# Nonlinear Forward-Backward Splitting with Projection Correction

Pontus Giselsson

# This talk

- We consider monotone inclusion problems of the form

$$0 \in Bx + Dx$$

  where
  - $B$ and $D$ are maximally monotone operators
  - $D$ is Lipschitz continuous
- Will give new interpretation of forward-backward-forward splitting

$$\begin{aligned} \hat{x}_k &:= (\mathrm{Id} + \gamma B)^{-1}(\mathrm{Id} - \gamma D)x_k \\ x_{k+1} &:= \hat{x}_k - \gamma(D\hat{x}_k - Dx_k) \end{aligned}$$

  where
  - first step is forward-backward step on $B$ and $D$
  - second step is a correction step that needs extra evaluation of $D$

# Proximal gradient method

- Consider convex optimization problems of the form

$$\text{minimize}\, f(x) + g(x)$$

  where
  - $f : \mathbb{R}^n \to \mathbb{R}$ is convex and smooth
  - $g : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is proper closed convex
- Since $f$ finite-valued, this is equivalent to solving

$$0 \in \partial g(x) + \nabla f(x)$$

  where
  - $\partial g$ and $\nabla f$ are maximally monotone
  - $\nabla f$ is Lipschitz continuous
- Proximal gradient method (forward-backward splitting)

$$x_{k+1} = (\mathrm{Id} + \gamma \partial g)^{-1}(\mathrm{Id} - \gamma \nabla f)x_k = \mathrm{prox}_{\gamma g}(x_k - \gamma \nabla f(x_k))$$

  does not need correction, why?

# Cocoercivity

**Baillon-Haddad theorem**

Let $D$ be Lipschitz continuous and gradient of convex function.

Then $D$ is cocoercive.

- In general: $\beta$-Lipschitz continuity $\Leftarrow \frac{1}{\beta}$-cocoercivity

  $$\|Dx - Dy\| \leq \beta\|x - y\| \quad \Leftarrow \quad \langle Dx - Dy, x - y\rangle \geq \tfrac{1}{\beta}\|Dx - Dy\|^2$$

  Cauchy-Schwarz on cocoercivity scalar product gives Lipschitz
- There exist Lipschitz operators that are not cocoercive
- Need correction step in forward-backward if not cocoercive

# Lipschitz but not cocoercive operators – Skew

- Skew-symmetric operators

$$K = \begin{bmatrix} 0 & L^* \\ -L & 0 \end{bmatrix}$$

are Lipschitz but not coocericve since $\langle Kx - Ky, x - y \rangle = 0$

- Arise when solving primal dual formulations of $\min g(x) + f(Lx)$:

$$0 \in \partial g(x) + L^* \underbrace{\partial f(Lx)}_{\mu} \quad \Leftrightarrow \quad 0 \in \begin{bmatrix} \partial g(x) + L^* \mu \\ \partial f(Lx) - \mu \end{bmatrix}$$

$$\Leftrightarrow \quad 0 \in \underbrace{\begin{bmatrix} \partial g(x) \\ \partial f^*(\mu) \end{bmatrix}}_{B(x,\mu)} + \underbrace{\begin{bmatrix} L^* \mu \\ -Lx \end{bmatrix}}_{K(x,\mu)}$$

where $K$ is skew, monotone, and Lipschitz, but not cocoercive

- Solvable by forward-backward forward, but not forward-backward

# Lipschitz not cocoercive – Min-max problems

- Convex-concave min-max problems

$$\min_x \max_\mu (h(x,\mu) + f(x) - g^*(\mu))$$

  where
  - $f$ is convex and $h$ is convex w.r.t. $x$
  - $g^*$ is concave and $h$ is concave w.r.t. $\mu$
  - $h$ is differentiable and with Lipschitz gradient

- Optimality condition

$$0 \in \underbrace{\begin{bmatrix} \partial f(x) \\ \partial g^*(\mu) \end{bmatrix}}_{B(x,\mu)} + \underbrace{\begin{bmatrix} \nabla_x h(x,\mu) \\ -\nabla_\mu h(x,\mu) \end{bmatrix}}_{D(x,\mu)}$$

  where $D$ is monotone and Lipschitz, but not cocoercive

- Solvable by forward-backward forward, but not forward-backward
- Motivation from training of GANs, although not convex-concave

# New interpretation of FBF

- FBF is special case of new algorithm called NOFOB
- NOFOB is a separate and project method:
  - "Create separating hyperplane and project onto it"
  - Separating hyperplane from nonlinear forward-backward map

