The Douglas–Rachford algorithm for inconsistent optimization problems: the complete story

Speaker: Walaa M. Moursi

Department of Combinatorics and Optimization, University of Waterloo
Waterloo, ON, Canada

Based on joint works with

Heinz H. Bauschke
Mathematics, University of British Columbia
Kelowna, B.C., Canada

The 7$^{\text{th}}$ Annual Loma Linda workshop 2021
on Particle Imaging and Radiation Therapy.
Monday August 2, 2021 via Zoom
11:00–11:30 am EDT

The setting

Throughout this talk

X is a real Hilbert space

with inner product $\langle \cdot \mid \cdot \rangle$, and induced norm $\|\cdot\|$, e.g., $\mathbb{R}^n$, $\mathbb{S}^n$ or ℓ^2.

- Recall that an operator $A\colon X \rightrightarrows X$ 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.,

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

The problem

Throughout the talk we assume that

A and B are maximally monotone operators on X.

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

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

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

$$T_{A,B} = \tfrac{1}{2}(\mathrm{Id} + \mathrm{R}_B \mathrm{R}_A).$$

• $\mathrm{Id}\colon X \to X\colon x \mapsto x$. • $\mathrm{R}_A := 2\mathrm{J}_A - \mathrm{Id} = 2(\mathrm{Id} + A)^{-1} - \mathrm{Id}$.

Motivation

The problem:

(P) Find $x \in \mathbb{R}^n$ such that x minimizes $f + g$.

- Suppose that f and g are smooth. Then (P) is equivalent to

$$\text{find } x \in X \text{ such that } 0 = \nabla(f+g)(x) = \nabla f(x) + \nabla g(x).$$

- If we drop the assumption of smoothness, (P) reduces to

$$\text{find } x \in X \text{ such that } 0 \in \partial(f+g)(x) = \partial f(x) + \partial g(x),$$

where $\partial f(x) = \left\{ u \in \mathbb{R}^n \;\middle|\; (\forall y)\ \langle u, y - x \,|\, + \rangle f(x) \le f(y) \right\}$.

Example of Constraint Qualifications (CQs): • $\operatorname{dom} f \cap \operatorname{int} \operatorname{dom} g \neq \varnothing$.

Examples

The problem:

(P) Find $x \in \mathbb{R}^n$ such that x minimizes $f + g$.

Let U be a nonempty closed convex subset of X. Recall that the indicator function of U, denoted by ι_U, is defined by

$$\iota_U(x) = \begin{cases} 0, & x \in U; \\ +\infty, & \text{otherwise.} \end{cases}$$

- Constrained convex optimization problem:

$$\left.\begin{array}{l} \text{minimize } f(x) \\ \text{subject to } x \in U \end{array}\right\} \longrightarrow \text{find } x \in \mathbb{R}^n \text{ such that } x \text{ minimizes } f + \iota_U.$$

- Convex feasibility problem:

find x such $x \in U \cap V \longrightarrow$ find $x \in \mathbb{R}^n$ such that x minimizes $\iota_U + \iota_V$.

Classical convergence results

Let $x_0 \in X$. Recall that when

$$\operatorname{zer}(A+B) = (A+B)^{-1}(0) \neq \varnothing$$

we have:

- Lions–Mercier (1979)

$$x_n = T^n x_0 \xrightarrow{\textit{weakly}} \text{some point } \overline{x} = T\overline{x} \in \operatorname{Fix} T \neq \operatorname{zer}(A+B).$$

- Combettes (2004) $\mathrm{J}_A(\operatorname{Fix} T) = \operatorname{zer}(A+B)$. Consequently,

$$\operatorname{Fix} T \neq \varnothing \Leftrightarrow \operatorname{zer}(A+B) \neq \varnothing.$$

- Svaiter (2009)

$$\mathrm{J}_A T^n x_0 \xrightarrow{\textit{weakly}} \mathrm{J}_A \overline{x} \in \operatorname{zer}(A+B).$$

• $\mathrm{J}_A := (\operatorname{Id} + A)^{-1}$. • $\mathrm{R}_A := 2\mathrm{J}_A - \operatorname{Id}$. • $T := \operatorname{Id} - \mathrm{J}_A + \mathrm{J}_B \mathrm{R}_A$.

