Douglas-Rachford algorithm for nonmonotone multioperator inclusion problems

Jan Harold Alcantara
RIKEN AIP, Tokyo, Japan

MATRIX Research Program
Splitting Algorithms - Advances, Challenges and Opportunities

This is a joint work with Prof. Akiko Takeda (The University of Tokyo)

Preview

Multioperator Inclusion Problem

$$\text{Find } x \in \mathcal{H} \text{ such that } 0 \in A_1(x) + A_2(x) + \cdots + A_m(x) \qquad \text{(InclProb)}$$

- $\mathcal{H}$ is a real Hilbert space
- $A_i : \mathcal{H} \rightrightarrows \mathcal{H}$ is a set valued operator, $m \geq 3$

Goals

- Develop a **Douglas-Rachford (DR) algorithm** for **nonmonotone** (InclProb).
- Establish **convergence guarantees** for the DR algorithm under **generalized monotonicity** conditions.

Outline

I. (Review) Douglas-Rachford when $m = 2$
II. Product space reformulations
III. Convergence results under generalized monotonicity
IV. (If time permits) Extensions to comonotone functions

Alcantara and Takeda, Douglas-Rachford algorithm for nonmonotone multioperator inclusion problems, arXiv:2501.02752 (2025). Preprint

DR for monotone case when $m = 2$

$$\text{Find } x \in \mathcal{H} \text{ such that } 0 \in A(x) + B(x) \qquad \text{(2operator-problem)}$$

- Often investigated under the assumption of *monotonicity*
- An operator $A : \mathcal{H} \rightrightarrows \mathcal{H}$ is monotone if

$$\langle x - y, u - v \rangle \geq 0 \quad \forall (x, u), (y, v) \in \operatorname{gra}(A).$$ [1]

A is maximal monotone if it is monotone and there is no monotone operator whose graph properly contains $\operatorname{gra}(A)$.

Douglas-Rachford (DR) algorithm

$$x^{k+1} = T_{A,B}(x^k) := x^k + \kappa(J_{\gamma B} \circ (2J_{\gamma A} - \mathrm{Id}) - J_{\gamma A})(x^k), \quad \kappa \in (0, 2).$$

- Note: The resolvent $J_{\gamma A} = (\mathrm{Id} + \gamma A)^{-1}$ is a single-valued mapping on $\mathcal{H}$ when A is maximal monotone.

[1] $\operatorname{gra}(A) := \{(x, u) : u \in A(x)\}$

Convergence under monotonicity assumption

- We can also write DR algorithm as

$$\begin{aligned} z^k &= J_{\gamma A}(x^k) \qquad \text{(shadow sequence)} \\ y^k &= J_{\gamma B}(2z^k - x^k) \\ x^{k+1} &= x^k + \kappa(y^k - z^k) \end{aligned}$$

Theorem (Lions and Mercier, 1979)

Suppose that A and B are maximal monotone such that $\operatorname{zer}(A+B) \neq \emptyset$, and let $\gamma > 0$. Then

(i) $\exists \bar{x} \in \operatorname{Fix}(T_{A,B})$ *s.t.* $x^k \rightharpoonup \bar{x}$, *with* $\bar{z} := J_{\gamma A}(\bar{x}) \in \operatorname{zer}(A+B)$.

(ii) $y^k - z^k \to 0$

(iii) $z^k \rightharpoonup \bar{z}$.

What if A and B are NOT maximal monotone?

The case of optimization: Nonconvex objectives

$$\Omega := \arg\min_{x \in \mathcal{H}} \; f(x) + g(x) \tag{Opt}$$

- We can apply DR to (2operator-problem) with $A = \partial f$ and $B = \partial g$.

	f	g	Remarks
Guo, Han, & Yuan (2017)	α-convex	β-convex	$\alpha + \beta > 0$
Themelis & Patrinos (2020)	L-smooth	proper lsc	*

Table: Existing works

- Convergence holds for *sufficiently small* stepsizes, assuming $\Omega \neq \emptyset$.
- Applies to finite-dimensional cases only.

[1] Given $\alpha \in \mathbb{R}$, f is α-convex if $f - \frac{\alpha}{2}\|\cdot\|^2$ is convex.

Generalization to operators

- Let $A : \mathcal{H} \rightrightarrows \mathcal{H}$ and let $\sigma \in \mathbb{R}$. We say that A is σ-monotone if

$$\langle x - y, u - v \rangle \geq \sigma \|x - y\|^2 \quad \forall (x, u), (y, v) \in \operatorname{gra}(A).$$

- A is maximal σ-monotone if A is σ-monotone and there is no σ-monotone operator whose graph properly contains $\operatorname{gra}(A)$.

Theorem (Dao & Phan, 2019)

Suppose that A and B are maximal α-monotone and maximal β-monotone, respectively, such that $\alpha + \beta > 0$ and $\operatorname{zer}(A + B) \neq \emptyset$, and suppose

$$1 + \gamma \frac{\alpha\beta}{\alpha + \beta} > \frac{\kappa}{2}.$$

Then

(i) $\exists \bar{x} \in \operatorname{Fix}(T_{A,B})$ s.t. $x^k \rightharpoonup \bar{x}$, with $\bar{z} := J_{\gamma A}(\bar{x}) \in \operatorname{zer}(A + B)$.

