A Generalized Forward-Backward Splitting

Jalal Fadili

Normandie Université, ENSICAEN, GREYC CNRS UMR 6072

Joint work with

Hugo Raguet, Gabriel Peyré and Jingwei Liang

SIAM IS'16

Class of composite 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 : $(0, \cdots, 0) \in \mathrm{sri}(\{(x - y_1, \cdots, x - y_n) : x \in \mathcal{H} \text{ and } y_i \in \mathrm{dom}(g_i)\})$;
  - Set of minimizers $\mathcal{M}^\star \neq \emptyset$.

Class of composite problems

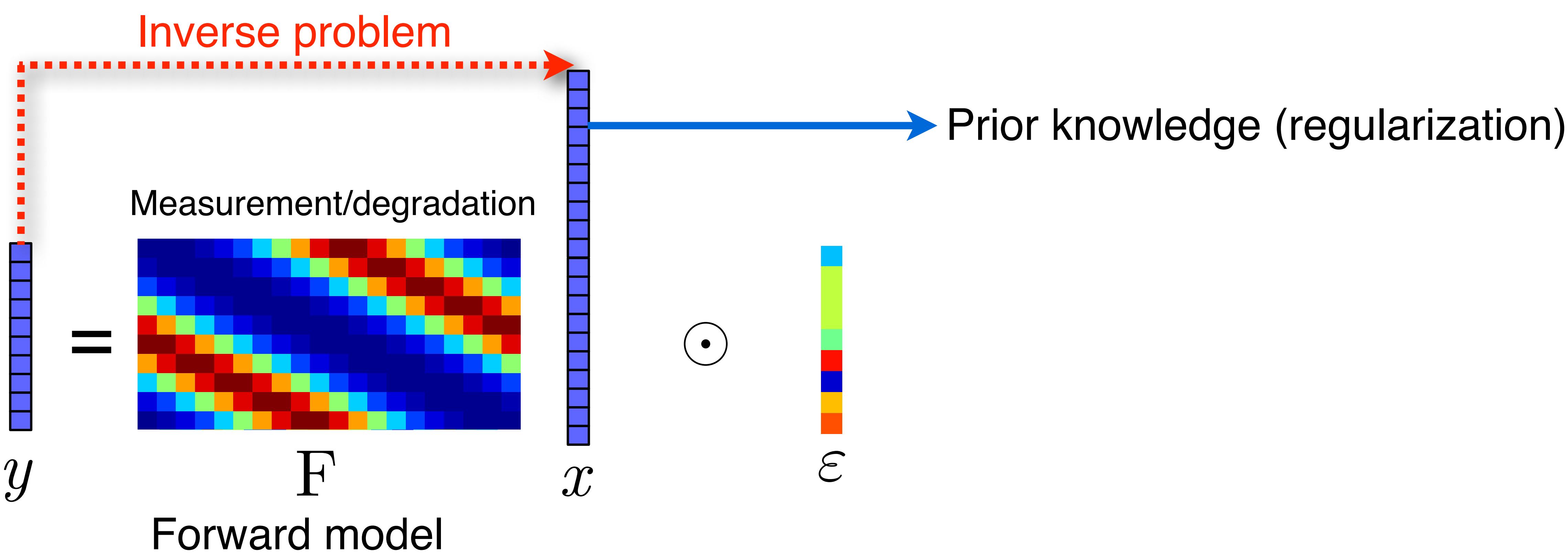
$$\min_{x\in\mathcal{H}} f(x) + \sum_{i=1}^{n} g_i(x)$$

- 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 : $(0, \cdots, 0) \in \mathrm{sri}(\{(x - y_1, \cdots, x - y_n) : x \in \mathcal{H} \text{ and } y_i \in \mathrm{dom}(g_i)\})$;
  - 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 (proximity operator easy to compute) and f smooth.
  - Deal with large scale data.
  - Avoid nested algorithms.