Classical convergence results: function version

Let $x_0 \in X$. Recall that when

$$\mathrm{zer}(\partial f + \partial g) = (\partial f + \partial g)^{-1}(0) \neq \varnothing$$

we have:

- Lions–Mercier (1979)

$$x_n = T^n x_0 \xrightarrow{\textit{weakly}} \text{some point } \overline{x} = T\overline{x} \in \mathrm{Fix}\, T \neq \mathrm{zer}(\partial f + \partial g).$$

- Combettes (2004) $\mathrm{Prox}_f(\mathrm{Fix}\, T) = \mathrm{zer}(\partial f + \partial g)$. Consequently,

$$\mathrm{Fix}\, T \neq \varnothing \Leftrightarrow \mathrm{zer}(\partial f + \partial g) \neq \varnothing.$$

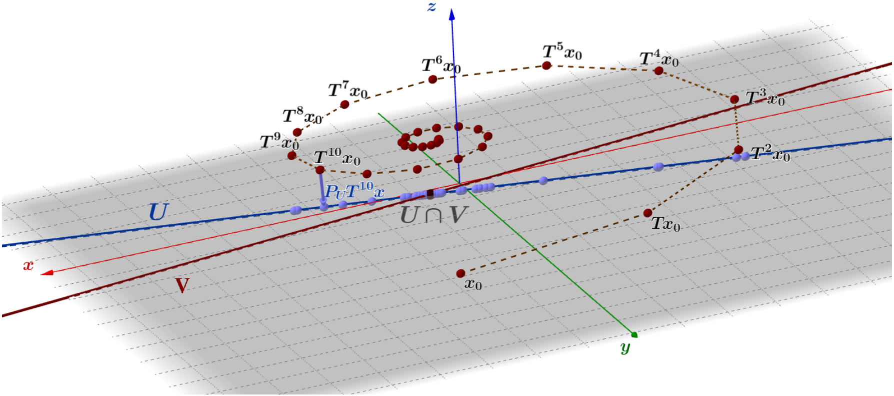
- Lions–Mercier–Svaiter

$$\mathrm{Prox}_f\, T^n x \xrightarrow{\textit{weakly}} \text{some point in } \mathrm{argmin}(f + g).$$

$\mathrm{Prox}_f(x) = \mathrm{argmin}_{y \in X} \left(f(y) + \frac{1}{2}\|x - y\|^2 \right).$

DR for two lines in $\mathbb{R}^3$

$$f = \iota_U,\ g = \iota_V \text{ and } T = \frac{1}{2}\Big(\mathrm{Id} + (2P_V - \mathrm{Id}) \circ (2P_U - \mathrm{Id})\Big).$$



U = the blue line,
V = the red line,
$(T^n x_0)_{n\in\mathbb{N}}$ = the red sequence,
$(P_U T^n x_0)_{n\in\mathbb{N}}$ = the blue sequence.

Convergence results: what if?

Let $x_0 \in X$. Recall that when

$$\operatorname{zer}(\partial f + \partial g) = (\partial f + \partial g)^{-1}(0) \neq \varnothing$$

we have:

- Lions–Mercier (1979)

$$x_n = T^n x_0 \xrightarrow{\textit{weakly}} \text{some point } \overline{x} = T\overline{x} \in \operatorname{Fix} T \neq \operatorname{zer}(\partial f + \partial g).$$

- Combettes (2004) $\operatorname{Prox}_f(\operatorname{Fix} T) = \operatorname{zer}(\partial f + \partial g)$. Consequently,

$$\operatorname{Fix} T \neq \varnothing \Leftrightarrow \operatorname{zer}(\partial f + \partial g) \neq \varnothing.$$

- Lions–Mercier–Svaiter

$$\operatorname{Prox}_f T^n x \xrightarrow{\textit{weakly}} \text{some point in } \operatorname{argmin}(f + g).$$

- Question: What happens when $\operatorname{zer}(\partial f + \partial g) = \varnothing$?