(ii) $\left\|x^k - x^{k+1}\right\| = o(1/\sqrt{k})$

(iii) $z^k \to \bar{z}$ and $\operatorname{zer}(A + B) = \{\bar{z}\}$.

Remarks

- Giselsson & Moursi (2021) obtained the same result when $\alpha\beta < 0$ and $\kappa = 1$.
- Dao & Phan (2019) actually proved the convergence of a more general algorithm.

Adaptive DR algorithm

$$z^k = J_{\gamma A}(x^k)$$
$$y^k = J_{\delta B}((1-\lambda)x^k + \lambda z^k)$$
$$x^{k+1} = x^k + \frac{\kappa}{2}\mu(z^k - y^k).$$

where

$$(\lambda - 1)(\mu - 1) = 1 \quad \text{and} \quad \delta = \gamma(\lambda - 1).$$

Can we extend the result to $m > 2$?

Pierra's product space reformulation (1976)

Find $\mathbf{x} = (x_1, \ldots, x_m) \in \mathcal{H}^m$ such that $0 \in \mathbf{F}(\mathbf{x}) + \mathbf{G}(\mathbf{x})$, (Pierra)

- $\mathbf{F}(\mathbf{x}) \coloneqq A_1(x_1) \times \cdots \times A_m(x_m)$
- $\mathbf{G} \coloneqq N_{\mathbf{D}_m}$ (maximal monotone)

$$\text{where } \mathbf{D}_m \coloneqq \{(x_1, \ldots, x_m) \in \mathcal{H}^m : x_1 = \cdots = x_m\}.$$

Recall that

$$N_{\mathbf{D}_m}(\mathbf{x}) = \begin{cases} \mathbf{D}_m^\perp = \{\mathbf{w} = (w_1, \ldots, w_m) : \sum_{i=1}^m w_i = 0\} & \text{if } \mathbf{x} \in \mathbf{D}_m, \\ \emptyset & \text{otherwise,} \end{cases}$$

- Nice property: Reformulation is valid for arbitrary operators!
- Easy resolvents too!

DR based on Pierra's product space reformulation

$$z_i^k \in J_{\gamma A_i}(x_i^k) \quad i = 1, \ldots, m.$$

$$y^k = \frac{1}{m}\sum_{i=1}^{m}(2z_i^k - x_i^k)$$

$$x_i^{k+1} = x_i^k + \kappa(y^k - z_i^k) \quad i = 1, \ldots, m$$

- Incompatible with generalized monotone operator framework! 😔

Campoy (2022)/Kruger's (1985) product space reformulation

Find $\mathbf{x} = (x_1, \ldots, x_{m-1}) \in \mathcal{H}^m$ such that $0 \in \mathbf{F}(\mathbf{x}) + \mathbf{G}(\mathbf{x})$, (Campoy/Kruger)

$$\mathbf{F}(\mathbf{x}) := A_1(x_1) \times \cdots \times A_{m-1}(x_{m-1}),$$
$$\mathbf{G}(\mathbf{x}) := \mathbf{K}(\mathbf{x}) + N_{\mathbf{D}_{m-1}}(\mathbf{x}),$$

where

$$\mathbf{K}(\mathbf{x}) := \frac{1}{m-1} A_m(x_1) \times \cdots \frac{1}{m-1} A_m(x_{m-1}).$$

- Reformulation is not valid for arbitrary operators.
- Convex-valuedness of $A_m(\cdot)$ is important!

$$\mathbf{F}(\mathbf{x}) := A_1(x_1) \times \cdots \times A_{m-1}(x_{m-1}),$$
$$\mathbf{G}(\mathbf{x}) := \mathbf{K}(\mathbf{x}) + N_{\mathbf{D}_{m-1}}(\mathbf{x}),$$
$$\mathbf{K}(\mathbf{x}) := \frac{1}{m-1} A_m(x_1) \times \cdots \frac{1}{m-1} A_m(x_{m-1}).$$

$$0 \in \mathbf{F}(\mathbf{x}) + \mathbf{K}(\mathbf{x}) + N_{\mathbf{D}_{m-1}}(\mathbf{x})$$
$$\iff \mathbf{x} \in \mathbf{D}_{m-1}, \quad \exists \mathbf{u} \in \mathbf{F}(\mathbf{x}), \quad \exists \mathbf{v} \in \mathbf{K}(\mathbf{x}), \quad \text{s.t. } -(\mathbf{u} + \mathbf{v}) \in N_{\mathbf{D}_{m-1}}(\mathbf{x})$$

- $\mathbf{x} = (x, \ldots, x)$
- $\mathbf{u} = (u_1, \ldots, u_{m-1}) \in A_1(x) \times \cdots \times A_{m-1}(x)$
- $\mathbf{v} = \frac{1}{m-1}(v_1, \ldots, v_{m-1})$ where $v_i \in A_m(x)$.

