (x_k - y_k) \\ (y_k - x_k)^\top r(y_k) \geq R(y_k) - R(x_k). \end{cases}$$

$$\Rightarrow \begin{cases} R(y_k) + (y_k - x_k)^\top \nabla F(x_k) + \gamma_k^{-1} \|y_k - x_k\|_{A_k}^2 \leq R(x_k), \\ \|\nabla F(x_k) + r(y_k)\| = \gamma_k^{-1} \|A_k (y_k - x_k)\| \leq \gamma_k^{-1} \sqrt{\overline{\nu}} \|y_k - x_k\|_{A_k} \\ \qquad \leq \underline{\eta}^{-1} \sqrt{\overline{\nu}} \|y_k - x_k\|_{A_k} \\ \qquad \rightsquigarrow \tau = \underline{\eta}^{-1} \sqrt{\overline{\nu}} \end{cases}$$

Assumptions

- R proper lsc convex and continuous on dom R, F differentiable, ∇F L-Lipschitz on dom R, G is coercive.

- G satisfies the *Kurdyka-Łojasiewicz* inequality:
For every $\xi \in \mathbb{R}$, for every bounded $E \subset \mathbb{R}^N$, there exist $\kappa, \zeta > 0$ and $\theta \in [0, 1)$ such that, for every $x \in E$ such that $|G(x) - \xi| \leq \zeta$,
$$(\forall r(x) \in \partial R(x)) \qquad \|\nabla F(x) + r(x)\| \geq \kappa |G(x) - \xi|^{\theta}.$$

- $(A_k)_{k\in\mathbb{N}}$ satisfies the *majorization conditions*.

- $(\lambda_k)_{k\in\mathbb{N}}$ and $(\gamma_k)_{k\in\mathbb{N}}$ bounded.

- Decreasing assumption: There exists $\underline{\alpha} \in (0, 1]$ such that
$$(\forall k \in \mathbb{N}) \quad G(x_{k+1}) \leq (1 - \underline{\alpha}) G(x_k) + \underline{\alpha} G(y_k).$$

Convergence results

Descent Property

There exists $\mu \in (0, +\infty)$, such that

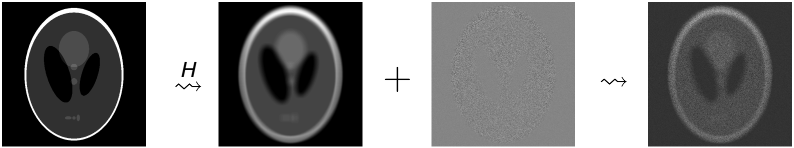
$$(\forall k \in \mathbb{N}) \quad G(x_{k+1}) \leq G(x_k) - \frac{\mu}{2}\|x_{k+1} - x_k\|^2.$$

Convergence theorem

Let $(x_k)_{k\in\mathbb{N}}$ and $(y_k)_{k\in\mathbb{N}}$ be sequences generated by the (inexact) VMFB algorithm.

- Global convergence:
  $\rightsquigarrow$ $\lim x_k = \lim y_k = \hat{x}$, where $\hat{x}$ is a critical point of G.
  $\rightsquigarrow$ $\lim G(x_k) = \lim G(y_k) = G(\hat{x})$.
- Local convergence:
  If $(\exists \upsilon > 0)$ such that $G(x_0) \leq \inf_{x\in\mathbb{R}^N} G(x) + \upsilon$,
  then $\lim x_k = \lim y_k = \hat{x}$, where $\hat{x}$ is a solution to Problem (1).

Image reconstruction under signal-dependent noise



$\overline{x} \in [0,+\infty)^N$ $\quad H\overline{x}$ $\quad w(H\overline{x})$ $\quad z \in [0,+\infty)^M$

- **Observation matrix:** $H \in [0,+\infty)^{M\times N}$.
- **Signal-dependent noise:** $w(H\bar{x}) = \left(w^{(m)}([H\bar{x}]^{(m)})\right)_{1\leq m\leq M}$, with $(\forall m \in \{1,\dots,M\})$
  - $a^{(m)} \in [0,+\infty)$, $b^{(m)} \in (0,+\infty)$,
  - $w^{(m)}([H\bar{x}]^{(m)})$ realization of $W^{(m)} \sim \mathcal{N}\left(0, a^{(m)}[H\bar{x}]^{(m)} + b^{(m)}\right)$.

