Proximal methods: tools for solving inverse problems on a large scale

Emilie Chouzenoux
LIGM, UPEM

Journée du Labex Bézout:
Data Science and Massive Data Analysis

12 June 2014
Joint work with

J.-C. Pesquet  A. Repetti  F. Abboud  A. Florescu
LIGM, University Paris East  Dunarea de Jos University Galati, Romania

J.-H. Chenot  L. Laborelli  P. Ciuciu
Institut National Audiovisuel (INA)  CEA, NeuroSpin center
Inverse problems and large scale optimization

[Microscopy, ISBI Challenge 2013, F. Soulez]
Inverse problems and large scale optimization

[Microscopy, ISBI Challenge 2013, F. Soulez]

Original image \( \bar{x} \in \mathbb{R}^N \)

Degraded image \( z = D(H\bar{x}) \in \mathbb{R}^M \)

- \( H \in \mathbb{R}^{M \times N} \): matrix associated with the degradation operator.
- \( D: \mathbb{R}^M \rightarrow \mathbb{R}^M \): noise degradation.

Inverse problem:
Find a good estimate of \( \bar{x} \) from the observations \( z \), using some a priori knowledge on \( \bar{x} \) and on the noise characteristics.
Inverse problems and large scale optimization

Inverse problem:
Find an estimate \( \hat{x} \) close to \( x \) from the observations \( z = D(Hx) \).

- Inverse filtering (if \( M = N \) and \( H \) is invertible)

\[
\hat{x} = H^{-1}z = H^{-1}(Hx + b) \quad \leftarrow \quad \text{if } b \in \mathbb{R}^M \text{ is an additive noise}
\]

\[
= \bar{x} + H^{-1}b
\]

→ Closed form expression, but amplification of the noise if \( H \) is ill-conditioned (ill-posed problem).
Inverse problems and large scale optimization

Inverse problem:
Find an estimate $\hat{x}$ close to $\bar{x}$ from the observations $z = D(H\bar{x})$.

- **Inverse filtering**
- **Variational approach**

$$\hat{x} \in \text{Argmin}_{x \in \mathbb{R}^N} f_1(x) + f_2(x)$$

Data fidelity term + Regularization term
Inverse problems and large scale optimization

Inverse problem:
Find an estimate \( \hat{x} \) close to \( x \) from the observations \( z = D(H\bar{x}) \).

- Inverse filtering
- Variational approach

\[ \hat{x} \in \text{Argmin}_{x \in \mathbb{R}^N} f_1(x) + f_2(x) \]

Data fidelity term
Regularization term

Examples of data fidelity term

- Gaussian noise
  \[ (\forall x \in \mathbb{R}^N) \quad f_1(x) = \frac{1}{\sigma^2} \| Hx - z \|^2 \]

- Poisson noise
  \[ (\forall x \in \mathbb{R}^N) \quad f_1(x) = \sum_{m=1}^{M} \left( [Hx]^{(m)} - z^{(m)} \log([Hx]^{(m)}) \right) \]
Examples of regularization terms (1)

- Admissibility constraints

\[
\text{Find } x \in C = \bigcap_{m=1}^{M} C_m
\]

where \((\forall m \in \{1, \ldots, M\}) \ C_m \subset \mathbb{R}^N\).
Examples of regularization terms (1)

- Admissibility constraints

Find \[ x \in C = \bigcap_{m=1}^{M} C_m \]

where \((\forall m \in \{1, \ldots, M\})\) \(C_m \subset \mathbb{R}^N\).

