A Generalized Forward-Backward Splitting

Jalal Fadili
GREYC, CNRS-ENSICAEN-Université de Caen,
http://www.greyc.ensicaen.fr/~jfadili
Joint work with Hugo Raguet and Gabriel Peyré

SIAM IS'12

Outline

- Class of problems and motivations.
- A GFB splitting algorithm.
- Stylized applications.
- Conclusion and future work.

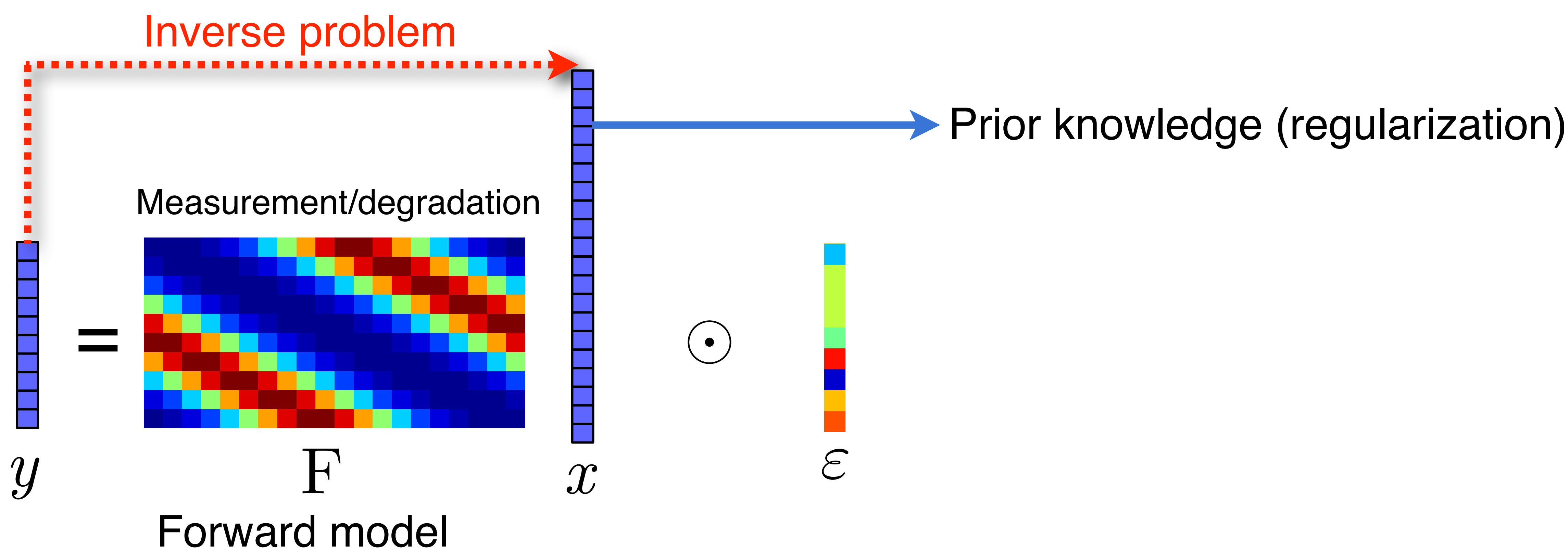
Outline

- Class of problems and motivations.
- A GFB splitting algorithm.
- Stylized applications.
- Conclusion and future work.

Motivations

- Inverse problems with mixed regularization, e.g. :

$$\min_{x \in \mathcal{H}} \underbrace{f(x)}_{\text{Data fidelity}} + \underbrace{g_1(x) + \cdots + g_n(x)}_{\text{Regularization, constraints}}$$



Typical models

Smooth, piecewise-smooth, sparse, cartoon, etc..