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..

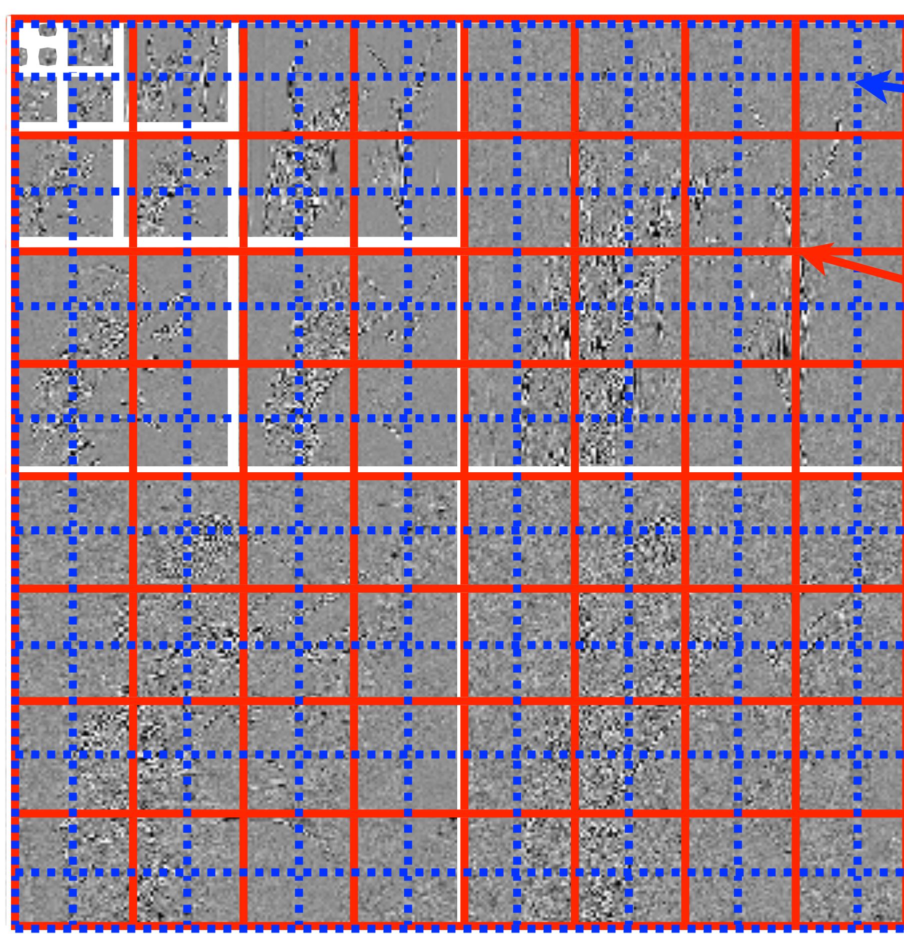
Motivations

- 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..

Monotone operator splitting

Find the zeros of a maximal monotone operator :

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

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

Monotone operator splitting

Find the zeros of a maximal monotone operator :

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

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

Example (Convex programming) : $A = \nabla f$, $B_i = \partial g_i$.

Outline

- A GFB splitting algorithm.
- Convergence.
- Stylized applications.
- Conclusion and future work.

Outline

- A GFB splitting algorithm.
- Convergence.
- Stylized applications.
- Conclusion and future work.

Monotone operator splitting

Find the zeros of a maximal monotone operator :

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

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

$\rightharpoonup$ and $\to$ are weak and strong convergence in $\mathcal{H}$.

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 $\sum_i w_i \langle x_i, y_i \rangle$.