- Variational formulation

\[ (\forall x \in \mathbb{R}^N) \quad f_2(x) = \sum_{m=1}^{M} \iota_{C_m}(x) \]

where, for all \(m \in \{1, \ldots, M\}\), \(\iota_{C_m}\) is the indicator function of \(C_m\):

\[ (\forall x \in \mathbb{R}^N) \quad \iota_{C_m}(x) = \begin{cases} 0 & \text{if } x \in C_m \\ +\infty & \text{otherwise.} \end{cases} \]
Examples of regularization terms (2)

- $\ell_1$ norm (analysis approach)

$$f_2(x) = \sum_{k=1}^{K} |[Fx]^{(k)}| = \|Fx\|_1$$

$F \in \mathbb{R}^{K \times N}$: Frame decomposition operator ($K \geq N$)

**signal $x$** \hspace{1cm} $F$ \hspace{1cm} **frame coefficients**
Examples of regularization terms (2)

- **$\ell_1$ norm (analysis approach)**

  \[(\forall x \in \mathbb{R}^N) \quad f_2(x) = \sum_{k=1}^{K} |[Fx]^{(k)}| = \|Fx\|_1\]

- **Total variation**

  \[(\forall x = (x^{(i_1,i_2)})_{1 \leq i_1 \leq N_1, 1 \leq i_2 \leq N_2} \in \mathbb{R}^{N_1 \times N_2}) \quad f_2(x) = tv(x) = \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \|\nabla x^{(i_1,i_2)}\|_2 \]

  $\nabla x^{(i_1,i_2)}$ : discrete gradient at pixel $(i_1, i_2)$. 
Inverse problems and large scale optimization

Inverse problem:
Find an estimate $\hat{x}$ close to $x$ from the observations $z = D(Hx)$.

- Inverse filtering
- Variational approach (more general context)

$\hat{x} \in \text{Argmin}_{x \in \mathbb{R}^N} \sum_{i=1}^{m} f_i(x)$

where $f_i$ may denote a data fidelity term / a (hybrid) regularization term / constraint.
Inverse problems and large scale optimization

**Inverse problem:**
Find an estimate $\hat{x}$ close to $x$ from the observations $z = D(Hx)$.

- **Inverse filtering**
- Variational approach (more general context)

$\hat{x} \in \text{Argmin}_{x \in \mathbb{R}^N} \sum_{i=1}^{m} f_i(x)$

where $f_i$ may denote a data fidelity term / a (hybrid) regularization term / constraint.

→ Often no closed form expression or solution expensive to compute (especially in large scale context).

- **Need for an efficient iterative minimization strategy!**
Outline

1. Proximal-based algorithms
   - Proximity operator
   - Forward-Backward algorithm
   - Acceleration via metric change
   - Acceleration via block alternation

2. Applications
   - Parallel magnetic resonance imaging
   - Phase retrieval
   - Blind deconvolution of television video
   - Multi-channel image restoration
Proximal-based algorithms
Gradient and subgradient algorithms

Optimization problem: Minimization of function $f \in \Gamma_0(\mathbb{R}^N)$ on $\mathbb{R}^N$.

- If $f$ has a $\beta$-Lipschitz gradient with $\beta \in ]0, +\infty[$

$$ (\forall \ell \in \mathbb{N}) \quad x_{\ell+1} = x_\ell - \gamma_\ell \nabla f(x_\ell) \quad \text{explicit step} $$

with $0 < \inf_{\ell \in \mathbb{N}} \gamma_\ell$ and $\sup_{\ell \in \mathbb{N}} \gamma_\ell < 2\beta^{-1}$. 
Gradient and subgradient algorithms

Optimization problem: Minimization of function $f \in \Gamma_0(\mathbb{R}^N)$ on $\mathbb{R}^N$.

- If $f$ has a $\beta$-Lipschitz gradient with $\beta \in ]0, +\infty[$
  
  $$ (\forall \ell \in \mathbb{N}) \quad x_{\ell+1} = x_\ell - \gamma_\ell \nabla f(x_\ell) \quad \text{explicit step} $$

  with $0 < \inf_{\ell \in \mathbb{N}} \gamma_\ell$ and $\sup_{\ell \in \mathbb{N}} \gamma_\ell < 2\beta^{-1}$.

