Nonlinear Forward-Backward Splitting with Projection Correction

Pontus Giselsson
Talk Message

• New algorithm called nonlinear forward-backward splitting
• Versatile algorithm with many special cases, e.g.:
  • Forward-backward splitting
  • Forward-backward-(half-)forward splitting – FB(H)F
  • Chambolle-Pock
  • Vu-Condat
  • Douglas-Rachford, ADMM, and proximal ADMM
  • Synchronous projective splitting
  • Asymmetric forward-backward adjoint splitting (AFBA)
  • A novel four operator splitting method
• FB(H)F is conservative special case
• FBF special case of backward method (without forward step)
• We propose new long-step FB(H)F variations
• Synchronous projective splitting is long-step FBF on specific problem
Nonlinear Forward-Backward Splitting (NOFOB)

• Solves maximal monotone inclusion problems of the form

\[ 0 \in Ax + Cx, \]

\( A \) is maximally monotone and \( C \) is \( \frac{1}{\beta} \)-cocoercive w.r.t. \( \| \cdot \|_P \)

• Proposed algorithm (NOFOB)

\[ \hat{x}_k := (M_k + A)^{-1}(M_k - C)x_k \]

\[ H_k := \{ z : \langle M_k x_k - M_k \hat{x}_k, z - \hat{x}_k \rangle \leq \frac{\beta}{4} \| x_k - \hat{x}_k \|_P^2 \} \]

\[ x_{k+1} := (1 - \theta_k)x_k + \theta_k \Pi_{H_k}^S(x_k) \]

where

• \( M_k \) is single-valued and strongly monotone
• \( P, S \) are linear self-adjoint positive definite operators
• \( H_k \) is a halfspace that contains \( \text{zer}(A + C) \) but not \( x_k \) (strictly)
• \( \Pi_{H_k}^S \) is projection onto \( H_k \) in metric \( \| \cdot \|_S \)
• \( \theta_k \in (\epsilon, 2 - \epsilon) \) is relaxation parameter

• Algorithm is of separate and project type
• Steps explained in following slides
• First step requires one \( M_k \) application, \( H_k \) construction another
Algorithm Steps

\[ \hat{x}_k := (M_k + A)^{-1}(M_k - C)x_k \]
\[ H_k := \{ z : \langle M_k x_k - M_k \hat{x}_k, z - \hat{x}_k \rangle \leq \frac{\beta}{4} \| x_k - \hat{x}_k \|_P^2 \} \]
\[ x_{k+1} := (1 - \theta_k)x_k + \theta_k \Pi_{H_k}^S(x_k) \]

1. Nonlinear forward-backward step\(^1\) on \( A + C \) with kernel \( M_k \)
2. Construction of \( H_k \) that contains solution set but not \( x_k \)
3. Projection from \( x_k \) onto separating hyperplane

\(^1\)Proposed at same time by Combettes’ group without \( C \), i.e., \( (M_k + A)^{-1} \circ M_k \) called warped resolvent.
Nonlinear FB Map – Special cases

• First step in algorithm is nonlinear FB evaluation

\[ \hat{x}_k = (M_k + A)^{-1}(M_k - C)x_k \]

• Special cases:
  • \( M_k = \gamma^{-1} \text{Id} \) gives standard FB step:

\[ \hat{x}_k = (\gamma^{-1} \text{Id} + A)^{-1}(\gamma^{-1} \text{Id} - C)x_k = (\text{Id} + \gamma A)^{-1}(x_k - \gamma Cx_k) \]

  • \( M_k = \nabla g \) with \( g \) strictly convex gives Bregman FB step
Nonlinear FB Map – Properties

Let $T_{FB} := (M + A)^{-1}(M - C)$

(i) Fixed-point set of $T_{FB}$ equals $\text{zer}(A + C)$

(ii) Define the affine function $\psi_x$ for each $x$ as:

$$
\psi_x(z) := \langle Mx - MT_{FB}x, z - T_{FB}x \rangle - \frac{\beta}{2} \|x - T_{FB}x\|_P^2
$$

Then

- $\psi_x(z) \leq 0$ for all $z \in \text{zer}(A + C)$
- $\psi_x(x) > 0$ for all points $x \notin \text{zer}(A + C)$
- $\psi_x(x) \geq \sigma \|x - T_{FB}x\|^2$ for some $\sigma > 0$ if $M_k$ strongly monotone