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

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

A GFB for monotone operator splitting

Initialization : Choose $(z_{i,0})_{1 \le i \le n} \in \boldsymbol{\mathcal{H}}$, $\gamma_k \in [\epsilon, 2\beta - \epsilon]$, a sequence $(\lambda_k)_k$ in $[\epsilon, 1/\alpha]$, 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 resolvent points (in parallel if desired) :

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

$$z_{i,k+1} = z_{i,k} + \lambda_k \Big(\overbrace{J_{\gamma_k / w_i B_i}}^{\text{Implicit step}} \big(2x_k - z_{i,k} - \gamma_k (\overbrace{A(x_k)}^{\text{Explicit step}} + \underbrace{e_{2,k}}_{\text{Errors}})\big) + \underbrace{e_{1,i,k}}_{\text{Errors}} - x_k \Big) \ .$$

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.

Resolvent : $J_{\mu B_i} = (\mathrm{Id} + \mu B_i)^{-1}$

A GFB for convex optimization

Initialization : Choose $(z_{i,0})_{1 \leq i \leq n} \in \boldsymbol{\mathcal{H}}$, $\gamma_k \in [\epsilon, 2\beta - \epsilon]$, a sequence $(\lambda_k)_k$ in $[\epsilon, 1/\alpha]$, 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 resolvent points (in parallel if desired) :

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

   Implicit step — Explicit step

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

   Errors

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.

Proximity operator : $\operatorname{prox}_{\mu g_i} = (\mathrm{Id} + \mu \partial g_i)^{-1}$

Outline

- A GFB splitting algorithm.
- Convergence.
- Stylized applications.
- Conclusion and future work.

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* $z = (z_1, \cdots, z_n) \in \operatorname{Fix}(T)$, *and*

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

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* $\alpha = \frac{2\beta}{4\beta - \gamma}$.

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 $\alpha = \frac{2\beta}{4\beta-\gamma}$, and $\sum_{t\in\mathbb{N}} \lambda_k(\|e_{1,k}\|_{\boldsymbol{\mathcal{H}}} + \|e_{2,k}\|_{\boldsymbol{\mathcal{H}}}) < +\infty$.

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

A.3 $(\gamma_k)_{k\in\mathbb{N}}$ s.t. $0 < \underline{\gamma} \leq \gamma_k \leq \overline{\gamma} < 2\beta$, $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, and $(\lambda_k\,|\gamma_k - \gamma|)_{k\in\mathbb{N}}$ is summable.

Theorem *Let $\gamma \in]0,2\beta[$. Let $z_0 \in \boldsymbol{\mathcal{H}}$. Define*

$$z_{k+1} = z_k + \lambda_k(T_{1,\gamma}(T_{2,\gamma}z_k + e_{2,k}) + e_{1,k} - z_k)\ .$$

(i) Assume that either (A.1 or A.2) and A.3 hold. Then,

- $z_k \rightharpoonup z \in \mathrm{Fix}(T_{1,\gamma} \circ T_{2,\gamma})$*, and* $x_k \rightharpoonup x \in \mathrm{zer}(A + \sum_i B_i)$.
- $T_{1,\gamma} \circ T_{2,\gamma} z_k - z_k \to 0$.

(ii) Moreover, if A.1, A.2 and A.3 hold, then

- $A(x_k) \to A(x)$.
- $x_k \to x$ *if one of the following holds :*
  - A *is uniformly monotone,*
  - $\times_{i=1}^n w_i^{-1} B_i$ *is uniformly monotone. The latter is true for instance if* $\forall i \in \{1,\ldots,n\}$, B_i *is uniformly monotone with its modulus being also subadditive or convex.*

Convergence: Rates

$$u_{i,k+1} = J_{\frac{\gamma}{\omega_i} B_i}(2x_k - z_{i,k} - \gamma_k A x_k),\ i \in \{1,\dots,n\}$$