- When $f$ is nonsmooth, replace gradient with subgradient

  $\partial f(x) = \left\{ t \in \mathbb{R}^N | (\forall y \in \mathbb{R}^N) \ f(y) \geq f(x) + \langle t | y - x \rangle \right\}$

  $t \in \partial f(x)$: subgradient at $x \in \mathbb{R}^N$

  $\partial f: \mathbb{R}^N \to 2^{\mathbb{R}^N}$: subdifferential
Subdifferential

\[ f(y) = f(x) + \langle y - x | t \rangle \]

\[ t \in \partial f(x) \]
Subdifferential

\[
f(y) = f(x) + \langle y - x | t \rangle
\]

\[t \in \partial f(x)\]
Subdifferential

\[ f(y) = f(x) + \langle y - x | t \rangle \]

\[ t \in \partial f(x) \]
Subdifferential

\[ f(y) = f(x) + \langle y - x | t \rangle \]

\[ t \in \partial f(x) \]
Subdifferential

\[ f(y) = f(x) + \langle y - x | t \rangle \]

\[ t \in \partial f(x) \]
Subdifferential

\[ f(y) = f(x) + \langle y - x \mid t \rangle \]

\[ t \in \partial f(x) \]
Subdifferential

\[
f(y) = f(x) + \langle y - x | t \rangle
\]

\[
t \in \partial f(x)
\]
Subdifferential

\[ f(y) = f(x) + \langle y - x | t \rangle \]

\[ t \in \partial f(x) \]
Example of subdifferential

Example:

- If $f$ is differentiable at $x \in \mathbb{R}^N$ then $\partial f(x) = \{ \nabla f(x) \}$.
- If $f = | \cdot |$ then

$$
(\forall x \in \mathbb{R}) \quad \partial f(x) = \begin{cases} 
\{ \text{sign}(x) \} & \text{if } x \neq 0 \\
[-1, 1] & \text{if } x = 0
\end{cases}
$$
From the subgradient algorithm ...

**Optimization problem:** Minimization of function $f \in \Gamma_0(\mathbb{R}^N)$ on $\mathbb{R}^N$.

**Subgradient algorithm** [Shor, 1979]

$$(\forall \ell \in \mathbb{N}) \quad x_{\ell+1} = x_\ell - \gamma_\ell t_\ell, \quad t_\ell \in \partial f(x_\ell)$$

where $(\forall \ell \in \mathbb{N}) \gamma_\ell \in ]0, +\infty[$ such that $\sum_{\ell=0}^{+\infty} \gamma_\ell^2 < +\infty$ and $\sum_{\ell=0}^{+\infty} \gamma_\ell = +\infty$. 

From the subgradient algorithm ...

Optimization problem: Minimization of function $f \in \Gamma_0(\mathbb{R}^N)$ on $\mathbb{R}^N$.

**Subgradient algorithm** [Shor, 1979]

$$(\forall \ell \in \mathbb{N}) \quad x_{\ell+1} = x_{\ell} - \gamma_{\ell} t_{\ell}, \quad t_{\ell} \in \partial f(x_{\ell})$$

where $(\forall \ell \in \mathbb{N}) \gamma_{\ell} \in ]0, +\infty[$ such that $\sum_{\ell=0}^{+\infty} \gamma_{\ell}^2 < +\infty$ and $\sum_{\ell=0}^{+\infty} \gamma_{\ell} = +\infty$.

**Implicit form**

$$(\forall \ell \in \mathbb{N}) \quad x_{\ell+1} = x_{\ell} - \gamma_{\ell} t'_{\ell}, \quad t'_{\ell} \in \partial f(x_{\ell+1})$$

$\Leftrightarrow$ $x_{\ell} - x_{\ell+1} \in \gamma_{\ell} \partial f(x_{\ell+1})$
... to the origins of the proximity operator!

**Property**

Let $\varphi \in \Gamma_0(\mathbb{R}^N)$. For all $x \in \mathbb{R}^N$, there exists a unique vector $\hat{x} \in \mathbb{R}^N$ such that $x - \hat{x} \in \partial \varphi(\hat{x})$.