Therefore, $H_k$ in the second step of the algorithm:

$$
H_k := \{z : \psi_{x_k}(z) \leq 0\}
$$

$$
= \{z : \langle M_kx_k - M_k\hat{x}_k, z - \hat{x}_k \rangle \leq \frac{\beta}{2} \|x_k - \hat{x}_k\|_P^2\}
$$

satisfies $\text{fix}T_{FB}^k \subseteq H_k$ and $x_k \notin H_k$, i.e., strict separation
The Projection

The third (last) step is relaxed projection in metric $\| \cdot \|_S$ onto $H_k$

$$x_{k+1} := (1 - \theta_k)x_k + \theta_k \Pi_{H_k}^S(x_k)$$

where

- projection is from previous point $x_k$
- linear projection metric operator $S$ is fixed
- $\theta_k$ is relaxation parameter
Convergence

• Consequences of separate and project principle:
  • $\| \cdot \|_S$-distance to fixed-point set decreasing (Fejer monotone)
  • Projection step length converges strongly to 0: $x_{k+1} - x_k \to 0$
• Convergence of algorithm if cuts are deep enough
• Weak convergence of method follows by standard arguments if

$$x_{k+1} - x_k \to 0 \implies T_{FB}^k x_k - x_k = \hat{x}_k - x_k \to 0$$

which holds if
• $M_k$ strongly monotone (easy to show)
• $M_k$ strictly monotone with some more assumptions and $C = 0$
**NOFOB with Explicit Projection**

- Projection onto separating hyperplane $H_k$ is
  
  $$z = x_k - \frac{\langle M_k x_k - M_k \hat{x}_k, x_k - \hat{x}_k \rangle - \frac{\beta}{4} \|x_k - \hat{x}_k\|_P^2}{\|M_k x_k - M_k \hat{x}_k\|_{S_{-1}}^2} S^{-1}(M_k x_k - M_k \hat{x}_k)$$

- Inserting into algorithm gives equivalent, more explicit, method

  $$\hat{x}_k := (M_k + A)^{-1}(M_k - C)x_k$$

  $$\mu_k := \frac{\langle M_k x_k - M_k \hat{x}_k, x_k - \hat{x}_k \rangle - \frac{\beta}{4} \|x_k - \hat{x}_k\|_P^2}{\|M_k x_k - M_k \hat{x}_k\|_{S_{-1}}^2}$$

  $$x_{k+1} := x_k - \theta_k \mu_k S^{-1}(M_k x_k - M_k \hat{x}_k)$$

- Algorithm converges with $\mu_k$ replaced by any $\hat{\mu}_k \in (0, \mu_k]$ (equivalent to algorithm with smaller relaxation parameter)
Constant-$\mu_k$ Variation

- Suppose that there exists $\mu$ such that for all $M_k$ and $x, y \in \mathcal{H}$:

$$
\mu \leq \frac{\langle M_k x - M_k y, x - y \rangle - \frac{\beta}{4} \| x - y \|^2}{\| M_k x - M_k y \|^2_{S^{-1}}}
$$

- $\mu_k$ in algorithm is exact local version with $x_k$ and $\hat{x}_k$:

$$
\mu_k := \frac{\langle M_k x_k - M_k \hat{x}_k, x_k - \hat{x}_k \rangle - \frac{\beta}{4} \| x_k - \hat{x}_k \|^2}{\| M_k x_k - M_k \hat{x}_k \|^2_{S^{-1}}}
$$

- Hence $\mu \in (0, \mu_k]$ and conservative special case of method is:

$$
\hat{x}_k := (M_k + A)^{-1}(M_k - C)x_k
$$

$$
x_{k+1} := x_k - \theta_k \mu S^{-1}(M_k x_k - M_k \hat{x}_k)
$$

where $\mu_k$ replaced by $\mu$ (alt. actual relaxation parameter is $\theta_k \frac{\mu}{\mu_k}$)

- If $C = 0$, $\mu$ is cocoercivity parameter that holds for all $M_k$
Forward-Backward splitting

- Let $M_k$ be linear symmetric and equal to projection kernel $S$
- Algorithm becomes (since $\mu = 1$ can be chosen)

\[ x_{k+1} := (1 - \theta_k)x_k + \theta_k(S + A)^{-1}(S - C)x_k \]

i.e., relaxed forward-backward splitting with kernel $S$
- If no relaxation, i.e., $\theta_k = 1$, we get forward-backward splitting

\[ x_{k+1} := (S + A)^{-1}(S - C)x_k \]