Theorem *(i) Suppose that A.1-A.2 hold, that* $0 < \inf_{k\in\mathbb{N}} \lambda_k \le \sup_{k\in\mathbb{N}} \lambda_k < \frac{4\beta-\gamma}{2\beta}$*,* $((k+1)\left\|e_{1,k}\right\|)_{k\in\mathbb{N}} \in \ell^1_+$ *and* $\forall i$*,* $((k+1)\left\|e_{2,i,k}\right\|)_{k\in\mathbb{N}} \in \ell^1_+$*, then*

(a)

$$\operatorname{dist}\left(0, \sum_i B_i u_{i,k+1} + A\left(\sum_i \omega_i u_{i,k+1}\right)\right) = O(\sqrt{1/k}).$$

(b) If moreover $e_{j,i,k} \equiv 0$*, then*

$$\operatorname{dist}\left(0, \sum_i B_i u_{i,k+1} + A\left(\sum_i \omega_i u_{i,k+1}\right)\right) = o(\sqrt{1/k}).$$

(ii) If A *is strongly monotone, then* x_k *converges to* $x^\star$ *linearly.*

Special instances

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

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

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

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

    Strong convergence appears new.
  - $f = \frac{1}{2}\|y - .\|^2$, $y \in \mathrm{Im}(\mathrm{I} + \sum_i \partial g_i) \Rightarrow$ 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].

Extensions

- Find $x \in \mathcal{H}$ such that

$$\min_{x\in\mathcal{H}} f(x) + g_i \circ L_i(x) \iff 0 \in \nabla f(x) + \sum_i L_i^* \circ \partial g_i(L_i x) \ ,$$

$\nabla f \in \beta\text{-}\mathrm{Lip}(\mathcal{H})$, and $\forall i$, g_i simple and L_i bounded linear operator on $\mathcal{H}$.

Approach 1

- Decompose $g_i \circ L_i = \sum_k g_{i,k} \circ L_{i,k}$, $L_{i,k}$ a tight frame (e.g. overlapping block sparsity, convolution with a FIR, etc.) ;
- Apply the GFB.

Approach 2 $L_i^* \circ L_i$ (or $L_i \circ L_i^*$) easily diagonalized :

- Introduce auxiliary variables : find $x \in \mathcal{H}$ and $(u_1, \cdots, u_n) \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_n$

$$\min_{x,(u_i)_i} f(x) + \sum_i \Big(g_i(u_i) + \imath_{\ker\,[\mathrm{I}_{\mathcal{K}_i} \ -L_i]}(u_i, x) \Big) \ .$$

- Apply the GFB.

Outline

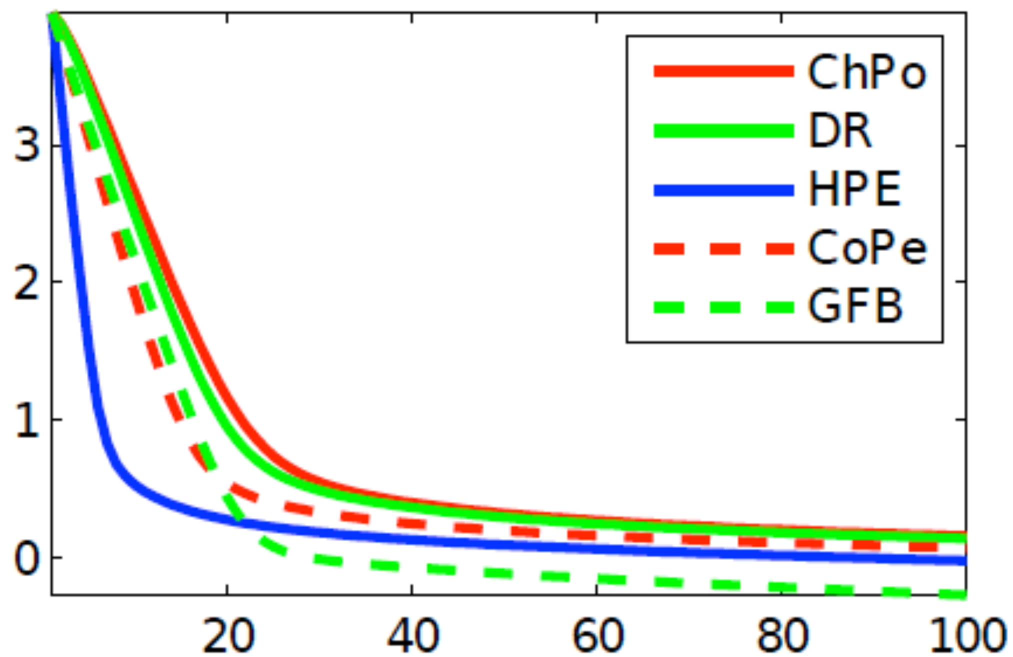
- A GFB splitting algorithm.
- Convergence.
- Stylized applications.
- Conclusion and future work.

