

The averagedness of Douglas–Rachford and forward-backward operators

Speaker: Walaa M. Moursi

Electrical Engineering, Stanford University, USA

Based partially on joint works with

Heinz H. Bauschke

University of British Columbia
Canada

Xianfu Wang

University of British Columbia
Canada

2nd Workshop on Operator Splitting Methods in Data Analysis
Center for Computational Mathematics, Faltiron Institute
New York, USA
Thursday, March 21, 2019
2:00 pm–2:25 pm

Introduction

Throughout this talk

X is a real Hilbert space

with inner product $\langle \cdot | \cdot \rangle$, and induced norm $\|\cdot\|$.

- ▶ Let $f: X \rightarrow]-\infty, +\infty]$ and $g: X \rightarrow]-\infty, +\infty]$ be proper and lower semicontinuous.
- ▶ Suppose that $f + g$ is convex such that $\operatorname{argmin}(f + g) \neq \emptyset$.

Introduction

Throughout this talk

X is a real Hilbert space

with inner product $\langle \cdot | \cdot \rangle$, and induced norm $\|\cdot\|$.

- ▶ Let $f: X \rightarrow]-\infty, +\infty]$ and $g: X \rightarrow]-\infty, +\infty]$ be proper and lower semicontinuous.
- ▶ Suppose that $f + g$ is convex such that $\operatorname{argmin}(f + g) \neq \emptyset$.
- ▶ When f and g are nice enough, splitting methods (e.g., Douglas–Rachford and forward-backward) can be used to solve the problem:

Find $x \in X$ such that x minimizes $f + g$.

Introduction

Throughout this talk

X is a real Hilbert space

with inner product $\langle \cdot | \cdot \rangle$, and induced norm $\|\cdot\|$.

- ▶ Let $f: X \rightarrow]-\infty, +\infty]$ and $g: X \rightarrow]-\infty, +\infty]$ be proper and lower semicontinuous.
- ▶ Suppose that $f + g$ is convex such that $\operatorname{argmin}(f + g) \neq \emptyset$.
- ▶ When f and g are **nice** enough, splitting methods (e.g., Douglas–Rachford and forward-backward) can be used to solve the problem:

Find $x \in X$ such that x minimizes $f + g$.

- ▶ **Nice** classes of functions in the case of forward-backward method include functions that satisfy the Kurdyka–Łojasiewicz (KL) property (Attouch et al, Bolte et al, Boţ et al, etc ...).
- ▶ **Nice** classes of functions in the case of Douglas–Rachford method include hypoconvex functions (Yuan et al, Dao and Phan).
- ▶ The proofs use Fejér monotonicity of the sequence of iterates w.r.t. the set of critical points of $f + g$.

Nonexpansiveness and related concepts

Let $T: X \rightarrow X$ and let $(x, y) \in X \times X$. Recall that

- ▶ T is **nonexpansive** if $\|Tx - Ty\| \leq \|x - y\|$.
- ▶ T is **α -averaged** if $T = (1 - \alpha)\text{Id} + \alpha N$, $\alpha \in]0, 1[$, N is nonexpansive.
- ▶ T is **firmly nonexpansive** if
$$\|Tx - Ty\|^2 + \|(\text{Id} - T)x - (\text{Id} - T)y\|^2 \leq \|x - y\|^2.$$
- ▶ Let $\beta > 0$. Then T is **cocoercive** if βT is firmly nonexpansive.

Monotone operators

- ▶ Recall that an operator $A: X \rightrightarrows X$ is **monotone** if

$$\{(x, u), (y, v)\} \subseteq \operatorname{gr} A \Rightarrow \langle x - y \mid u - v \rangle \geq 0.$$

- ▶ Recall also that a monotone operator A is **maximally monotone** if A cannot be properly extended without destroying monotonicity.
- ▶ **Examples:** Matrices with positive semidefinite parts, **subdifferential operators** ∂f of convex functions and skew symmetric operators, e.g.,

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

ρ -monotonicity

Let $A: X \rightrightarrows X$ and let $\rho \in \mathbb{R}$. Then

(i) A is ρ -monotone if $(\forall (x, u) \in \text{gr } A) (\forall (y, v) \in \text{gr } A)$ we have

$$\langle x - y \mid u - v \rangle \geq \rho \|x - y\|^2.$$

ρ -monotonicity

Let $A: X \rightrightarrows X$ and let $\rho \in \mathbb{R}$. Then

(i) A is ρ -monotone if $(\forall (x, u) \in \text{gr } A) (\forall (y, v) \in \text{gr } A)$ we have

$$\langle x - y \mid u - v \rangle \geq \rho \|x - y\|^2.$$

(ii) A is *maximally ρ -monotone* if A is ρ -monotone and there is no ρ -monotone operator $B: X \rightrightarrows X$ such that $\text{gr } B$ properly contains $\text{gr } A$, i.e., for every $(x, u) \in X \times X$,

$$(x, u) \in \text{gr } A \Leftrightarrow (\forall (y, v) \in \text{gr } A) \langle x - y \mid u - v \rangle \geq \rho \|x - y\|^2.$$

ρ -monotonicity: Resolvents

Proposition

Let $A: X \rightrightarrows X$ be *maximally* ρ -monotone where $\rho > -1$. Then the following hold:

(i) $\text{ran}(\text{Id} + A) = X = \text{dom } J_A$.

ρ -monotonicity: Resolvents

Proposition

Let $A: X \rightrightarrows X$ be *maximally* ρ -monotone where $\rho > -1$. Then the following hold:

- (i) $\text{ran}(\text{Id} + A) = X = \text{dom } J_A$.
- (ii) J_A is single-valued.

ρ -monotonicity: Resolvents

Proposition

Let $A: X \rightrightarrows X$ be *maximally* ρ -monotone where $\rho > -1$. Then the following hold:

- (i) $\text{ran}(\text{Id} + A) = X = \text{dom } J_A$.
- (ii) J_A is single-valued.
- (iii) J_A is $(1 + \rho)$ -cocoercive.

$\rho > -1$ is critical!

Example

Suppose that $X \neq \{0\}$.

- ▶ Let C be a nonempty closed convex subset of X ,
- ▶ let $r \in \mathbb{R}_+$,
- ▶ set $A = -\text{Id} - rP_C$, and set $\rho = -(1+r) \leq -1$.

Then the following hold:

- (i) $A - \rho \text{Id} = r(\text{Id} - P_C)$ is maximally monotone.
- (ii) A is maximally ρ -monotone.
- (iii) $\text{ran}(\text{Id} + A) = \text{ran}(-rP_C) = -rC = (\rho + 1)C$.
- (iv) $\text{Id} + A$ is surjective $\Leftrightarrow [C = X \text{ and } r > 0]$.
- (v) $J_A = (-rP_C)^{-1}$ is at most single-valued $\Leftrightarrow [C = X \text{ and } r > 0]$.

Minty parametrization

Theorem (**Minty parametrization**)

Let $A: X \rightrightarrows X$ be ρ -monotone where $\rho > -1$. Then

$$\text{gr } A = \{ (J_A x, (\text{Id} - J_A)x) \mid x \in \text{ran}(\text{Id} + A) \}.$$

Minty parametrization

Theorem (Minty parametrization)

Let $A: X \rightrightarrows X$ be ρ -monotone where $\rho > -1$. Then

$$\text{gr } A = \{(J_A x, (\text{Id} - J_A)x) \mid x \in \text{ran}(\text{Id} + A)\}.$$

Moreover, A is maximally ρ -monotone $\Leftrightarrow \text{ran}(\text{Id} + A) = X$, in which case

$$\text{gr } A = \{(J_A x, (\text{Id} - J_A)x) \mid x \in X\}.$$

Hypoconvex functions (a.k.a. weakly convex functions)

Let $\lambda > 0$. Recall that f is **λ -hypoconvex** if for all $(x, y) \in X \times X$ and $\tau \in]0, 1[$,

$$f((1 - \tau)x + \tau y) \leq (1 - \tau)f(x) + \tau f(y) + \frac{\lambda}{2} \tau(1 - \tau) \|x - y\|^2,$$

or, equivalently,

$$f + \frac{\lambda}{2} \|\cdot\|^2 \text{ is convex.}$$

Hypoconvex functions (a.k.a. weakly convex functions)

Let $\lambda > 0$. Recall that f is **λ -hypoconvex** if for all $(x, y) \in X \times X$ and $\tau \in]0, 1[$,

$$f((1 - \tau)x + \tau y) \leq (1 - \tau)f(x) + \tau f(y) + \frac{\lambda}{2} \tau(1 - \tau) \|x - y\|^2,$$

or, equivalently,

$$f + \frac{\lambda}{2} \|\cdot\|^2 \text{ is convex.}$$

An **abstract subdifferential** $\partial_{\#}$ associates a subset $\partial_{\#}f(x)$ of X to f at $x \in X$, and it satisfies the following properties:

- (i) $\partial_{\#}f = \partial f$ if f is a proper lower semicontinuous convex function;
- (ii) $\partial_{\#}f = \nabla f$ if f is continuously differentiable;
- (iii) $0 \in \partial_{\#}f(x)$ if f attains a local minimum at $x \in \text{dom } f$;
- (iv) for every $\beta \in \mathbb{R}$,

$$\partial_{\#}\left(f + \beta \frac{\|\cdot - x\|^2}{2}\right) = \partial_{\#}f + \beta(\text{Id} - x).$$

Hypoconvex functions

Proposition

Suppose that $f: X \rightarrow]-\infty, +\infty]$ is a proper lower semicontinuous λ -hypoconvex function. Then

$$\partial_{\#}f = \partial\left(f + \frac{\lambda}{2}\|\cdot\|^2\right) - \lambda \operatorname{Id}.$$

• For $\gamma > 0$, the *proximal mapping* $\operatorname{Prox}_{\gamma f}$ is defined at $x \in X$ by $\operatorname{Prox}_{\gamma f}(x) = \operatorname{argmin}_{y \in X} (f(y) + \frac{1}{2\gamma}\|x - y\|^2)$.