$\operatorname{Prox}_f(x) = \operatorname{argmin}_{y \in X} \left(f(y) + \frac{1}{2}\|x - y\|^2 \right).$

The case of infeasible affine subspaces: Example

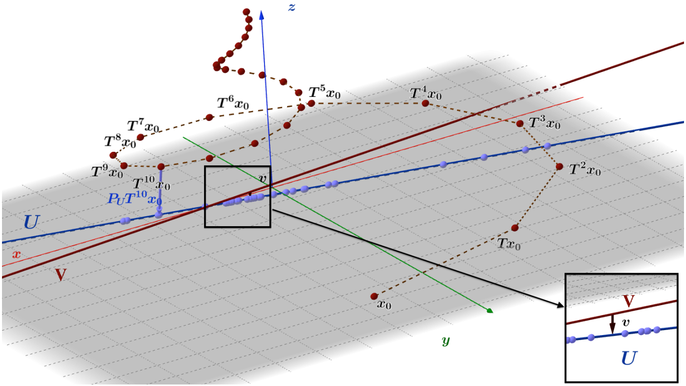


Figure: A GeoGebra snapshot. Two nonintersecting affine subspaces U (blue line) and V (red line) in $\mathbb{R}^3$. Shown are also the first few iterates of $(T^n x_0)_{n\in\mathbb{N}}$ (red points) and $(P_U T^n x_0)_{n\in\mathbb{N}}$ (blue points). In this case $\|T^n x_0\| \to +\infty$ but $(P_U T^n x_0)_{n\in\mathbb{N}}$ remains bounded!

The generalized framework of the normal problem: the right tools

- The minimal displacement vector

$$v := \mathrm{P}_{\overline{\mathrm{ran}}(\mathrm{Id} - T)}(0).$$

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

$$x \in \mathrm{zer}\big(-v + A + B(\cdot - v)\big)$$

- The generalized solution set or the *normal solutions*

$$Z = \big\{x \in X \bigm| 0 \in -v + Ax + B(x - v)\big\}.$$

Roots in linear algebra: least squares

- Suppose that $X = \mathbb{R}^n$, let $A \in \mathbb{R}^{n\times n}$ be such that $A + A^{\mathsf{T}}$ is positive semidefinite (A is maximally monotone!).
- Find $x \in \mathbb{R}^n$ such that $Ax = b$. Set $B \equiv -b$. The problem reduces to: Find $x \in \mathbb{R}^n$ such that
$$x \in \operatorname{zer}(A + B).$$
- If $b \notin \operatorname{ran} A$ then we $\operatorname{zer}(A + B) = \varnothing$.
- The minimal displacement vector is
$$v = -\mathrm{P}_{(\operatorname{ran} A)^{\perp}}(b).$$
- The normal solutions are the least squares solutions!