Deconvolution

Overlapping block-sparsity (TI-DWT)

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

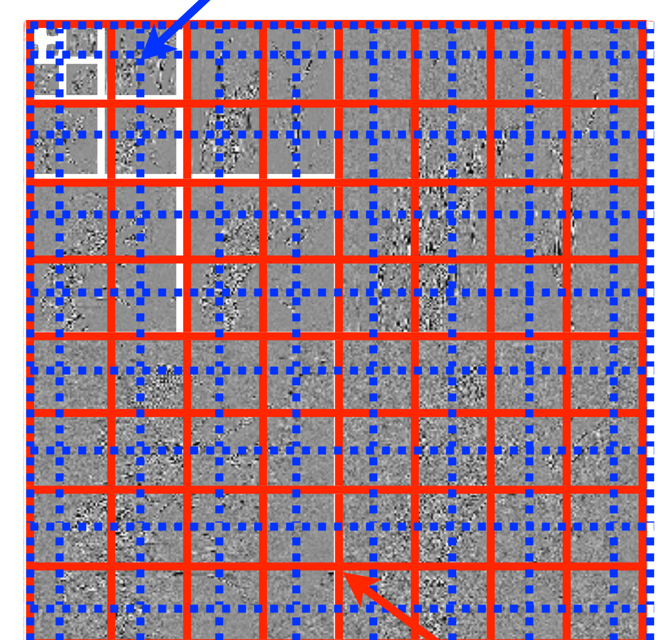
(a) $\log(\Psi - \Psi_{\min})$ vs. iteration #



(b) computing time

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

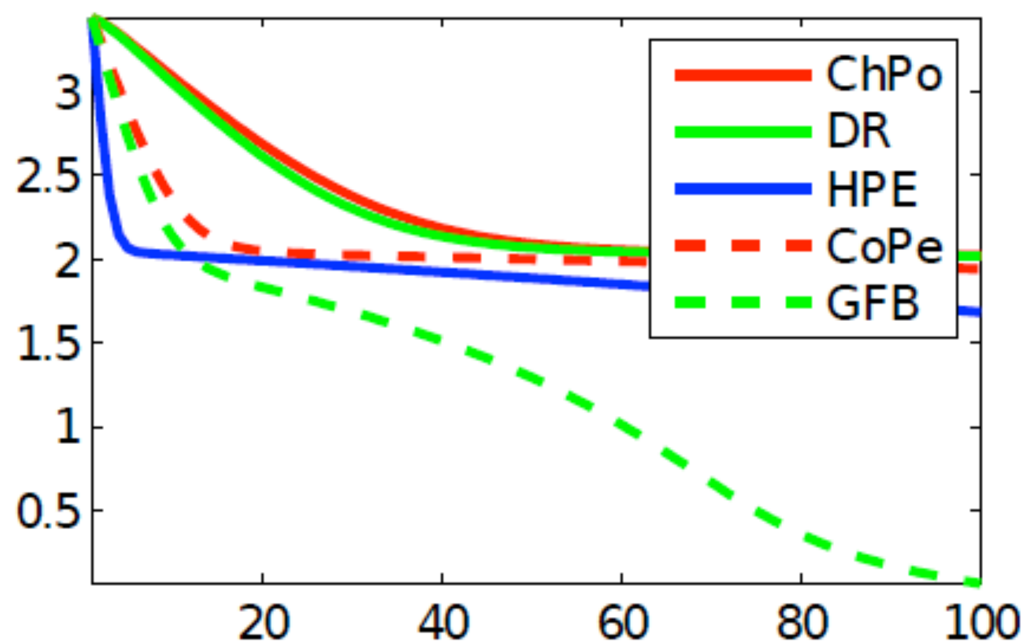
(e) $\widehat{y}_0 = W\widehat{x}$, 22.45 dB