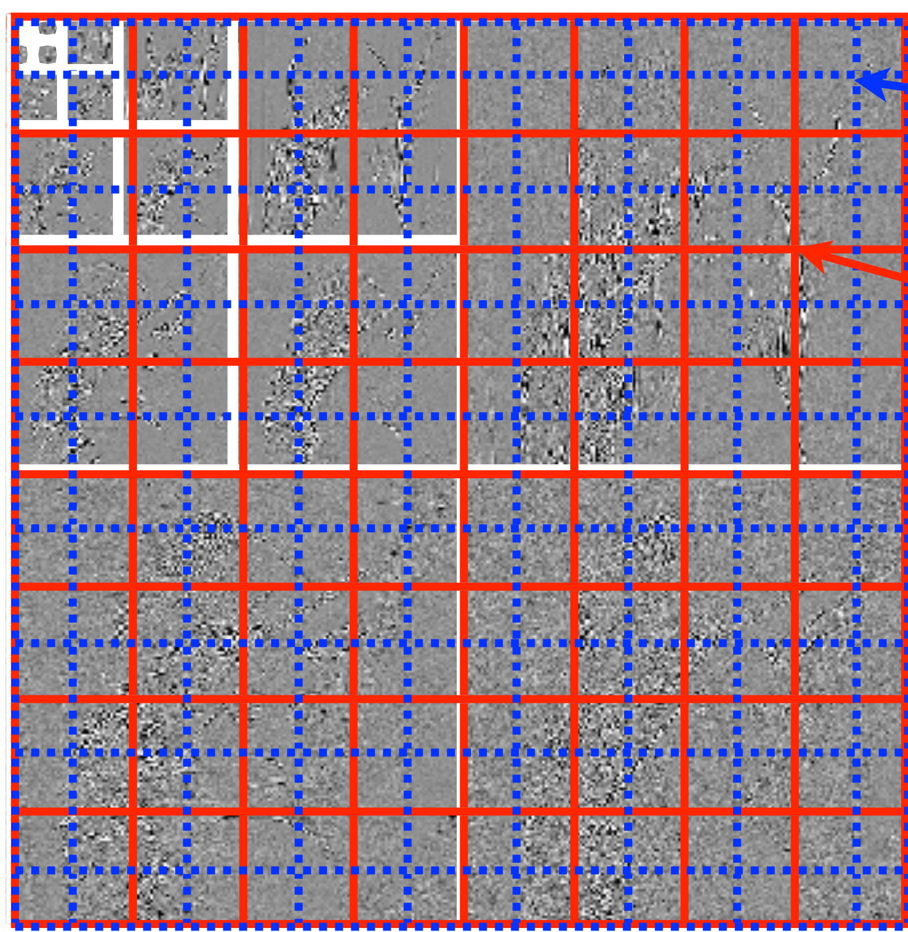
Motivations

- Inverse problems with mixed regularization, e.g. :

$$\min_{x\in\mathcal{H}} \underbrace{f(x)}_{\text{Data fidelity}} + \underbrace{g_1(x) + \cdots + g_n(x)}_{\text{Regularization, constraints}}$$

- Inverse problems with structured sparsity, e.g. :

$$\min_{x\in\mathcal{H}} \underbrace{f(x)}_{\text{Data fidelity}} + \underbrace{g_1(x) + \cdots + g_n(x)}_{\text{Structured sparsity (e.g. } \ell_p - \ell_q \text{ norm on overlapping blocks)}}$$



$$g_2(x) = \sum_{b\in\mathcal{B}_2} \|x_b\|_2$$

$$g_1(x) = \sum_{b\in\mathcal{B}_1} \|x_b\|_2$$

Motivations

- Inverse problems with mixed regularization, e.g. :

$$\min_{x\in\mathcal{H}} \underbrace{f(x)}_{\text{Data fidelity}} + \underbrace{g_1(x)+\cdots+g_n(x)}_{\text{Regularization, constraints}}$$

- Inverse problems with structured sparsity, e.g. :

$$\min_{x\in\mathcal{H}} \underbrace{f(x)}_{\text{Data fidelity}} + \underbrace{g_1(x)+\cdots+g_n(x)}_{\text{Structured sparsity (e.g. } \ell_p-\ell_q \text{ norm on overlapping blocks)}}$$

- Other potential applications : signal and image processing, machine learning, classification, statistical estimation, etc..

Class of problems

$$\min_{x \in \mathcal{H}} f(x) + \sum_{i=1}^{n} g_i(x)$$

- Assumptions :
  - $f, g_i : \mathcal{H} \to \mathbb{R} \cup \{+\infty\},\ f, g_i \in \Gamma_0(\mathcal{H})$;
  - f C^1 with β-Lipschitz gradient, all g_i's are simple.
  - Domain qualification condition : $(0, \cdots, 0) \in \mathrm{sri}(\{(x-y, x-y_1, \cdots, x-y_n) : x \in \mathcal{H} \text{ and } y \in \mathrm{dom}(f), y_i \in \mathrm{dom}(g_i)\})$;
  - Set of minimizers $\mathcal{M}^\star \neq \emptyset$.
- Requirements :
  - Exploit the (composite) additive structure of the objective.
  - Exploit the properties of the individual functions : g_i simple (closed-form proximity operator) and f smooth.
  - Deal with large scale data.
  - Avoid nested algorithms.