Hypoconvex functions

Proposition

Suppose that $f: X \rightarrow]-\infty, +\infty]$ is a proper lower semicontinuous λ -hypoconvex function. Then

$$\partial_{\#}f = \partial\left(f + \frac{\lambda}{2}\|\cdot\|^2\right) - \lambda \operatorname{Id}.$$

Moreover, we have:

- (i) The Clarke–Rockafellar, Mordukhovich, and Fréchet subdifferential operators of f all coincide, (easy)
- (ii) $\partial_{\#}f$ is maximally $(-\lambda)$ -monotone, (easy)
- (iii) $(\forall \mu \in]0, \lambda[) \operatorname{Prox}_{\mu f} = J_{\mu \partial_{\#}f} = (\operatorname{Id} + \mu \partial_{\#}f)^{-1},$

• For $\gamma > 0$, the proximal mapping $\operatorname{Prox}_{\gamma f}$ is defined at $x \in X$ by $\operatorname{Prox}_{\gamma f}(x) = \operatorname{argmin}_{y \in X} (f(y) + \frac{1}{2\gamma} \|x - y\|^2).$

Hypoconvex functions

Proposition

Suppose that $f: X \rightarrow]-\infty, +\infty]$ is a proper lower semicontinuous λ -hypoconvex function. Then

$$\partial_{\#}f = \partial\left(f + \frac{\lambda}{2}\|\cdot\|^2\right) - \lambda \operatorname{Id}.$$

Moreover, we have:

- (i) The Clarke–Rockafellar, Mordukhovich, and Fréchet subdifferential operators of f all coincide, (easy)
- (ii) $\partial_{\#}f$ is maximally $(-\lambda)$ -monotone, (easy)
- (iii) $(\forall \mu \in]0, \lambda[) \operatorname{Prox}_{\mu f} = J_{\mu \partial_{\#}f} = (\operatorname{Id} + \mu \partial_{\#}f)^{-1}$,
- (iv) $(\forall \mu \in]0, \lambda[) \operatorname{Prox}_{\mu f}$ is $\frac{\lambda - \mu}{\lambda}$ -cocoercive. (consequence of our earlier results)

• For $\gamma > 0$, the proximal mapping $\operatorname{Prox}_{\gamma f}$ is defined at $x \in X$ by $\operatorname{Prox}_{\gamma f}(x) = \operatorname{argmin}_{y \in X} (f(y) + \frac{1}{2\gamma} \|x - y\|^2)$.

To conclude this part:

- ▶ The class of hypomonotone operators (ρ -monotone, when $\rho < 0$) is a nice class: Indeed, we obtain **single-valuedness**, **full domain** and **cocoercivity** of the resolvents.
- ▶ **BUT** this is a special class.
- ▶ **Question:**
What are other possible/more general classes of operators that have “nice” resolvents?

PART II: On the averagedness of the Douglas–Rachford operator

The Douglas–Rachford operator

Suppose that

A and B are maximally monotone operators on X .

The problem:

Find $x \in X$ such that

$$x \in \text{zer}(A + B) = (A + B)^{-1}(0).$$

$$\bullet J_A = (\text{Id} + A)^{-1}. \bullet R_A = 2J_A - \text{Id}.$$

The Douglas–Rachford operator

Suppose that

A and B are maximally monotone operators on X .

The problem:

Find $x \in X$ such that

$$x \in \operatorname{zer}(A + B) = (A + B)^{-1}(0).$$

The Douglas–Rachford algorithm: One successful technique to find a zero of $A + B$ is via iterating the Douglas–Rachford operator $T_{A,B}$ defined for the **ordered** pair (A, B) by

$$T = T_{A,B} = \frac{1}{2}(\operatorname{Id} + R_B R_A).$$

$$\bullet J_A = (\operatorname{Id} + A)^{-1}. \bullet R_A = 2J_A - \operatorname{Id}.$$

Classical convergence results

Let $x_0 \in X$. Recall that $T_{A,B} = \frac{1}{2}(\text{Id} + R_B R_A)$. When

$$\text{zer}(A + B) \neq \emptyset$$

we have, for A and B maximally monotone:

- Lions–Mercier (1979) and Eckstein–Bertsekas (1992)

T is firmly nonexpansive.

- Krasnosel'skiĭ–Mann (1950s)

$x_n = T_{A,B}^n x_0 \xrightarrow{\text{weakly}} \text{some point } \bar{x} \in \text{Fix } T_{A,B} \neq \text{zer}(A + B) \text{ (in general).}$

Classical convergence results

Let $x_0 \in X$. Recall that $T_{A,B} = \frac{1}{2}(\text{Id} + R_B R_A)$. When

$$\text{zer}(A + B) \neq \emptyset$$

we have, for A and B maximally monotone:

- Lions–Mercier (1979) and Eckstein–Bertsekas (1992)

T is firmly nonexpansive.

- Krasnosel'skiĭ–Mann (1950s)

$$x_n = T_{A,B}^n x_0 \xrightarrow{\text{weakly}} \text{some point } \bar{x} \in \text{Fix } T_{A,B} \neq \text{zer}(A + B) \text{ (in general).}$$

- Combettes (2004) $J_A(\text{Fix } T_{A,B}) = \text{zer}(A + B)$.

Classical convergence results

Let $x_0 \in X$. Recall that $T_{A,B} = \frac{1}{2}(\text{Id} + R_B R_A)$. When

$$\text{zer}(A + B) \neq \emptyset$$

we have, for A and B maximally monotone:

- Lions–Mercier (1979) and Eckstein–Bertsekas (1992)

T is firmly nonexpansive.

- Krasnosel'skiĭ–Mann (1950s)

$$x_n = T_{A,B}^n x_0 \xrightarrow{\text{weakly}} \text{some point } \bar{x} \in \text{Fix } T_{A,B} \neq \text{zer}(A + B) \text{ (in general).}$$

- Combettes (2004) $J_A(\text{Fix } T_{A,B}) = \text{zer}(A + B)$.
- Lions–Mercier (1979) and Svaiter (2011)

$$J_A T_{A,B}^n x \xrightarrow{\text{weakly}} \text{some point in } \text{zer}(A + B).$$



In the absence of monotonicity

- ▶ Suppose for instance that A is **not monotone**. Then J_A (and, in turn, R_A and T) is **not** necessarily single-valued and/or does **not** necessarily have full domain. ☹

In the absence of monotonicity

- ▶ Suppose for instance that A is **not monotone**. Then J_A (and, in turn, R_A and T) is **not** necessarily single-valued and/or does **not** necessarily have full domain. ☹
- ▶ **BUT**, if A is ρ -monotone, and B is nicer than merely monotone, then there is hope for some fun. ☺

Can we show that T is “nice”?

Lemma

Let $\lambda \in]0, 1[$. Suppose that $T_1: X \rightarrow X$ and $T_2: X \rightarrow X$. Set

$$T_\lambda = (1 - \lambda) \text{Id} + \lambda(2T_2 - \text{Id}) \underbrace{(2T_1 - \text{Id})}_{R_1}.$$

Let $(x, y) \in X \times X$.

Can we show that T is “nice”?

Lemma

Let $\lambda \in]0, 1[$. Suppose that $T_1: X \rightarrow X$ and $T_2: X \rightarrow X$. Set

$$T_\lambda = (1 - \lambda) \text{Id} + \lambda(2T_2 - \text{Id}) \underbrace{(2T_1 - \text{Id})}_{R_1}.$$

Let $(x, y) \in X \times X$. Then

$$\begin{aligned} & \langle T_\lambda x - T_\lambda y \mid (\text{Id} - T_\lambda)x - (\text{Id} - T_\lambda)y \rangle \\ &= (1 - 2\lambda) \langle x - y \mid (\text{Id} - T_\lambda)x - (\text{Id} - T_\lambda)y \rangle \\ &+ 4\lambda^2 \langle T_1 x - T_1 y \mid (\text{Id} - T_1)x - (\text{Id} - T_1)y \rangle \\ &+ 4\lambda^2 \langle T_2 R_1 x - T_2 R_1 y \mid (\text{Id} - T_2)R_1 x - (\text{Id} - T_2)R_1 y \rangle. \end{aligned}$$

Can we show that T is “nice”?

Lemma

Let $\lambda \in]0, 1[$. Suppose that $T_1: X \rightarrow X$ and $T_2: X \rightarrow X$. Set

$$T_\lambda = (1 - \lambda) \text{Id} + \lambda (2T_2 - \text{Id}) \underbrace{(2T_1 - \text{Id})}_{R_1}.$$

Let $(x, y) \in X \times X$. Then

$$\begin{aligned} & \langle T_\lambda x - T_\lambda y \mid (\text{Id} - T_\lambda)x - (\text{Id} - T_\lambda)y \rangle \\ &= (1 - 2\lambda) \langle x - y \mid (\text{Id} - T_\lambda)x - (\text{Id} - T_\lambda)y \rangle \\ &+ 4\lambda^2 \langle T_1 x - T_1 y \mid (\text{Id} - T_1)x - (\text{Id} - T_1)y \rangle \\ &+ 4\lambda^2 \langle T_2 R_1 x - T_2 R_1 y \mid (\text{Id} - T_2)R_1 x - (\text{Id} - T_2)R_1 y \rangle. \end{aligned}$$

Corollary (Eckstein–Bertsekas)

Suppose that $T_1: X \rightarrow X$ and $T_2: X \rightarrow X$. Set

$$T = \frac{1}{2}(\text{Id} + (2T_2 - \text{Id})(2T_1 - \text{Id})).$$

Then

$$\begin{aligned} \langle T x - T y \mid (\text{Id} - T)x - (\text{Id} - T)y \rangle &= \langle T_1 x - T_1 y \mid (\text{Id} - T_1)x - (\text{Id} - T_1)y \rangle \\ &+ \langle T_2 R_1 x - T_2 R_1 y \mid (\text{Id} - T_2)R_1 x - (\text{Id} - T_2)R_1 y \rangle. \end{aligned}$$

Characterization of averaged operators

Lemma

Let $T: X \rightarrow X$, let $\alpha \in]0, 1[$ and let $(x, y) \in X \times X$.

Characterization of averaged operators

Lemma

Let $T: X \rightarrow X$, let $\alpha \in]0, 1[$ and let $(x, y) \in X \times X$. Then: T is α -averaged $\Leftrightarrow$

$$2\alpha \langle Tx - Ty \mid (\text{Id} - T)x - (\text{Id} - T)y \rangle \geq (1 - 2\alpha) \|(\text{Id} - T)x - (\text{Id} - T)y\|^2.$$

Averagedness of the classical Douglas–Rachford operator when $\mu > \omega \geq 0$

Theorem

Let $\mu > \omega \geq 0$ and let $\gamma \in]0, (\mu - \omega)/(2\mu\omega)[$. Suppose that one of the following holds:

- (i) A is maximally $(-\omega)$ -monotone and B is maximally μ -monotone.
- (ii) A is maximally μ -monotone and B is maximally $(-\omega)$ -monotone.

Set

$$T = \frac{1}{2}(\text{Id} + R_{\gamma B} R_{\gamma A}) = \text{Id} - J_{\gamma A} + J_{\gamma B} R_{\gamma A}, \quad \alpha = \frac{\mu - \omega}{2(\mu - \omega - \gamma\mu\omega)}.$$

Then $\alpha \in]0, 1[$ and T is α -averaged.

Averagedness of the classical Douglas–Rachford operator when $\mu > \omega \geq 0$

Theorem

Let $\mu > \omega \geq 0$ and let $\gamma \in]0, (\mu - \omega)/(2\mu\omega)[$. Suppose that one of the following holds:

- (i) A is maximally $(-\omega)$ -monotone and B is maximally μ -monotone.
- (ii) A is maximally μ -monotone and B is maximally $(-\omega)$ -monotone.