OBJECTIVE: Produce an estimate $\hat{x} \in [0,+\infty)^N$ of the target image $\bar{x}$ from the observed data z.

Optimization problem

Solve Problem (1): Find $\hat{x} \in \text{Argmin}\{G = F + R\}$ where

Data fidelity term (neg-log-likelihood of the data)

$$F(x) = \begin{cases} F_1(x) + F_2(x) & \text{if } x \in [0, +\infty)^N \\ +\infty & \text{otherwise} \end{cases}, \text{ where}$$

- $F_1(x) = \frac{1}{2} \sum_{m=1}^{M} \frac{([Hx]^{(m)} - z^{(m)})^2}{a^{(m)}[Hx]^{(m)} + b^{(m)}} \rightsquigarrow$ Convex function.
- $F_2(x) = \frac{1}{2} \sum_{m=1}^{M} \log\left(a^{(m)}[Hx]^{(m)} + b^{(m)}\right) \rightsquigarrow$ Concave function.

Penalization term

$(\forall x \in \mathbb{R}^N)\ R(x) = R_1(x) + R_2(x)$, where

- $R_1 \rightsquigarrow \iota_{[x_{min}, x_{max}]^N}(x) = \begin{cases} 0 & \text{if } x \in [x_{min}, x_{max}]^N \\ +\infty & \text{otherwise} \end{cases}.$
- $R_2 \rightsquigarrow$ Sparsity prior in analysis frame or Non Local Total Variation.

Majorization strategy for F_1

$F_1 \rightsquigarrow$ Convex and additive separable function.

$(\forall x \in [0,+\infty)^N)\ F_1(x) = \sum_{m=1}^{M} \rho_1^{(m)}([Hx]^{(m)})$,

where $(\forall m \in \{1,\dots,M\})\ (\forall u \in [0,+\infty))\ \rho_1^{(m)}(u) = \dfrac{1}{2}\dfrac{(u-z^{(m)})^2}{a^{(m)}u + b^{(m)}}$.

Then $(\forall k \in \mathbb{N})$ a majorant function of F_1 on $[0,+\infty)^N$ at x_k is given by

$$\begin{cases} Q_1(\cdot, x_k) = F_1(x_k) + (\cdot - x_k)^\top \nabla F_1(x_k) + (\cdot - x_k)^\top A_k(\cdot - x_k) \\ A_k = \mathrm{Diag}(P^\top \omega(Hx_k)) + \varepsilon\, \mathrm{I}_N \qquad \text{for } \varepsilon \geq 0. \end{cases}$$

with
- $\omega\colon (v^{(m)})_{1\leq m\leq M} \in [0,+\infty)^M \mapsto \left(\omega^{(m)}(v^{(m)})\right)_{1\leq m\leq M} \in \mathbb{R}^M$, where

$$(\forall m \in \{1,\dots,M\}) \quad \omega^{(m)}(u) = \begin{cases} \ddot{\rho}_1^{(m)}(0) & \text{if } u = 0, \\ 2\frac{\rho_1^{(m)}(0) - \rho_1^{(m)}(u) + u\dot{\rho}_1^{(m)}(u)}{u^2} & \text{if } u > 0. \end{cases}$$

- $(\forall m \in \{1,\dots,M\})\ (\forall n \in \{1,\dots,N\})\ P^{(m,n)} = H^{(m,n)} \sum_{p=1}^{N} H^{(m,p)}$.

$\rightsquigarrow$ Proof based on the concavity of $\dot{\rho}_1^{(m)}$ and Jensen's inequality ([Erdogan and Fessler - 1999]).