Earlier works

Let $x_0 \in X$. When $\operatorname{zer}(A+B) = \varnothing$ we always have

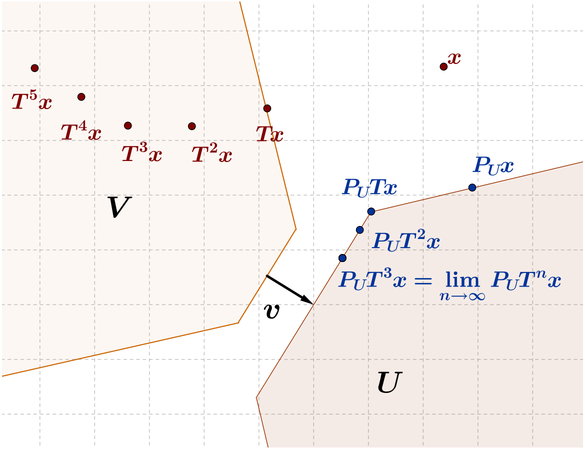
$$\|T^n x_0\| \to \infty.$$

Suppose that

$$v \in \operatorname{ran}(\operatorname{Id} - T).$$

- Bauschke–Combettes–Luke (2003) proved that when $(f, g) = (\iota_U, \iota_V)$, U, V nonempty closed convex subsets of X, then the shadow sequence $(\mathrm{P}_U T^n x)_{n\in\mathbb{N}}$ is bounded and its weak cluster points are minimizers of the function $\iota_U + \iota_V(\cdot - v)$ (i.e., normal solutions!).
- Bauschke–M (2015) proved the strong convergence of the shadow sequence with a linear rate and indetified the limit when U, V are closed affine subspaces.
- Bauschke–M (2016) proved the weak convergence of the shadow sequence to a normal solution when U, V nonempty closed convex subsets of X.
- Bauschke–M (2019) proved the weak convergence of the shadow sequence to a normal solution when f is convex lower semicontinuous and proper and $g = \iota_U$ where U is a closed affine subspace X under the assumption that $0 \in \operatorname{dom} f^* + U^\perp$.

Convex feasibility example



A GeoGebra snapshot. U and V are two nonintersecting sets in $\mathbb{R}^2$. Also, the first few iterates of the governing sequence $(T^n x)_{n\in\mathbb{N}}$ (red points) and the shadow sequence $(P_U T^n x)_{n\in\mathbb{N}}$ (blue points) are shown.

Related works

Of central importance of these results was the following fact:

- Bauschke–Hare–M (2014): Suppose X is finite-dimensional and A and B are nice, e.g., subdifferentials of convex functions f and g respectively. Then
$$\overline{\operatorname{ran}}(\mathrm{Id} - T) = \overline{\operatorname{dom} f - \operatorname{dom} g} \cap \overline{\operatorname{dom} f^* + \operatorname{dom} g^*}.$$
- Ryu–Lin–Yin (2017 and 2018 respectively) proposed a method based on the Douglas–Rachford algorithm that identifies, in certain situations, infeasible, unbounded, and pathological conic (and feasible and infeasible convex, respectively) optimization problems.
- Banjac–Goulart–Stellato–Boyd (2018) showed that for certain classes of convex optimization problems, ADMM can detect primal and dual infeasibility of the problem and they propose a termination criterion.
- Banjac–Lygeros and Banjac (2020) extended some of the geometric properties of the minimal displacement vector established in our 2019 work.

More generally ...

Let $x_0 \in X$.

- Can we learn more when A and B are nice maximally monotone operators?
- As a first step: Can we characterize when the shadows are bounded?
- Suppose the shadows are bounded. What can we say about the weak cluster points? full convergence??

Our assumptions: A1

We assume that

$$\overline{\operatorname{ran}}(\operatorname{Id} - T) = \overline{\operatorname{dom} A - \operatorname{dom} B} \cap \overline{\operatorname{ran} A + \operatorname{ran} B}.$$

True, e.g., when X is finite-dimensional and $(A, B) = (\partial f, \partial g)$.

A1 holds in the optimization settings when X is finite-dimensional. ✓

The beautiful geomerty and the vectors v_D and v_R!

Recall that $\overline{\operatorname{ran}}(\operatorname{Id}-T) = \overline{\operatorname{dom} A - \operatorname{dom} B} \cap \overline{\operatorname{ran} A + \operatorname{ran} B}$. We now introduce the vectors

$$v_D := \mathrm{P}_{\overline{\operatorname{dom} A - \operatorname{dom} B}}(0) \quad \text{and} \quad v_R := \mathrm{P}_{\overline{\operatorname{ran} A + \operatorname{ran} B}}(0)$$

We can conclude

(i) $v_D \in (-\operatorname{rec} \overline{\operatorname{dom}}\, A)^{\ominus} \cap (\operatorname{rec} \overline{\operatorname{dom}}\, B)^{\ominus}$.

(ii) $v_R \in (-\operatorname{rec} \overline{\operatorname{ran}}\, A)^{\ominus} \cap (-\operatorname{rec} \overline{\operatorname{ran}}\, B)^{\ominus}$.

• $\operatorname{rec} C = \{x \in X \mid x + C \subseteq C\}$. • $C^{\ominus} = \{u \in X \mid \sup\langle C \mid u\rangle \leq 0\}$.