Set

$$T = \frac{1}{2}(\text{Id} + R_{\gamma B} R_{\gamma A}) = \text{Id} - J_{\gamma A} + J_{\gamma B} R_{\gamma A}, \quad \alpha = \frac{\mu - \omega}{2(\mu - \omega - \gamma\mu\omega)}.$$

Then $\alpha \in]0, 1[$ and T is α -averaged.

Proof. To verify that $\alpha \in]0, 1[$ is easy ✓. We now verify that T is averaged assuming (i) holds.

Note that γA is maximally $(-\gamma\omega)$ -monotone and $-\gamma\omega > -1$. Therefore, $J_{\gamma A}$ (and in turn $R_{\gamma A}$ and T) is single-valued, full domain and, in fact, $J_{\gamma A}$ is cocoercive.

Proof continued

Our goal is show that T is averaged, i.e., in view of the earlier characterization of averaged operators:

-
- A is maximally $(-\omega)$ -monotone, B is maximally μ -monotone, $\mu > \omega \geq 0$.
 - $\text{Id} - T = J_{\gamma A} - J_{\gamma B} R_{\gamma A}$

Proof continued

Our goal is show that T is averaged, i.e., in view of the earlier characterization of averaged operators:

$$\langle Tx - Ty \mid (\text{Id} - T)x - (\text{Id} - T)y \rangle$$

$$> -\frac{1}{2} \|(\text{Id} - T)x - (\text{Id} - T)y\|^2.$$

-
- A is maximally $(-\omega)$ -monotone, B is maximally μ -monotone, $\mu > \omega \geq 0$.
 - $\text{Id} - T = J_{\gamma A} - J_{\gamma B} R_{\gamma A}$

Proof continued

Our goal is show that T is averaged, i.e., in view of the earlier characterization of averaged operators:

$$\begin{aligned} & \langle Tx - Ty \mid (\text{Id} - T)x - (\text{Id} - T)y \rangle \\ &= \langle J_{\gamma A}x - J_{\gamma A}y \mid (\text{Id} - J_{\gamma A})x - (\text{Id} - J_{\gamma A})y \rangle \\ & \quad + \langle J_{\gamma B}R_{\gamma A}x - J_{\gamma B}R_{\gamma A}y \mid (\text{Id} - J_{\gamma B})R_{\gamma A}x - (\text{Id} - J_{\gamma B})R_{\gamma A}y \rangle \end{aligned}$$

$$> -\frac{1}{2} \|(\text{Id} - T)x - (\text{Id} - T)y\|^2.$$

-
- A is maximally $(-\omega)$ -monotone, B is maximally μ -monotone, $\mu > \omega \geq 0$.
 - $\text{Id} - T = J_{\gamma A} - J_{\gamma B}R_{\gamma A}$

Proof continued

Our goal is show that T is averaged, i.e., in view of the earlier characterization of averaged operators:

$$\begin{aligned} & \langle Tx - Ty \mid (\text{Id} - T)x - (\text{Id} - T)y \rangle \\ &= \langle J_{\gamma A}x - J_{\gamma A}y \mid (\text{Id} - J_{\gamma A})x - (\text{Id} - J_{\gamma A})y \rangle \\ & \quad + \langle J_{\gamma B}R_{\gamma A}x - J_{\gamma B}R_{\gamma A}y \mid (\text{Id} - J_{\gamma B})R_{\gamma A}x - (\text{Id} - J_{\gamma B})R_{\gamma A}y \rangle \\ & \geq \gamma\mu \|J_{\gamma B}R_{\gamma A}x - J_{\gamma B}R_{\gamma A}y\|^2 - \gamma\omega \|J_{\gamma A}x - J_{\gamma A}y\|^2 \end{aligned}$$

$$> -\frac{1}{2} \|(\text{Id} - T)x - (\text{Id} - T)y\|^2.$$

-
- A is maximally $(-\omega)$ -monotone, B is maximally μ -monotone, $\mu > \omega \geq 0$.
 - $\text{Id} - T = J_{\gamma A} - J_{\gamma B}R_{\gamma A}$

Proof continued

Our goal is show that T is averaged, i.e., in view of the earlier characterization of averaged operators:

$$\begin{aligned} & \langle Tx - Ty \mid (\text{Id} - T)x - (\text{Id} - T)y \rangle \\ &= \langle J_{\gamma A}x - J_{\gamma A}y \mid (\text{Id} - J_{\gamma A})x - (\text{Id} - J_{\gamma A})y \rangle \\ & \quad + \langle J_{\gamma B}R_{\gamma A}x - J_{\gamma B}R_{\gamma A}y \mid (\text{Id} - J_{\gamma B})R_{\gamma A}x - (\text{Id} - J_{\gamma B})R_{\gamma A}y \rangle \\ &\geq \gamma\mu \|J_{\gamma B}R_{\gamma A}x - J_{\gamma B}R_{\gamma A}y\|^2 - \gamma\omega \|J_{\gamma A}x - J_{\gamma A}y\|^2 \\ &= \gamma\mu (\|J_{\gamma B}R_{\gamma A}x - J_{\gamma B}R_{\gamma A}y\|^2 - \frac{\omega}{\mu} \|J_{\gamma A}x - J_{\gamma A}y\|^2) \end{aligned}$$

$$> -\frac{1}{2} \|(\text{Id} - T)x - (\text{Id} - T)y\|^2.$$

-
- A is maximally $(-\omega)$ -monotone, B is maximally μ -monotone, $\mu > \omega \geq 0$.
 - $\text{Id} - T = J_{\gamma A} - J_{\gamma B}R_{\gamma A}$

Proof continued

Our goal is show that T is averaged, i.e., in view of the earlier characterization of averaged operators:

$$\begin{aligned} & \langle Tx - Ty \mid (\text{Id} - T)x - (\text{Id} - T)y \rangle \\ &= \langle J_{\gamma A}x - J_{\gamma A}y \mid (\text{Id} - J_{\gamma A})x - (\text{Id} - J_{\gamma A})y \rangle \\ & \quad + \langle J_{\gamma B}R_{\gamma A}x - J_{\gamma B}R_{\gamma A}y \mid (\text{Id} - J_{\gamma B})R_{\gamma A}x - (\text{Id} - J_{\gamma B})R_{\gamma A}y \rangle \\ &\geq \gamma\mu \|J_{\gamma B}R_{\gamma A}x - J_{\gamma B}R_{\gamma A}y\|^2 - \gamma\omega \|J_{\gamma A}x - J_{\gamma A}y\|^2 \\ &= \gamma\mu (\|J_{\gamma B}R_{\gamma A}x - J_{\gamma B}R_{\gamma A}y\|^2 - \frac{\omega}{\mu} \|J_{\gamma A}x - J_{\gamma A}y\|^2) \\ &\geq -\frac{\gamma\mu(\omega/\mu)}{1-(\omega/\mu)} \|J_{\gamma A}x - J_{\gamma A}y - (J_{\gamma B}R_{\gamma A}x - J_{\gamma B}R_{\gamma A}y)\|^2 \\ &> -\frac{1}{2} \|(\text{Id} - T)x - (\text{Id} - T)y\|^2. \end{aligned}$$

-
- A is maximally $(-\omega)$ -monotone, B is maximally μ -monotone, $\mu > \omega \geq 0$.
 - $\text{Id} - T = J_{\gamma A} - J_{\gamma B}R_{\gamma A}$

Proof continued

Our goal is show that T is averaged, i.e., in view of the earlier characterization of averaged operators:

$$\begin{aligned}
 & \langle Tx - Ty \mid (\text{Id} - T)x - (\text{Id} - T)y \rangle \\
 &= \langle J_{\gamma A}x - J_{\gamma A}y \mid (\text{Id} - J_{\gamma A})x - (\text{Id} - J_{\gamma A})y \rangle \\
 & \quad + \langle J_{\gamma B}R_{\gamma A}x - J_{\gamma B}R_{\gamma A}y \mid (\text{Id} - J_{\gamma B})R_{\gamma A}x - (\text{Id} - J_{\gamma B})R_{\gamma A}y \rangle \\
 &\geq \gamma\mu \|J_{\gamma B}R_{\gamma A}x - J_{\gamma B}R_{\gamma A}y\|^2 - \gamma\omega \|J_{\gamma A}x - J_{\gamma A}y\|^2 \\
 &= \gamma\mu (\|J_{\gamma B}R_{\gamma A}x - J_{\gamma B}R_{\gamma A}y\|^2 - \frac{\omega}{\mu} \|J_{\gamma A}x - J_{\gamma A}y\|^2) \\
 &\geq -\frac{\gamma\mu(\omega/\mu)}{1-(\omega/\mu)} \|J_{\gamma A}x - J_{\gamma A}y - (J_{\gamma B}R_{\gamma A}x - J_{\gamma B}R_{\gamma A}y)\|^2 \\
 &= -\frac{\gamma\mu\omega}{\mu-\omega} \|J_{\gamma A}x - J_{\gamma B}R_{\gamma A}x - (J_{\gamma A}y - J_{\gamma B}R_{\gamma A}y)\|^2 \\
 &> -\frac{1}{2} \|(\text{Id} - T)x - (\text{Id} - T)y\|^2.
 \end{aligned}$$

-
- A is maximally $(-\omega)$ -monotone, B is maximally μ -monotone, $\mu > \omega \geq 0$.
 - $\text{Id} - T = J_{\gamma A} - J_{\gamma B}R_{\gamma A}$

Proof continued

Our goal is show that T is averaged, i.e., in view of the earlier characterization of averaged operators:

$$\begin{aligned}
 & \langle Tx - Ty \mid (\text{Id} - T)x - (\text{Id} - T)y \rangle \\
 &= \langle J_{\gamma A}x - J_{\gamma A}y \mid (\text{Id} - J_{\gamma A})x - (\text{Id} - J_{\gamma A})y \rangle \\
 & \quad + \langle J_{\gamma B}R_{\gamma A}x - J_{\gamma B}R_{\gamma A}y \mid (\text{Id} - J_{\gamma B})R_{\gamma A}x - (\text{Id} - J_{\gamma B})R_{\gamma A}y \rangle \\
 &\geq \gamma\mu \|J_{\gamma B}R_{\gamma A}x - J_{\gamma B}R_{\gamma A}y\|^2 - \gamma\omega \|J_{\gamma A}x - J_{\gamma A}y\|^2 \\
 &= \gamma\mu (\|J_{\gamma B}R_{\gamma A}x - J_{\gamma B}R_{\gamma A}y\|^2 - \frac{\omega}{\mu} \|J_{\gamma A}x - J_{\gamma A}y\|^2) \\
 &\geq -\frac{\gamma\mu(\omega/\mu)}{1-(\omega/\mu)} \|J_{\gamma A}x - J_{\gamma A}y - (J_{\gamma B}R_{\gamma A}x - J_{\gamma B}R_{\gamma A}y)\|^2 \\
 &= -\frac{\gamma\mu\omega}{\mu-\omega} \|J_{\gamma A}x - J_{\gamma B}R_{\gamma A}x - (J_{\gamma A}y - J_{\gamma B}R_{\gamma A}y)\|^2 \\
 &= -\frac{\gamma\mu\omega}{\mu-\omega} \|(\text{Id} - T)x - (\text{Id} - T)y\|^2 \\
 &> -\frac{1}{2} \|(\text{Id} - T)x - (\text{Id} - T)y\|^2.
 \end{aligned}$$

-
- A is maximally $(-\omega)$ -monotone, B is maximally μ -monotone, $\mu > \omega \geq 0$.
 - $\text{Id} - T = J_{\gamma A} - J_{\gamma B}R_{\gamma A}$



Averagedness of the relaxed Douglas–Rachford operator

Theorem

Let $\mu > \omega \geq 0$, let $\lambda \in]0, 1[$ and let $\gamma \in]0, (1 - \lambda)(\mu - \omega) / (\mu\omega)[$.
Suppose that one of the following holds:

- (i) A is maximally $(-\omega)$ -monotone and B is maximally μ -monotone.
- (ii) A is maximally μ -monotone and B is maximally $(-\omega)$ -monotone.