Inpainting

Overlapping block-sparsity (TI-DWT)

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

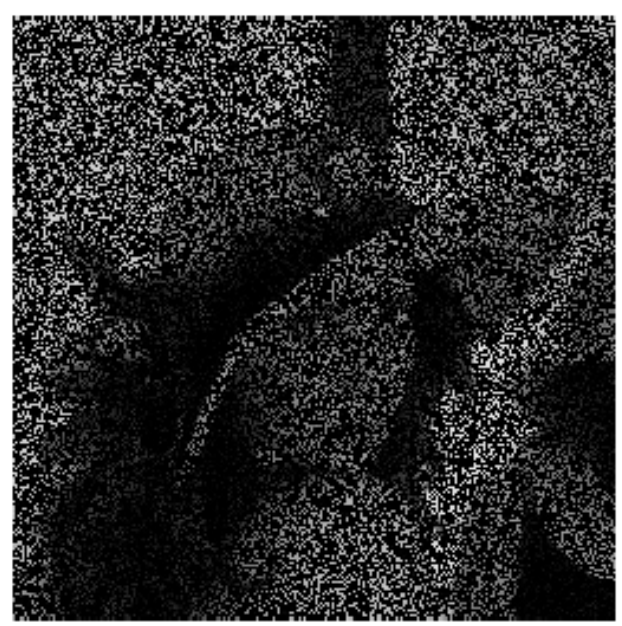
(a) $\log(\Psi - \Psi_{\min})$ vs. iteration #



(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

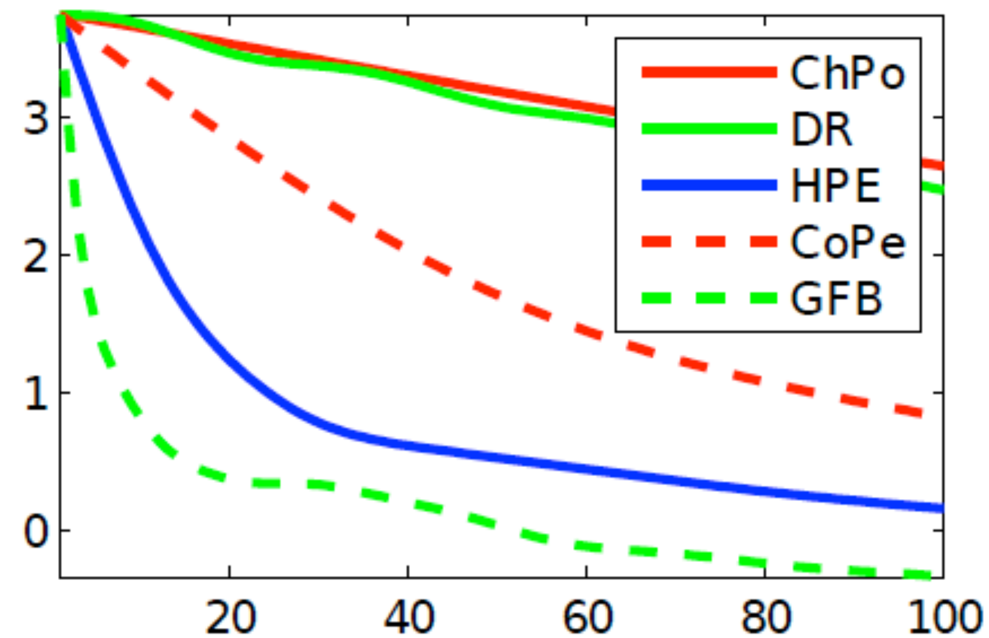
(e) $\widehat{y}_0 = W\widehat{x}$, 21.66 dB