Fact

Let U and V be nonempty closed convex subsets of X. Then

$$\mathrm{P}_{\overline{U-V}}(0) \in \overline{(\mathrm{P}_U - \operatorname{Id})(V)} \cap \overline{(\operatorname{Id} - \mathrm{P}_V)(U)} \subseteq (-\operatorname{rec} U)^{\ominus} \cap (\operatorname{rec} V)^{\ominus}.$$

The beautiful geomerty

The following lemma is of crucial importance in our work.

Lemma

The following hold for A and B:

(i) $(\operatorname{rec}\overline{\operatorname{dom}}A)^{\ominus} \subseteq \operatorname{rec}(\overline{\operatorname{ran}}A)$ *and* $(\operatorname{rec}\overline{\operatorname{dom}}B)^{\ominus} \subseteq \operatorname{rec}(\overline{\operatorname{ran}}B)$.

(ii) $(\operatorname{rec}\overline{\operatorname{ran}}A)^{\ominus} \subseteq \operatorname{rec}(\overline{\operatorname{dom}}A)$ *and* $(\operatorname{rec}\overline{\operatorname{ran}}B)^{\ominus} \subseteq \operatorname{rec}(\overline{\operatorname{dom}}B)$.

Proof.

Using the celebrated Brezis–Haraux theorem

$$\overline{\operatorname{ran}}\,A + \overline{\operatorname{ran}}\,\mathrm{N}_{\overline{\operatorname{dom}}A} \subseteq \overline{\operatorname{ran}\,A + \operatorname{ran}\,\mathrm{N}_{\overline{\operatorname{dom}}A}} = \overline{\operatorname{ran}}(A + \mathrm{N}_{\overline{\operatorname{dom}}A}) = \overline{\operatorname{ran}}\,A$$

and we conclude that

$$\overline{\operatorname{ran}}\,\mathrm{N}_{\overline{\operatorname{dom}}A} \subseteq \operatorname{rec}\overline{\operatorname{ran}}A.$$

On the other hand, using a result by Zarantonello we have

$$\overline{\operatorname{ran}}\,\mathrm{N}_{\overline{\operatorname{dom}}A} = \overline{\operatorname{ran}}\,(\operatorname{Id} - \mathrm{P}_{\overline{\operatorname{dom}}A}) = (\operatorname{rec}\overline{\operatorname{dom}}A)^{\ominus}.$$

$\square$

• $\operatorname{rec} C = \{x \in X \mid x + C \subseteq C\}$. • $C^{\ominus} = \{u \in X \mid \sup\langle C \mid u\rangle \leq 0\}$.

The beautiful geomerty and locating v_D and v_R!

Proposition

The following hold:

(i) $v_D \in (-\operatorname{rec}\overline{\operatorname{dom}}\,A)^{\ominus} \cap (\operatorname{rec}\overline{\operatorname{dom}}\,B)^{\ominus}$.

(ii) $v_D \in (-\operatorname{rec}\overline{\operatorname{ran}}\,A) \cap (\operatorname{rec}\overline{\operatorname{ran}}\,B)$.

(iii) $v_R \in (-\operatorname{rec}\overline{\operatorname{ran}}\,A)^{\ominus} \cap (-\operatorname{rec}\overline{\operatorname{ran}}\,B)^{\ominus}$.

(iv) $v_R \in (-\operatorname{rec}\overline{\operatorname{dom}}\,A) \cap (-\operatorname{rec}\overline{\operatorname{dom}}\,B)$.

(v) $\langle v_D \mid v_R \rangle = 0$.

(vi) $v_D + v_R \in \overline{\operatorname{dom} A - \operatorname{dom} B} \cap \overline{\operatorname{ran} A + \operatorname{ran} B}$.

(vii) $v = v_D + v_R$.

(viii) $\|v\|^2 = \|v_D\|^2 + \|v_R\|^2 = \|(v_R, v_D)\|^2$.

• $v_D := P_{\overline{\operatorname{dom} A - \operatorname{dom} B}}(0)$. • $v_R := P_{\overline{\operatorname{ran} A + \operatorname{ran} B}}(0)$.