Set

$$T = (1 - \lambda) \text{Id} + \lambda R_{\gamma B} R_{\gamma A}, \quad \text{and} \quad \alpha = \frac{\lambda(\mu - \omega)}{\mu - \omega - \gamma\mu\omega}.$$

Averagedness of the relaxed Douglas–Rachford operator

Theorem

Let $\mu > \omega \geq 0$, let $\lambda \in]0, 1[$ and let $\gamma \in]0, (1 - \lambda)(\mu - \omega) / (\mu\omega)[$.
Suppose that one of the following holds:

- (i) A is maximally $(-\omega)$ -monotone and B is maximally μ -monotone.
- (ii) A is maximally μ -monotone and B is maximally $(-\omega)$ -monotone.

Set

$$T = (1 - \lambda) \text{Id} + \lambda R_{\gamma B} R_{\gamma A}, \quad \text{and} \quad \alpha = \frac{\lambda(\mu - \omega)}{\mu - \omega - \gamma\mu\omega}.$$

Then $\alpha \in]0, 1[$ and T is α -averaged.

Averagedness of the relaxed Douglas–Rachford operator

Theorem

Let $\mu > \omega \geq 0$, let $\lambda \in]0, 1[$ and let $\gamma \in]0, (1 - \lambda)(\mu - \omega) / (\mu\omega)[$.
Suppose that one of the following holds:

- (i) A is maximally $(-\omega)$ -monotone and B is maximally μ -monotone.
- (ii) A is maximally μ -monotone and B is maximally $(-\omega)$ -monotone.

Set

$$T = (1 - \lambda) \text{Id} + \lambda R_{\gamma B} R_{\gamma A}, \quad \text{and} \quad \alpha = \frac{\lambda(\mu - \omega)}{\mu - \omega - \gamma\mu\omega}.$$

Then $\alpha \in]0, 1[$ and T is α -averaged.

Proof.

- $J_{\gamma A}$ is single-valued and full domain. (easy)

Averagedness of the relaxed Douglas–Rachford operator

Theorem

Let $\mu > \omega \geq 0$, let $\lambda \in]0, 1[$ and let $\gamma \in]0, (1 - \lambda)(\mu - \omega) / (\mu\omega)[$.
Suppose that one of the following holds:

- (i) A is maximally $(-\omega)$ -monotone and B is maximally μ -monotone.
- (ii) A is maximally μ -monotone and B is maximally $(-\omega)$ -monotone.

Set

$$T = (1 - \lambda) \text{Id} + \lambda R_{\gamma B} R_{\gamma A}, \quad \text{and} \quad \alpha = \frac{\lambda(\mu - \omega)}{\mu - \omega - \gamma\mu\omega}.$$

Then $\alpha \in]0, 1[$ and T is α -averaged.

Proof.

- ▶ $J_{\gamma A}$ is single-valued and full domain. (easy)
- ▶ $R_{\gamma B} R_{\gamma A} = (1 - 2\delta) \text{Id} + 2\delta N$, N is nonexpansive, $\delta = \frac{\mu - \omega}{2(\mu - \omega - \gamma\mu\omega)}$.
(proof omitted, it uses the previous result)

Averagedness of the relaxed Douglas–Rachford operator

Theorem

Let $\mu > \omega \geq 0$, let $\lambda \in]0, 1[$ and let $\gamma \in]0, (1 - \lambda)(\mu - \omega) / (\mu\omega)[$.
Suppose that one of the following holds:

- (i) A is maximally $(-\omega)$ -monotone and B is maximally μ -monotone.
- (ii) A is maximally μ -monotone and B is maximally $(-\omega)$ -monotone.

Set

$$T = (1 - \lambda) \text{Id} + \lambda R_{\gamma B} R_{\gamma A}, \quad \text{and} \quad \alpha = \frac{\lambda(\mu - \omega)}{\mu - \omega - \gamma\mu\omega}.$$

Then $\alpha \in]0, 1[$ and T is α -averaged.

Proof.

- ▶ $J_{\gamma A}$ is single-valued and full domain. (easy)
- ▶ $R_{\gamma B} R_{\gamma A} = (1 - 2\delta) \text{Id} + 2\delta N$, N is nonexpansive, $\delta = \frac{\mu - \omega}{2(\mu - \omega - \gamma\mu\omega)}$.
(proof omitted, it uses the previous result)
- ▶ Altogether

$$\begin{aligned} T &= (1 - \lambda) \text{Id} + \lambda R_{\gamma B} R_{\gamma A} = (1 - \delta) \text{Id} + \lambda((1 - 2\delta) \text{Id} + 2\delta N) \\ &= (1 - 2\lambda\delta) \text{Id} + 2\lambda\delta N = (1 - \alpha) \text{Id} + \alpha N. \end{aligned}$$



Averagedness of the relaxed Douglas–Rachford operator

Theorem

Let $\mu > \omega \geq 0$, let $\lambda \in]0, 1[$ and let $\gamma \in]0, (1 - \lambda)(\mu - \omega) / (\mu\omega)[$.
Suppose that one of the following holds:

- (i) A is maximally $(-\omega)$ -monotone and B is maximally μ -monotone.
- (ii) A is maximally μ -monotone and B is maximally $(-\omega)$ -monotone.

Set

$$T = (1 - \lambda) \text{Id} + \lambda R_{\gamma B} R_{\gamma A}, \quad \text{and} \quad \alpha = \frac{\lambda(\mu - \omega)}{\mu - \omega - \gamma\mu\omega}.$$

Then $\alpha \in]0, 1[$ and T is α -averaged.

Proof.

- ▶ $J_{\gamma A}$ is single-valued and full domain. (easy)
- ▶ $R_{\gamma B} R_{\gamma A} = (1 - 2\delta) \text{Id} + 2\delta N$, N is nonexpansive, $\delta = \frac{\mu - \omega}{2(\mu - \omega - \gamma\mu\omega)}$.
(proof omitted, it uses the previous result)
- ▶ Altogether

$$\begin{aligned} T &= (1 - \lambda) \text{Id} + \lambda R_{\gamma B} R_{\gamma A} = (1 - \delta) \text{Id} + \lambda((1 - 2\delta) \text{Id} + 2\delta N) \\ &= (1 - 2\lambda\delta) \text{Id} + 2\lambda\delta N = (1 - \alpha) \text{Id} + \alpha N. \end{aligned}$$



Convergence of the shadow sequence

Theorem

Let $\mu > \omega \geq 0$, let $\lambda \in]0, 1[$ and let $\gamma \in]0, (1 - \lambda)(\mu - \omega) / (\mu\omega)[$.
Suppose that one of the following holds:

- (i) A is maximally μ -monotone and B is maximally $(-\omega)$ -monotone.
- (ii) A is maximally $(-\omega)$ -monotone and B is maximally μ -monotone.

Set

$$T = (1 - \lambda) \text{Id} + \lambda R_{\gamma B} R_{\gamma A},$$

and let $x_0 \in X$.

Convergence of the shadow sequence

Theorem

Let $\mu > \omega \geq 0$, let $\lambda \in]0, 1[$ and let $\gamma \in]0, (1 - \lambda)(\mu - \omega) / (\mu\omega)[$.
Suppose that one of the following holds:

- (i) A is maximally μ -monotone and B is maximally $(-\omega)$ -monotone.
- (ii) A is maximally $(-\omega)$ -monotone and B is maximally μ -monotone.

Set

$$T = (1 - \lambda) \text{Id} + \lambda R_{\gamma B} R_{\gamma A},$$

and let $x_0 \in X$. Then $\text{zer}(A + B) \neq \emptyset$.

Convergence of the shadow sequence

Theorem

Let $\mu > \omega \geq 0$, let $\lambda \in]0, 1[$ and let $\gamma \in]0, (1 - \lambda)(\mu - \omega) / (\mu\omega)[$.
Suppose that one of the following holds:

- (i) A is maximally μ -monotone and B is maximally $(-\omega)$ -monotone.
- (ii) A is maximally $(-\omega)$ -monotone and B is maximally μ -monotone.

Set

$$T = (1 - \lambda) \text{Id} + \lambda R_{\gamma B} R_{\gamma A},$$

and let $x_0 \in X$. Then $\text{zer}(A + B) \neq \emptyset$. Moreover, there exists $\bar{x} \in \text{Fix } T = \text{Fix } R_{\gamma B} R_{\gamma A}$, $\text{zer}(A + B) = \{J_{\gamma A} \bar{x}\} = \{J_{\gamma B} R_{\gamma A} \bar{x}\}$,

Convergence of the shadow sequence

Theorem

Let $\mu > \omega \geq 0$, let $\lambda \in]0, 1[$ and let $\gamma \in]0, (1 - \lambda)(\mu - \omega) / (\mu\omega)[$.
Suppose that one of the following holds:

- (i) A is maximally μ -monotone and B is maximally $(-\omega)$ -monotone.
- (ii) A is maximally $(-\omega)$ -monotone and B is maximally μ -monotone.

Set

$$T = (1 - \lambda) \text{Id} + \lambda R_{\gamma B} R_{\gamma A},$$

and let $x_0 \in X$. Then $\text{zer}(A + B) \neq \emptyset$. Moreover, there exists $\bar{x} \in \text{Fix } T = \text{Fix } R_{\gamma B} R_{\gamma A}$, $\text{zer}(A + B) = \{J_{\gamma A} \bar{x}\} = \{J_{\gamma B} R_{\gamma A} \bar{x}\}$, $T^n x_0 \rightarrow \bar{x}$,

Convergence of the shadow sequence

Theorem

Let $\mu > \omega \geq 0$, let $\lambda \in]0, 1[$ and let $\gamma \in]0, (1 - \lambda)(\mu - \omega) / (\mu\omega)[$. Suppose that one of the following holds:

- (i) A is maximally μ -monotone and B is maximally $(-\omega)$ -monotone.
- (ii) A is maximally $(-\omega)$ -monotone and B is maximally μ -monotone.