Thus,

$$u_1 + \cdots + u_{m-1} + \underbrace{\frac{1}{m-1} \sum_{i=1}^{m-1} v_i}_{\text{should be in } A_m(x)} = 0$$

Alternative reformulation

Find $\mathbf{x} = (x_1, \ldots, x_{m-1}) \in \mathcal{H}^m$ such that $0 \in \mathbf{F}(\mathbf{x}) + \mathbf{G}(\mathbf{x})$,

$$\mathbf{F}(\mathbf{x}) := A_1(x_1) \times \cdots \times A_{m-1}(x_{m-1}),$$
$$\mathbf{G}(\mathbf{x}) := \mathbf{K}(\mathbf{x}) + N_{\mathbf{D}_{m-1}}(\mathbf{x}),$$

where

$$\mathbf{K}(\mathbf{x}) := \begin{cases} \{(\lambda_1 v, \ldots, \lambda_{m-1} v) : v \in A_m(x_1)\} & \text{if } \mathbf{x} = (x_1, \ldots, x_{m-1}) \in \mathbf{D}_{m-1} \\ \emptyset & \text{otherwise.} \end{cases}$$

- If $\sum_{i=1}^{m-1} \lambda_i = 1$, then

$$\boxed{\operatorname{zer}(\mathbf{F} + \mathbf{G}) = \mathbf{\Delta}_{m-1}\left(\operatorname{zer}\left(\sum_{i=1}^{m} A_i\right)\right)}$$

where $\mathbf{\Delta}_{m-1}(x) := (x, \ldots, x)$.

Convex-valued case

Proposition

Suppose $\sum_{i=1}^{m-1} \lambda_i = 1$, and let

$$\mathbf{G} := \mathbf{K} + N_{\mathbf{D}_{m-1}},$$
$$\mathbf{G}_{\text{Campoy}} := \mathbf{K}_{\text{Campoy}} + N_{\mathbf{D}_{m-1}},$$

where

$$\mathbf{K}(\mathbf{x}) := \begin{cases} \{(\lambda_1 v, \dots, \lambda_{m-1} v) : v \in A_m(x_1)\} & \text{if } \mathbf{x} \in \mathbf{D}_{m-1} \\ \emptyset & \text{otherwise.} \end{cases}$$
$$\mathbf{K}_{\text{Campoy}}(\mathbf{x}) := \lambda_1 A_m(x_1) \times \cdots \times \lambda_{m-1} A_m(x_{m-1}).$$

$$\text{gra}(\mathbf{G}) = \text{gra}(\mathbf{G}_{\text{Campoy}}).$$

How about the resolvents?

Define the $\boldsymbol{\Lambda}$-resolvent of $\mathbf{F}$ as

$$J_{\lambda \mathbf{F}}^{\boldsymbol{\Lambda}}(\mathbf{x}) := (\mathbf{Id} + \lambda \boldsymbol{\Lambda}^{-1} \circ \mathbf{F})^{-1},$$

where $\boldsymbol{\Lambda}$ is the diagonal operator

$$\boldsymbol{\Lambda}(\mathbf{x}) = (\lambda_1 x_1, \ldots, \lambda_{m-1} x_{m-1}).$$

By direct calculations

$$J_{\lambda \mathbf{F}}^{\boldsymbol{\Lambda}}(\mathbf{x}) = J_{\frac{\lambda}{\lambda_1} A_1}(x_1) \times \cdots \times J_{\frac{\lambda}{\lambda_{m-1}} A_{m-1}}(x_{m-1}),$$
$$J_{\lambda \mathbf{G}}^{\boldsymbol{\Lambda}}(\mathbf{x}) = \boldsymbol{\Delta}_{m-1}\left(J_{\lambda A_m}\left(\sum_{i=1}^{m-1} \lambda_i x_i \right)\right)$$

DR for multioperator inclusion

Define the DR iterates as

$$\mathbf{x}^{k+1} \in T(\mathbf{x}^k),$$

where

$$T(\mathbf{x}) := \{\mathbf{x} + \kappa(\mathbf{y} - \mathbf{z}) : \mathbf{z} \in J^{\mathbf{\Lambda}}_{\lambda \mathbf{F}}(\mathbf{x}),\ \mathbf{y} \in J^{\mathbf{\Lambda}}_{\lambda \mathbf{G}}(2\mathbf{z} - \mathbf{x})\}$$
$$= \frac{(2-\kappa)\,\mathbf{Id} + \kappa R^{\mathbf{\Lambda}}_{\lambda \mathbf{G}} R^{\mathbf{\Lambda}}_{\lambda \mathbf{F}}}{2}, \qquad R^{\mathbf{\Lambda}}_{\lambda \#} := 2J^{\mathbf{\Lambda}}_{\lambda \#} - \mathbf{Id}$$