Dynamic consequences

Known: Let $x \in X$. Then

$$\mathrm{J}_A T^n x - \mathrm{J}_B \mathrm{R}_A T^n x = \mathrm{J}_{A^{-1}} T^n x + \mathrm{J}_{B^{-1}} \mathrm{R}_A T^n x = T^n x - T^{n+1} x \to v.$$

Proposition

Let $x \in X$. Then the following hold:

(i) $\mathrm{J}_A T^n x - \mathrm{J}_A T^{n+1} x \to v_R$.

(ii) $\mathrm{J}_{A^{-1}} T^n x - \mathrm{J}_{A^{-1}} T^{n+1} x \to v_D$.

Proposition (**shadow convergence: necessary condition**)

Let $x \in X$. Then the following hold:

(i) $(\mathrm{J}_A T^n x)_{n \in \mathbb{N}}$ *is asymptotically regular* $\Leftrightarrow v_R = 0$.

(ii) $(\mathrm{J}_{A^{-1}} T^n x)_{n \in \mathbb{N}}$ *is asymptotically regular* $\Leftrightarrow v_D = 0$.

• $\mathrm{Id} - T = \mathrm{J}_A - \mathrm{J}_B \mathrm{R}_A = \mathrm{J}_{A^{-1}} + \mathrm{J}_{B^{-1}} \mathrm{R}_A$. • $v_D := \mathrm{P}_{\overline{\mathrm{dom}\, A - \mathrm{dom}\, B}}(0)$. • $v_R := \mathrm{P}_{\overline{\mathrm{ran}\, A + \mathrm{ran}\, B}}(0)$. • $v = v_D + v_R$.

Our assumptions: A2

We assume that

$$v \in \operatorname{ran}(\mathrm{Id} - T).$$

Equivalently (proof omitted),

$$Z = \{x \in X \mid 0 \in -v + Ax + B(x - v)\} \neq \varnothing.$$

And finally we see Fejér monotonicity!

Working in $X \times X$ we state the following key result:

Theorem

Suppose that $v \in \operatorname{ran}(\mathrm{Id} - T)$, let $x \in X$, and let $n \in \mathbb{N}$. Then the following hold:

(i) *Suppose that A and B are paramonotone (true when $(A, B) = (\partial f, \partial g)$). Then the sequence*

$$\big((0, -v) + (\mathrm{J}_A T^n x + n v_R, \mathrm{J}_{A^{-1}} T^n x + n v_D)\big)_{n \in \mathbb{N}}$$

is Fejér monotone with respect to $Z \times K$.

(ii) *The sequence $(\mathrm{J}_A T^n x + n v_R, \mathrm{J}_{A^{-1}} T^n x + n v_D)_{n \in \mathbb{N}}$ is bounded.*

(iii) *The sequence $(\mathrm{J}_A T^n x)_{n \in \mathbb{N}}$ is bounded $\Leftrightarrow$ $v_R = 0$.*

(iv) *The sequence $(\mathrm{J}_{A^{-1}} T^n x)_{n \in \mathbb{N}}$ is bounded $\Leftrightarrow$ $v_D = 0$.*

• $Z := \operatorname{zer}(-v + A + B(\cdot - v))$. • $K := \operatorname{zer}((-v + A)^{-1} + (B(\cdot - v))^{-\circledcirc})$.
• $A^{-\circledcirc} = (-\mathrm{Id}) \circ A^{-1} \circ (-\mathrm{Id})$.

The optimization setting

From now on we assume that

$$f \text{ and } g \text{ are proper lsc convex functions on } X$$

and that $(A, B) = (\partial f, \partial g)$. And finally we assume (A3):

$$v_R = 0 \quad \Leftrightarrow \quad v = \mathrm{P}_{\overline{\operatorname{ran}}(\mathrm{Id}-T)}(0) = \mathrm{P}_{\overline{\operatorname{dom} f - \operatorname{dom} g}}(0) = v_D.$$

We use the abbreviations

$$\big(P_f, P_{f^*}, P_g, R_f\big) = \big(\operatorname{Prox}_f, \operatorname{Prox}_{f^*}, \operatorname{Prox}_g, 2\operatorname{Prox}_f - \mathrm{Id}\big).$$

Hence

$$T = T_{(\partial f, \partial g)} = \mathrm{Id} - P_f + P_g R_f.$$

• $v_D := \mathrm{P}_{\overline{\operatorname{dom} A - \operatorname{dom} B}}(0)$. • $v_R := \mathrm{P}_{\overline{\operatorname{ran} A + \operatorname{ran} B}}(0)$. • $v = v_D + v_R$. • $\mathrm{J}_{\partial f} = P_f$. • $\mathrm{J}_{(\partial f)^{-1}} = \mathrm{J}_{\partial f^*} = P_{f^*}$.