Set

$$T = (1 - \lambda) \text{Id} + \lambda R_{\gamma B} R_{\gamma A},$$

and let $x_0 \in X$. Then $\text{zer}(A + B) \neq \emptyset$. Moreover, there exists $\bar{x} \in \text{Fix } T = \text{Fix } R_{\gamma B} R_{\gamma A}$, $\text{zer}(A + B) = \{J_{\gamma A} \bar{x}\} = \{J_{\gamma B} R_{\gamma A} \bar{x}\}$, $T^n x_0 \rightarrow \bar{x}$, $J_{\gamma A} T^n x_0 \rightarrow J_{\gamma A} \bar{x}$,

Convergence of the shadow sequence

Theorem

Let $\mu > \omega \geq 0$, let $\lambda \in]0, 1[$ and let $\gamma \in]0, (1 - \lambda)(\mu - \omega) / (\mu\omega)[$. Suppose that one of the following holds:

- (i) A is maximally μ -monotone and B is maximally $(-\omega)$ -monotone.
- (ii) A is maximally $(-\omega)$ -monotone and B is maximally μ -monotone.

Set

$$T = (1 - \lambda) \text{Id} + \lambda R_{\gamma B} R_{\gamma A},$$

and let $x_0 \in X$. Then $\text{zer}(A + B) \neq \emptyset$. Moreover, there exists $\bar{x} \in \text{Fix } T = \text{Fix } R_{\gamma B} R_{\gamma A}$, $\text{zer}(A + B) = \{J_{\gamma A} \bar{x}\} = \{J_{\gamma B} R_{\gamma A} \bar{x}\}$, $T^n x_0 \rightarrow \bar{x}$, $J_{\gamma A} T^n x_0 \rightarrow J_{\gamma A} \bar{x}$, and $J_{\gamma B} R_{\gamma A} T^n x_0 \rightarrow J_{\gamma B} R_{\gamma A} \bar{x}$.

Convergence of the shadow sequence: proof continued

► Claim 1:

$$\|J_{\gamma A} T^n x_0 - J_{\gamma A} \bar{x}\|^2 - \|J_{\gamma B} R_{\gamma A} T^n x_0 - J_{\gamma B} R_{\gamma A} \bar{x}\|^2$$

$\rightarrow 0.$

$$\bullet \bar{x} - T\bar{x} = 0 \Rightarrow J_{\gamma A} \bar{x} = J_{\gamma B} R_{\gamma B} \bar{x}$$

Convergence of the shadow sequence: proof continued

► Claim 1:

$$\|J_{\gamma A} T^n x_0 - J_{\gamma A} \bar{x}\|^2 - \|J_{\gamma B} R_{\gamma A} T^n x_0 - J_{\gamma B} R_{\gamma A} \bar{x}\|^2 \rightarrow 0.$$

► Claim 2:

$$\|J_{\gamma A} T^n x_0 - J_{\gamma A} \bar{x}\|^2 - \frac{\omega}{\mu} \|J_{\gamma B} R_{\gamma A} T^n x_0 - J_{\gamma B} R_{\gamma A} \bar{x}\|^2 \rightarrow 0.$$

$$\bullet \bar{x} - T\bar{x} = 0 \Rightarrow J_{\gamma A} \bar{x} = J_{\gamma B} R_{\gamma B} \bar{x}$$

Convergence of the shadow sequence: proof continued

► Claim 1:

$$\begin{aligned} & \|J_{\gamma A} T^n x_0 - J_{\gamma A} \bar{x}\|^2 - \|J_{\gamma B} R_{\gamma A} T^n x_0 - J_{\gamma B} R_{\gamma A} \bar{x}\|^2 \\ &= \langle J_{\gamma A} T^n x_0 - J_{\gamma B} R_{\gamma A} T^n x_0 \mid J_{\gamma A} T^n x_0 + J_{\gamma B} R_{\gamma A} T^n x_0 - J_{\gamma A} \bar{x} - J_{\gamma B} R_{\gamma A} \bar{x} \rangle \\ &\quad \rightarrow 0. \end{aligned}$$

► Claim 2:

$$\begin{aligned} & \|J_{\gamma A} T^n x_0 - J_{\gamma A} \bar{x}\|^2 - \frac{\omega}{\mu} \|J_{\gamma B} R_{\gamma A} T^n x_0 - J_{\gamma B} R_{\gamma A} \bar{x}\|^2 \\ &\quad \rightarrow 0. \end{aligned}$$

$$\bullet \bar{x} - T\bar{x} = 0 \Rightarrow J_{\gamma A} \bar{x} = J_{\gamma B} R_{\gamma B} \bar{x}$$

Convergence of the shadow sequence: proof continued

► Claim 1:

$$\begin{aligned} & \|J_{\gamma A} T^n x_0 - J_{\gamma A} \bar{x}\|^2 - \|J_{\gamma B} R_{\gamma A} T^n x_0 - J_{\gamma B} R_{\gamma A} \bar{x}\|^2 \\ &= \langle J_{\gamma A} T^n x_0 - J_{\gamma B} R_{\gamma A} T^n x_0 \mid J_{\gamma A} T^n x_0 + J_{\gamma B} R_{\gamma A} T^n x_0 - J_{\gamma A} \bar{x} - J_{\gamma B} R_{\gamma A} \bar{x} \rangle \\ &= \langle \underbrace{T^n x_0 - T^{n+1} x_0}_{\rightarrow 0} \mid \underbrace{J_{\gamma A} T^n x_0 + J_{\gamma B} R_{\gamma A} T^n x_0 - J_{\gamma A} \bar{x} - J_{\gamma B} R_{\gamma A} \bar{x}}_{\text{bounded}} \rangle \rightarrow 0. \end{aligned}$$

► Claim 2:

$$\|J_{\gamma A} T^n x_0 - J_{\gamma A} \bar{x}\|^2 - \frac{\omega}{\mu} \|J_{\gamma B} R_{\gamma A} T^n x_0 - J_{\gamma B} R_{\gamma A} \bar{x}\|^2 \rightarrow 0.$$

$$\bullet \bar{x} - T\bar{x} = 0 \Rightarrow J_{\gamma A} \bar{x} = J_{\gamma B} R_{\gamma A} \bar{x}$$

Convergence of the shadow sequence: proof continued

► Claim 1:

$$\begin{aligned}
 & \|J_{\gamma A} T^n x_0 - J_{\gamma A} \bar{x}\|^2 - \|J_{\gamma B} R_{\gamma A} T^n x_0 - J_{\gamma B} R_{\gamma A} \bar{x}\|^2 \\
 &= \langle J_{\gamma A} T^n x_0 - J_{\gamma B} R_{\gamma A} T^n x_0 \mid J_{\gamma A} T^n x_0 + J_{\gamma B} R_{\gamma A} T^n x_0 - J_{\gamma A} \bar{x} - J_{\gamma B} R_{\gamma A} \bar{x} \rangle \\
 &= \langle \underbrace{T^n x_0 - T^{n+1} x_0}_{\rightarrow 0} \mid \underbrace{J_{\gamma A} T^n x_0 + J_{\gamma B} R_{\gamma A} T^n x_0 - J_{\gamma A} \bar{x} - J_{\gamma B} R_{\gamma A} \bar{x}}_{\text{bounded}} \rangle \rightarrow 0.
 \end{aligned}$$

► Claim 2:

$$\begin{aligned}
 0 &\leftarrow \langle T^{n+1} x_0 - \bar{x} \mid T^n x_0 - T^{n+1} x_0 \rangle - (1 - 2\lambda) \langle T^n x_0 - \bar{x} \mid T^n x_0 - T^{n+1} x_0 \rangle \\
 &\geq 4\gamma\mu\lambda^2 \left(\|J_{\gamma A} T^n x_0 - J_{\gamma A} \bar{x}\|^2 - \frac{\omega}{\mu} \|J_{\gamma B} R_{\gamma A} T^n x_0 - J_{\gamma B} R_{\gamma A} \bar{x}\|^2 \right) \\
 &\hspace{15em} \rightarrow 0.
 \end{aligned}$$

• $\bar{x} - T\bar{x} = 0 \Rightarrow J_{\gamma A} \bar{x} = J_{\gamma B} R_{\gamma A} \bar{x}$

Convergence of the shadow sequence: proof continued

► Claim 1:

$$\begin{aligned}
 & \|J_{\gamma A} T^n x_0 - J_{\gamma A} \bar{x}\|^2 - \|J_{\gamma B} R_{\gamma A} T^n x_0 - J_{\gamma B} R_{\gamma A} \bar{x}\|^2 \\
 &= \langle J_{\gamma A} T^n x_0 - J_{\gamma B} R_{\gamma A} T^n x_0 \mid J_{\gamma A} T^n x_0 + J_{\gamma B} R_{\gamma A} T^n x_0 - J_{\gamma A} \bar{x} - J_{\gamma B} R_{\gamma A} \bar{x} \rangle \\
 &= \langle \underbrace{T^n x_0 - T^{n+1} x_0}_{\rightarrow 0} \mid \underbrace{J_{\gamma A} T^n x_0 + J_{\gamma B} R_{\gamma A} T^n x_0 - J_{\gamma A} \bar{x} - J_{\gamma B} R_{\gamma A} \bar{x}}_{\text{bounded}} \rangle \rightarrow 0.
 \end{aligned}$$

► Claim 2:

$$\begin{aligned}
 0 &\leftarrow \langle T^{n+1} x_0 - \bar{x} \mid T^n x_0 - T^{n+1} x_0 \rangle - (1 - 2\lambda) \langle T^n x_0 - \bar{x} \mid T^n x_0 - T^{n+1} x_0 \rangle \\
 &\geq 4\gamma\mu\lambda^2 \left(\|J_{\gamma A} T^n x_0 - J_{\gamma A} \bar{x}\|^2 - \frac{\omega}{\mu} \|J_{\gamma B} R_{\gamma A} T^n x_0 - J_{\gamma B} R_{\gamma A} \bar{x}\|^2 \right) \\
 &\geq (1 - 2\lambda) \langle T^n x_0 - \bar{x} \mid T^n x_0 - T^{n+1} x_0 \rangle - \frac{\gamma\mu\omega}{\mu - \omega} \|T^n x_0 - T^{n+1} x_0\|^2 \rightarrow 0.
 \end{aligned}$$

$$\bullet \bar{x} - T\bar{x} = 0 \Rightarrow J_{\gamma A} \bar{x} = J_{\gamma B} R_{\gamma A} \bar{x}$$

Application to the optimization problems

Theorem

Let $\mu > \omega \geq 0$, let $\lambda \in]0, 1[$ and let $\gamma \in]0, (1 - \lambda)(\mu - \omega) / (\mu\omega)[$.
Suppose that one of the following holds:

- (i) f is μ -strongly convex, g is ω -hypoconvex.
 - (ii) f is ω -hypoconvex, and g is μ -strongly convex.
- and that $\text{zer}(\partial f + \partial_{\#} g) \neq \emptyset$ (require sufficient conditions).