Douglas-Rachford algorithm for *m*-operator inclusion

$$z_i^k \in J_{\frac{\lambda}{\lambda_i} A_i}(x_i^k), \qquad (i = 1, \ldots, m-1)$$
$$y^k \in J_{\lambda A_m}\left(\sum_{i=1}^{m-1} \lambda_i (2z_i^k - x_i^k)\right)$$
$$x_i^{k+1} = x_i^k + \kappa(y^k - z_i^k) \qquad (i = 1, \ldots, m-1).$$

Fixed points of the DR map

Recall that

$$T(\mathbf{x}) = \{\mathbf{x} + \kappa(\mathbf{y} - \mathbf{z}) : \mathbf{z} \in J^{\mathbf{\Lambda}}_{\lambda\mathbf{F}}(\mathbf{x}),\ \mathbf{y} \in J^{\mathbf{\Lambda}}_{\lambda\mathbf{G}}(2\mathbf{z} - \mathbf{x})\}$$
$$= \frac{(2-\kappa)\,\mathbf{Id} + \kappa R^{\mathbf{\Lambda}}_{\lambda\mathbf{G}} R^{\mathbf{\Lambda}}_{\lambda\mathbf{F}}}{2}, \qquad R^{\mathbf{\Lambda}}_{\lambda\#} := 2J^{\mathbf{\Lambda}}_{\lambda\#} - \mathbf{Id}$$

Proposition

We have

$$\mathbf{x} \in \mathrm{Fix}(T) \iff \exists \mathbf{z} \in J^{\mathbf{\Lambda}}_{\lambda\mathbf{F}}(\mathbf{x}) \cap \mathrm{zer}(\mathbf{F} + \mathbf{G})$$

In particular, if $J^{\mathbf{\Lambda}}_{\lambda\mathbf{F}}$ is single-valued, then

$$J^{\mathbf{\Lambda}}_{\lambda\mathbf{F}}(\mathrm{Fix}(T)) = \mathrm{zer}(\mathbf{F} + \mathbf{G}).$$

Quick summary

- Any multioperator inclusion problem can be equivalently reformulated as a two-operator problem

$$\text{Find } \mathbf{x} = (x_1, \ldots, x_{m-1}) \in \mathcal{H}^m \text{ such that } 0 \in \mathbf{F}(\mathbf{x}) + \mathbf{G}(\mathbf{x}), \tag{P}$$

- Douglas-Rachford algorithm for (P):

$$x^{k+1} \in T(x^k) \tag{DR}$$

- We have the relation:

$$\mathbf{x} \in \operatorname{Fix}(T) \iff \exists \mathbf{z} \in J_{\lambda \mathbf{F}}^{\mathbf{\Lambda}}(\mathbf{x}) \cap \operatorname{zer}(\mathbf{F} + \mathbf{G})$$

What can we say about the convergence of (DR)?

Nonmonotone cases we consider

I. Generalized monotone operators

II. $A_i = \partial f_i$ with nonconvex f_i's.

III. (If time permits) Comonotone operators

I. Generalized monotone operators

Assumption: A_i is maximal σ_i-monotone for all i, and $\operatorname{zer}(A_1 + \cdots + A_m) \neq \emptyset$.

Conjecture: We get convergence if $\sigma_1 + \cdots + \sigma_m > 0$.

$$\mathbf{F}(\mathbf{x}) \coloneqq A_1(x_1) \times \cdots \times A_{m-1}(x_{m-1}),$$

$$\mathbf{G}(\mathbf{x}) \coloneqq \mathbf{K}(\mathbf{x}) + N_{\mathbf{D}_{m-1}}(\mathbf{x})$$

$$\mathbf{K}(\mathbf{x}) \coloneqq \begin{cases} \{(\lambda_1 v, \ldots, \lambda_{m-1} v) : v \in A_m(x_1)\} & \text{if } \mathbf{x} \in \mathbf{D}_{m-1} \\ \emptyset & \text{otherwise.} \end{cases}$$

Proposition

- $\mathbf{F}$ is maximal $\sigma_{\mathbf{F}}$-monotone with

$$\sigma_{\mathbf{F}} \coloneqq \min_{i=1,\ldots,m-1} \sigma_i.$$

- $\mathbf{G}$ is maximal $\sigma_{\mathbf{G}}$-monotone with

$$\sigma_{\mathbf{G}} \coloneqq \frac{\sigma_m}{m-1}.$$

A convergence result

Direct application of Dao & Phan's theorem leads to the following result.

Theorem

Suppose that $\begin{cases} \sigma_{\mathbf{F}} + \sigma_{\mathbf{G}} > 0, \\ 1 + \gamma \frac{\sigma_{\mathbf{F}}\sigma_{\mathbf{G}}}{\sigma_{\mathbf{F}}+\sigma_{\mathbf{G}}} > \frac{\kappa}{2}, \end{cases}$ or equivalently,

$$\begin{cases} \sigma_i + \frac{\sigma_m}{m-1} > 0 & (\forall i = 1, \dots, m-1) \\ 1 + \gamma \dfrac{\left(\min\limits_{i \leq m-1} \sigma_i\right) \cdot \sigma_m}{(m-1) \min\limits_{i \leq m-1} \sigma_i + \sigma_m} > \frac{\kappa}{2}. \end{cases}$$

Then

(i) $\exists \bar{\mathbf{x}} \in \operatorname{Fix}(T_{\mathbf{F},\mathbf{G}})$ s.t. $\mathbf{x}^k \rightharpoonup \bar{\mathbf{x}}$, with $\bar{\mathbf{z}} := J_{\gamma \mathbf{F}}(\bar{\mathbf{x}}) \in \operatorname{zer}(\mathbf{F} + \mathbf{G})$.