Proximity operator

Definition (Proximity operator [J.-J. Moreau 1962]) *Let $f \in \Gamma_0(\mathcal{H})$. Then, for every $x \in \mathcal{H}$, the function $z \mapsto \frac{1}{2}\|x - z\|^2 + f(z)$ achieves its infimum at a unique point denoted by $\operatorname{prox}_f x$. The uniquely-valued operator $\operatorname{prox}_f : \mathcal{H} \to \mathcal{H}$ thus defined is the proximity operator of f. $\operatorname{rprox}_f = 2\operatorname{prox}_f - \mathrm{I}$ is the reflection proximity operator.*

- $\operatorname{prox}_f(x)$ is single-valued.
- $\operatorname{prox}_{\gamma f}$ is the resolvent of the subdifferential of f : $\operatorname{prox}_{\gamma f} = (\mathrm{I} + \gamma\partial f)^{-1}$, $\gamma > 0$.
- The proximity operator is firmly nonexpansive. Hence its is nonexpansive and so is its reflection operator.
- Many useful properties and a whole calculus framework.

Outline

- Class of problems and motivations.
- A GFB splitting algorithm.
- Stylized applications.
- Conclusion and future work.

Class of problems

$$\min_{x\in\mathcal{H}} f(x) + \sum_{i=1}^{n} g_i(x)$$

- $f, g_i : \mathcal{H} \to \mathbb{R} \cup \{+\infty\},\ f, g_i \in \Gamma_0(\mathcal{H})$;
- f C^1 with β-Lipschitz gradient, all g_i's are simple.
- Domain qualification condition : $(0, \cdots, 0) \in \operatorname{sri}(\{(x-y, x-x_1, \cdots, x-x_n) : x \in \mathcal{H} \text{ and } y \in \operatorname{dom}(f), x_i \in \operatorname{dom}(g_i)\})$;
- Set of minimizers $\mathcal{M}^\star \neq \emptyset$.

For $w_i \in]0,1]$ with $\sum_{i=1}^n w_i = 1$, let $\boldsymbol{\mathcal{H}}$ be the real Hilbert space obtained by endowing the Cartesian product $\mathcal{H}^n$ with the scalar product $\sum_i w_i \langle x_i, y_i \rangle$.

$$\mathcal{S} = \{(x_1, \cdots, x_n) \in \boldsymbol{\mathcal{H}} : x_1 = x_2 = \cdots = x_n\},$$

$$\boldsymbol{\Pi} : \mathcal{H} \to \mathcal{S}, x \mapsto (x, \cdots, x) \quad \text{(canonical isometry)} .$$

A Generalized Forward-Backward Algorithm

Initialization : Choose $(z_{i,0})_{1\le i\le n} \in \boldsymbol{\mathcal{H}}$, $\gamma \in]0, 2\beta[$, a sequence $(\lambda_k)_k$ in $]0,1]$, weights $w_i \in]0,1]$ (e.g. $1/n$). Let $x_0 = \sum_{i=1}^n w_i z_{i,k}$.

Main iteration :

repeat

1. Compute the resolvant points (in parallel if desired) :
   for $i = 1$ *to* n **do**
   $$z_{i,k+1} = z_{i,k} + \lambda_k \left(\operatorname{prox}_{\gamma/w_i g_i}(2x_k - z_{i,k} - \gamma(\nabla f(x_k) + e_{2,k})) + e_{1,i,k} - x_k \right) .$$
2. Update by averaging :
   $$x_{k+1} = \sum_{i=1}^n w_i z_{i,k} .$$
3. $k \leftarrow k+1$.

until *Convergence* ;

Output : x_k.

Convergence

Theorem *Let* $\gamma \in]0, 2\beta[$, *let* $(\lambda_k)_{k\in\mathbb{N}}$ *be a sequence in* $]0, 1]$. *Let* $z_0 \in \mathcal{H}$. *Define the sequence of the above GFB algorithm. Assume that (i)* $\liminf_t \lambda_k > 0$, *and (ii) that the proximity operators* $\mathrm{prox}_{\gamma/w_i g_i}$ *and* ∇f *are computed with errors that are summable in* $\mathcal{H}$. *Then,*

- $x_k \rightharpoonup x \in \mathcal{M}^\star$.
- $\nabla f(x_k) \to \nabla f(x)$.
- *Suppose that* $\mathrm{int}(\mathcal{M}^\star) \neq \emptyset$. *Then,* $x_k \to x \in \mathcal{M}^\star$.
- *If* f *is uniformly convex then* $x_k \to x$, *the unique global minimizer.*

- Convergence of the non-stationary version, i.e. varying sequence $(\gamma_k)_{k\in\mathbb{N}}$, can also be established if $(\gamma_k - \gamma)_{k\in\mathbb{N}}$ is absolutely summable.

Monotone operator splitting

Find the zeros of a maximal monotone operator :

$$0 \in Ax + \sum_{i=1}^{n} B_i x$$

- $A, B_i : \mathcal{H} \to 2^{\mathcal{H}}$ are maximal monotone ;
- A single-valued with $\beta A \in \mathcal{A}(\frac{1}{2})$, B_i simple $\forall i$;
- $\mathrm{zer}(A + \sum_i B_i) \neq \emptyset$.