Implementation

Construction of the majorant

$$F(x) = \begin{cases} F_1(x) + F_2(x) & \text{if } x \in [0, +\infty)^N \\ +\infty & \text{otherwise} \end{cases}, \text{ where}$$

- $F_1 \rightsquigarrow$ Convex function.
  $\rightsquigarrow$ Majorized at x_k by $Q_1(\cdot, x_k)$.
- $F_2 \rightsquigarrow$ Concave function.
  $\rightsquigarrow$ Majorized at x_k by $Q_2(\cdot, x_k) = F_2(x_k) + (\cdot - x_k)^\top \nabla F_2(x_k)$.

Implementation

Construction of the majorant

$$F(x) = \begin{cases} F_1(x) + F_2(x) & \text{if } x \in [0, +\infty)^N \\ +\infty & \text{otherwise} \end{cases}, \text{ where}$$

- $F_1 \rightsquigarrow$ Convex function.
  $\rightsquigarrow$ Majorized at x_k by $Q_1(\cdot, x_k)$.
- $F_2 \rightsquigarrow$ Concave function.
  $\rightsquigarrow$ Majorized at x_k by $Q_2(\cdot, x_k) = F_2(x_k) + (\cdot - x_k)^\top \nabla F_2(x_k)$.

Backward step

$$y_k = \underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \left\{ R(x) + \frac{1}{2} \|x - \bar{y}_k\|^2_{\gamma_k^{-1} A_k} \right\} \text{ with } \bar{y}_k = x_k - \gamma_k A_k^{-1} \nabla F(x_k)$$

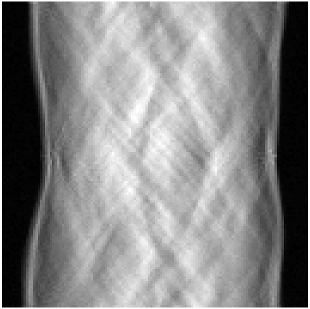
$$\Leftrightarrow y_k = \gamma_k^{1/2} A_k^{-1/2} \underset{x \in \mathbb{R}^N}{\operatorname{argmin}} \left\{ R(\gamma_k^{1/2} A_k^{-1/2} x) + \frac{1}{2} \|x - \gamma_k^{-1/2} A_k^{1/2} \bar{y}_k\|^2 \right\}$$

$\rightsquigarrow$ Dual Forward-Backward Algorithm [Combettes *et al.* - 2011]

Reconstruction with sparsity prior

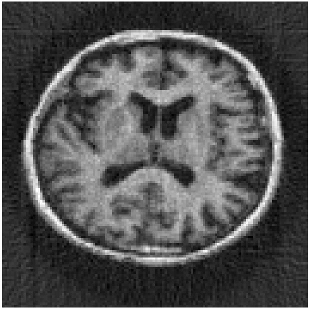
- H: Radon matrix modeling $M = 16384$ parallel projections from 128 acquisitions lines and 128 angles.
- $(\forall m \in \{1, \ldots, M\})\ a^{(m)} = 0.01$ and $b^{(m)} = 0.1$

Original image *Zubal*



Degraded sinogram

Results: Restored images



FBP: SNR=7 dB

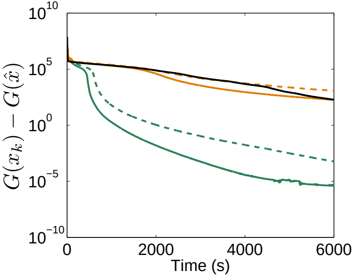
VMFB: SNR=18.9 dB

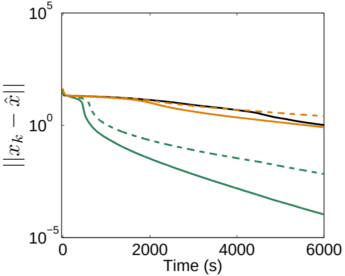
Results

VMFB Algorithm with • $\lambda_k \equiv 1$ and $\gamma_k \equiv 1.9$ (solid line)
• $\lambda_k \equiv 1$ and $\gamma_k \equiv 1$ (dashed line)

FB Algorithm with • $\lambda_k \equiv 1$ and $\gamma_k \equiv 1.9$ (solid line)
• $\lambda_k \equiv 1$ and $\gamma_k \equiv 1$ (dashed line)

