Nonlinear Forward-Backward Splitting with Projection or Momentum Correction

Pontus Giselsson

Joint work with Martin Morin and Sebastian Banert

Proximal point algorithm

- Consider the problem

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax$$

  where $A : \mathcal{H} \to 2^{\mathcal{H}}$ is maximally monotone
- Proximal point algorithm (PPA) solves it by iterating resolvent

$$x_{k+1} = J_{\gamma_k A} x_k$$

  where
  - $J_{\gamma_k A} := (\mathrm{Id} + \gamma_k A)^{-1}$ is resolvent
  - Uniformly upper bounded $\gamma_k \geq \epsilon > 0$ is a step-size parameter

Conceptual algorithm

- In general as expensive to take one step of PPA as solving problem
- Clever choice of space $\mathcal{H}$ and/or A gives important special cases
  - The Chambolle–Pock method
  - Douglas–Rachford splitting
  - ADMM (with dual step-size 1)

Unified convergence analysis

- PPA provides unified convergence analysis for all special cases
- PPA convergence analysis for *maximally monotone* A
  - $J_{\gamma_k A}$ has full domain (Minty) $\Rightarrow$ algorithm defined for all inputs
  - $J_{\gamma_k A}$ firmly nonexpansive $\Rightarrow$ single-valuedness and convergence

  which is often easier than directly proving special cases

Adding cocoercive operator

- We can add $\frac{1}{\beta}$-cocoercive operator $C : \mathcal{H} \to \mathcal{H}$ to get problem

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + Cx$$

- Can be solved using forward-backward splitting

$$x_{k+1} = J_{\gamma_k A}(\mathrm{Id} - \gamma_k C)x_k$$

  which generalizes PPA
- Algorithm analysis similar (composition averaged if $\gamma_k \in [\epsilon, \frac{2-\epsilon}{\beta}]$)
- Special cases:
  - Proximal gradient method
  - Condat–Vũ

More operator splitting methods

- Many more methods exist that are not special cases of FB, e.g.,:
  - Tseng's forward-backward-forward splitting [1]
  - Forward-backward-half-forward splitting [2]
  - Solodov and Tseng [3]
  - (Synchronous) projective splitting [4]
  - Asymmetric forward-backward-adjoint splitting [5]
  - Briceño-Arias and Combettes (error-free version) [6]
  - Proximal alternating predictor corrector [7]
  - He and Yuan [8]
  - Malitsky–Tam [9]
  - Forward-reflected-Douglas–Rachford [10]
  - ...
- Is there a unifying framework for these and previous methods?

[1] A Modified Forward-Backward Splitting Method for Maximal Monotone Mappings, P. Tseng
[2] Forward-Backward-Half Forward Algorithm for Solving Monotone Inclusions, L. M. Briceño-Arias and D. Davis
[3] Modified Projection-type Methods for Monotone Variational Inequalities, M. V. Solodov, and P. Tseng
[4] Asynchronous Block-Iterative Primal-Dual Decomposition Methods for Monotone Inclusions, P. L. Combettes and J. Eckstein
[5] Asymmetric Forward-Backward-Adjoint Splitting for Solving Monotone Inclusions Involving Three Operators, P. Latafat and P. Patrinos
[6] A Monotone + Skew Splitting Model for Composite Monotone Inclusions in Duality, L. M. Briceño-Arias and P. L. Combettes
[7] A Simple Algorithm for a Class of Nonsmooth Convex-Concave Saddle-Point Problems, Y. Drori, S. Sabach, M. Teboulle
[8] Convergence Analysis of Primal-Dual Algorithms for a Saddle-Point Problem: From Contraction Perspective, He and Yuan
[9] Forward-Backward Splitting Method for Monotone Inclusions Without Cocoercivity, Y. Malitsky and M. K. Tam
[10] Finding the Forward-Douglas-Rachford-Forward Method, E. K. Ryu and B. C. Vũ

YES – Such a framework exists!

- Will present such an algorithmic framework based on
  - *Nonlinear FB map* (special case: *nonlinear resolvent*[1])
  - Projection or momentum correction
- Algorithm solves monotone inclusion $0 \in Ax + Cx$ where
  - $A : \mathcal{H} \to 2^{\mathcal{H}}$ is maximally monotone
  - $C : \mathcal{H} \to \mathcal{H}$ is cocoercive

---

[1] Also known as *warped resolvent* (Bùi, Combettes) or *F-resolvent* (Bauschke, Wang, Yao)

Nonlinear forward-backward map

- Let $M : \mathcal{H} \to \mathcal{H}$ be maximally monotone
- *Nonlinear forward-backward map* is

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

  and

  - if $C = 0$ reduces to *nonlinear resolvent* $(M + A)^{-1} \circ M$
  - M is called a kernel

Nonlinear forward-backward map

- Let $M : \mathcal{H} \to \mathcal{H}$ be maximally monotone
- *Nonlinear forward-backward map* is

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

  and
  - if $C = 0$ reduces to *nonlinear resolvent* $(M + A)^{-1} \circ M$
  - M is called a kernel
- Special cases with different kernels:
  - $M = \gamma^{-1}\mathrm{Id}$ gives standard FB step:

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

  - $M = \gamma^{-1}P$ with $P \in \mathcal{P}(\mathcal{H})$[1] gives preconditioned FB
  - $M = \nabla g$ with g convex gives Bregman FB step

[1] $\mathcal{P}(\mathcal{H})$ set of bounded linear self-adjoint strongly positive operators on $\mathcal{H}$

Iterating FB map – Convergence?