A GFB for monotone operator splitting

Initialization : Choose $(z_{i,0})_{1\leq i\leq n} \in \boldsymbol{\mathcal{H}}$, $\gamma \in]0, 2\beta[$, a sequence $(\lambda_k)_k$ in $]0,1]$, weights $w_i \in]0,1]$ (e.g. $1/n$). Let $x_0 = \sum_{i=1}^n w_i z_{i,k}$.

Main iteration :

repeat

1. Compute the resolvant points (in parallel if desired) :

   for $i = 1$ *to* n **do**

   $$z_{i,k+1} = z_{i,k} + \lambda_k \left(\operatorname{prox}_{\gamma/w_i g_i} (2x_k - z_{i,k} - \gamma(\nabla f(x_k) + e_{2,k})) + e_{1,i,k} - x_k \right) .$$

2. Update by averaging :

   $$x_{k+1} = \sum_{i=1}^n w_i z_{i,k} .$$

3. $k \leftarrow k+1$.

until *Convergence* ;

Output : x_k.

The key: a fixed point equation

Theorem *Let* $\gamma > 0$*. Then* $x \in \operatorname{zer}(A + \sum_i B_i)$ *if and only if* $x = \sum_{i=1}^n w_i z_i$*, where* $\boldsymbol{\mathcal{H}} \ni z = (z_1, \cdots, z_n) \in \operatorname{Fix}(T)$*, and*

$$T = \underbrace{\left(\frac{\mathrm{I}_{\boldsymbol{\mathcal{H}}} + R_{\gamma\mathbf{B}} \circ R_{N_{\mathcal{S}}}}{2}\right)}_{T_1 \in \mathcal{A}\left(\frac{1}{2}\right)} \circ \underbrace{(\mathrm{I}_{\boldsymbol{\mathcal{H}}} - \gamma\mathbf{A} \circ J_{N_{\mathcal{S}}})}_{T_2 \in \mathcal{A}\left(\frac{\gamma}{2\beta}\right)}$$

where $\mathcal{S} = \{(y_1, \cdots, y_n) \in \boldsymbol{\mathcal{H}} : y_1 = y_2 = \cdots = y_n\}$*,* $N_{\mathcal{S}}$ *its normal cone,*

$$J_{N_{\mathcal{S}}} = \operatorname{proj}_{\mathcal{S}} : \boldsymbol{\mathcal{H}} \to \mathcal{S}, z \mapsto \mathbf{\Pi}\left(\sum_i w_i z_i\right),$$

$$\mathbf{\Pi} : \mathcal{H} \to \mathcal{S}, x \mapsto (x, \cdots, x) \quad (\textit{canonical isometry}),$$

$$\mathbf{A}, \mathbf{B} : \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}}, \quad \mathbf{B}z = \times_{i=1}^n w_i^{-1} B_i z_i, \quad \mathbf{A}(z) = \times_{i=1}^n A z_i \ .$$

Moreover, for $\gamma \in]0, 2\beta[$*,* $T \in \mathcal{A}(\alpha)$*, with* $\alpha = \max(2/3, 2/(1 + 2\beta/\gamma))$*.*

Convergence

A.1 $(\lambda_k)_{k\in\mathbb{N}}$ is a sequence in $]0,1/\alpha[$ such that $\sum_k \lambda_k(1-\alpha\lambda_k) = +\infty$, where $\alpha = \max(2/3, 2/(1+2\beta/\gamma))$, and $\sum_{t\in\mathbb{N}} \lambda_k(\|e_{1,t}\|_{\boldsymbol{\mathcal{H}}} + \|e_{2,t}\|_{\boldsymbol{\mathcal{H}}}) < +\infty$.

A.2 $\lambda_k \in]0,1]$ with $\liminf_k \lambda_k > 0$, and the errors are summable.

Theorem *Let $\gamma \in]0, 2\beta[$. Let $z^{(0)} \in \boldsymbol{\mathcal{H}}$. Define*

$$z_{k+1} = z_k + \lambda_k(T_1(T_2 z_k + e_{2,t}) + e_{1,t} - z_k) \ .$$

(i) Assume that either A.1 or A.2 hold. Then,

- $z_k \rightharpoonup z \in \mathrm{Fix}(T_1 \circ T_2)$*, and* $x_k \rightharpoonup x \in \mathrm{zer}(A + \sum_i B_i)$*.*
- $T_1 \circ T_2 z_k - z_k \to 0$*.*

(ii) Moreover, if A.2 holds, then

- $A(x_k) \to A(x)$*.*
- *If* $\mathrm{int}(\mathrm{Fix}(T_1 \circ T_2)) \neq \emptyset$ *(i.e.* $\mathrm{int}(\mathrm{zer}(A + \sum_i B_i)) \neq \emptyset$*). Then,* $x_k \to x \in \mathrm{zer}(A + \sum_i B_i)$*.*
- *Strong convergence holds if A is uniformly monotone.*