- Let $\hat{x} = \text{prox}_\varphi(x)$.
- $\text{prox}_\varphi : \mathbb{R}^N \to \mathbb{R}^N$ : proximity operator

**Proximal point algorithm**

\[
(\forall \ell \in \mathbb{N}) \quad x_\ell - x_{\ell+1} \in \gamma_\ell \partial f(x_{\ell+1}) \\
\Leftrightarrow \quad x_{\ell+1} = \text{prox}_{\gamma_\ell f}(x_\ell)
\]

where $\inf_{\ell \in \mathbb{N}} \gamma_\ell > 0$ such that $\sum_{\ell=0}^{+\infty} \gamma_\ell = +\infty$. 
Another definition of the proximity operator

**Property**

Let $f \in \Gamma_0(\mathbb{R}^N)$. For all $x \in \mathbb{R}^N$, $\text{prox}_f(x)$ is the unique minimizer of

$$y \mapsto f(y) + \frac{1}{2} \|x - y\|^2.$$

**Example:**

Let $C$ a closed non empty subset of $\mathbb{R}^N$. Then, $\text{prox}_{\iota_C}$ reduces to the projection operator on the set $C$. 
Some other examples

- Explicit form for objective functions associated to the usual log-concave probability densities [Chaux et al. - 2007]
  - Laplace
  - Generalized gaussian
  - maximum entropy
  - gamma
  - uniform
  - Weibull
  - Generalized inverse gaussian
  - Gaussian
  - Huber
  - Smoothed Laplace
  - chi
  - triangular
  - Pearson type I
  - And many other functions! [Combettes, Pesquet - 2010]
Forward-backward algorithm

Optimization problem:
Minimization of $f + g$ on $\mathbb{R}^N$, assuming that $g$ has a $\beta$-Lipschitz gradient.

Forward-backward algorithm

$$(\forall \ell \in \mathbb{N}) \quad x_{\ell+1} = x_\ell - \gamma_\ell (t'_\ell + \nabla g(x_\ell)), \quad t'_\ell \in \partial f(x_{\ell+1})$$

$\iff x_{\ell+1} = \text{prox}_{\gamma_\ell f}(x_\ell - \gamma_\ell \nabla g(x_\ell))$$
Forward-backward algorithm

Optimization problem:
Minimization of $f + g$ on $\mathbb{R}^N$, assuming that $g$ has a $\beta$-Lipschitz gradient.

Forward-backward algorithm

$$(\forall \ell \in \mathbb{N}) \quad x_{\ell+1} = x_\ell + \lambda_\ell \left( \text{prox}_{\gamma_\ell f}(x_\ell - \gamma_\ell \nabla g(x_\ell)) - x_\ell \right)$$

Convergence of $(x_\ell)_{\ell \in \mathbb{N}}$ if $0 < \inf_{\ell \in \mathbb{N}} \gamma_\ell$, $\sup_{\ell \in \mathbb{N}} \gamma_\ell < 2\beta^{-1}$,

$0 < \inf_{\ell \in \mathbb{N}} \lambda_\ell$ and $\sup_{\ell \in \mathbb{N}} \lambda_\ell \leq 1$.

- $f$ and $g$ convex [Chen, Rockafellar, 1997][Combettes, Wajs, 2005]
- $f$ and $g$ nonconvex (under Kurdyka-Łojasiewicz assumption) [Attouch et al. - 2011]
How to make the forward-backward algorithm efficient for big data optimization?
First trick: Majoration-Minimization strategy
MM point of view

**Majorize-Minimize Assumption**

- For every $\ell \in \mathbb{N}$, there exists a symmetric positive definite (SPD) matrix $A_\ell(x_\ell) \in \mathbb{R}^{N \times N}$ such that for every $x \in \mathbb{R}^N$

$$Q(x, x_\ell) = g(x_\ell) + (x - x_\ell)^T \nabla g(x_\ell) + \frac{1}{2}(x - x_\ell)^T A_\ell(x_\ell)(x - x_\ell),$$

is a majorant function of $g$ at $x_\ell$ on $\text{dom } f$, i.e.,

$$g(x_\ell) = Q(x_\ell, x_\ell) \quad \text{and} \quad (\forall x \in \text{dom } f) \quad g(x) \leq Q(x, x_\ell).$$
MM point of view