- An algorithm candidate is to iterate the nonlinear FB-map

$$x_{k+1} = (M + A)^{-1} \circ (M - C)x_k$$

  since fixed-point set equals solution set $\mathrm{zer}(A + C)$
- However, may not converge under following assumptions on M:
  - Strongly monotone (if linear: strongly positive)
  - Lipschitz continuous (if linear: bounded)

  but if M also linear self-adjoint, it converges (if M large enough)

Counter-example

- Problem: $C = 0$ and A skew-symmetric (and monotone):

$$A : \mathbb{R}^2 \to \mathbb{R}^2 : (x, y) \mapsto (-y, x)$$

  which is a 90° rotation

- Kernel $M = \gamma^{-1}\mathrm{Id} - A$ with $\gamma > 0$ is
  - bounded linear strongly positive
  - but not self-adjoint

  and gives iteration

$$\begin{aligned} x_{k+1} &= (M + A)^{-1} M x_k = (\gamma^{-1}\mathrm{Id} - A + A)^{-1}(\gamma^{-1}\mathrm{Id} - A) x_k \\ &= (\mathrm{Id} - \gamma A) x_k = \begin{bmatrix} 1 & \gamma \\ -\gamma & 1 \end{bmatrix} x_k \end{aligned}$$

  which diverges for all $\gamma \neq 0$ (rotation with gain $\sqrt{1 + \gamma^2} > 1$)

Counter-example

- Problem: $C = 0$ and A skew-symmetric (and monotone):

$$A : \mathbb{R}^2 \to \mathbb{R}^2 : (x, y) \mapsto (-y, x)$$

  which is a 90° rotation

- Kernel $M = \gamma^{-1}\mathrm{Id} - A$ with $\gamma > 0$ is
  - bounded linear strongly positive
  - but not self-adjoint

  and gives iteration

$$\begin{aligned} x_{k+1} &= (M + A)^{-1} M x_k = (\gamma^{-1}\mathrm{Id} - A + A)^{-1}(\gamma^{-1}\mathrm{Id} - A)x_k \\ &= (\mathrm{Id} - \gamma A)x_k = \begin{bmatrix} 1 & \gamma \\ -\gamma & 1 \end{bmatrix} x_k \end{aligned}$$

  which diverges for all $\gamma \neq 0$ (rotation with gain $\sqrt{1 + \gamma^2} > 1$)

- Need correction to use nonlinear FB map in algorithm

Nonlinear FB map creates separating hyperplane

- Assume
  - $A : \mathcal{H} \to 2^{\mathcal{H}}$ maximally monotone
  - $C : \mathcal{H} \to \mathcal{H}$ is $\frac{1}{\ell}$-cocoercive with $\ell \in [0, 4)$ w.r.t. $P \in \mathcal{P}(\mathcal{H})$[1]
  - $M : \mathcal{H} \to \mathcal{H}$ is 1-strongly monotone w.r.t. $P \in \mathcal{P}(\mathcal{H})$[2]
- Define the affine function ψ_x for each x with $\hat{x} = T_{\mathrm{FB}}x$ as:

$$\psi_x(z) := \langle Mx - M\hat{x}, z - \hat{x} \rangle - \tfrac{\ell}{2}\|x - \hat{x}\|_P^2$$

  Then
  - $\psi_x(z) \leq 0$ for all $z \in \mathrm{zer}(A + C)$
  - $\psi_x(x) \geq (1 - \frac{\ell}{4})\|x - T_{\mathrm{FB}}x\|^2$ for all $x \in \mathcal{H}$
  - $\psi_x(x) > 0$ for all points $x \notin \mathrm{zer}(A + C)$ (since $\ell \in [0, 4)$)

Nonlinear FB map creates separating hyperplane

- Assume
  - $A : \mathcal{H} \to 2^{\mathcal{H}}$ maximally monotone
  - $C : \mathcal{H} \to \mathcal{H}$ is $\frac{1}{\ell}$-cocoercive with $\ell \in [0, 4)$ w.r.t. $P \in \mathcal{P}(\mathcal{H})$[1]
  - $M : \mathcal{H} \to \mathcal{H}$ is 1-strongly monotone w.r.t. $P \in \mathcal{P}(\mathcal{H})$[2]
- Define the affine function ψ_x for each x with $\hat{x} = T_{\mathrm{FB}}x$ as:

$$\psi_x(z) := \langle Mx - M\hat{x}, z - \hat{x} \rangle - \frac{\ell}{2}\|x - \hat{x}\|_P^2$$

  Then
  - $\psi_x(z) \leq 0$ for all $z \in \mathrm{zer}(A + C)$
  - $\psi_x(x) \geq (1 - \frac{\ell}{4})\|x - T_{\mathrm{FB}}x\|^2$ for all $x \in \mathcal{H}$
  - $\psi_x(x) > 0$ for all points $x \notin \mathrm{zer}(A + C)$ (since $\ell \in [0, 4)$)
- Nonlinear FB map output $\hat{x}$ helps define halfspace

$$H := \{z : \psi_x(z) \leq 0\}$$

  that (strictly) separates $\mathrm{zer}(A + C) \subseteq H$ and $x \notin H$

---

[1] $C : \mathcal{H} \to \mathcal{H}$ is ℓ^{-1}-cocoercive w.r.t. P if $\forall x, y \in \mathcal{H}$ we have $\langle Cx - Cy, x - y \rangle \geq \ell^{-1}\|Cx - Cy\|_{P^{-1}}^2$

[2] $M : \mathcal{H} \to \mathcal{H}$ is 1-strongly monotone w.r.t. P if $\forall x, y \in \mathcal{H}$ we have $\langle Mx - My, x - y \rangle \geq \|x - y\|_P^2$

NOFOB with projection correction

- Nonlinear forward-backward splitting with projection correction

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

  which converges weakly to a solution if

  - M_k is Lipschitz continuous and 1-strongly monotone w.r.t. P
  - P, S are bounded linear self-adjoint strongly positive operators
  - 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