Special instances

- GFB encompasses many special cases :
  - $n = 1 \Rightarrow$ Classical Forward-Backward splitting ($\lambda_k \in]0,1]$) [Combettes04] :

$$x_{k+1} = x_k + \lambda_k (\operatorname{prox}_{\gamma g_1} \circ (\mathrm{I} - \gamma \nabla f)(x_k) - x_k) \ .$$

  - $f = 0 \Rightarrow$ Spingarn method [Spingarn83] and also parallel DR splitting on a product space [Combettes09] ($\lambda_k \in]0,2[$) :

$$z_{k+1} = \left(1 - \frac{\lambda_k}{2}\right) z_k + \frac{\lambda_k}{2} (\operatorname{rprox}_{\gamma / w_i g_i})_i \circ \operatorname{rproj}_{\mathcal{S}}(z_k) \ .$$

  - $f = \frac{1}{2} \|y - .\|^2$, $y \in \operatorname{Im}(\mathrm{I} + \sum_i \partial g_i) \Rightarrow$ Proximity operator of $\sum_i g_i$ by absorbing f in the g_i's and use proximal calculus.
- Non-relaxed stationary GFB can be derived from BD-HPE [MonteiroSvaiter10].

Outline

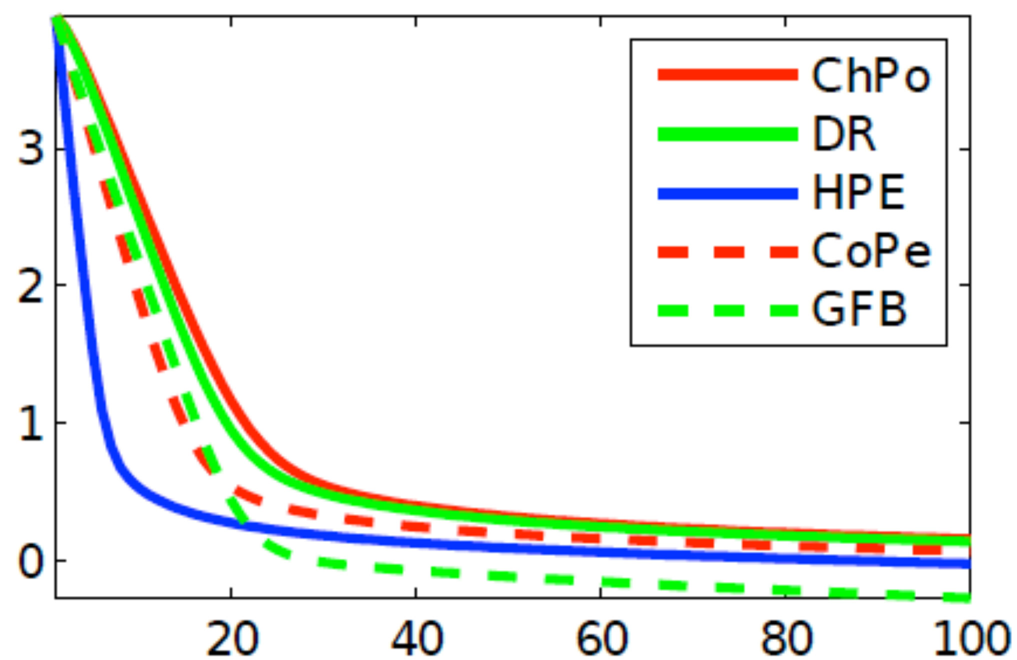
- Class of problems and motivations.
- A GFB splitting algorithm.
- Stylized applications.
- Conclusion and future work.

Deconvolution

Overlapping block-sparsity (TI-DWT)

$$\min_{x \in \mathbb{R}^P} \frac{1}{2} \|y - \mathrm{H}\Phi x\|_2^2 + \lambda \sum_{k=1}^{4} \|x_{\mathcal{B}_k}\|_{2,1}$$

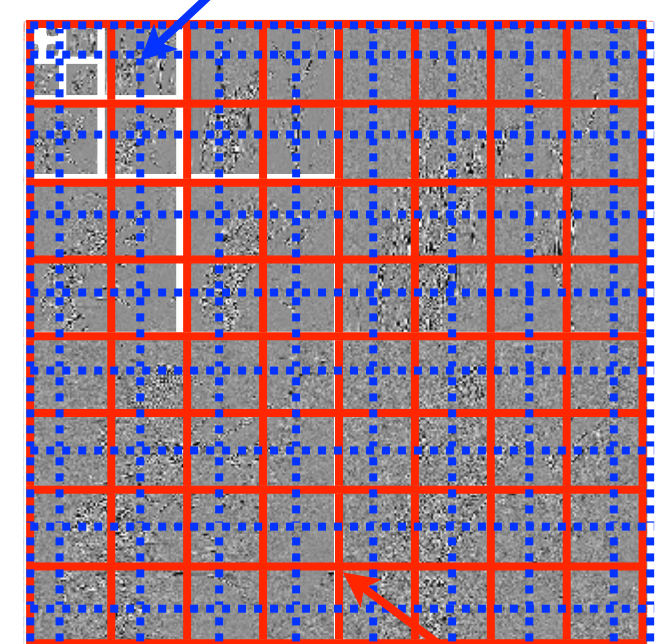
(a) $\log(\Psi - \Psi_{\min})$ vs. iteration #