Convergence proof in a nutshell

- Step 1: refining Z.
  Because $v_R = 0$ we learn that $Z = \{x \in X \mid 0 \in \partial f(x) + \partial g(x - v)\}$.
  Because $Z \neq \varnothing$ (recalling (A2) $v \in \operatorname{ran}(\operatorname{Id} - T)$) we prove that $Z = \operatorname{argmin}(f + g(\cdot - v))$.
- Step 2: boundedness of the shadows. This is a consequnce of Fejér monotonicity and the assumption $v_R = 0$.
- Step 3: locating the weak cluster points of the shadows. We show that the weak cluster points are minimizers of $f + g(\cdot - v)$.
- Step 4: full weak convergence of the shadows. We combine Step 2, Step 3 and properties of Fejér monotone sequences.

Step 1: refining Z

Proposition

Recalling that $Z = \{x \in X \mid 0 \in -v + \partial f(x) + \partial g(x - v)\}$, *and that* $v_R = 0$ *we have:*

(i) $Z = \{x \in X \mid 0 \in \partial f(x) + \partial g(x - v)\}$.

(ii) $Z \neq \varnothing \Rightarrow Z = \operatorname{argmin}_{x \in X}(f(x) + g(x - v))$.

• $v_D := P_{\overline{\mathrm{dom}\, A - \mathrm{dom}\, B}}(0)$. • $v_R := P_{\overline{\mathrm{ran}\, A + \mathrm{ran}\, B}}(0)$. • $v = v_D + v_R$.

Step 2: boundedness of the shadows

We proved earlier that: The sequence

$$((0, -v) + (\mathrm{J}_A T^n x + n v_R, \mathrm{J}_{A^{-1}} T^n x + n v_D))_{n\in\mathbb{N}}$$

is Fejér monotone with respect to $Z \times K$.
Using that $(A, B, v_R) = (\partial f, \partial g, 0)$ we have $(\mathrm{J}_A, \mathrm{J}_{A^{-1}}) = (P_f, P_{f^*})$ and therefore the sequence

$$((0, -v) + (P_f T^n x, P_{f^*} T^n x + n v))_{n\in\mathbb{N}}$$

is Fejér monotone with respect to $Z \times K$.

• $v_D := \mathrm{P}_{\overline{\mathrm{dom}\, A - \mathrm{dom}\, B}}(0)$. • $v_R := \mathrm{P}_{\overline{\mathrm{ran}\, A + \mathrm{ran}\, B}}(0)$. • $v = v_D + v_R$. •
$Z := \mathrm{zer}(\partial f + \partial g(\cdot - v))$. • $K := \mathrm{zer}((\partial f)^{-1} + (\partial g(\cdot - v))^{-\circledcirc})$.

Step 3: locating the weak cluster points of the shadows.

Proposition

Set $\mu := \min_{x\in X}(f(x)+g(x-v))$ and let $x \in X$. Then the following hold:

(i) $(P_f T^n x)_{n\in\mathbb{N}}$ *is bounded and its weak cluster points are minimizers of* $f+g(\cdot - v)$.

(ii) $(P_g R_f T^n x)_{n\in\mathbb{N}}$ *is bounded and its weak cluster points are minimizers of* $f(\cdot + v)+g$.

Now let $\overline{z}$ be a weak cluster point of $(P_f T^n x)_{n\in\mathbb{N}}$. Then:

(iii) $f(P_f T^n x) \to f(\overline{z})$. *(value convergence ✓)*

(iv) $g(P_g R_f T^n x) \to g(\overline{z} - v)$.

(v) $f(P_f T^n x) + g(P_g R_f T^n x) \to \mu$.