NOFOB with projection correction

- Nonlinear forward-backward splitting with projection correction

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

  which converges weakly to a solution if

  - M_k is Lipschitz continuous and 1-strongly monotone w.r.t. P
  - P, S are bounded linear self-adjoint strongly positive operators
  - 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

- Note: algorithm requires two forward evaluations of M_k:
  - T_{FB} evaluation (first step) requires $M_k x_k$
  - H_k creation requires $M_k x_k$ (already computed) and $M_k \hat{x}_k$

NOFOB with explicit projection

- Stating projection explicitly gives equivalent more explicit method

$$\begin{aligned} \hat{x}_k &:= (M_k + A)^{-1}(M_k - C)x_k \\ \mu_k &:= \frac{\langle M_k x_k - M_k \hat{x}_k, x_k - \hat{x}_k \rangle - \frac{\ell}{4}\|x_k - \hat{x}_k\|_P^2}{\|M_k x_k - M_k \hat{x}_k\|_{S^{-1}}^2} \\ x_{k+1} &:= x_k - \theta_k \mu_k S^{-1}(M_k x_k - M_k \hat{x}_k) \end{aligned}$$

where $\mu_k \geq \epsilon$ (unless $x \in \mathrm{zer}(A + C)$, in which case $\mu_k = \frac{0}{0} = 0$)

NOFOB with explicit projection

- Stating projection explicitly gives equivalent more explicit method

$$\begin{aligned}\hat{x}_k &:= (M_k + A)^{-1}(M_k - C)x_k \\ \mu_k &:= \frac{\langle M_k x_k - M_k \hat{x}_k, x_k - \hat{x}_k\rangle - \frac{\ell}{4}\|x_k - \hat{x}_k\|_P^2}{\|M_k x_k - M_k \hat{x}_k\|_{S^{-1}}^2} \\ x_{k+1} &:= x_k - \theta_k \mu_k S^{-1}(M_k x_k - M_k \hat{x}_k)\end{aligned}$$

  where $\mu_k \geq \epsilon$ (unless $x \in \mathrm{zer}(A + C)$, in which case $\mu_k = \frac{0}{0} = 0$)
- Algorithm converges with μ_k replaced by any $\hat{\mu}_k \in [\epsilon, \mu_k]$
  - Equivalent to algorithm with smaller relaxation parameter $\theta_k \frac{\hat{\mu}_k}{\mu_k}$
  - Gives shorter step-lengths

Special case – Forward-backward splitting

- Suppose
  - $M_k = \gamma_k^{-1} M$ with $M \in \mathcal{P}(\mathcal{H})$ and $P = M$
  - projection metric $S = M$
  - C is $\frac{1}{\beta}$-cocoercive w.r.t. M (and P)

  then $\mu_k = \gamma_k(1 - \frac{\gamma_k \beta}{4})$
- Let $\lambda_k = \theta_k(1 - \frac{\gamma_k \beta}{4})$ to get relaxed preconditioned FB splitting

$$\begin{aligned} \hat{x}_k &:= (M + \gamma_k A)^{-1}(M - \gamma_k C)x_k \\ x_{k+1} &:= x_k - \lambda_k(x_k - \hat{x}_k) \end{aligned}$$

- Note that:
  - second evaluation of M not needed (since $S^{-1}M = \gamma_k^{-1}\mathrm{Id}$)
  - projection correction only kicks in if needed

Convergence and special cases

- Relaxed preconditioned FB splitting (with $\lambda_k = \theta_k(1 - \frac{\gamma_k \beta}{4})$)

$$\begin{aligned} \hat{x}_k &:= (M + \gamma_k A)^{-1}(M - \gamma_k C)x_k \\ x_{k+1} &:= x_k - \lambda_k(x_k - \hat{x}_k) \end{aligned}$$

- Converges if $\gamma_k \in [\epsilon, \frac{4-\epsilon}{\beta}]$ (extended range) and $\theta_k \in [\epsilon, 2 - \epsilon]$
  - $\gamma_k \geq \frac{2}{\beta}$ possible $\Rightarrow \lambda_k < 1$ (under-relaxation)
  - $\gamma_k \in [\epsilon, \frac{2-\epsilon}{\beta}]$: $\lambda_k = 1$ possible, but also $\lambda_k > 1$ (over-relaxation)
- Since FB is special case of NOFOB, it has special cases:
  - Chambolle–Pock
  - Vũ–Condat
  - Douglas–Rachford, ADMM (with dual step-size 1)
  - Proximal gradient method

Other special cases

- These are special cases of NOFOB with projection correction
  - Nonlinear resolvent step:
    - Tseng's forward-backward-forward splitting (M nonlinear)
    - Solodov and Tseng (M nonlinear)
    - (Synchronous) projective splitting (M not self-adjoint)
    - Briceño-Arias/Combettes (error-free version) (M not self-adjoint)
    - He and Yuan (M not self-adjoint)
  - Nonlinear FB step:
    - Forward-backward-half-forward splitting (M nonlinear)
    - AFBA (M not self-adjoint)
    - Proximal alternating predictor corrector (M not self-adjoint)
- Can add cocoercive term in methods based on resolvent

Drawback of projection correction

- In general, two evaluations of M_k is needed in every iteration
- Exception, e.g., standard FB splitting that has $S^{-1}M_k = \gamma_k^{-1}\mathrm{Id}$