(b) computing time

$t_{\text{ChPo}} = 153$ s
$t_{\text{DR}} = 95$ s
$t_{\text{HPE}} = 148$ s
$t_{\text{CoPe}} = 235$ s
$t_{\text{GFB}} = 73$ s

$$\|x_{\mathcal{B}_1}\|_{2,1} = \sum_{b \in \mathcal{B}_1} \|x_b\|_2$$



$$\|x_{\mathcal{B}_2}\|_{2,1} = \sum_{b \in \mathcal{B}_2} \|x_b\|_2$$

(c) LaBoute y_0

(d) $y = K y_0 + w$, 19.63 dB

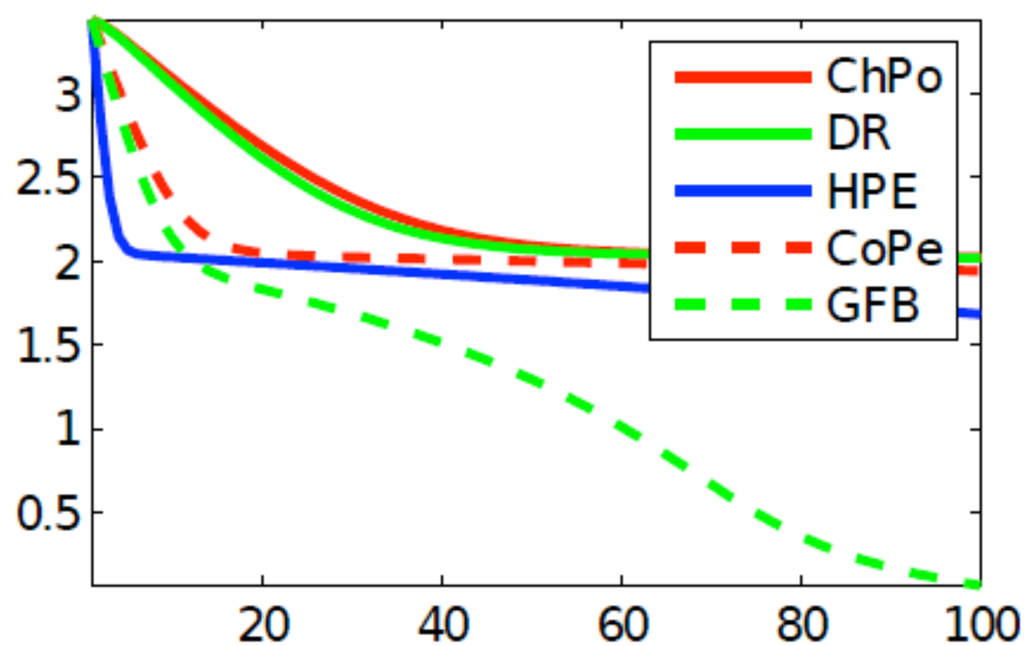
(e) $\widehat{y}_0 = W\widehat{x}$, 22.45 dB

Inpainting

Overlapping block-sparsity (TI-DWT)

$$\min_{x \in \mathbb{R}^P} \frac{1}{2} \|y - \mathrm{M}\Phi x\|_2^2 + \lambda \sum_{k=1}^{4} \|x_{\mathcal{B}_k}\|_{2,1}$$

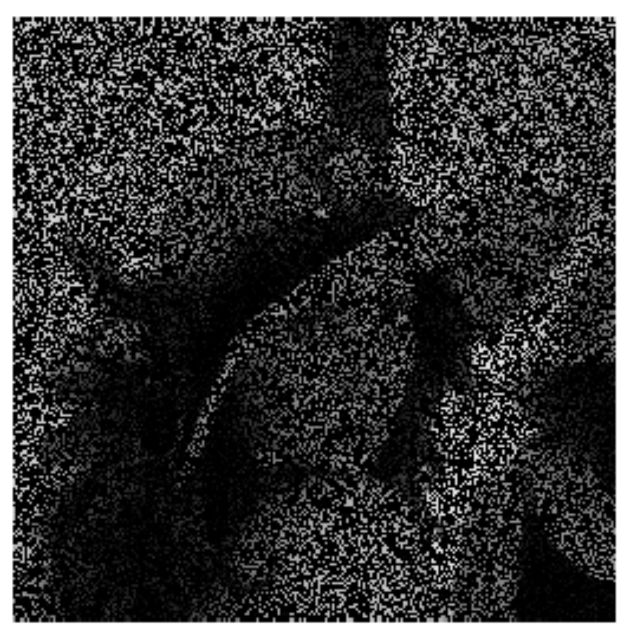
(a) $\log(\Psi - \Psi_{\min})$ vs. iteration #