(ii) $\left\|\mathbf{x}^k - \mathbf{x}^{k+1}\right\| = o(1/\sqrt{k})$

(iii) $\mathbf{z}^k \to \bar{\mathbf{z}}$ and $\operatorname{zer}(\mathbf{F} + \mathbf{G}) = \{\bar{\mathbf{z}}\}$.

Remarks

- The theorem reduces to Dao & Phan's convergence result when $m = 2$.
- Stricter than our initial conjecture for $m \geq 3$.

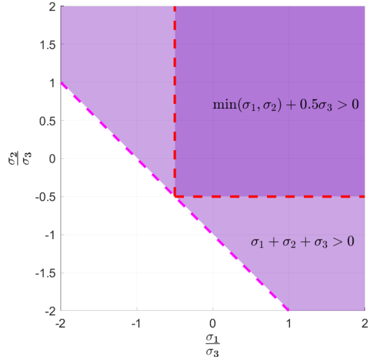

Figure: $\boxed{\min\{\sigma_1, \sigma_2\} + \frac{\sigma_3}{2} > 0}$ vs. $\boxed{\sigma_1 + \sigma_2 + \sigma_3 > 0}$ when $\sigma_3 > 0$.
Conjecture

[1] Dark region represents the intersection.

Main tool to improve convergence result

Let $\mathbf{R} := \mathrm{Id} - T$ and $U(\mathbf{x}) := J_{\lambda A_m}\left(\sum_{i=1}^{m-1} \lambda_i R_{\frac{\lambda}{\lambda_i} A_i}(x_i)\right)$.

Lemma

$$
\begin{aligned}
&\| T(\mathbf{x}) - T(\mathbf{y})\|_{\mathbf{\Lambda}}^2 \\
&\leq \|\mathbf{x} - \mathbf{y}\|_{\mathbf{\Lambda}}^2 - \frac{2}{\kappa} \sum_{i=1}^{m-1} \lambda_i \kappa_i \|R_i(\mathbf{x}) - R_i(\mathbf{y})\|^2 \\
&- 2\kappa\lambda \sum_{i\in\mathcal{I}} \theta_i \left\| \sigma_i \left(J_{\frac{\lambda}{\lambda_i} A_i}(x_i) - J_{\frac{\lambda}{\lambda_i} A_i}(y_i) \right) + \sigma_m \delta_i (U(\mathbf{x}) - U(\mathbf{y})) \right\|^2
\end{aligned}
$$

where

$$
\kappa_i := \begin{cases} 1 + \frac{\lambda}{\lambda_i} \frac{\sigma_i \sigma_m \delta_i}{\sigma_i + \sigma_m \delta_i} - \frac{\kappa}{2} & \text{if } i \in \mathcal{I} \\ 1 - \frac{\kappa}{2} & \text{if } i \notin \mathcal{I} \end{cases}, \qquad \theta_i := \frac{1}{\sigma_i + \sigma_m \delta_i}, \qquad \sum_{i\in\mathcal{I}} \delta_i = 1,\ \delta_i \in \mathbb{R},
$$

$$
\mathcal{I} := \{ i \in \{1, \ldots, m-1\} : \sigma_i \neq 0 \}, \qquad \mathcal{I}^- := \{ i \in \mathcal{I} : \sigma_i < 0 \}, \qquad \mathcal{I}^+ := \mathcal{I} \setminus \mathcal{I}^-.
$$

Proof sketch (1/2)

- Recall $T = \frac{(2-\kappa)}{2}\,\mathbf{Id} + \frac{\kappa}{2} R^{\mathbf{\Lambda}}_{\lambda \mathbf{G}} R^{\mathbf{\Lambda}}_{\lambda \mathbf{F}}$.
- Apply the "cute identity"[2] to expand $\|T(\mathbf{x}) - T(\mathbf{y})\|^2_{\mathbf{\Lambda}}$:

$$\|T(\mathbf{x}) - T(\mathbf{y})\|^2_{\mathbf{\Lambda}} = \frac{2-\kappa}{2}\|\mathbf{x}-\mathbf{y}\|^2_{\mathbf{\Lambda}} + \frac{\kappa}{2}\left\|R^{\mathbf{\Lambda}}_{\lambda \mathbf{G}} R^{\mathbf{\Lambda}}_{\lambda \mathbf{F}}(\mathbf{x}) - R^{\mathbf{\Lambda}}_{\lambda \mathbf{G}} R^{\mathbf{\Lambda}}_{\lambda \mathbf{F}}(\mathbf{y})\right\|^2_{\mathbf{\Lambda}} - \frac{2-\kappa}{\kappa}\sum_{i=1}^{m-1}\lambda_i \|R_i(\mathbf{x}) - R_i(\mathbf{y})\|^2$$