NOFOB with momentum correction

- Consider the same problem problem

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + Cx$$

  where
  - $A : \mathcal{H} \to 2^{\mathcal{H}}$ maximally monotone
  - $C : \mathcal{H} \to \mathcal{H}$ is $\frac{1}{\ell}$-cocoercive w.r.t. $S \in \mathcal{P}(\mathcal{H})$
- Nonlinear forward-backward splitting with *momentum* correction

$$\begin{aligned} x_{k+1} &= (M_k + A)^{-1}(M_k x_k - Cx_k + \gamma_k^{-1} u_k) \\ u_{k+1} &= (\gamma_k M_k - S)x_{k+1} - (\gamma_k M_k - S)x_k \end{aligned}$$

  where $S \in \mathcal{P}(\mathcal{H})$ and M_k possibly nonlinear
- Momentum term is in the $\gamma_k M_k - S$ operator

M_k evaluations

- Nonlinear forward-backward splitting with *momentum* correction

$$x_{k+1} = (M_k + A)^{-1}(M_k x_k - C x_k + \gamma_k^{-1} u_k)$$
$$u_{k+1} = (\gamma_k M_k - S)x_{k+1} - (\gamma_k M_k - S)x_k$$

- Comparison to projection correction in terms of M_k evaluations
  - Need to evaluate $M_{k-1}x_k$ and $M_k x_k \Rightarrow$ in general no savings
  - If $M_k = \alpha_k M_{k-1}$ (with M_k still nonlinear) $\Rightarrow$ we save one[1]
  - If $M_k = \alpha_k^{-1}\mathrm{Id} - D$ (with D nonlinear) $\Rightarrow$ we save one D-eval.

---

[1] Recall: To save one M_k evaluation with projection correction $M_k = M = \alpha_k S$, which gives standard FB splitting.

Restrictions on M_k

- Nonlinear forward-backward splitting with *momentum* correction

$$x_{k+1} = (M_k + A)^{-1}(M_k x_k - C x_k + \gamma_k^{-1} u_k)$$
$$u_{k+1} = (\gamma_k M_k - S)x_{k+1} - (\gamma_k M_k - S)x_k$$

- Letting $M_k = \gamma_k^{-1} S \in \mathcal{P}(\mathcal{H})$ gives standard FB splitting ($u_k = 0$)
- M_k can deviate from $\gamma_k^{-1} S$, we assume

$\gamma_k M_k - S$ is L_k-Lipschitz continuous w.r.t. S

and we have weak convergence if all $\gamma_k \geq \epsilon$ and

$$1 - L_{k-1} - L_k - \tfrac{\gamma_k \ell}{2} \geq \epsilon > 0$$

Convergence – Lyapunov analysis

- Let $z \in \mathrm{zer}(A + C)$ and define

$$\mathcal{V}_k = \|x_k + S^{-1}u_k - z\|_S^2 + (1 - L_{k-1})L_{k-1}\|x_k - x_{k-1}\|_S^2$$

- Assume that $L_k < 1$ (Lipschitz constant of $\gamma_k M_k - S$), then

$$\mathcal{V}_{k+1} \leq \mathcal{V}_k - (1 - L_{k-1} - L_k - \tfrac{\gamma_k \ell}{2})\|x_{k+1} - x_k\|_S^2$$

- Convergence condition

$$1 - L_{k-1} - L_k - \tfrac{\gamma_k \ell}{2} \geq \epsilon > 0$$

comes from having residual coefficient strictly positive

Special cases

- These are special cases of NOFOB with momentum correction
  - Nonlinear resolvent
    - Malitsky–Tam (forward-reflected-backward) (M nonlinear)
    - Forward-reflected-Douglas–Rachford (M nonlinear)
  - Nonlinear forward–backward map
    - Malitsky–Tam ("three-operator splitting") (M nonlinear)
- Can add cocoercive term in methods based on resolvent

Momentum instead of projection correction

- Methods with projection correction
  - Tseng's forward-backward-forward splitting
  - Solodov and Tseng
  - (Synchronous) projective splitting
  - Briceno-Arias/Combettes (error-free version)
  - He and Yuan
  - Forward-backward-half-forward splitting
  - Asymmetric forward-backward-adjoint splitting
  - Proximal alternating predictor corrector
- Can derive methods based on momentum correction for the above
  - Comes at the cost or more restrictive parameter requirements
  - Gives Malitsky–Tam methods if done for FB(H)F

Polyak Momentum

- Equivalent formulation with Polyak momentum with $\theta < 1$

$$x_{k+1} = (M_k + A)^{-1}(M_k x_k - C x_k + \gamma_k^{-1} u_k + \gamma_k^{-1}\theta S(x_k - x_{k-1})),$$
$$u_{k+1} = (\gamma_k M_k - S)x_{k+1} - (\gamma_k M_k - S)x_k,$$

- Denote by $\hat{\gamma}_k$ and $\hat{u}_k$ original algorithm parameters, and let

$$\gamma_k = (1-\theta)\hat{\gamma}_k \qquad u_k = (1-\theta)\hat{u}_k - \theta S(x_k - x_{k-1})$$

  to get Polyak momentum method

- Translated requirements for convergence

$$1 - \theta - 2|\theta| - L_{k-1} - L_k - \gamma_k \tfrac{\ell}{2} \geq \varepsilon$$

- Can add Polyak momentum (interpretation) to all special cases

Polyak momentum in FB setting

- General requirements for Polyak momentum convergence

$$1 - \theta - 2|\theta| - L_{k-1} - L_k - \gamma_k \tfrac{\ell}{2} \geq \varepsilon$$

- Assume $M_k = \gamma_k^{-1} S$ ($L_k = 0$), C $\frac{1}{\beta}$-cocoercive w.r.t. S ($\beta = \ell$)
- This gives standard forward-backward setting, if $\gamma_k = \gamma$, we allow

$$\theta \in (\tfrac{-2+\gamma\beta}{2}, \tfrac{2-\gamma\beta}{6})$$

which implies
  - if $\gamma = \frac{1}{\beta}$: $\theta \in (-\frac{1}{2}, \frac{1}{6})$
  - if $C = 0$ ($\beta = 0$): $\theta \in (-1, \frac{1}{3})$

note that we allow for negative momentum (more than positive)