Application to the optimization problems

Theorem

Let $\mu > \omega \geq 0$, let $\lambda \in]0, 1[$ and let $\gamma \in]0, (1 - \lambda)(\mu - \omega) / (\mu\omega)[$.
Suppose that one of the following holds:

- (i) f is μ -strongly convex, g is ω -hypoconvex.
- (ii) f is ω -hypoconvex, and g is μ -strongly convex.

and that $\text{zer}(\partial f + \partial_{\#}g) \neq \emptyset$ (require sufficient conditions). Set

$$T = (1 - \lambda) \text{Id} + \lambda(2 \text{Prox}_{\gamma g} - \text{Id})(2 \text{Prox}_{\gamma f} - \text{Id}), \quad \text{and} \quad \alpha = \frac{\lambda(\mu - \omega)}{\mu - \omega - \gamma\mu\omega},$$

and let $x_0 \in X$.

Application to the optimization problems

Theorem

Let $\mu > \omega \geq 0$, let $\lambda \in]0, 1[$ and let $\gamma \in]0, (1 - \lambda)(\mu - \omega) / (\mu\omega)[$.
Suppose that one of the following holds:

- (i) f is μ -strongly convex, g is ω -hypoconvex.
- (ii) f is ω -hypoconvex, and g is μ -strongly convex.

and that $\text{zer}(\partial f + \partial_{\#} g) \neq \emptyset$ (require sufficient conditions). Set

$$T = (1 - \lambda) \text{Id} + \lambda(2 \text{Prox}_{\gamma g} - \text{Id})(2 \text{Prox}_{\gamma f} - \text{Id}), \quad \text{and} \quad \alpha = \frac{\lambda(\mu - \omega)}{\mu - \omega - \gamma\mu\omega},$$

and let $x_0 \in X$. Then $\alpha \in]0, 1[$, and T is α -averaged.

Application to the optimization problems

Theorem

Let $\mu > \omega \geq 0$, let $\lambda \in]0, 1[$ and let $\gamma \in]0, (1 - \lambda)(\mu - \omega)/(\mu\omega)[$.
Suppose that one of the following holds:

- (i) f is μ -strongly convex, g is ω -hypoconvex.
- (ii) f is ω -hypoconvex, and g is μ -strongly convex.

and that $\text{zer}(\partial f + \partial_{\#}g) \neq \emptyset$ (require sufficient conditions). Set

$$T = (1 - \lambda) \text{Id} + \lambda(2 \text{Prox}_{\gamma g} - \text{Id})(2 \text{Prox}_{\gamma f} - \text{Id}), \quad \text{and} \quad \alpha = \frac{\lambda(\mu - \omega)}{\mu - \omega - \gamma\mu\omega},$$

and let $x_0 \in X$. Then $\alpha \in]0, 1[$, and T is α -averaged. Moreover,
($\exists \bar{x} \in \text{Fix } T$) such that $T^n x_0 \rightarrow \bar{x}$, $\text{argmin}(f + g) = \{\text{Prox}_f \bar{x}\}$, and
 $\text{Prox}_f T^n x_0 \rightarrow \text{Prox}_f \bar{x}$.

To conclude this part:

► **Question:**

What does the algorithm do when $A + B$ is just monotone?

- Recently, Dao and Phan introduced an adaptation of the Douglas–Rachford algorithm to deal with the case when $A + B$ is monotone. They show that the governing sequence converges to a fixed point of T . However, the behaviour of the shadow sequence is not clear. There is a room for progress in this direction.

PART III: On the averagedness of the forward-backward operator

The forward-backward operator

Let $\beta > 0$. Suppose that

$A: X \rightarrow X$ is β -cocoercive,

and that

$B: X \rightrightarrows X$ is maximally monotone.

The problem: Find $x \in X$ such that

$$x \in \text{zer}(\textcolor{green}{A} + \textcolor{brown}{B}) = (A + B)^{-1}(0).$$

- $J_{\gamma} \textcolor{green}{A} = (\text{Id} + \gamma A)^{-1}$

The forward-backward operator

Let $\beta > 0$. Suppose that

$A: X \rightarrow X$ is β -cocoercive,

and that

$B: X \rightrightarrows X$ is maximally monotone.

The problem: Find $x \in X$ such that

$$x \in \text{zer}(A + B) = (A + B)^{-1}(0).$$

The forward-backward algorithm: One successful technique to find a zero of $A + B$ is via iterating the forward-backward operator T defined by

$$T = J_{\gamma B}(\text{Id} - \gamma A),$$

$$\gamma \in]0, 2\beta[.$$

- $J_{\gamma A} = (\text{Id} + \gamma A)^{-1}$

Classical convergence results

Let $x_0 \in X$. Recall that $T = A + B$.

Classical convergence results

Let $x_0 \in X$. Recall that $T = A + B$. When

$$\text{zer}(A + B) \neq \emptyset$$

we have, for A cocoercive and B maximally monotone:



T is averaged.

Classical convergence results

Let $x_0 \in X$. Recall that $T = A + B$. When

$$\text{zer}(A + B) \neq \emptyset$$

we have, for A cocoercive and B maximally monotone:



T is averaged.

► Krasnosel'skiĭ–Mann (1950s)

$$x_n = T^n x_0 \xrightarrow[\text{weakly}]{} \text{some point } \bar{x} \in \text{Fix } T = \text{zer}(A + B).$$

In the absence of monotonicity

- ▶ Suppose for instance that A is **not monotone**. Then J_A (and, in turn, R_A and T) is **not** necessarily single-valued and/or does **not** necessarily has full domain. ☹

In the absence of monotonicity

- ▶ Suppose for instance that A is **not monotone**. Then J_A (and, in turn, R_A and T) is **not** necessarily single-valued and/or does **not** necessarily has full domain. ☹
- ▶ **BUT**, if A is Lipschitz continuous + “nice” and $A + B$ is monotone, then there is more to say. ☺

Let $\mu \geq 0$, $\omega \geq 0$, and $\beta > 0$. We prove that the forward-backward operator is averaged, in each of the following situations:

- (i) A is maximally μ -monotone, $A - \mu \text{Id}$ is β -cocoercive, B is maximally $(-\omega)$ -monotone, and $\mu \geq \omega$.
- (ii) A is maximally $(-\omega)$ -monotone, $A + \omega \text{Id}$ is β -cocoercive, B is maximally μ -monotone, and $\mu \geq \omega$.
- (iii) A is $(1/\beta)$ -Lipschitz continuous, B is maximally μ -monotone, and $\mu \geq 1/\beta$.

Theorem

Let $\mu \geq \omega \geq 0$, and let $\beta > 0$.

Theorem

Let $\mu \geq \omega \geq 0$, and let $\beta > 0$. Suppose that

- ▶ A is *maximally μ -monotone*, $A - \mu \text{Id}$ is *β -cocoercive*,

Theorem

Let $\mu \geq \omega \geq 0$, and let $\beta > 0$. Suppose that

- ▶ A is *maximally μ -monotone*, $A - \mu \text{Id}$ is *β -cocoercive*,
- ▶ B is *maximally $(-\omega)$ -monotone*
- ▶ such that $\text{zer}(A + B) \neq \emptyset$.

Theorem

Let $\mu \geq \omega \geq 0$, and let $\beta > 0$. Suppose that

- ▶ A is *maximally μ -monotone*, $A - \mu \text{Id}$ is β -cocoercive,
- ▶ B is *maximally $(-\omega)$ -monotone*
- ▶ such that $\text{zer}(A + B) \neq \emptyset$.

Let $\gamma \in]0, \min\{1/\mu, 2\beta/(1 + 2\mu\beta)\}[$.

Theorem

Let $\mu \geq \omega \geq 0$, and let $\beta > 0$. Suppose that

- ▶ A is *maximally μ -monotone*, $A - \mu \text{Id}$ is β -cocoercive,
- ▶ B is *maximally $(-\omega)$ -monotone*
- ▶ such that $\text{zer}(A + B) \neq \emptyset$.

Let $\gamma \in]0, \min\{1/\mu, 2\beta/(1 + 2\mu\beta)\}[$. Set

$$T = J_{\gamma B}(\text{Id} - \gamma A),$$

Theorem

Let $\mu \geq \omega \geq 0$, and let $\beta > 0$. Suppose that

- ▶ A is *maximally μ -monotone*, $A - \mu \text{Id}$ is β -cocoercive,
- ▶ B is *maximally $(-\omega)$ -monotone*
- ▶ such that $\text{zer}(A + B) \neq \emptyset$.

Let $\gamma \in]0, \min\{1/\mu, 2\beta/(1 + 2\mu\beta)\}[$. Set

$$T = J_{\gamma B}(\text{Id} - \gamma A),$$

set $\nu = \gamma/(2\beta(1 - \gamma\mu))$ and set $\delta = (1 - \gamma\mu)/(1 - \gamma\omega)$.

Theorem

Let $\mu \geq \omega \geq 0$, and let $\beta > 0$. Suppose that

- ▶ A is *maximally μ -monotone*, $A - \mu \text{Id}$ is β -cocoercive,
- ▶ B is *maximally $(-\omega)$ -monotone*
- ▶ such that $\text{zer}(A + B) \neq \emptyset$.

Let $\gamma \in]0, \min\{1/\mu, 2\beta/(1 + 2\mu\beta)\}[$. Set

$$T = J_{\gamma B}(\text{Id} - \gamma A),$$

set $\nu = \gamma/(2\beta(1 - \gamma\mu))$ and set $\delta = (1 - \gamma\mu)/(1 - \gamma\omega)$. *Then*
 $\{\nu, \delta\} \subseteq]0, 1[$.

Theorem

Let $\mu \geq \omega \geq 0$, and let $\beta > 0$. Suppose that

- ▶ A is *maximally μ -monotone*, $A - \mu \text{Id}$ is *β -cocoercive*,
- ▶ B is *maximally $(-\omega)$ -monotone*
- ▶ such that $\text{zer}(A + B) \neq \emptyset$.

Let $\gamma \in]0, \min\{1/\mu, 2\beta/(1 + 2\mu\beta)\}[$. Set

$$T = J_{\gamma B}(\text{Id} - \gamma A),$$

set $\nu = \gamma/(2\beta(1 - \gamma\mu))$ and set $\delta = (1 - \gamma\mu)/(1 - \gamma\omega)$. *Then*
 $\{\nu, \delta\} \subseteq]0, 1[$. Moreover, the following hold:

- (i) *T is $(1 - \delta(1 - 1/(2 - \nu)))$ -averaged.*
- (ii) *There exists $\bar{x} \in \text{Fix } T = \text{zer}(A + B)$ such that $T^n x_0 \rightharpoonup \bar{x}$.*

Theorem

Let $\mu \geq \omega \geq 0$, and let $\beta > 0$. Suppose that

- ▶ A is *maximally μ -monotone*, $A - \mu \text{Id}$ is *β -cocoercive*,
- ▶ B is *maximally $(-\omega)$ -monotone*
- ▶ such that $\text{zer}(A + B) \neq \emptyset$.