- Note that second application of $M_k$ is not needed anymore!
- Projection point is result of FB step – $\hat{x}_k$
- Since FB is special case, has the following special cases:
  - Chambolle-Pock
  - Vu-Condat
Symmetry and linearity of $M_k$

- If $M_k$ symmetric and linear (and the same for all $k$)
  - can avoid second application of $M_k$ by letting $S = M_k$
  - reason: projection point is given by $\hat{x}_k$ that is already known
  - projection is there, but already computed
- If $M_k$ is not symmetric or not linear
  - algorithm without projection can diverge
  - need (e.g.) projection to guarantee convergence
Special Cases
Forward-Backward-Forward Splitting (FBF)

• Solves monotone inclusion problems of the form

\[ 0 \in Bx + Dx \]

where \( B + D \) is maximally monotone and \( D \) is \( L \)-Lipschitz

• Algorithm:

\[ \hat{x}_k := (\text{Id} + \gamma B)^{-1}(\text{Id} - \gamma D)x_k \]

\[ x_{k+1} := \hat{x}_k - \gamma(D\hat{x}_k - Dx_k) \]

• Algorithm needs second application of \( D \), at \( \hat{x}_k \)

• Will show special case of NOFOB with \( C = 0 \)
Arriving at FBF from Resolvent Method (1/2)

• Nonlinear resolvent method with constant $\mu_k = \mu$

$$\hat{x}_k := (M_k + A)^{-1}M_kx_k$$

$$x_{k+1} := x_k - \theta_k \mu S^{-1}(M_kx_k - M_k\hat{x}_k)$$

• The trick: Let $M_k = \gamma^{-1}\text{Id} - D$ and $A = B + D$, then

$$\hat{x}_k = (M_k + A)^{-1}M_kx_k = (\gamma^{-1}\text{Id} - D + B + D)^{-1}(\gamma^{-1}\text{Id} - D)$$

$$= (\gamma^{-1}\text{Id} + B)^{-1}(\gamma^{-1}\text{Id} - D)$$

$$= (\text{Id} + \gamma B)^{-1}(\text{Id} - \gamma D)$$

resolvent of $B + D$ in $M_k$ evaluated as forward-backward step:

$$(M_k + A)^{-1} \circ M_k = (\text{Id} + \gamma B)^{-1} \circ (\text{Id} - \gamma D)$$
Arriving at FBF from Resolvent Method (2/2)

- Nonlinear resolvent method

\[
\hat{x}_k = (\text{Id} + \gamma B)^{-1}(\text{Id} - \gamma D)x_k \\
x_{k+1} := x_k - \theta_k \mu S^{-1}((\gamma^{-1}\text{Id} - D)x_k - (\gamma^{-1}\text{Id} - D)\hat{x}_k)
\]

- Now use:
  - Projection metric \( S = \text{Id} \)
  - \( \mu = 1/(L + \gamma^{-1}) \) since \( M_k \) is \( 1/(L + \gamma^{-1}) \)-cocoercive
  - Relaxation \( \theta_k = (L + \gamma^{-1})/\gamma^{-1} \in (1, 2) \), for \( \gamma \in (0, \frac{1}{L}) \)

   to get resulting algorithm (FBF):

\[
\hat{x}_k := (\text{Id} + \gamma B)^{-1}(\text{Id} - \gamma D)x_k \\
x_{k+1} := \hat{x}_k - \gamma(D\hat{x}_k - Dx_k)
\]
Convergence of FBF

- Requirement: $M_k = \gamma^{-1}\text{Id} - D$ strongly monotone
- Satisfied if $\gamma^{-1} - L > 0$, where $L$ Lipschitz constant of $D$
- Gives standard step-length requirement of FBF: $\gamma \in (0, \frac{1}{L})$
- Shows that relaxation $\theta = (L + \gamma^{-1})/\gamma^{-1} \in (1, 2)$
Summary of FBF derivation

- FBF is specific nonlinear resolvent method
- $\mu_k$ is global instead of local cocoercivity constant $\Rightarrow$ conservative
- Relaxation parameter fixed function of $\gamma$ and $L \Rightarrow$ restrictive
A Long-step FBF