Summary

- Many methods are special cases of presented NOFOB framework
- Can select projection or momentum correction
- Can add cocoercive term to those that do not have
- Can avoid one M_k application by using momentum correction
- Can add Polyak momentum to many methods
- Easy to design and prove convergence of new methods

Special Cases and New Algorithms

Special cases and new algorithms – Outline

- FB(H)F and Malitsky–Tam
- Solodov and Tseng
- Novel four-operator splitting method
  - Special case: AFBA
  - Two novel four-operator splitting primal-dual methods
- Four-operator splitting primal-dual method with different kernel
- Extension to multi-operator setting
  - Synchronous projective splitting

FBF and Malitsky–Tam

- Consider monotone inclusion problem of the form

$$0 \in Bx + Dx$$

where $B + D$ is maximally monotone and D is δ-Lipschitz

- Forward-backward-forward splitting

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

needs second application of D (at $\hat{x}_k$)

FBF and Malitsky–Tam

- Consider monotone inclusion problem of the form

$$0 \in Bx + Dx$$

where $B + D$ is maximally monotone and D is δ-Lipschitz

- Forward-backward-forward splitting

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

needs second application of D (at $\hat{x}_k$)

- Malitsky–Tam

$$\begin{aligned} x_{k+1} &:= (\mathrm{Id} + \gamma_k B)^{-1}(x_k - \gamma_k D x_k + u_k) \\ u_{k+1} &:= \gamma_k (Dx_k - Dx_{k+1}) \end{aligned}$$

avoids second application of D (or rather, it can be reused)

Derivation from NOFOB

- Let $A = B + D$, $C = 0$, and $M_k = \gamma_k^{-1}\mathrm{Id} - D$, then

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

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

Derivation from NOFOB

- Let $A = B + D$, $C = 0$, and $M_k = \gamma_k^{-1}\mathrm{Id} - D$, then

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

  resolvent of $B + D$ in M_k evaluated as forward-backward step
- Projection correction with
  - Projection metric $S = \mathrm{Id}$ and step-size $\gamma_k \in [\epsilon, \frac{1}{\delta} - \epsilon]$
  - Conservative $\hat{\mu}_k = \frac{1}{\gamma_k^{-1}+\delta}$ (since M_k is $\frac{1}{\gamma_k^{-1}+\delta}$-cocoercive)
  - Relaxation $\theta_k = 1 + \delta\gamma_k \in [1 + \varepsilon, 2 - \varepsilon]$

  gives FBF (and convergence conditions agree)

Derivation from NOFOB

- Let $A = B + D$, $C = 0$, and $M_k = \gamma_k^{-1}\mathrm{Id} - D$, then

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

  resolvent of $B + D$ in M_k evaluated as forward-backward step
- Projection correction with
  - Projection metric $S = \mathrm{Id}$ and step-size $\gamma_k \in [\epsilon, \frac{1}{\delta} - \epsilon]$
  - Conservative $\hat{\mu}_k = \frac{1}{\gamma_k^{-1}+\delta}$ (since M_k is $\frac{1}{\gamma_k^{-1}+\delta}$-cocoercive)
  - Relaxation $\theta_k = 1 + \delta\gamma_k \in [1 + \varepsilon, 2 - \varepsilon]$

  gives FBF (and convergence conditions agree)
- Momentum correction with $S = \mathrm{Id}$ gives Malitsky–Tam
  - Lipschitz constant for $\gamma_k M_k - S = \gamma_k D$ is $L_k = \gamma_k\delta$
  - Convergence condition: $L_k + L_{k-1} \leq 1 - \epsilon$
  - Satisfied if all $\gamma_k \in [\epsilon, \frac{1-\epsilon}{2\delta}]$ (which is condition in Malitsky–Tam)

Extensions

- Extensions with cocoercive term exist
    - Forward-backward-half-forward (projection correction)
    - Three-operator-splitting in Malitsky–Tam (momentum correction)
- Polyak momentum extension also in Malitsky–Tam paper

Solodov and Tseng

- Solves

$$0 \in Dx + N_X x$$

where (in Theorem 3.1)

  - D is maximally monotone and δ-Lipschitz continuous
  - N_X is normal cone operator to nonempty closed convex set X

- Let $A = D + N_X$, $C = 0$, $M_k = \gamma_k^{-1}\mathrm{Id} - D$, projection correction

$$\begin{aligned}
\hat{x}_k &= (\mathrm{Id} + \gamma_k N_X)^{-1}(x_k - \gamma_k D x_k) = \Pi_X(x_k - \gamma_k D x_k) \\
\mu_k &= \gamma_k \frac{\langle x_k - \hat{x}_k - \gamma_k D x_k + \gamma_k D\hat{x}_k, x_k - \hat{x}_k \rangle}{\|x_k - \hat{x}_k - \gamma_k D x_k + \gamma_k D\hat{x}_k\|_{S^{-1}}^2} \\
x_{k+1} &= x_k - \tfrac{\theta_k \mu_k}{\gamma_k} S^{-1}(x_k - \gamma_k D x_k - (\hat{x}_k - \gamma_k D\hat{x}_k))
\end{aligned}$$