**Majorize-Minimize Assumption**

- For every $\ell \in \mathbb{N}$, there exists a symmetric positive definite (SPD) matrix $A_\ell(x_\ell) \in \mathbb{R}^{N \times N}$ such that for every $x \in \mathbb{R}^N$

  $$Q(x, x_\ell) = g(x_\ell) + (x - x_\ell)^\top \nabla g(x_\ell) + \frac{1}{2}(x - x_\ell)^\top A_\ell(x_\ell)(x - x_\ell),$$

  is a **majorant function** of $g$ at $x_\ell$ on $\text{dom } f$, i.e.,

  $$g(x_\ell) = Q(x_\ell, x_\ell) \quad \text{and} \quad (\forall x \in \text{dom } f) \quad g(x) \leq Q(x, x_\ell).$$

\[ g \text{ is differentiable with a } \beta\text{-Lipschitzian gradient on a convex subset of } \mathbb{R}^N \quad \implies \quad A_\ell(x_\ell) \equiv \beta \text{ Id} \text{ satisfies the above assumption} \]

[Bertsekas - 1999]
MM Algorithm [Jacobson and Fessler - 2007]

\[ x_{\ell+1} \in \text{Argmin}_{x \in \mathbb{R}^N} f(x) + Q(x, x_\ell) \]
**MM algorithm** [Jacobson and Fessler - 2007]

\[
x_{\ell+1} \in \text{Argmin}_{x \in \mathbb{R}^N} f(x) + Q(x, x_\ell)
\]
**MM algorithm** [Jacobson and Fessler - 2007]

\[ x_{\ell+1} \in \text{Argmin} \ f(x) + Q(x, x_\ell) \]

\[ x \in \mathbb{R}^N \]

\[ g + f \]

\[ Q(\cdot, x_{\ell+2}) + f \]

\[ x_{\ell+2} x_{\ell+3} \cdots \]
**MM algorithm** [Jacobson and Fessler - 2007]

\[ x_{\ell+1} \in \text{Argmin}_{x \in \mathbb{R}^N} f(x) + Q(x, x_\ell) \]

\[ \iff \text{Forward-backward algorithm with} \]
\[ \begin{align*}
  \triangleright & \quad A_\ell(x_\ell) \equiv \beta \text{Id} \\
  \triangleright & \quad \lambda_\ell \equiv 1 \\
  \triangleright & \quad \gamma_\ell \equiv 1
\end{align*} \]

\[ \rightsquigarrow \text{Why not trying more sophisticated matrices} (A_\ell)_{\ell \in \mathbb{N}} ? \]

\[ \quad \begin{align*}
  \triangleright & \quad \text{Variable metric forward-backward algorithm !}
\end{align*} \]
Acceleration via metric change

Definition

Let $x \in \mathbb{R}^N$. Let $A$ be a SPD matrix. The proximity operator relative to the metric induced by $A$ is defined by

$$\text{prox}_{\gamma^{-1}A,f}(x) = \text{Argmin}_{y \in \mathbb{R}^N} f(y) + \frac{1}{2\gamma} \|y - x\|_A^2.$$ 

Variable metric forward-backward algorithm

$$(\forall \ell \in \mathbb{N}) \quad x_{\ell+1} = \text{prox}_{\gamma^{-1}_{\ell}A_{\ell}(x_{\ell})} \left( x_{\ell} - \gamma_{\ell} A_{\ell}(x_{\ell})^{-1} \nabla g(x_{\ell}) \right).$$

Convergence of $(x_{\ell})_{\ell \in \mathbb{N}}$

- $f$ and $g$ convex [Combettes et al. - 2012]
- $f$ and $g$ nonconvex [Chouzenoux et al. - 2013]

► Significant acceleration in practice!
Second trick: Block alternation
Acceleration via block alternation

Assumption: $f$ is an additively block separable function.
Acceleration via block alternation

Assumption: $f$ is an additively block separable function.