(Note: $\mathbf{Id} - R^{\mathbf{\Lambda}}_{\lambda \mathbf{G}} R^{\mathbf{\Lambda}}_{\lambda \mathbf{F}} = \frac{2}{\kappa}\mathbf{R}$)

- Meanwhile,

$$\left\|R^{\mathbf{\Lambda}}_{\lambda \mathbf{G}} R^{\mathbf{\Lambda}}_{\lambda \mathbf{F}}(\mathbf{x}) - R^{\mathbf{\Lambda}}_{\lambda \mathbf{G}} R^{\mathbf{\Lambda}}_{\lambda \mathbf{F}}(\mathbf{y})\right\|^2_{\mathbf{\Lambda}} \leq \|\mathbf{x}-\mathbf{y}\|^2_{\mathbf{\Lambda}} - 4\lambda \sum_{i=1}^{m-1}\sigma_i \left\|J_{\frac{\lambda}{\lambda_i}A_i}(x_i) - J_{\frac{\lambda}{\lambda_i}A_i}(y_i)\right\|^2 - 4\lambda\sigma_m \|U(\mathbf{x}) - U(\mathbf{y})\|^2.$$

[2] Named after Heinz's playful terminology during his invited talk at this workshop.

Proof sketch (2/2)

- Combining, we get

$$\begin{aligned}&\|T(\mathbf{x}) - T(\mathbf{y})\|_{\boldsymbol{\Lambda}}^2 \\ &\quad \le \|\mathbf{x} - \mathbf{y}\|_{\boldsymbol{\Lambda}}^2 - \frac{2-\kappa}{\kappa}\sum_{i=1}^{m-1}\lambda_i\|R_i(\mathbf{x}) - R_i(\mathbf{y})\|^2 \\ &\quad\quad - 2\lambda\kappa\underbrace{\left(\sum_{i\in\mathcal{I}}\sigma_i\left\|J_{\frac{\lambda}{\lambda_i}A_i}(x_i) - J_{\frac{\lambda}{\lambda_i}A_i}(y_i)\right\|^2 + \overbrace{\sigma_m\|U(\mathbf{x}) - U(\mathbf{y})\|^2}^{(\clubsuit)}\right)}_{(\star)}\end{aligned}$$

- Write $(\clubsuit) = \sum_{i\in\mathcal{I}}\delta_i\sigma_m\|U(\mathbf{x}) - U(\mathbf{y})\|^2$ where $\delta_i \in \mathbb{R}$ with $\sum_{i\in\mathcal{I}}\delta_i = 1$.
- Apply to $(\star)$ the "cuter(?) identity"

$$\alpha\|x\|^2 + \beta\|y\|^2 = \tfrac{\alpha\beta}{\alpha+\beta}\|x-y\|^2 + \tfrac{1}{\alpha+\beta}\|\alpha x + \beta y\|^2, \quad \alpha+\beta \neq 0.$$

Main tool to improve convergence result

Let $\mathbf{R} := \mathrm{Id} - T$ and $U(\mathbf{x}) := J_{\lambda A_m}\left(\sum_{i=1}^{m-1} \lambda_i R_{\frac{\lambda}{\lambda_i}A_i}(x_i)\right)$.

Lemma

$$\begin{aligned}
&\|T(\mathbf{x}) - T(\mathbf{y})\|_{\mathbf{\Lambda}}^2 \\
&\leq \|\mathbf{x} - \mathbf{y}\|_{\mathbf{\Lambda}}^2 - \frac{2}{\kappa}\sum_{i=1}^{m-1} \lambda_i \kappa_i \|R_i(\mathbf{x}) - R_i(\mathbf{y})\|^2 \\
&- 2\kappa\lambda \sum_{i\in\mathcal{I}} \theta_i \left\| \sigma_i \left(J_{\frac{\lambda}{\lambda_i}A_i}(x_i) - J_{\frac{\lambda}{\lambda_i}A_i}(y_i)\right) + \sigma_m \delta_i (U(\mathbf{x}) - U(\mathbf{y}))\right\|^2
\end{aligned}$$

where

$$\kappa_i := \begin{cases} 1 + \frac{\lambda}{\lambda_i}\frac{\sigma_i\sigma_m\delta_i}{\sigma_i+\sigma_m\delta_i} - \frac{\kappa}{2} & \text{if } i \in \mathcal{I} \\ 1 - \frac{\kappa}{2} & \text{if } i \notin \mathcal{I} \end{cases}, \qquad \theta_i := \frac{1}{\sigma_i + \sigma_m\delta_i}, \qquad \sum_{i\in\mathcal{I}} \delta_i = 1,\ \delta_i \in \mathbb{R},$$

$$\mathcal{I} := \{i \in \{1,\dots,m-1\} : \sigma_i \neq 0\}, \qquad \mathcal{I}^- := \{i \in \mathcal{I} : \sigma_i < 0\}, \qquad \mathcal{I}^+ := \mathcal{I} \setminus \mathcal{I}^-.$$