Deconvolution and inpainting

Overlapping block-sparsity and TV

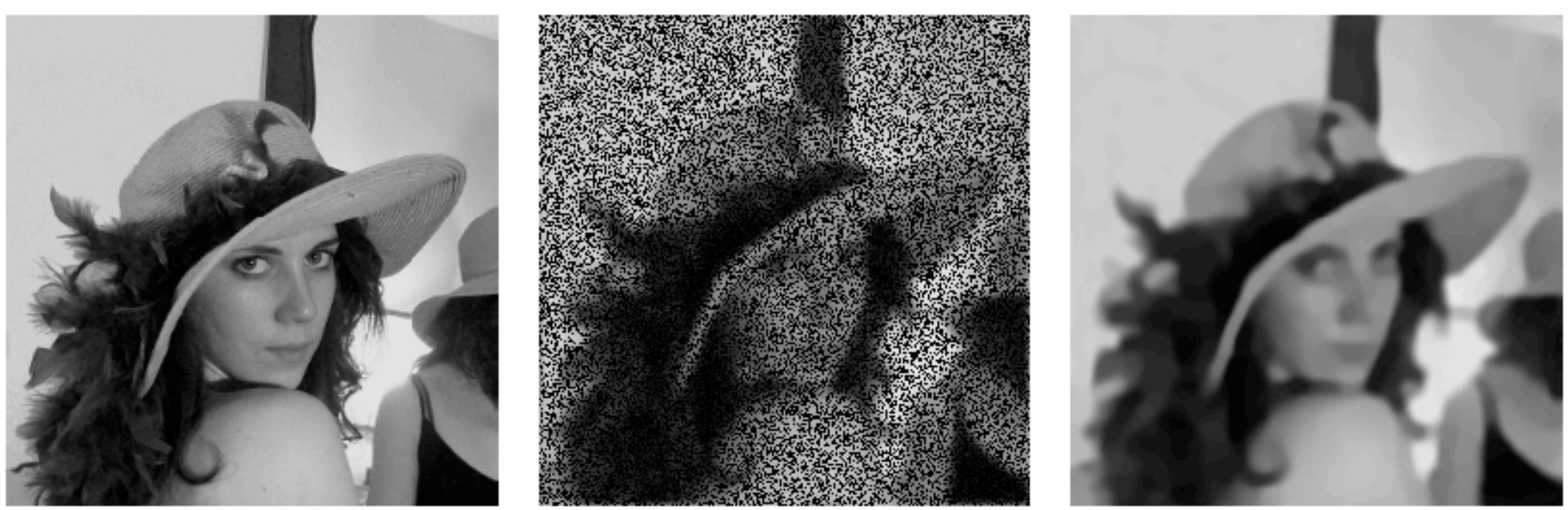
$$\min_{x\in\mathbb{R}^P} \frac{1}{2}\|y - \mathrm{MH}\Phi x\|_2^2 + \lambda \sum_{k=1}^{4} \|x_{\mathcal{B}_k}\|_{2,1} + \mu \|\Phi x\|_{\mathrm{TV}}$$

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



(b) computing time

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

- A GFB splitting algorithm.
- Convergence.
- 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, rates).
- Acceleration (inertial, variable metric): done.

Preprints on arxiv and papers on

https://fadili.users.greyc.fr/

Thanks
Any questions ?