(b) computing time

$t_{\text{ChPo}} = 229$ s
$t_{\text{DR}} = 219$ s
$t_{\text{HPE}} = 352$ s
$t_{\text{CoPe}} = 340$ s
$t_{\text{GFB}} = 203$ s

(c) LaBoute y_0



(d) $y = My_0 + w$, 1.54 dB

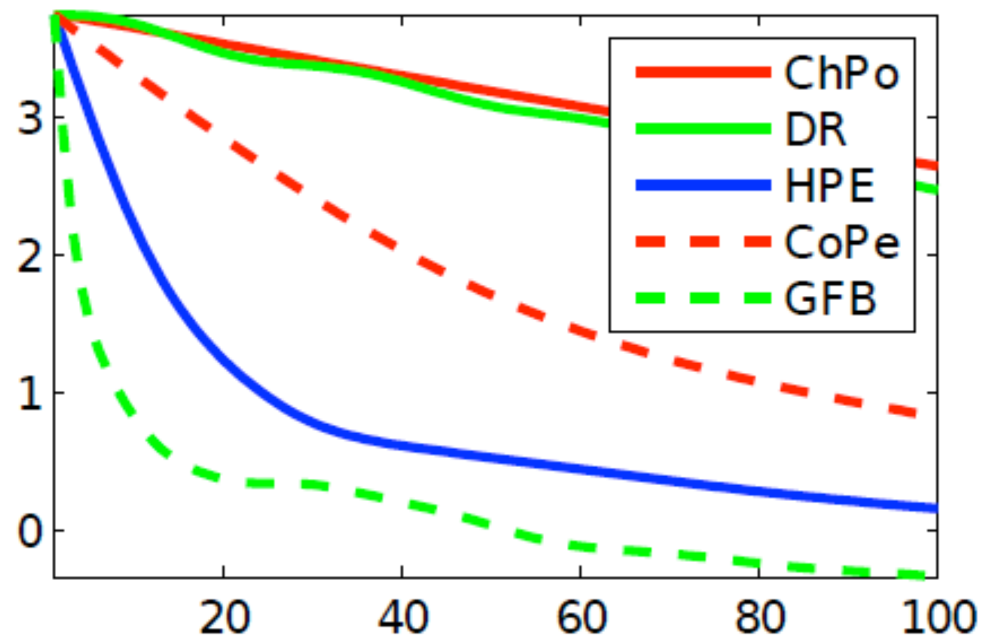
(e) $\widehat{y}_0 = W\widehat{x}$, 21.66 dB

Deconvolution and inpainting

Overlapping block-sparsity and TV

$$\min_{x \in \mathbb{R}^P} \frac{1}{2} \|y - \mathrm{MH}\Phi x\|_2^2 + \lambda \sum_{k=1}^{4} \|x_{\mathcal{B}_k}\|_{2,1} + \mu \|\Phi x\|_{\mathrm{TV}}$$

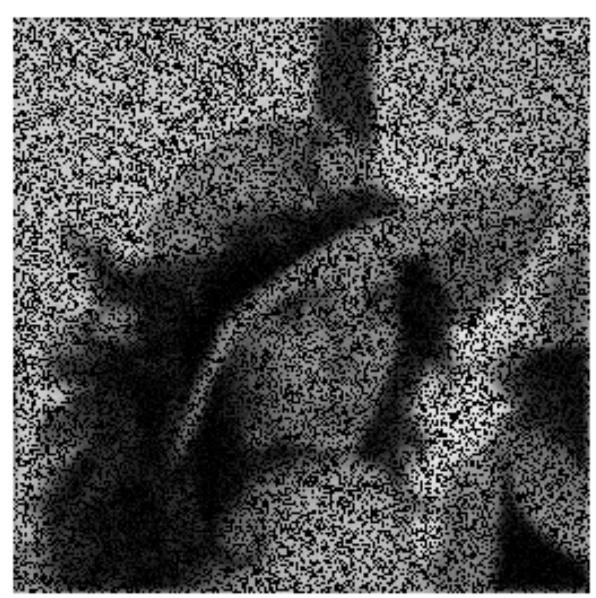
(a) $\log(\Psi - \Psi_{\min})$ vs. iteration #