Abstract convergence result

Theorem

Suppose $\mathcal{I} \neq \emptyset$ and the following holds:

(a) $\exists(\delta_i)_{i\in\mathcal{I}}$ with $\delta_i \in \mathbb{R}$ s.t.

$$\begin{cases} \sigma_i + \sigma_m\delta_i > 0 \quad (\forall i \in \mathcal{I}) \\ \sum_{i\in\mathcal{I}} \delta_i = 1 \end{cases}$$

(b) $\lambda \in (0, +\infty)$ is chosen such that

$$1 + \frac{\lambda}{\lambda_i}\frac{\sigma_i\sigma_m\delta_i}{\sigma_i + \sigma_m\delta_i} > \frac{\mu}{2} \quad (\forall i \in \mathcal{I}).$$

Then

(i) $\exists \bar{\mathbf{x}} \in \mathrm{Fix}(T_{\mathbf{F},\mathbf{G}})$ s.t. $\mathbf{x}^k \rightharpoonup \bar{\mathbf{x}}$, with $\bar{\mathbf{z}} := J_{\lambda\mathbf{F}}^{\mathbf{\Lambda}}(\bar{\mathbf{x}}) \in \mathrm{zer}(\mathbf{F}+\mathbf{G})$.

(ii) $\left\|\mathbf{x}^k - \mathbf{x}^{k+1}\right\| = o(1/\sqrt{k})$

(iii) $\mathbf{z}^k \to \bar{\mathbf{z}}$ and $\mathrm{zer}(\mathbf{F}+\mathbf{G}) = \{\bar{\mathbf{z}}\}$.

When can we guarantee (a)?

Proposition

Suppose $\mathcal{I} \neq \emptyset$ and $\sigma_m \neq 0$. Denote

$$X := \prod_{i \in \mathcal{I}} X_i \quad \text{where} \quad X_i := \{\delta_i \in \mathbb{R} : \sigma_i + \sigma_m \delta_i \geq 0\}$$

$$S = \{\delta = (\delta_i)_{i \in \mathcal{I}} : \sum_{i \in \mathcal{I}} \delta_i = 1\}$$

Then

(i) $X \cap S$ is compact;

Moreover, if $\sum_{i=1}^{m} \sigma_i > 0$, then the following hold:

(ii) $\text{int}(X) \cap S \neq \emptyset$;

(iii) $N_{X \cap S}(\delta) = N_X(\delta) + N_S(\delta)$ for any $\delta \in X \cap S$;

Is there an optimal stepsize?

Proposition

Consider

$$\bar{\lambda}^* := \max_{\delta \in \mathbb{R}^{|\mathcal{I}|}, \bar{\lambda} \geq 0} \quad \bar{\lambda}$$
$$\text{s.t.} \quad 1 + \frac{\bar{\lambda}}{\lambda_i} \frac{\sigma_i \sigma_m \delta_i}{\sigma_i + \sigma_m \delta_i} - \frac{\mu}{2} \geq 0 \quad i \in \mathcal{I}, \qquad \text{(MaxStep)}$$
$$\delta = (\delta_i)_{i \in \mathcal{I}} \in X \cap S.$$

If $\sum_{i=1}^{m} \sigma_i > 0$ and $\mathcal{I} \neq \emptyset$, then the following holds:

(i) If either $\mathcal{I}^- \neq \emptyset$ and $\sigma_m \neq 0$, or $\mathcal{I}^- = \emptyset$ and $\sigma_m < 0$, then (MaxStep) has a solution, say $(\delta^*, \bar{\lambda}^*) \in S \times \mathbb{R}_+$, which satisfies

$$-\frac{\lambda_i(\sigma_i + \sigma_m \delta_i^*)}{\sigma_i \sigma_m \delta_i^*} = -\frac{\lambda_j(\sigma_j + \sigma_m \delta_j^*)}{\sigma_j \sigma_m \delta_j^*} > 0 \quad \forall i, j \in \mathcal{I}, \tag{1}$$

and $\bar{\lambda}^* = -\left(1 - \frac{\kappa}{2}\right)\left(\frac{\lambda_i(\sigma_i + \sigma_m \delta_i^*)}{\sigma_i \sigma_m \delta_i^*}\right)$.

(ii) If $\mathcal{I}^- = \emptyset$ and $\sigma_m \geq 0$, then $\bar{\lambda}^* = +\infty$.

Main convergence result

Theorem

Suppose $\sum_{i=1}^{m} \sigma_i > 0$, $\mathcal{I} \neq \emptyset$, and let $\lambda \in (0, \bar{\lambda}^*)$ (given in previous proposition). Then

(i) $\exists \bar{\mathbf{x}} \in \mathrm{Fix}(T_{\mathbf{F},\mathbf{G}})$ s.t. $\mathbf{x}^k \rightharpoonup \bar{\mathbf{x}}$, with $\bar{\mathbf{z}} := J_{\lambda \mathbf{F}}^{\mathbf{\Lambda}}(\bar{\mathbf{x}}) \in \mathrm{zer}(\mathbf{F} + \mathbf{G})$.