$\begin{bmatrix}
  x(1) \\
  x(2) \\
  \vdots \\
  x(J)
\end{bmatrix} \in \mathbb{R}^{N_1} \times \mathbb{R}^{N_2} \times \cdots \times \mathbb{R}^{N_J}
$

$N = \sum_{j=1}^{J} N_j$
Acceleration via block alternation

- Assumption: \( f \) is an additively block separable function.

\[
\begin{aligned}
f(x) &= f(x^{(1)}) + f(x^{(2)}) + \cdots + f(x^{(J)}) \\
&= \sum_{j=1}^{J} f_j(x^{(j)})
\end{aligned}
\]
Acceleration via block alternation

**Block coordinate forward-backward algorithm**

$(\forall \ell \in \mathbb{N})$, pick a block $j_{\ell} \in \{1, \ldots, J\}$, and update:

\[
\begin{align*}
    x^{(j_{\ell})}_{\ell+1} &= \text{prox}_{\gamma_{\ell} f_{j_{\ell}}} \left( x^{(j_{\ell})}_{\ell} - \gamma_{\ell} \nabla_{j_{\ell}} g(x_{\ell}) \right) \\
    x^{(\overline{j}_{\ell})}_{\ell+1} &= x^{(\overline{j}_{\ell})}_{\ell} \\
    x^{(j_{\ell})}_{\ell+1} &= x^{(j_{\ell})}_{\ell}
\end{align*}
\]

- Convergence of $(x_{\ell})_{\ell \in \mathbb{N}}$ (assuming a cyclic update rule) established in [Bolte et al. - 2013] for possibly nonconvex functions $f$ and $g$ verifying Kurdyka-Łojasiewicz assumption.
Acceleration via block alternation

Block coordinate forward-backward algorithm

$(\forall \ell \in \mathbb{N})$, pick a block $j_\ell \in \{1, \ldots, J\}$, and update:

\[
\begin{cases}
    x^{(j_\ell)}_{\ell+1} = \text{prox}_{\gamma_\ell} f_{j_\ell} \left( x^{(j_\ell)}_{\ell} - \gamma_\ell \nabla_{j_\ell} g(x_\ell) \right) \\
    x^{(\bar{j}_\ell)}_{\ell+1} = x^{(\bar{j}_\ell)}_{\ell}
\end{cases}
\]

- Convergence of $(x_\ell)_{\ell \in \mathbb{N}}$ (assuming a cyclic update rule) established in [Bolte et al. - 2013] for possibly nonconvex functions $f$ and $g$ verifying Kurdyka-Łojasiewicz assumption.

- Block alternation presents several advantages:
  
  ✓ more flexibility,
  
  ✓ reduced computational cost at each iteration,
  
  ✓ reduced memory requirement.
Combining first and second trick ...
Acceleration via block alternation and metric change

**Block coordinate variable metric forward-backward algorithm**

\( \forall \ell \in \mathbb{N} \), pick a block \( j_\ell \in \{1, \ldots, J\} \), and update

\[
\begin{align*}
x^{(j_\ell)}_{\ell+1} &= \text{prox}_{\frac{1}{\gamma_\ell} A_{j_\ell}(x_\ell)} \left( x^{(j_\ell)}_{\ell} - \gamma_\ell A_{j_\ell}(x_\ell)^{-1} \nabla_{j_\ell} g(x_\ell) \right) \\
x^{(\overline{j}_\ell)}_{\ell+1} &= x^{(\overline{j}_\ell)}_{\ell}
\end{align*}
\]

- Convergence of \( (x_\ell)_{\ell \in \mathbb{N}} \) (assuming a quasi cyclic update rule) established in [Chouzenoux et al. - 2013] for nonconvex functions \( f \) and \( g \) verifying Kurdyka-Łojasiewicz assumption.

- **Benefits from the advantages of both acceleration techniques!**
<table>
<thead>
<tr>
<th>Introduction</th>
<th>Proximal-based algorithms</th>
<th>Applications</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Applications</td>
<td></td>
</tr>
</tbody>
</table>

Proximal methods: tools for solving inverse problems on a large scale
Parallel Magnetic Resonance Imaging [Florescu et al. - 2014]

**Challenges:**

- Parallel acquisition and compressive sensing
- Complex-valued signals

**Results:**

Original  Proposed method
Parallel Magnetic Resonance Imaging [Florescu et al. - 2014]

**Challenges:**

- Parallel acquisition and compressive sensing
- Complex-valued signals

**Results:**

![Original vs Proposed Method](image-url)
Parallel Magnetic Resonance Imaging [Florescu et al. - 2014]