- Algorithm uses two evaluations of D

Solodov and Tseng

- The NOFOB algorithm:

$$\hat{x}_k = (\mathrm{Id} + \gamma_k N_X)^{-1}(x_k - \gamma_k D x_k) = \Pi_X(x_k - \gamma_k D x_k)$$
$$\mu_k = \gamma_k \frac{\langle x_k - \hat{x}_k - \gamma_k D x_k + \gamma_k D \hat{x}_k, x_k - \hat{x}_k \rangle}{\|x_k - \hat{x}_k - \gamma_k D x_k + \gamma_k D \hat{x}_k\|_{S^{-1}}^2}$$
$$x_{k+1} = x_k - \tfrac{\theta_k \mu_k}{\gamma_k} S^{-1}(x_k - \gamma_k D x_k - (\hat{x}_k - \gamma_k D \hat{x}_k))$$

- Solodov and Tseng obtained by conservative $\hat{\mu}_k$:

$$\hat{\mu}_k := \gamma_k \frac{(1 - \gamma_k \delta)\|x_k - \hat{x}_k\|^2}{\|x_k - \hat{x}_k - \gamma_k D x_k + \gamma_k D \hat{x}_k\|_{P^{-1}}^2} \leq \mu_k$$

by Cauchy–Scharz and δ-Lipschitz continuity of D in numerator

Solodov and Tseng

- The NOFOB algorithm:

$$\hat{x}_k = (\mathrm{Id} + \gamma_k N_X)^{-1}(x_k - \gamma_k D x_k) = \Pi_X(x_k - \gamma_k D x_k)$$

$$\mu_k = \gamma_k \frac{\langle x_k - \hat{x}_k - \gamma_k D x_k + \gamma_k D\hat{x}_k, x_k - \hat{x}_k \rangle}{\|x_k - \hat{x}_k - \gamma_k D x_k + \gamma_k D\hat{x}_k\|_{S^{-1}}^2}$$

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

- Solodov and Tseng obtained by conservative $\hat{\mu}_k$:

$$\hat{\mu}_k := \gamma_k \frac{(1 - \gamma_k \delta)\|x_k - \hat{x}_k\|^2}{\|x_k - \hat{x}_k - \gamma_k D x_k + \gamma_k D\hat{x}_k\|_{P^{-1}}^2} \leq \mu_k$$

  by Cauchy–Scharz and δ-Lipschitz continuity of D in numerator

- Extensions
  - use μ_k instead of $\hat{\mu}_k$
  - add a cocoercive term
  - use momentum correction instead to avoid one D evaluation

Novel four operator splitting methods

- Solves monotone inclusions

$$0 \in Bx + Cx + Dx + Kx$$

  where

  - $B + D$ maximally monotone and D is δ-Lipschitz continuous
  - C is $\frac{1}{\ell}$-cocoercive (w.r.t. P or S)
  - 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 in D, K, and C, resolvent in B

- Use projection correction or momentum correction

Asymmetric forward-backward-adjoint splitting (AFBA)

- If $D = 0$, $Q_k = P$ and projection correction is used, we get AFBA
- Special cases, e.g.,:
  - Proximal alternating predictor corrector
  - Primal dual method of He and Yuan
  - Primal dual method of Briceño-Arias and Combettes

Primal-dual framework

- Problem

$$0 \in B_1 y + (V^* \circ B_2 \circ V)y + Ey + Fy$$

  - $B_1 \colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B_2 \colon \mathcal{K} \to 2^{\mathcal{K}}$ are maximally monotone
  - $E \colon \mathcal{H} \to \mathcal{H}$ is monotone and δ-Lipschitz continuous
  - $F \colon \mathcal{H} \to \mathcal{H}$ is β^{-1}-cocoercive
  - $V \colon \mathcal{H} \to \mathcal{K}$ is linear and bounded

- Four-operator splitting primal-dual formulation

$$0 \in Bx + Cx + Dx + Kx$$

with $x = (y, z) \in \mathcal{H} \times \mathcal{K}$ and

$$B = \begin{bmatrix} B_1 & 0 \\ 0 & B_2^{-1} \end{bmatrix}, \quad D = \begin{bmatrix} E & 0 \\ 0 & 0 \end{bmatrix}, \quad K = \begin{bmatrix} 0 & V^* \\ -V & 0 \end{bmatrix}, \quad C = \begin{bmatrix} F & 0 \\ 0 & 0 \end{bmatrix}$$

which satisfies four-operator splitting assumptions

Primal-dual kernel

- We use the following kernel in NOFOB

$$M_k = \underbrace{\begin{bmatrix} \tau^{-1}\mathrm{Id} & 0 \\ -\lambda_k V & \sigma^{-1}\mathrm{Id} \end{bmatrix}}_{Q_k} - \underbrace{\begin{bmatrix} E & 0 \\ 0 & 0 \end{bmatrix}}_{D} - \underbrace{\begin{bmatrix} 0 & V^* \\ -V & 0 \end{bmatrix}}_{K}$$

- This gives nonlinear forward-backward step

$$\begin{aligned} x_{k+1} &= (M_k + A)^{-1}(M_k - C)x_k \\ &= (Q_k + B)^{-1}(Q_k - D - K - C)x_k \\ &= \begin{bmatrix} (\mathrm{Id} + \tau B_1)^{-1}(y_k - \tau E y_k - \tau V^* z_k - \tau F y_k) \\ (\mathrm{Id} + \sigma B_2^{-1})^{-1}(z_k + \sigma V(\lambda_k y_{k+1} - (\lambda_k - 1)y_k)) \end{bmatrix} \end{aligned}$$

- If $\lambda_k = 2$ and $E = 0$, we get Condat–Vũ and $M_k = M \in \mathcal{P}(\mathcal{H})$
- In general, $M_k \notin \mathcal{P}(\mathcal{H})$ and we need correction

A primal-dual method with projection correction

- Use projection correction with metric operator S

$$S = \begin{bmatrix} \tau^{-1}\mathrm{Id} & 0 \\ 0 & \sigma^{-1}\mathrm{Id} \end{bmatrix}$$

  comments

  - S is diagonal for cheap evaluation of $S^{-1}D$ (in $S^{-1}M_k$)
  - if $D = 0$: S that includes V, V^* possible