• We propose long-step FBF method (NOFOB with full projection)

\[
\hat{x}_k := (\text{Id} + \gamma B)^{-1}(\text{Id} - \gamma D)x_k \\
\mu_k := \frac{\langle (\text{Id} - \gamma D)x_k - (\text{Id} - \gamma D)\hat{x}_k, x_k - \hat{x}_k \rangle}{\| (\text{Id} - \gamma D)x_k - (\text{Id} - \gamma D)\hat{x}_k \|^2} \\
x_{k+1} := x_k - \theta_k \mu_k((\text{Id} - \gamma D)x_k - (\text{Id} - \gamma D)\hat{x}_k)
\]

• Essentially same computational cost as FBF, longer steps
• Local, not global, cocoercivity constant \( \hat{\mu}_k \) of \( M_k = \gamma^{-1}\text{Id} - D \)
• Convergence for \( \gamma \in (0, \frac{1}{L}) \) and \( \theta_k \in (0, 2) \)

Variations:

• If \( D \) linear skew adjoint, all \( \gamma > 0 \) OK (as in standard FBF)
• Can make all step-sizes \( \gamma \) depend on iteration
Projective splitting

- Solves monotone inclusion problems of the form

\[ 0 \in \sum_{i=1}^{n-1} L_i^* B_i(L_i x) + B_n(x) \]

- Primal dual condition (monotone+skew)

\[
0 \in \begin{bmatrix}
B_1^{-1}(w_1) \\
\vdots \\
B_{n-1}^{-1}(w_{n-1}) \\
B_n(x)
\end{bmatrix}
+ \begin{bmatrix}
-L_1 \\
\vdots \\
-L_{n-1}
\end{bmatrix}
\begin{bmatrix}
w_1 \\
\vdots \\
w_{n-1}
\end{bmatrix}
+ \begin{bmatrix}
L_1^* & \cdots & L_{n-1}^*
\end{bmatrix}
\begin{bmatrix}
K
\end{bmatrix}
\begin{bmatrix}
p
\end{bmatrix}
\]

- Full splitting method: resolvents on \( B_i \), forward evaluations on \( L_i \)
Algorithm 1 Synchronous Projective Splitting  Combettes, Eckstein 2018

1: **Input:** $x_0 \in \mathcal{H}$ and $w_{i,0} \in \mathcal{G}_i$ for $i = 1, \ldots, n - 1$

2: **for** $k = 0, 1, \ldots$ **do**

3: $\hat{x}_k := J_{\tau_{n,k}} B_i (x_k - \tau_{n,k} \sum_{i=1}^{n-1} L_i^* w_{i,k})$

4: $\hat{y}_k := (\tau_{n,k}^{-1} x_k - \sum_{i=1}^{n-1} L_i^* w_{i,k}) - \tau_{n,k}^{-1} \hat{x}_k$

5: **for** $i = 1, \ldots, n - 1$ **do**

6: $\hat{v}_{i,k} := J_{\tau_{i,k}} B_i (L_i x_k + \tau_{i,k} w_{i,k})$

7: $\hat{w}_{i,k} := w_{i,k} + \tau_{i,k}^{-1} L_i x_k - \tau_{i,k}^{-1} \hat{v}_{i,k}$

8: **end for**

9: $t^*_k := \hat{y}_k + \sum_{i=1}^{n-1} L_i^* \hat{w}_{i,k}$

10: $t_{i,k} := \hat{v}_{i,k} - L \hat{x}_k$

11: $\mu_k := \frac{\left( \sum_{i=1}^{n-1} \langle t_{i,k}, w_{i,k} \rangle - \langle \hat{v}_{i,k}, \hat{w}_{i,k} \rangle \right) + \langle t^*_k, x_k \rangle - \langle \hat{y}_k, \hat{x}_k \rangle}{\sum_{i=1}^{n-1} \| t_{i,k} \|^2 + \| t^*_k \|^2}$

12: **for** $i = 1, \ldots, n - 1$ **do**

13: $w_{i,k+1} = w_{i,k} - \theta_k \mu_k t_{i,k}$

14: **end for**

15: $x_{k+1} := x_k - \theta_k \mu_k t^*_k$

16: **end for**
Projective splitting in our framework

Apply NOFOB to primal dual condition $0 \in Bp + Kp$ with

- **Kernel**

$$ M_k = \begin{bmatrix} \sigma_1^{-1} \text{Id} \\ \vdots \\ \sigma_{n-1}^{-1} \text{Id} \\ \tau \text{Id} \end{bmatrix} - \begin{bmatrix} -L_1 \\ \vdots \\ -L_{n-1} \end{bmatrix} $$

that subtracts the skew symmetric operator $K$, and

- $\sigma_i$ and $\tau$ become individual resolvent parameters for $B_i$
- $M_k$ strongly monotone for all $\sigma_i, \tau > 0$ – no step-size restrictions!

