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:

1. \(f, g_i : \mathcal{H} \rightarrow \mathbb{R} \cup \{+\infty\}, f, g_i \in \Gamma_0(\mathcal{H})\);
2. \(f \in C^1\) with \(\beta\)-Lipschitz gradient, all \(g_i\)'s are simple.
3. Domain qualification condition: \((0, \cdots, 0) \in \text{sri}\left(\{(x - y_1, \cdots, x - y_n) : x \in \mathcal{H} \text{ and } y_i \in \text{dom}(g_i)\}\right)\);
4. Set of minimizers \(\mathcal{M}^* \neq \emptyset\).
Class of composite problems

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

**Assumptions:**
- \( f, g_i : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}, f, g_i \in \Gamma_0(\mathcal{H}) \);
- \( f \in C^1 \) with \( \beta \)-Lipschitz gradient, all \( g_i \)'s are simple.
- Domain qualification condition : \((0, \cdots, 0) \in \text{sri}(\{(x - y_1, \cdots, x - y_n) : x \in \mathcal{H} \text{ and } y_i \in \text{dom}(g_i)\}) \);
- Set of minimizers \( \mathcal{M}^* \neq \emptyset \).

**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.:

\[
\min_{x \in \mathcal{H}} \left( f(x) + g_1(x) + \cdots + g_n(x) \right)
\]

Data fidelity  Regularization, constraints

Inverse problem

Measurement/degradation

Forward model

Prior knowledge (regularization)

Typical models

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

Inverse problems with mixed regularization, e.g.:

\[
\min_{x \in \mathcal{H}} \underbrace{f(x)} + \underbrace{g_1(x) + \cdots + g_n(x)}
\]

- Data fidelity
- Regularization, constraints

Inverse problems with structured sparsity, e.g.:

\[
\min_{x \in \mathcal{H}} \underbrace{f(x)} + \underbrace{g_1(x) + \cdots + g_n(x)}
\]

- Data fidelity
- Structured sparsity (e.g. \(\ell_p - \ell_q\) norm on overlapping blocks)

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

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

Inverse problems with mixed regularization, e.g.:

\[
\min_{x \in \mathcal{H}} \ f(x) + g_1(x) + \cdots + g_n(x)
\]

Data fidelity  Regularization, constraints

Inverse problems with structured sparsity, e.g.:

\[
\min_{x \in \mathcal{H}} \ f(x) + g_1(x) + \cdots + g_n(x)
\]

Data fidelity  Structured sparsity (e.g. \(l_p - l_q\) 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 \( \beta A \in A(\frac{1}{2}) \), \( B_i \) simple \( \forall i \);
- \( \text{zer}(A + \sum_i B_i) \neq \emptyset \).
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 \( \beta A \in A(\frac{1}{2}) \), \( B_i \) simple \( \forall i \);
- \( \text{zer}(A + \sum_i B_i) \neq \emptyset \).

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:

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

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

\( \rightharpoonup \) and \( \rightharpoonup \) are weak and strong convergence in \( \mathcal{H} \).

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

\[ S = \{(x_1, \cdots, x_n) \in \mathcal{H} : x_1 = x_2 = \cdots = x_n \}, \]
\[ \Pi : \mathcal{H} \rightarrow S, x \mapsto (x, \cdots, x) \quad \text{(canonical isometry)} \].
A GFB for monotone operator splitting

Initialization: Choose \((z_{i,0})_{1 \leq i \leq n} \in \mathcal{H}, \gamma_k \in [\epsilon, 2\beta - \epsilon]\), 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
     \[
     z_{i,k+1} = z_{i,k} + \lambda_k \left( J_{\gamma_k/w_i B_i} \left( 2x_k - z_{i,k} - \gamma_k (A(x_k) + e_{2,k}) \right) + e_{1,i,k} - x_k \right).
     \]
   Implicit step
   Explicit step
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: \(J_{\mu B_i} = (\text{Id} + \mu B_i)^{-1}\)
A GFB for convex optimization

Initialization: Choose \((z_i, 0)_{1 \leq i \leq n} \in \mathcal{H}, \gamma_k \in [\epsilon, 2\beta - \epsilon], \) a sequence \((\lambda_k)_k\) in \([\epsilon, 1/\alpha]\), weights \(w_i \in ]0, 1]\) (e.g. \(1/n\)). Let \(x_0 = \sum_{i=1}^n w_iz_{i,k}\).

Main iteration:

repeat

1. Compute the resolvent points (in parallel if desired):
   for \(i = 1\) to \(n\) do
     \[ z_{i,k+1} = z_{i,k} + \lambda_k \left( \text{prox}_{\gamma_k/w_ig_i} \left( 2x_k - z_{i,k} - \gamma_k \left( \nabla f(x_k) + e_{2,k} \right) \right) + e_{1,i,k} - x_k \right). \]

2. Update by averaging:
   \[ x_{k+1} = \sum_{i=1}^n w_iz_{i,k}. \]

3. \(k \leftarrow k + 1.\)

until Convergence;

Output: \(x_k\).