- Set $\lambda_k = \lambda \in \mathbb{R}$ in M_k and use projection correction

$$\begin{aligned}
\hat{y}_k &= (\mathrm{Id} + \tau B_1)^{-1}(y_k - \tau E y_k - \tau V^* z_k - \tau F y_k) \\
\hat{z}_k &= (\mathrm{Id} + \sigma B_2^{-1})^{-1}(z_k + \sigma V(\lambda y_{k+1} - (\lambda - 1) y_k)) \\
y_{k+1} &= y_k - \theta_k \mu_k (y_k - \hat{y}_k - \tau V^*(z_k - \hat{z}_k) - \tau(E y_k - E\hat{y}_k)) \\
z_{k+1} &= z_k - \theta_k \mu_k (z_k - \hat{z}_k + (1 - \lambda) V(y_k - \hat{y}_k))
\end{aligned}$$

Comments

- The primal-dual algorithm

$$
\begin{aligned}
\hat{y}_k &= (\mathrm{Id} + \tau B_1)^{-1}(y_k - \tau E y_k - \tau V^* z_k - \tau F y_k) \\
\hat{z}_k &= (\mathrm{Id} + \sigma B_2^{-1})^{-1}(z_k + \sigma V(\lambda y_{k+1} - (\lambda - 1) y_k)) \\
y_{k+1} &= y_k - \theta_k \mu_k (y_k - \hat{y}_k - \tau V^*(z_k - \hat{z}_k) - \tau(E y_k - E \hat{y}_k)) \\
z_{k+1} &= z_k - \theta_k \mu_k (z_k - \hat{z}_k + (1 - \lambda) V(y_k - \hat{y}_k))
\end{aligned}
$$

- Comments
  - Evaluations
    - Two for V, V^* (unless $\lambda = 1$) and E
    - One for remaining operators
  - If $D = 0$ and $\lambda_k = 2$, then $M_k \in \mathcal{P}(\mathcal{H})$
    - Choice of S gives $S^{-1} M_k \neq \alpha \mathrm{Id}$ for any $\alpha \in \mathbb{R}$
    - Algorithm does not give Condat-Vũ (but different S does)
  - Convergence with specific P (used to compute μ_k) if

$$1 - \tfrac{\tau\sigma\lambda^2}{4}\|V\|^2 - \tau\delta - \tfrac{\tau\beta}{4} > 0$$

Convergence proof

- Let $(\tau^{-1} - \delta - \frac{\sigma\lambda^2}{4}\|V\|^2) > 0$ and set

$$P = \begin{bmatrix} \tau^{-1}\mathrm{Id} - \delta\mathrm{Id} & -\frac{\lambda}{2}L^* \\ -\frac{\lambda}{2}L & \sigma^{-1}\mathrm{Id} \end{bmatrix} \in \mathcal{P}(\mathcal{H})$$

- The kernel M_k is 1-strongly monotone w.r.t. P since

$$M_k = P + \begin{bmatrix} 0 & \frac{\lambda}{2}L^* \\ -\frac{\lambda}{2}L & 0 \end{bmatrix} + \begin{bmatrix} \delta\mathrm{Id} - E & 0 \\ 0 & 0 \end{bmatrix} - K$$

which implies

$$\begin{aligned} \langle M_k x - M_k x', x - x' \rangle &= \|x - x'\|_P^2 + \langle \delta y - Ey - (\delta y' - Ey'), y - y' \rangle \\ &\geq \|x - x'\|_P^2 \end{aligned}$$

- C is $(\tau^{-1} - \delta - \frac{\sigma\lambda^2}{4}\|V\|^2)/\beta$-cocoercive w.r.t. P

$$\begin{aligned} \|Cx - Cx'\|_{P^{-1}} &\leq (\tau^{-1} - \delta - \tfrac{\sigma\lambda^2}{4}\|V\|^2)^{-1}\|Fy - Fy'\|^2 \\ &\leq (\tau^{-1} - \delta - \tfrac{\sigma\lambda^2}{4}\|V\|^2)^{-1}\langle Fy - Fy', y - y' \rangle \\ &= \beta(\tau^{-1} - \delta - \tfrac{\sigma\lambda^2}{4}\|V\|^2)^{-1}\langle Cx - Cx', x - x' \rangle \end{aligned}$$

i.e., $\beta/(\tau^{-1} - \delta - \frac{\sigma\lambda^2}{4}\|V\|^2) \in [0, 4)$, upper bound gives condition

A primal-dual method with momentum correction

- Momentum correction and algorithm design parameters

$$S = \begin{bmatrix} \mathrm{Id} & -\tau V^* \\ -\tau V & \sigma^{-1}\tau \mathrm{Id} \end{bmatrix}, \quad M_k = \begin{bmatrix} \tau^{-1}\mathrm{Id} & 0 \\ -\lambda_k V & \sigma^{-1}\mathrm{Id} \end{bmatrix} - D - K, \quad \gamma_k = \tau$$

- Gives a lower block-triangular update and algorithm

$$\begin{aligned} y_{k+1} &= (\mathrm{Id} + \tau B_1)^{-1}(y_k - \tau V^* z_k - \tau(2Ey_k - Ey_{k-1}) - \tau F y_k), \\ v_{k+1} &= \lambda_k(y_{k+1} - y_k) + (2 - \lambda_{k-1})(y_k - y_{k-1}), \\ z_{k+1} &= (\mathrm{Id} + \sigma B_2^{-1})^{-1}(z_k + \sigma V(y_k + v_{k+1})), \end{aligned}$$