- $A = B + K$ and $C = 0$ (NOFOB solves $0 \in Ax + Cx$)
- Induced projection metric norm
Projective splitting in our framework

- Kernel

\[ M_k = P - K \]

not symmetric, need to compute projection

- Backward-step in NOFOB on \( A = B + K \) \((C = 0)\):

\[ \hat{p}_k = (M_k + A)^{-1} M_k p_k = (P + K + B - K)^{-1} (P - K)p_k = (P + B)^{-1} (P - K)p_k \]

same as in FBF

- Since full projection, algorithm is special case of long-step FBF
Chambolle-Pock

• Solves monotone inclusion problems of the form

\[ 0 \in L^* B_1(Lx) + B_2(x) \]

via primal dual optimality condition (monotone+skew)

\[ 0 \in \begin{bmatrix} B_1^{-1}(w) \\ B_2(x) \end{bmatrix} + \begin{bmatrix} 0 & -L \\ L^* & 0 \end{bmatrix} \begin{bmatrix} w \\ x \end{bmatrix} \]

• Well known to be resolvent method

• Cast in our algorithm format by setting (linear and symmetric)

\[ M_k = \begin{bmatrix} \sigma^{-1} \text{Id} & \tau^{-1} \text{Id} \\ \tau^{-1} \text{Id} & \sigma^{-1} \text{Id} \end{bmatrix} + \begin{bmatrix} 0 & L \\ L^* & 0 \end{bmatrix}, \]

\[ S = M_k, \text{ gives } \mu_k = 1 \text{ (no restriction) and } \theta_k = 1 \]

• Projection step redundant since \( M_k = S \) is symmetric!

• Standard step-size restriction from \( M_k \) strongly monotone
Projective splitting vs Chambolle Pock

- If two summands, projective splitting and Chambolle Pock solves
  \[ 0 \in L^* B_1(Lx) + B_2(x) \]
- Projective splitting with two summands is NOFOB with kernel
  \[ M_k = \begin{bmatrix} \sigma^{-1} \text{Id} & \tau^{-1} \text{Id} \\ \text{Id} & \text{Id} \end{bmatrix} + \begin{bmatrix} 0 & L \\ -L^* & 0 \end{bmatrix} \]
  not symmetric – projection needed, no step-size restrictions
- Chambolle-Pock is NOFOB with linear symmetric kernel
  \[ M_k = \begin{bmatrix} \sigma^{-1} \text{Id} & \tau^{-1} \text{Id} \\ \text{Id} & \text{Id} \end{bmatrix} + \begin{bmatrix} 0 & L \\ L^* & 0 \end{bmatrix} \]
  symmetry of \( M_k \) – no projection needed, but step-size restrictions
- Difference between \( M_k \) in the two algorithms is
  \[ \begin{bmatrix} 0 & 0 \\ -2L^* & 0 \end{bmatrix} \]
  (but projection kernels \( S \) differ more)
A novel four operator splitting method

• Solves monotone inclusions

\[ 0 \in Bx + Cx + Dx + Kx \]

where

• \( B + D \) maximally monotone, \( D \) Lipschitz
• \( C \) cocoercive
• \( K \) linear skew-adjoint

• Let \( A = B + D + K \) and \( M_k = Q_k - D - K \) to get FB map

\[
(M_k + A)^{-1}(M_k - C) = (Q_k + B)^{-1}(Q_k - D - K - C)
\]

that is forward evaluation on \( D, K, \) and \( C \), resolvent on \( B \)

• Then create separating hyperplane and project as in NOFOB

• Special cases
  • \( K = C = 0 \): FBF
  • \( K = 0 \): FBHF
  • \( C = D = 0 \): Projective splitting, Chambolle Pock
  • \( K = D = 0 \): FB, Vu-Condat
  • \( D = 0, Q_k \) PD+skew linear: AFBA
NOFOB Variation

- NOFOB creates separating hyperplane then projects
- Variation: collect sequence of hyperplanes before projection
- Convergence analysis is identical
Summary

• We have proposed nonlinear forward-backward splitting
• It has many special cases, have focused on
  • FBF
  • Chambolle-Pock
  • Projective splitting
  • Novel four operator splitting
• New interpretation of FBF as separate and project method
Thank you

Preprint available on arXiv:1908.07449