# 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)

$$\begin{aligned}
\hat{x}_k &:= (M_k + A)^{-1}(M_k - C)x_k \\
H_k &:= \{z : \langle M_k x_k - M_k \hat{x}_k, z - \hat{x}_k \rangle \leq \tfrac{\beta}{4}\|x_k - \hat{x}_k\|_P^2\} \\
x_{k+1} &:= (1-\theta_k)x_k + \theta_k \Pi_{H_k}^S(x_k)
\end{aligned}$$

where

  - $M_k$ is Lipschitz and strongly monotone (can be relaxed if $C = 0$)
  - $H_k$ is a halfspace that contains $\mathrm{zer}(A + C)$ 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
  - $P$ and $S$ are linear self-adjoint positive definite operators

- First step requires one $M_k$ application, $H_k$ construction another

# Convergence

- Consequences of separate and project principle:
  - $\| \cdot \|_S$-distance to fixed-point set nonincreasing (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

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

which holds under stated assumptions on $M_k$

# 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 $\Pi^S_{H_k}(x_k) = \hat{x}_k$ that is already known
  - projection is in algorithm, but already computed
- If $M_k$ is not symmetric or not linear
  - algorithm without projection can diverge
  - need (e.g.) projection to guarantee convergence

# 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:

$$\begin{aligned} \hat{x}_k &:= (\mathrm{Id} + \gamma B)^{-1}(\mathrm{Id} - \gamma D)x_k \\ x_{k+1} &:= \hat{x}_k - \gamma(D\hat{x}_k - Dx_k) \end{aligned}$$

- 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 first step:

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

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

$$\begin{aligned}\hat{x}_k &= (M_k + A)^{-1} M_k x_k = (\gamma^{-1}\mathrm{Id} - D + B + D)^{-1}(\gamma^{-1}\mathrm{Id} - D)\\ &= (\gamma^{-1}\mathrm{Id} + B)^{-1}(\gamma^{-1}\mathrm{Id} - D)\\ &= (\mathrm{Id} + \gamma B)^{-1}(\mathrm{Id} - \gamma D)\end{aligned}$$

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

$$(M_k + A)^{-1} \circ M_k = (\mathrm{Id} + \gamma B)^{-1} \circ (\mathrm{Id} - \gamma D)$$

# Arriving at FBF from Resolvent Method (2/2)

- Nonlinear resolvent method

$$\begin{aligned} \hat{x}_k &:= (M_k + A)^{-1} M_k x_k \\ H_k &:= \{z : \langle M_k x_k - M_k \hat{x}_k, z - \hat{x}_k \rangle \leq 0\} \\ x_{k+1} &:= (1 - \theta_k) x_k + \theta_k \Pi^S_{H_k}(x_k) \end{aligned}$$

- Use in projection step:
  - Projection metric $S = \mathrm{Id}$
  - Relaxation parameter $\theta_k = \gamma \frac{\|M_k x_k - M_k \hat{x}_k\|_2^2}{\langle M_k x_k - M_k \hat{x}_k, x_k - \hat{x}_k \rangle}$

  to get resulting algorithm (FBF):

$$\begin{aligned} \hat{x}_k &:= (\mathrm{Id} + \gamma B)^{-1} (\mathrm{Id} - \gamma D) x_k \\ x_{k+1} &:= \hat{x}_k - \gamma (D \hat{x}_k - D x_k) \end{aligned}$$

- If $\gamma \in (0, \frac{1}{L})$, relaxation parameter $\theta_k \in [\epsilon, 2 - \epsilon]$ but often small

# A Long-step FBF

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

$$\begin{aligned}\hat{x}_k &:= (\mathrm{Id} + \gamma B)^{-1}(\mathrm{Id} - \gamma D)x_k \\ \mu_k &:= \frac{\langle (\mathrm{Id} - \gamma D)x_k - (\mathrm{Id} - \gamma D)\hat{x}_k, x_k - \hat{x}_k \rangle}{\|(\mathrm{Id} - \gamma D)x_k - (\mathrm{Id} - \gamma D)\hat{x}_k\|^2} \\ x_{k+1} &:= x_k - \theta_k \mu_k((\mathrm{Id} - \gamma D)x_k - (\mathrm{Id} - \gamma D)\hat{x}_k)\end{aligned}$$

- Essentially same computational cost as FBF, longer steps
- Arbitrary relaxation parameter $\theta_k$
- 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

# Summary

- We have proposed nonlinear forward-backward splitting (NOFOB)
- Shown that forward-backward forward is special case
- NOFOB has many more special cases:
  - Forward-backward splitting
  - Forward-backward-half-forward splitting
  - Chambolle-Pock
  - Vu-Condat
  - Douglas-Rachford, ADMM, and proximal ADMM
  - Synchronous projective splitting
  - Asymmetric forward-backward adjoint splitting (AFBA)
- NOFOB also gives rise to novel four operator splitting method

# Thank you

Preprint available on arXiv:1908.07449