Challenges:

- Parallel acquisition and compressive sensing
- Complex-valued signals

Results:

Convergence speed of several proximal-based algorithms
Phase retrieval [Repetti et al. - ICIP 2014]

**Challenges:**
- Only the modulus of the observed data is available
- Non-Fourier measurements
- Nonconvex data fidelity term

**Results:**

SparseFienup  
Proposed method
Phase retrieval [Repetti et al. - ICIP 2014]

Challenges:
- Only the modulus of the observed data is available
- Non-Fourier measurements
- Nonconvex data fidelity term

Results:

SparseFienup

Proposed method
Phase retrieval [Repetti et al. - ICIP 2014]

**Challenges:**

- Only the modulus of the observed data is available
- Non-Fourier measurements
- Nonconvex data fidelity term

**Results:**

Influence of the variable metric strategy
Blind deconvolution of video [Abboud et al. - EUSIPCO 2014]

**Challenges:**
- The degradation blur operator is unknown
- Nonconvex data fidelity term

**Results:**

Observed

Restored
Blind deconvolution of video [Abboud et al. - EUSIPCO 2014]

**Challenges:**
- The degradation blur operator is unknown
- Nonconvex data fidelity term

**Results:**

![Observed](image1.png)  ![Restored](image2.png)
Blind deconvolution of video [Abboud et al. - EUSIPCO 2014]

**Challenges:**
- The degradation blur operator is unknown
- Nonconvex data fidelity term

**Results:**

![Estimated blur kernel graph]

...
Multi-channel image restoration [Chierchia et al. - 2014]

**Challenges:**
- Deal with images having a large number of components
- Circumvent the choice of regularization parameters by introducing suitable nonlocal constraints
- Develop epigraphical techniques to address these constraints efficiently
Multi-channel image restoration [Chierchia et al. - 2014]

**Challenges:**

- Deal with images having a large number of components
- Circumvent the choice of regularization parameters by introducing suitable nonlocal constraints
- Develop epigraphical techniques to address these constraints efficiently

![Graph showing comparison between constrained formulation and variational formulation](image-url)
Conclusion

✓ Proximal-based algorithms: An efficient tool for solving large scale possibly difficult optimization problem;
Conclusion

✓ Proximal-based algorithms: An efficient tool for solving large scale possibly difficult optimization problem;

✓ Two recipes for accelerating the algorithms:
  ▶ Majoration-Minimization strategy
  ▶ Block alternation
Conclusion

✓ Proximal-based algorithms: An efficient tool for solving large scale possibly difficult optimization problem;

✓ Two recipes for accelerating the algorithms:
  ▶ Majoration-Minimization strategy
  ▶ Block alternation

✓ No need to invert large size matrices through primal-dual forward-backward based methods;
Conclusion

✓ Proximal-based algorithms: An efficient tool for solving large scale possibly difficult optimization problem;

✓ Two recipes for accelerating the algorithms:
  ▶ Majoration-Minimization strategy
  ▶ Block alternation

✓ No need to invert large size matrices through primal-dual forward-backward based methods;

✓ Parallel implementations possible thanks to splitting techniques.
Conclusion

✓ Proximal-based algorithms: An efficient tool for solving large scale possibly difficult optimization problem;

✓ Two recipes for accelerating the algorithms:
  ▶ Majoration-Minimization strategy
  ▶ Block alternation

✓ No need to invert large size matrices through primal-dual forward-backward based methods;

✓ Parallel implementations possible thanks to splitting techniques.

Future challenges: Find more tricks!
Thank you ! Questions ?

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.*  

E. Chouzenoux, J.-C. Pesquet and A. Repetti.  
*A Block Coordinate Variable Metric Forward-Backward Algorithm.*  
http://www.optimization-online.org/DB_HTML/2013/12/4178.html.

*A Majorize-Minimize Memory Gradient Method for Complex-Valued Inverse Problems.*  

*A Hybrid Alternating Proximal Method for Blind Video Restoration.*  
Accepted to EUSIPCO 2014.