- Comments
  - Each resolvent, forward step, and V and V^* evaluated once
  - if $F = 0$, $\lambda_k = 2$, $V = \mathrm{Id} \Rightarrow$ forward-reflected-Douglas-Rachford
  - if $E = 0$, $\lambda_k = 2 \Rightarrow$ Condat–Vũ (by choice of S)
  - If $B_2 = 0$ ($V = 0$) we get Malitsky–Tam three-operator splitting

Convergence

- Convergence condition

$$\tau\sigma\|V\|^2 + (|2-\lambda_k| + |2-\lambda_{k+1}|)\sqrt{\tau\sigma}\|V\| + \tau(2\delta + \tfrac{1}{2}\beta) < 1-\epsilon$$

- Proof
  - C is $\frac{1}{\ell}$-cocoercive w.r.t. S with $\ell = \frac{\beta}{1-\tau\sigma\|V\|^2}$
  - $\gamma_k M_k - S$ is L_k-Lipschitz continuous w.r.t. S where

$$L_k = \frac{1}{1-\tau\sigma\|V\|^2}(|2-\lambda_k|\sqrt{\tau\sigma}\|V\| + \tau\delta)$$

  - Insert into general convergence condition

$$1 - L_{k-1} - L_k - \frac{\tau\ell}{2} \geq \epsilon > 0$$

    to get result

- Reduces to Malitsky–Tam condition if $V = 0$

A primal-dual method with resolvent in kernel

- Let T_a be translation by a and use

$$M_k = \begin{bmatrix} \tau^{-1}\mathrm{Id} - V^* \circ (\mathrm{Id} + \sigma B_2^{-1})^{-1} \circ T_{-z_k} \circ \sigma V & 0 \\ 0 & \sigma^{-1}\mathrm{Id} \end{bmatrix} - D,$$

$$S = \begin{bmatrix} \mathrm{Id} & 0 \\ 0 & \tau\sigma^{-1}\mathrm{Id} \end{bmatrix}, \quad \gamma_k = \tau$$

- With momentum correction and after some algebra, we get

$$\nu_{k+1} = (\mathrm{Id} + \sigma B_2^{-1})^{-1}(z_k + \sigma V y_k)$$
$$y_{k+1} = (\mathrm{Id} + \tau B_1)^{-1}(y_k - \tau V^*(z_k + \nu_{k+1} - \nu_k) - \tau(2Ey_k - Ey_{k-1}) - \tau F y_k)$$
$$z_{k+1} = (\mathrm{Id} + \sigma B_2^{-1})^{-1}(z_k + \sigma V y_{k+1})$$

- Comments
  - B_2^{-1} resolvent evaluated twice, remaining operators once
  - Not aware of other methods that require extra B_2^{-1} resolvent
  - If $B_2 = 0$ ($V = 0$) we get Malitsky–Tam three-operator splitting

Convergence

- Algorithm converges if

$$2\tau\sigma\|V\|^2 + \tau(2\delta + \tfrac{\beta}{2}) < 1,$$

- Proof
  - C is $\frac{1}{\ell}$-cocoercive w.r.t. S with $\ell = \beta$
  - $\gamma_k M_k - S$ is $L_k = (\tau\delta + \tau\sigma\|V\|^2)$-Lipschitz continuous w.r.t. S
  - Insert into general convergence condition

$$1 - L_{k-1} - L_k - \tfrac{\tau\ell}{2} \geq \epsilon > 0$$

    to get result

- Reduces to Malitsky–Tam condition if $V = 0$

Extension to multi-operator problems

- Consider 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 formulation (monotone+skew)

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

fits in four-operator splitting framework (with $C = D = 0$)

- Can be extended by Lipschitz and cocoercive operators in last row

Projective splitting – How it usually looks

Algorithm Synchronous Projective Splitting Combettes, Eckstein 2018

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

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

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

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

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

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

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

$\quad$ **end for**

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

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

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

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

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

$\quad$ **end for**

$\quad x_{k+1} := x_k - \theta_k \mu_k t_k^*$

end for

Projective splitting from NOFOB

- Let $A = B + K$, $C = 0$ and subtract skew linear K in M_k

$$M_k = \underbrace{\begin{bmatrix} \sigma_1^{-1}\mathrm{Id} & & & \\ & \ddots & & \\ & & \sigma_{n-1}^{-1}\mathrm{Id} & \\ & & & \tau\mathrm{Id} \end{bmatrix}}_{P} - \underbrace{\begin{bmatrix} & & & -L_1 \\ & & & \vdots \\ & & & -L_{n-1} \\ L_1^* & \cdots & L_{n-1}^* & \end{bmatrix}}_{K}$$

- Projective splitting: backward-step on $A = B + K$ $(C = 0)$

$$\begin{aligned}\hat{p}_k &= (M_k + A)^{-1} M_k p_k \\ &= (P + K + B - K)^{-1}(P - K)p_k = (P + B)^{-1}(P - K)p_k\end{aligned}$$

  with $p_k = (w_{1,k}, \ldots, w_{n-1,k}, x_k)$ and project (M_k not symmetric)
- σ_i and τ are individual resolvent parameters for B_i
- $P = \epsilon\mathrm{Id}$: M_k is 1-strongly monotone w.r.t. P for all $\sigma_i, \tau > 0$
  $\Rightarrow$ no step-size restrictions but projection needed!

Summary

- We have presented NOFOB framework
- Can use projection or momentum correction
- Many existing operator splitting methods are special cases
- Easy to design and prove convergence of new methods

Thank you

Based on:

[1] P. Giselsson, *Nonlinear Forward-Backward Splitting with Projection Correction*, SIAM Journal on Optimization, 2021.

[2] M. Morin, S. Banert. P. Giselsson, *Nonlinear Forward-Backward Splitting with Momentum Correction*. Submitted (available: arXiv:2112.00481), 2021.