Let $\gamma \in]0, \min\{1/\mu, 2\beta/(1 + 2\mu\beta)\}[$. Set

$$T = J_{\gamma B}(\text{Id} - \gamma A),$$

set $\nu = \gamma/(2\beta(1 - \gamma\mu))$ and set $\delta = (1 - \gamma\mu)/(1 - \gamma\omega)$. *Then*
 $\{\nu, \delta\} \subseteq]0, 1[$. Moreover, the following hold:

- (i) *T is $(1 - \delta(1 - 1/(2 - \nu)))$ -averaged.*
- (ii) *There exists $\bar{x} \in \text{Fix } T = \text{zer}(A + B)$ such that $T^n x_0 \rightarrow \bar{x}$.*

Suppose that $\mu > \omega$. Then we additionally have:

- (iii) *T is Banach contraction with constant $\delta = (1 - \gamma\mu)/(1 - \gamma\omega) < 1$.*
- (iv) *$\text{zer}(A + B) = \{\bar{x}\}$ and $T^n x_0 \rightarrow \bar{x}$ with a linear rate $\delta < 1$.*

Sketch of the proof

There exist a **nonexpansive mapping** N ,
such that

$$A - \mu \text{Id} = \frac{1}{2\beta} \text{Id} + \frac{1}{2\beta} N,$$

Hence, T is averaged as claimed.



Theorem Let $\mu \geq \omega \geq 0$, and let $\beta > 0$. Suppose that

- ▶ A is **maximally μ -monotone**, $A - \mu \text{Id}$ is **β -cocoercive**,
- ▶ B is **maximally $(-\omega)$ -monotone**
- ▶ such that $\text{zer}(A + B) \neq \emptyset$.

Set $\nu = \gamma/(2\beta(1 - \gamma\mu))$ and set $\delta = (1 - \gamma\mu)/(1 - \gamma\omega)$.

Then T is $(1 - \delta(1 - 1/(2 - \nu)))$ -averaged.

Sketch of the proof

There exist a **nonexpansive mapping** N , and a **firmly nonexpansive mapping** $T_1: X \rightarrow X$, such that

$$A - \mu \text{Id} = \frac{1}{2\beta} \text{Id} + \frac{1}{2\beta} N, \quad \text{and} \quad J_{\gamma B} = \frac{1}{1-\gamma\omega} T_1.$$

Hence, T is averaged as claimed. □

Theorem Let $\mu \geq \omega \geq 0$, and let $\beta > 0$. Suppose that

- ▶ A is **maximally μ -monotone**, $A - \mu \text{Id}$ is **β -cocoercive**,
- ▶ B is **maximally $(-\omega)$ -monotone**
- ▶ such that $\text{zer}(A + B) \neq \emptyset$.

Set $\nu = \gamma/(2\beta(1 - \gamma\mu))$ and set $\delta = (1 - \gamma\mu)/(1 - \gamma\omega)$.

Then T is $(1 - \delta(1 - 1/(2 - \nu)))$ -averaged.

Sketch of the proof

There exist a **nonexpansive mapping** N , and a **firmly nonexpansive mapping** $T_1: X \rightarrow X$, such that

$$A - \mu \text{Id} = \frac{1}{2\beta} \text{Id} + \frac{1}{2\beta} N, \quad \text{and} \quad J_{\gamma B} = \frac{1}{1-\gamma\omega} T_1.$$

T

$$= \delta(T_3 \circ T_2).$$

Hence, T is averaged as claimed.



Theorem Let $\mu \geq \omega \geq 0$, and let $\beta > 0$. Suppose that

- ▶ A is **maximally μ -monotone**, $A - \mu \text{Id}$ is **β -cocoercive**,
- ▶ B is **maximally $(-\omega)$ -monotone**
- ▶ such that $\text{zer}(A + B) \neq \emptyset$.

Set $\nu = \gamma/(2\beta(1 - \gamma\mu))$ and set $\delta = (1 - \gamma\mu)/(1 - \gamma\omega)$.

Then T is $(1 - \delta(1 - 1/(2 - \nu)))$ -averaged.

Sketch of the proof

There exist a **nonexpansive mapping** N , and a **firmly nonexpansive mapping** $T_1: X \rightarrow X$, such that

$$A - \mu \text{Id} = \frac{1}{2\beta} \text{Id} + \frac{1}{2\beta} N, \quad \text{and} \quad J_{\gamma B} = \frac{1}{1-\gamma\omega} T_1.$$

$$T = J_{\gamma B}(\text{Id} - \gamma A) = J_{\gamma B}(\text{Id} - \gamma(A - \mu \text{Id} + \mu \text{Id}))$$

$$= \delta(T_3 \circ T_2).$$

Hence, T is averaged as claimed.



Theorem Let $\mu \geq \omega \geq 0$, and let $\beta > 0$. Suppose that

- ▶ A is **maximally μ -monotone**, $A - \mu \text{Id}$ is **β -cocoercive**,
- ▶ B is **maximally $(-\omega)$ -monotone**
- ▶ such that $\text{zer}(A + B) \neq \emptyset$.

Set $\nu = \gamma/(2\beta(1 - \gamma\mu))$ and set $\delta = (1 - \gamma\mu)/(1 - \gamma\omega)$.

Then T is $(1 - \delta(1 - 1/(2 - \nu)))$ -averaged.

Sketch of the proof

There exist a **nonexpansive mapping** N , and a **firmly nonexpansive mapping** $T_1: X \rightarrow X$, such that

$$A - \mu \text{Id} = \frac{1}{2\beta} \text{Id} + \frac{1}{2\beta} N, \quad \text{and} \quad J_{\gamma B} = \frac{1}{1-\gamma\omega} T_1.$$

$$\begin{aligned} T &= J_{\gamma B}(\text{Id} - \gamma A) = J_{\gamma B}(\text{Id} - \gamma(A - \mu \text{Id} + \mu \text{Id})) \\ &= \frac{1}{1-\gamma\omega} T_1 \circ (1 - \gamma\mu) \text{Id} \circ (\text{Id} - \frac{\gamma}{1-\gamma\mu} (\frac{1}{2\beta} \text{Id} + \frac{1}{2\beta} N)) \end{aligned}$$

$$= \delta(T_3 \circ T_2).$$

Hence, T is averaged as claimed. □

Theorem Let $\mu \geq \omega \geq 0$, and let $\beta > 0$. Suppose that

- ▶ A is **maximally μ -monotone**, $A - \mu \text{Id}$ is **β -cocoercive**,
- ▶ B is **maximally $(-\omega)$ -monotone**
- ▶ such that $\text{zer}(A + B) \neq \emptyset$.

Set $\nu = \gamma/(2\beta(1 - \gamma\mu))$ and set $\delta = (1 - \gamma\mu)/(1 - \gamma\omega)$.

Then T is $(1 - \delta(1 - 1/(2 - \nu)))$ -averaged.

Sketch of the proof

There exist a **nonexpansive mapping** N , and a **firmly nonexpansive mapping** $T_1: X \rightarrow X$, such that

$$\begin{aligned} A - \mu \text{Id} &= \frac{1}{2\beta} \text{Id} + \frac{1}{2\beta} N, \quad \text{and} \quad J_{\gamma B} = \frac{1}{1-\gamma\omega} T_1. \\ T &= J_{\gamma B}(\text{Id} - \gamma A) = J_{\gamma B}(\text{Id} - \gamma(A - \mu \text{Id} + \mu \text{Id})) \\ &= \frac{1}{1-\gamma\omega} T_1 \circ (1 - \gamma\mu) \text{Id} \circ (\text{Id} - \frac{\gamma}{1-\gamma\mu} (\frac{1}{2\beta} \text{Id} + \frac{1}{2\beta} N)) \\ &= \frac{1-\gamma\mu}{1-\gamma\omega} \underbrace{\frac{1}{1-\gamma\mu} T_1 \circ (1 - \gamma\mu) \text{Id}}_{=T_3 \text{ (f.n.e.)}} \circ \underbrace{((1-\nu) \text{Id} + \nu(-N))}_{=T_2 \text{ (averaged)}} \\ &= \delta(T_3 \circ T_2). \end{aligned}$$

Hence, T is averaged as claimed. □

Theorem Let $\mu \geq \omega \geq 0$, and let $\beta > 0$. Suppose that

- ▶ A is **maximally μ -monotone**, $A - \mu \text{Id}$ is **β -cocoercive**,
- ▶ B is **maximally $(-\omega)$ -monotone**
- ▶ such that $\text{zer}(A + B) \neq \emptyset$.

Set $\nu = \gamma/(2\beta(1-\gamma\mu))$ and set $\delta = (1-\gamma\mu)/(1-\gamma\omega)$.

Then T is $(1-\delta(1-1/(2-\nu)))$ -averaged.

Sketch of the proof

There exist a **nonexpansive mapping** N , and a **firmly nonexpansive mapping** $T_1: X \rightarrow X$, such that

$$\begin{aligned} A - \mu \text{Id} &= \frac{1}{2\beta} \text{Id} + \frac{1}{2\beta} N, \quad \text{and} \quad J_{\gamma B} = \frac{1}{1-\gamma\omega} T_1. \\ T &= J_{\gamma B}(\text{Id} - \gamma A) = J_{\gamma B}(\text{Id} - \gamma(A - \mu \text{Id} + \mu \text{Id})) \\ &= \frac{1}{1-\gamma\omega} T_1 \circ (1 - \gamma\mu) \text{Id} \circ \left(\text{Id} - \frac{\gamma}{1-\gamma\mu} \left(\frac{1}{2\beta} \text{Id} + \frac{1}{2\beta} N \right) \right) \\ &= \frac{1-\gamma\mu}{1-\gamma\omega} \underbrace{\frac{1}{1-\gamma\mu} T_1 \circ (1 - \gamma\mu) \text{Id}}_{= T_3 \text{ (f.n.e.)}} \circ \underbrace{\left((1 - \nu) \text{Id} + \nu(-N) \right)}_{= T_2 \text{ (averaged)}} \\ &= \frac{1-\gamma\mu}{1-\gamma\omega} T_3 \circ T_2 = \delta(T_3 \circ T_2). \end{aligned}$$

Hence, T is averaged as claimed. □

Theorem Let $\mu \geq \omega \geq 0$, and let $\beta > 0$. Suppose that

- ▶ A is **maximally μ -monotone**, $A - \mu \text{Id}$ is **β -cocoercive**,
- ▶ B is **maximally $(-\omega)$ -monotone**
- ▶ such that $\text{zer}(A + B) \neq \emptyset$.

Set $\nu = \gamma/(2\beta(1 - \gamma\mu))$ and set $\delta = (1 - \gamma\mu)/(1 - \gamma\omega)$.

Then T is $(1 - \delta(1 - 1/(2 - \nu)))$ -averaged.