Step 4: full weak convergence of the shadows.

Proposition

Let $x \in X$. Then the following hold:

(i) *The sequence $(P_f T^n x)_{n\in\mathbb{N}}$ converges weakly to a minimizer of $f + g(\cdot - v)$.*

(ii) *The sequence $(P_g R_f T^n x)_{n\in\mathbb{N}}$ converges weakly to a minimizer of $f(\cdot + v) + g$.*

Proof.

(i) We showed that the sequence $(P_f T^n x, -v + P_{f^*} T^n x + nv)_{n\in\mathbb{N}}$ is Fejér monotone with respect to $Z \times K$. Now let z_1 and z_2 be two weak cluster points of $(P_f T^n x)_{n\in\mathbb{N}}$. On the one hand,

$$\{z_1, z_2\} \subseteq \operatorname*{argmin}_{x\in X}(f + g(\cdot - v)) = Z; \quad \text{hence,} \quad z_1 - z_2 \in Z - Z.$$

On the other hand, $z_1 - z_2 \in (Z - Z)^\perp$ (proof omitted). Altogether we conclude that $z_1 - z_2 \in (Z - Z) \cap (Z - Z)^\perp = \{0\}$. Hence, $z_1 = z_2$.✓
(ii) A direct consequence of (i) and earlier result. □

How critical are our assumptions?

- (A1)
$$\overline{\operatorname{ran}}(\mathrm{Id} - T) = \overline{\operatorname{dom} A - \operatorname{dom} B} \cap \overline{\operatorname{ran} A + \operatorname{ran} B}.$$
True, e.g., when X is finite-dimensional and $(A, B) = (\partial f, \partial g)$.
A1 holds in the optimization settings when X is finite-dimensional. .✓
- (A3) $v_R = 0$. We have proved that it is a necessary and sufficient condition for convergence.✓
- (A2) $v \in \operatorname{ran}(\mathrm{Id} - T)$.

A2 fails and the shadows converge in one step!

- Suppose that $X = \mathbb{R}$.
- Set $(f, g) = (\iota_{\{0\}}, -\sqrt{\cdot})$.
- Clearly, $\operatorname{dom}\partial f = \operatorname{dom} N_{\{0\}} = \{0\}$, $\operatorname{dom}\partial g =]0, +\infty[$. Moreover, $\operatorname{ran}\partial f = \mathbb{R} = \operatorname{ran}\partial f + \operatorname{ran}\partial g$.
- Hence, $Z = \varnothing$.
- $\overline{\operatorname{ran}}(\operatorname{Id} - T) = \overline{\operatorname{dom}\partial f - \operatorname{dom}\partial g} = [0, +\infty[$ and $v = 0 \notin \operatorname{ran}(\operatorname{Id} - T)$.
- $(\forall n \in \mathbb{N})\ P_f T^n x = P_{\{0\}} T^n x = 0$. ✓

A2 fails and the shadows are unbounded.

We revisit an example by Ryu–Liu–Yin (2019).

- Suppose that $X = \mathbb{R}^3$.
- Let $K = \left\{(x_1, x_2, x_3) \mid \sqrt{x_1^2 + x_2^2} \leq |x_3|\right\}$
- Set $(f, g) = (\iota_K, \langle e_1 \mid \cdot\rangle + \iota_{\{x_2 = x_3\}})$.
- Let $x \in \mathbb{R} \times \mathbb{R} \times \{0\}$.
- After filling a lot of details ....
- $v = 0 \notin \operatorname{ran}(\mathrm{Id} - T)$.
- $Z = \varnothing$.
- $\operatorname{argmin}(f + g) = K \cap (\mathbb{R} \cdot (0, 1, 1)) \neq \varnothing$.
- $(\forall n \in \mathbb{N})\ \|P_f T^n x_0\| = \|P_K T^n x\| \to +\infty.$ ✓

References

H.H. Bauschke and W.M. Moursi (2021). On the Douglas–Rachford algorithm for solving possibly inconsistent optimization problems, https://arxiv.org/pdf/2106.11547.pdf

THANK YOU!!

- email: walaa.moursi@uwaterloo.ca