Proximity operator: \(\text{prox}_{\mu g_i} = (\text{Id} + \mu \partial g_i)^{-1}\)
Outline

- A GFB splitting algorithm.
- Convergence.
- Stylized applications.
- Conclusion and future work.
The key: a fixed point equation

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

$$T = \left( \frac{I_{\mathcal{H}} + R_{\gamma B} \circ R_{NS}}{2} \right) \circ \left( I_{\mathcal{H}} - \gamma A \circ J_{NS} \right)$$

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

$$J_{NS} = \text{proj}_S : \mathcal{H} \to S, z \mapsto \Pi \left( \sum_i w_i z_i \right),$$

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

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

Moreover, for $\gamma \in ]0, 2\beta[$, $T \in A(\alpha)$, 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 \(\sum_{t \in \mathbb{N}} \lambda_k (\|e_{1,k}\|_\mathcal{H} + \|e_{2,k}\|_\mathcal{H}) < +\infty\).

A.2 \(\lambda_k \in ]0, 1]\) with \(\lim \inf_k \lambda_k > 0\), and the errors are summable.

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

**Theorem** Let \(\gamma \in ]0, 2\beta[\). Let \(z_0 \in \mathcal{H}\). 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,
- \(z_k \rightharpoonup z \in \text{Fix}(T_{1,\gamma} \circ T_{2,\gamma})\), and \(x_k \rightharpoonup x \in \text{zer}(A + \sum_i B_i)\).
- \(T_{1,\gamma} \circ T_{2,\gamma}z_k - z_k \rightharpoonup 0\).

(ii) Moreover, if A.1, A.2 and A.3 hold, then
- \(A(x_k) \rightharpoonup A(x)\).
- \(x_k \rightharpoonup x\) 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

\[ u_{i,k+1} = J\frac{\gamma}{\omega_i} B_i (2x_k - z_{i,k} - \gamma_k A x_k), \quad i \in \{1, \ldots, n\} \]

**Theorem**

(i) Suppose that A.1-A.2 hold, that 
\[
0 < \inf_{k \in \mathbb{N}} \lambda_k \leq \sup_{k \in \mathbb{N}} \lambda_k < \frac{4\beta - \gamma}{2\beta}, \quad ((k + 1) \| e_{1,k} \|)_{k \in \mathbb{N}} \in \ell_+^1 \quad \text{and} \quad \forall i, \quad ((k + 1) \| e_{2,i,k} \|)_{k \in \mathbb{N}} \in \ell_+^1, \quad \text{then}
\]

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

(b) If moreover \( e_{j,i,k} \equiv 0 \), then
\[
\text{dist} \left( 0, \sum_i B_i u_{i,k+1} + A \left( \sum_i \omega_i u_{i,k+1} \right) \right) = o(\sqrt{1/k}).
\]

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

GFB encompasses many special cases:

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

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

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

$$z_{k+1} = \left(1 - \frac{\lambda_k}{2}\right) z_k + \frac{\lambda_k}{2} (r\text{prox}_{\gamma / w_i g_i})_i \circ r\text{proj}_S(z_k).$$

Strong convergence appears new.

- $f = \frac{1}{2} \|y - .\|^2, y \in \text{Im}(I + \sum_i \partial g_i) \Rightarrow$ 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

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

\( \nabla f \in \beta\text{-Lip}(\mathcal{H}) \), 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 \((u_1, \ldots, u_n) \in \mathcal{K}_1 \times \ldots \times \mathcal{K}_n\)

\[
\min_{x, (u_i)} f(x) + \sum_i \left( g_i(u_i) + \nu_{\ker [I_{\mathcal{K}_i} - L_i]}(u_i, x) \right).
\]
- Apply the GFB.
Outline

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

Overlapping block-sparsity (TI-DWT)

\[
\min_{x \in \mathbb{R}^P} \frac{1}{2} \| y - H\Phi x \|_2^2 + \lambda \sum_{k=1}^{4} \| x_{B_k} \|_{2,1}
\]

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

(b) computing time

\[
\begin{align*}
    t_{\text{ChPo}} &= 153 \text{ s} \\
    t_{\text{DR}} &= 95 \text{ s} \\
    t_{\text{HPE}} &= 148 \text{ s} \\
    t_{\text{CoPe}} &= 235 \text{ s} \\
    t_{\text{GFB}} &= 73 \text{ s}
\end{align*}
\]

(c) LaBoute $y_0$

(d) $y = Ky_0 + w$, 19.63 dB

(e) $\tilde{y}_0 = W\hat{x}$, 22.45 dB
**Inpainting**

**Overlapping block-sparsity (TI-DWT)**

\[
\min_{x \in \mathbb{R}^P} \frac{1}{2} \| y - M\Phi x \|_2^2 + \lambda \sum_{k=1}^{\frac{1}{4}} \| x_{B_k} \|_{2,1}
\]

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

(b) computing time

\[
\begin{align*}
    t_{\text{ChPo}} &= 229 \text{ s} \\
    t_{\text{DR}} &= 219 \text{ s} \\
    t_{\text{HPE}} &= 352 \text{ s} \\
    t_{\text{CoPe}} &= 340 \text{ s} \\
    t_{\text{GFB}} &= 203 \text{ s}
\end{align*}
\]

(c) LaBoute \( y_0 \)

(d) \( y = My_0 + w, \ 1.54 \text{ dB} \)

(e) \( \widehat{y}_0 = W\widehat{x}, \ 21.66 \text{ dB} \)
Deconvolution and inpainting

Overlapping block-sparsity and TV

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

(a) log($\Psi - \Psi_{\text{min}}$) 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) $\hat{y}_0 = W\hat{x}$, 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?