Lemma

Let $f_1: X \rightarrow \mathbb{R}$, $f_2: X \rightarrow \mathbb{R}$ be Fréchet differentiable convex functions and let $\delta > \beta > 0$. Suppose that ∇f_1 (respectively ∇f_2) is $\frac{1}{\beta}$ -Lipschitz continuous (respectively $\frac{1}{\delta}$ -Lipschitz continuous). Then the following hold:

- (i) $\nabla f_1 - \nabla f_2$ is $\frac{1}{\beta}$ -Lipschitz continuous.
- (ii) Suppose that $f_1 - f_2$ is convex. Then $\nabla f_1 - \nabla f_2$ is β -cocoercive.

Lemma

Let $f_1: X \rightarrow \mathbb{R}$, $f_2: X \rightarrow \mathbb{R}$ be Fréchet differentiable convex functions and let $\delta > \beta > 0$. Suppose that ∇f_1 (respectively ∇f_2) is $\frac{1}{\beta}$ -Lipschitz continuous (respectively $\frac{1}{\delta}$ -Lipschitz continuous). Then the following hold:

- (i) $\nabla f_1 - \nabla f_2$ is $\frac{1}{\beta}$ -Lipschitz continuous.
- (ii) Suppose that $f_1 - f_2$ is convex. Then $\nabla f_1 - \nabla f_2$ is β -cocoercive.

Lemma

Let $\mu \geq 0$, let $\beta > 0$ and let $f: X \rightarrow \mathbb{R}$ be a Fréchet differentiable function. Suppose that f is μ -strongly convex with a $\frac{1}{\beta}$ -Lipschitz continuous gradient. Then the following hold:

- (i) $f - \frac{\mu}{2} \|\cdot\|^2$ is convex.
- (ii) ∇f is maximally μ -monotone.
- (iii) $\nabla f - \mu \text{Id}$ is β -cocoercive.

Theorem (the forward-backward algorithm when f is μ -strongly convex)

Let $\mu \geq \omega \geq 0$, and let $\beta > 0$. Let f be μ -strongly convex and Fréchet differentiable with a $\frac{1}{\beta}$ -Lipschitz continuous gradient, and let g be ω -hypoconvex. Suppose that $\operatorname{argmin}(f + g) \neq \emptyset$. Let $\gamma \in]0, \min\{1/\mu, 2\beta/(1 + 2\mu\beta)\}[$, and set $\delta = (1 - \gamma\mu)/(1 - \gamma\omega)$. Set

$$T = \operatorname{Prox}_{\gamma g}(\operatorname{Id} - \gamma \nabla f),$$

and let $x_0 \in X$.

Theorem (the forward-backward algorithm when f is μ -strongly convex)

Let $\mu \geq \omega \geq 0$, and let $\beta > 0$. Let f be μ -strongly convex and Fréchet differentiable with a $\frac{1}{\beta}$ -Lipschitz continuous gradient, and let g be ω -hypoconvex. Suppose that $\operatorname{argmin}(f + g) \neq \emptyset$. Let $\gamma \in]0, \min\{1/\mu, 2\beta/(1 + 2\mu\beta)\}[$, and set $\delta = (1 - \gamma\mu)/(1 - \gamma\omega)$. Set

$$T = \operatorname{Prox}_{\gamma g}(\operatorname{Id} - \gamma \nabla f),$$

and let $x_0 \in X$. Then the following hold:

- (i) There exists $\bar{x} \in \operatorname{Fix} T = \operatorname{zer}(A + B) = \operatorname{argmin}(f + g)$ such that $T^n x_0 \rightarrow \bar{x}$.

Theorem (the forward-backward algorithm when f is μ -strongly convex)

Let $\mu \geq \omega \geq 0$, and let $\beta > 0$. Let f be μ -strongly convex and Fréchet differentiable with a $\frac{1}{\beta}$ -Lipschitz continuous gradient, and let g be ω -hypoconvex. Suppose that $\operatorname{argmin}(f + g) \neq \emptyset$. Let $\gamma \in]0, \min\{1/\mu, 2\beta/(1 + 2\mu\beta)\}[$, and set $\delta = (1 - \gamma\mu)/(1 - \gamma\omega)$. Set

$$T = \operatorname{Prox}_{\gamma g}(\operatorname{Id} - \gamma \nabla f),$$

and let $x_0 \in X$. Then the following hold:

- (i) There exists $\bar{x} \in \operatorname{Fix} T = \operatorname{zer}(A + B) = \operatorname{argmin}(f + g)$ such that $T^n x_0 \rightarrow \bar{x}$.

Suppose that $\mu > \omega$. Then, we additionally have:

- (ii) $\operatorname{Fix} T = \operatorname{argmin}(f + g) = \{\bar{x}\}$ and $T^n x_0 \rightarrow \bar{x}$ with a linear rate $\delta < 1$.

Theorem (the forward-backward algorithm when f is μ -strongly convex)

Let $\mu \geq \omega \geq 0$, and let $\beta > 0$. Let f be μ -strongly convex and Fréchet differentiable with a $\frac{1}{\beta}$ -Lipschitz continuous gradient, and let g be ω -hypoconvex. Suppose that $\operatorname{argmin}(f + g) \neq \emptyset$. Let $\gamma \in]0, \min\{1/\mu, 2\beta/(1 + 2\mu\beta)\}[$, and set $\delta = (1 - \gamma\mu)/(1 - \gamma\omega)$. Set

$$T = \operatorname{Prox}_{\gamma g}(\operatorname{Id} - \gamma \nabla f),$$

and let $x_0 \in X$. Then the following hold:

- (i) There exists $\bar{x} \in \operatorname{Fix} T = \operatorname{zer}(A + B) = \operatorname{argmin}(f + g)$ such that $T^n x_0 \rightarrow \bar{x}$.

Suppose that $\mu > \omega$. Then, we additionally have:

- (ii) $\operatorname{Fix} T = \operatorname{argmin}(f + g) = \{\bar{x}\}$ and $T^n x_0 \rightarrow \bar{x}$ with a linear rate $\delta < 1$.

Proof.

∇f is maximally μ -monotone,

Theorem (the forward-backward algorithm when f is μ -strongly convex)

Let $\mu \geq \omega \geq 0$, and let $\beta > 0$. Let f be μ -strongly convex and Fréchet differentiable with a $\frac{1}{\beta}$ -Lipschitz continuous gradient, and let g be ω -hypoconvex. Suppose that $\operatorname{argmin}(f + g) \neq \emptyset$. Let $\gamma \in]0, \min\{1/\mu, 2\beta/(1 + 2\mu\beta)\}[$, and set $\delta = (1 - \gamma\mu)/(1 - \gamma\omega)$. Set

$$T = \operatorname{Prox}_{\gamma g}(\operatorname{Id} - \gamma \nabla f),$$

and let $x_0 \in X$. Then the following hold:

- (i) There exists $\bar{x} \in \operatorname{Fix} T = \operatorname{zer}(A + B) = \operatorname{argmin}(f + g)$ such that $T^n x_0 \rightarrow \bar{x}$.

Suppose that $\mu > \omega$. Then, we additionally have:

- (ii) $\operatorname{Fix} T = \operatorname{argmin}(f + g) = \{\bar{x}\}$ and $T^n x_0 \rightarrow \bar{x}$ with a linear rate $\delta < 1$.

Proof.

∇f is maximally μ -monotone, $\nabla f - \mu \operatorname{Id}$ is $\frac{1}{\beta}$ -Lipschitz continuous by an earlier result,

Theorem (the forward-backward algorithm when f is μ -strongly convex)

Let $\mu \geq \omega \geq 0$, and let $\beta > 0$. Let f be μ -strongly convex and Fréchet differentiable with a $\frac{1}{\beta}$ -Lipschitz continuous gradient, and let g be ω -hypoconvex. Suppose that $\operatorname{argmin}(f + g) \neq \emptyset$. Let $\gamma \in]0, \min\{1/\mu, 2\beta/(1 + 2\mu\beta)\}[$, and set $\delta = (1 - \gamma\mu)/(1 - \gamma\omega)$. Set

$$T = \operatorname{Prox}_{\gamma g}(\operatorname{Id} - \gamma \nabla f),$$

and let $x_0 \in X$. Then the following hold:

- (i) There exists $\bar{x} \in \operatorname{Fix} T = \operatorname{zer}(A + B) = \operatorname{argmin}(f + g)$ such that $T^n x_0 \rightarrow \bar{x}$.

Suppose that $\mu > \omega$. Then, we additionally have:

- (ii) $\operatorname{Fix} T = \operatorname{argmin}(f + g) = \{\bar{x}\}$ and $T^n x_0 \rightarrow \bar{x}$ with a linear rate $\delta < 1$.

Proof.

∇f is maximally μ -monotone, $\nabla f - \mu \operatorname{Id}$ is $\frac{1}{\beta}$ -Lipschitz continuous by an earlier result, hence β -cocoercive.

Theorem (the forward-backward algorithm when f is μ -strongly convex)

Let $\mu \geq \omega \geq 0$, and let $\beta > 0$. Let f be μ -strongly convex and Fréchet differentiable with a $\frac{1}{\beta}$ -Lipschitz continuous gradient, and let g be ω -hypoconvex. Suppose that $\operatorname{argmin}(f + g) \neq \emptyset$. Let $\gamma \in]0, \min\{1/\mu, 2\beta/(1 + 2\mu\beta)\}[$, and set $\delta = (1 - \gamma\mu)/(1 - \gamma\omega)$. Set

$$T = \operatorname{Prox}_{\gamma g}(\operatorname{Id} - \gamma \nabla f),$$

and let $x_0 \in X$. Then the following hold:

- (i) There exists $\bar{x} \in \operatorname{Fix} T = \operatorname{zer}(A + B) = \operatorname{argmin}(f + g)$ such that $T^n x_0 \rightarrow \bar{x}$.

Suppose that $\mu > \omega$. Then, we additionally have:

- (ii) $\operatorname{Fix} T = \operatorname{argmin}(f + g) = \{\bar{x}\}$ and $T^n x_0 \rightarrow \bar{x}$ with a linear rate $\delta < 1$.

Proof.

∇f is maximally μ -monotone, $\nabla f - \mu \operatorname{Id}$ is $\frac{1}{\beta}$ -Lipschitz continuous by an earlier result, hence β -cocoercive. Apply the previous result. □

Questions:

- ▶ What are other possible/more general classes of functions/operators which have “nice” prox operators/resolvents?
- ▶ What can we say about splitting operators when applied to problems involving these classes, if any?

References

For references, please check the preprints below and the references therein.

-  H.H. Bauschke, W.M. Moursi, and X. Wang (2019). Generalized monotone operators and their averaged resolvents, <https://arxiv.org/abs/1902.09827>.
-  W.M. Moursi (2019). On the averagedness of the Douglas–Rachford and the forward-backward operators. Preprint to appear on arxiv in the next few days.

THANK YOU !!