FISTA

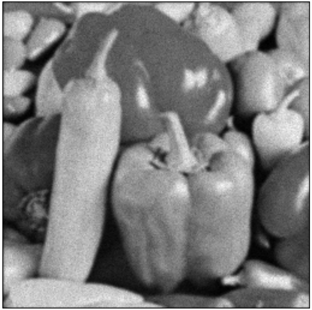




Deblurring with Non Local Total Variation

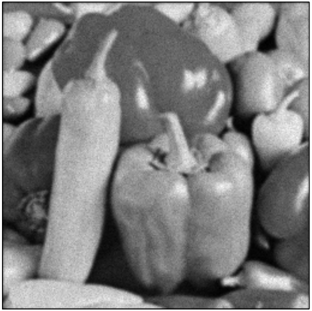
- H: Blur operator corresponding to a truncated Gaussian kernel of standard deviation 1 and size 7×7.
- $(\forall m \in \{1, \dots, M\})$, $a^{(m)} = 0.5$ and $b^{(m)} = 1$

Original image *Peppers*



Degraded image: SNR=21.85 dB

Results: Restored images



Degraded image: SNR=21.85 dB

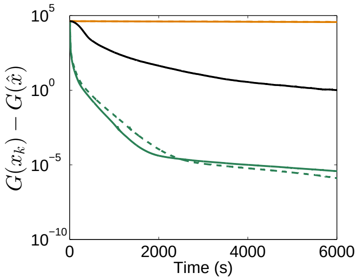
Restored image: SNR=27.11 dB

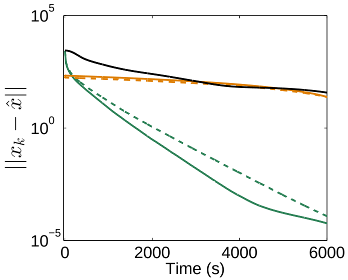
Results

VMFB Algorithm with • $\lambda_k \equiv 1$ and $\gamma_k \equiv 1.9$ (solid line)
• $\lambda_k \equiv 1$ and $\gamma_k \equiv 1$ (dashed line)

FB Algorithm with • $\lambda_k \equiv 1$ and $\gamma_k \equiv 1.9$ (solid line)
• $\lambda_k \equiv 1$ and $\gamma_k \equiv 1$ (dashed line)

FISTA





Conclusion

- Convergence of the VMFB algorithm for the sum of a non convex differentiable function F and a non smooth convex function R.
- Choice of variable metric $(A_k)_{k \in \mathbb{N}}$ based on MM principle.
- Inexact VMFB algorithm for the calculation of the proximity operator.

The variable metric strategy leads to a significant acceleration in terms of decay of both the objective function and the error on the iterates in each experiment.

Bibliography

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.
Math. Program., vol. 137, pp. 91-129, Feb. 2011.

J. Bolte, A. Daniilidis, O. Ley and L. Mazet.
Characterizations of Łojasiewicz inequalities and applications.
Trans. Amer. Math. Soc., vol.362, no. 6, pp. 3319-3363, 2010.

S. Becker and J. Fadili.
A quasi-Newton proximal splitting method.
Tech. Rep., 2012. Available on http://arxiv.org/abs/1206.1156

P. L. Combettes and J.-C. Pesquet.
Proximal thresholding algorithm for minimization over orthonormal bases.
SIAM J. Optim., vol. 18, no. 4, pp. 1351-1376, Nov. 2007.

P. L. Combettes and B. C. Vũ.
Variable metric forward-backward splitting with applications to monotone inclusions in duality.
to appear in Optimization, 2013.

H. Erdogan and J. A. Fessler.
Monotonic algorithms for transmission tomography.
IEEE Trans. Med. Imag., vol. 18, no. 9, pp. 801-814, Nov. 1999.

J. J. Moreau.
Proximité et Dualité dans un espace hilbertien.
Bull. Soc. Math. France, vol. 93, pp. 273-299, 1965.

Thank you !

E. Chouzenoux, J.-C. Pesquet and A. Repetti.
Variable Metric Forward-Backward algorithm for minimizing the sum of a differentiable function and a convex function.
Tech. Rep., 2013. Available on
http://www.optimization-online.org/DB_HTML/2013/01/3749.html