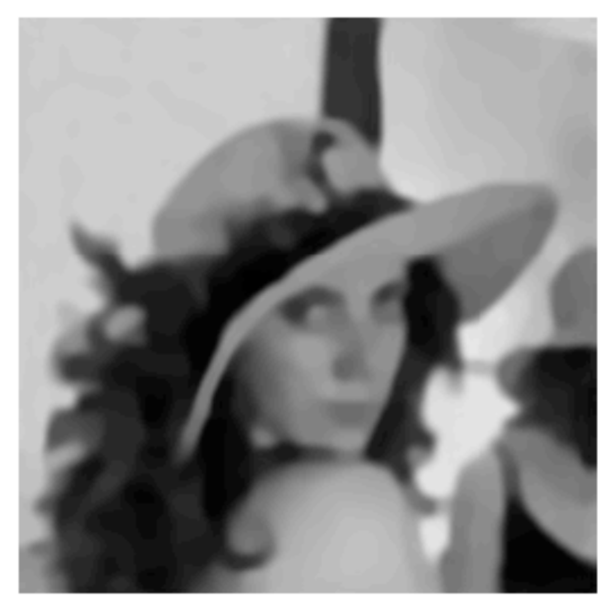
(b) computing time

$t_{\text{ChPo}} = 358$ s
$t_{\text{DR}} = 294$ s
$t_{\text{HPE}} = 409$ s
$t_{\text{CoPe}} = 441$ s
$t_{\text{GFB}} = 286$ s

(c) LaBoute y_0



(d) $y = MKy_0 + w$, 3.93 dB



(e) $\widehat{y}_0 = W\widehat{x}$, 22.48 dB

Many other problems can be solved within this framework.

Outline

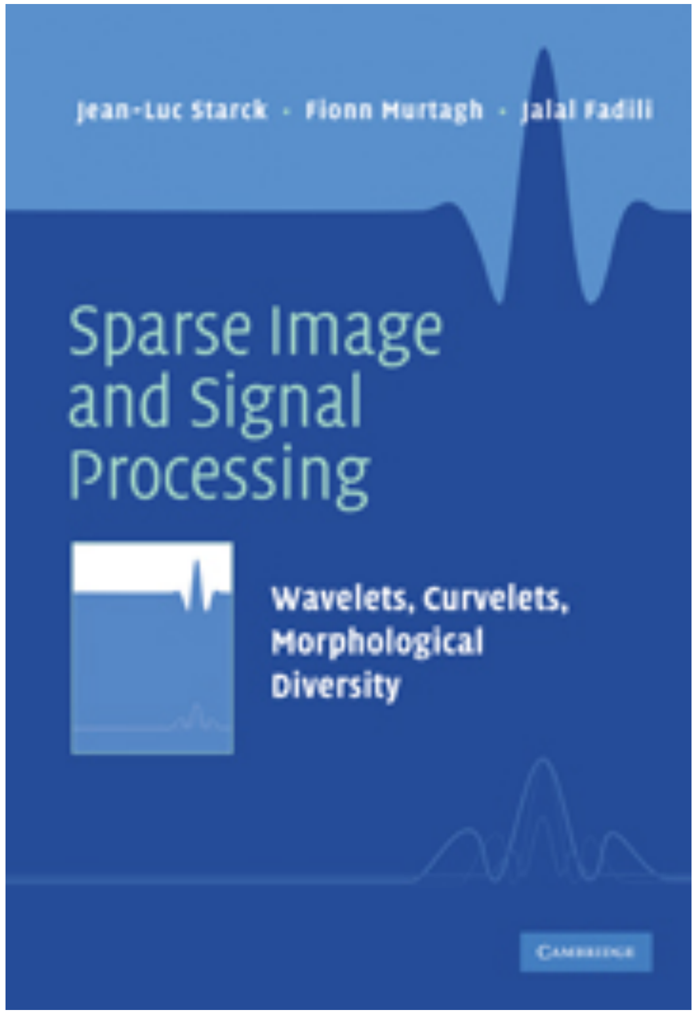
- Class of problems and motivations.
- A GFB splitting algorithm.
- Stylized applications.
- Conclusion and future work.

Take away messages

- Convex analysis and monotone operator splitting are a powerful framework for solving sparse recovery problems, non-necessarily smooth.
- A new splitting algorithm that exploits the structure of the problem (smoothness+simplicity).
- A fast solver for large-scale problems with theoretical guarantees (convergence, robustness).
- Convergence rate (for special instances).
- Acceleration.

Extended experiments, toolboxes available

http://www.greyc.ensicaen.fr/~jfadili
http://www.sparsesignalrecipes.info
http://www.ceremade.dauphine.fr/~peyre/numerical-tour/tours/



Thanks

Any questions ?