(ii) $\left\|\mathbf{x}^k - \mathbf{x}^{k+1}\right\| = o(1/\sqrt{k})$

(iii) $\mathbf{z}^k \to \bar{\mathbf{z}}$ and $\mathrm{zer}(\mathbf{F} + \mathbf{G}) = \{\bar{\mathbf{z}}\}$.

II. Optimization

Nonmonotone cases we consider

I. Generalized monotone operators $\longrightarrow$ DONE!

II. $A_i = \partial f_i$ $(i = 1, \ldots, m-1)$ with nonconvex f_i's

III. (If time permits) Comonotone operators

Nonconvex optimization

$$\Omega := \operatorname*{arg\,min}_{x \in \mathcal{H}} \; F(x) := f_1(x) + \cdots + f_m(x), \qquad \text{(Opt)}$$

Douglas-Rachford for sum-of-m-functions optimization

$$\begin{aligned}
z_i^k &\in \operatorname{prox}_{\frac{\lambda}{\lambda_i} f_i}(x_i^k), && (i = 1, \ldots, m-1) \\
y^k &\in \operatorname{prox}_{\lambda f_m}\left(\sum_{i=1}^{m-1} \lambda_i (2z_i^k - x_i^k)\right) \\
x_i^{k+1} &= x_i^k + \kappa(y^k - z_i^k) && (i = 1, \ldots, m-1).
\end{aligned}$$

Note: This is only an instance of DR. Indeed, we can only guarantee

$$\operatorname{prox}_{\gamma f}(\cdot) \subseteq J_{\gamma \partial f}(\cdot).$$

Assumptions

Recall of $m = 2$ case:

	f	g	Remarks
Guo, Han, & Yuan (2017)	α-convex	β-convex	$\alpha + \beta > 0$
Themelis & Patrinos (2020)	L-smooth	proper lsc	*

Table: Existing works

- Easy case: f_i is proper, closed and σ_i-convex, and $\sigma_1 + \cdots + \sigma_m > 0$
  - Just use the previous theorem!
- We consider another setting:

Assumption: The following holds:

(1) f_i is L_i-smooth, $i = 1, \ldots, m - 1$.

(2) f_m is proper and closed.

(3) F is coercive.

(4) $\mathcal{H}$ is finite-dimensional.

Main result

Note: Assumption (1) $\implies f_i$ is σ_i-convex with $\sigma_i \in [-L_i, 0]$.

Theorem

Denote

$$\bar{\gamma}_i := \begin{cases} \frac{1}{L_i} & \text{if } -2\sigma_i < (2-\kappa)L_i \\ -\frac{1}{\sigma_i}\left(1-\frac{\kappa}{2}\right) & \text{otherwise} \end{cases},$$

and let $\frac{\lambda}{\lambda_i} \in (0, \bar{\gamma}_i)$. Then

(i) $\{(x_1^k, \dots, x_{m-1}^k, z_1^k, \dots, z_{m-1}^k, y^k)\}$ is bounded;

(ii) $z_i^*, y^* \in \operatorname{zer}\left(\sum_{i=1}^m \partial f_i\right)$ if z_i^* and y^* are accu. points of $\{z_i^k\}$ and $\{y^k\}$.

Proof techniques make use of a merit function.

Summary and future works

- We proposed a DR algorithm suitable for arbitrary nonmonotone inclusion problem
- We established its convergence under suitable conditions.
- Future works
    - Linear rates of convergence
    - Extensions to comonotone operators (To be released soon!)
    - Extensions to other splitting methods (e.g., Malitsky-Tam)

Thank you for listening!

Main reference: Alcantara and Takeda, Douglas-Rachford algorithm for nonmonotone multioperator inclusion problems, arXiv:2501.02752 (2025). Preprint

References

- Heinz H. Bauschke and Patrick L. Combettes. Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, Berlin, 2nd edition, 2017.
- Rubén Campoy. A product space reformulation with reduced dimension for splitting algorithms. Computational Optimization and Applications, 83:319–348, 2022.
- Minh N. Dao and Hung M. Phan. Adaptive Douglas–Rachford splitting algorithm for the sum of two operators. SIAM Journal on Optimization, 29(4):2697–2724, 2019.
- Kai Guo, Dongdong Han, and Xiao Yuan. Convergence analysis of Douglas–Rachford splitting method for strongly + weakly convex programming. SIAM Journal on Numerical Analysis, 55 (4):1549–1577, 2017.
- Alexander Y. Kruger. Generalized differentials of nonsmooth functions, and necessary conditions for an extremum. Siberian Mathematical Journal, 26(3):370–379, 1985.
- Andreas Themelis and Panagiotis Patrinos. Douglas–Rachford splitting and ADMM for nonconvex optimization: Tight convergence results. SIAM Journal on Optimization, 30(1):149–181, 2020.