

Finding the Forward-Douglas-Rachford-Forward Method

Ernest K. Ryu · Bằng Công Vũ

Received: date / Accepted: date

Abstract We consider the monotone inclusion problem with a sum of 3 operators, in which 2 are monotone and 1 is monotone-Lipschitz. The classical Douglas–Rachford and Forward-backward-forward methods respectively solve the monotone inclusion problem with a sum of 2 monotone operators and a sum of 1 monotone and 1 monotone-Lipschitz operators. We first present a method that naturally combines Douglas–Rachford and Forward-backward-forward and show that it solves the 3 operator problem under further assumptions, but fails in general. We then present a method that naturally combines Douglas–Rachford and forward-reflected-backward, a recently proposed alternative to Forward-backward-forward by Malitsky and Tam [arXiv:1808.04162, 2018]. We show that this second method solves the 3 operator problem generally, without further assumptions.

Keywords Douglas–Rachford · Forward-backward-forward · Forward-reflected-backward · Monotone inclusion

Mathematics Subject Classification (2000) 47H05 · 47H09 · 90C25

1 Introduction

We consider the monotone inclusion problem of finding a zero of the sum of 2 maximal monotone and 1 monotone-Lipschitz operators. The classical

Communicated by Jalal Fadili.

Ernest K. Ryu, Corresponding author
University of California, Los Angeles
California, United States
eryu@math.ucla.edu

Bằng Công Vũ,
Department of Mathematics, Vietnam National University
Hanoi, Vietnam
bangcvn@gmail.com

Douglas–Rachford (DR) splitting by Lions and Mercier [1] solves the problem with a sum of 2 maximal monotone operators. The classical forward-backward-forward (FBF) splitting by Tseng [2] solves the problem with a sum of 1 maximal monotone and 1 monotone-Lipschitz operators. We consider the generalization of the setups of DR and FBF.

Recently, there has been much work developing splitting methods combining and unifying classical ones. Another classical method is forward-backward (FB) splitting [3, 4], which solves the problem with a sum of 1 monotone and 1 cocoercive operators. The effort of combining DR and FB was started by Raguet, Fadili, and Peyré [5, 6], extended by Briceño-Arias [7], and completed by Davis and Yin [8] as they proved convergence for the sum of 2 monotone and 1 cocoercive operators. FB and FBF were combined by Briceño-Arias and Davis [9] as they proved convergence for 1 monotone, 1 cocoercive, and 1 monotone-Lipschitz operators. These combined splitting methods can efficiently solve monotone inclusion problems with more complex structure.

On the other hand, DR and FBF have not been fully combined, to the best of our knowledge. Banert’s relaxed forward backward (in the thesis [10]) and Briceño-Arias’s forward–partial inverse–forward [11] combine DR and FBF in the setup where one operator is a normal cone operator with respect to a closed subspace. However, neither method applies to the general setup with 2 maximal monotone and 1 monotone-Lipschitz operators.

In this work, we first present a method that naturally combines and unifies DR and FBF. We prove convergence under further assumptions, and we prove, through a counterexample, that convergence cannot be established in full generality. We then propose a second method that naturally combines and unifies DR and forward-reflected-backward (FRB), a recently proposed alternative to FBF by Malitsky and Tam [12]. We show that this combination of DR and FRB does converge in full generality.

The paper is organized as follows. Section 2 states the problem formally. Section 3 reviews preliminary information and sets up the notation. Section 4 presents our first proposed method combining DR and FBF, proves convergence under certain further assumptions, and proves divergence in the fully general case. Section 5 presents our second proposed method combining DR and FRB and proves convergence in the fully general case. Section 6 compares our presented method with other similar and relevant methods.

2 Problem Statement, Contribution, and Prior Work

Consider the monotone inclusion problem

$$\text{find } x \in \mathcal{H} \quad \text{such that} \quad 0 \in Ax + Bx + Cx, \quad (1)$$

where \mathcal{H} is a real Hilbert space. Throughout, we assume for some $\mu \in]0, \infty[$:

- (A1) $A: \mathcal{H} \rightrightarrows \mathcal{H}$ and $B: \mathcal{H} \rightrightarrows \mathcal{H}$ are maximal monotone.
- (A2) $C: \mathcal{H} \rightarrow \mathcal{H}$ is monotone and μ -Lipschitz continuous.
- (A3) $\text{zer}(A + B + C)$ is not empty.

Let $J_{\gamma A}$, $J_{\gamma B}$, and $J_{\gamma C}$, respectively, denote the resolvents with respect to A , B , and C with parameters γ . We informally assume $J_{\gamma A}(x)$, $J_{\gamma B}(x)$, and $C(x)$ can be evaluated efficiently for any input $x \in \mathcal{H}$. However, $J_{\gamma C}(x)$ may be difficult to evaluate. Therefore, we restrict our attention to methods that activate C through direct evaluations, rather than through $J_{\gamma C}$. The monotone-Lipschitz operator C of Problem (1) arises as skew linear operators primal-dual optimization [13, 14] and saddle point problems [15].

When $C = 0$, we can use the classical Douglas–Rachford (DR) splitting presented by Lions and Mercier [1]:

$$\begin{cases} x_{n+1} = J_{\gamma B}(z_n) \\ y_{n+1} = J_{\gamma A}(2x_{n+1} - z_n) \\ z_{n+1} = z_n + y_{n+1} - x_{n+1}, \end{cases}$$

where the step size parameter satisfies $\gamma \in]0, +\infty[$. When $B = 0$, we can use the classical forward-backward-forward (FBF) splitting by Tseng [2]:

$$\begin{cases} y_{n+1} = J_{\gamma A}(x_n - \gamma Cx_n) \\ x_{n+1} = y_{n+1} - \gamma(Cy_{n+1} - Cx_n), \end{cases}$$

where the step size parameter satisfies $\gamma \in]0, 1/\mu[$. Recently, Malitsky and Tam [12] have proposed forward-reflected-backward (FRB) splitting, another method for the case $B = 0$:

$$\begin{cases} x_{n+1} = J_{\gamma A}(x_n - \gamma(2Cx_n - Cx_{n-1})), \end{cases}$$

where the step size parameter satisfies $\gamma \in]0, 1/(2\mu)[$.

The contribution of this work is the study of splitting methods combining DR with other methods to incorporate an additional monotone-Lipschitz operator. We characterize to what extent DR+FBF works and to what extent it fails. We then demonstrate that DR+FRB is a successful combination.

Several other 3-operator splitting methods have been presented in recent years. Combettes and Pesquet’s PPXA [13], Boţ–Hendrich [16], Latafat and Patrinos’s AFBA [17], and Ryu’s 3-operator resolvent-splitting [18] solve the problem with 3 or more monotone operators by activating the operators through their individual resolvents. Condat–Vũ [19, 20], FDR [5, 6, 7, 8], and Yan’s PD3O [21] solve the problem with 2 monotone and 1 cocoercive operators by activating the 2 monotone operators through their resolvents and the cocoercive operator through forward evaluations. FBHF [9] solves the problem with 1 monotone, 1 cocoercive, and 1 monotone-Lipschitz operators by activating the

monotone operator through its resolvent and the cocoercive and monotone-Lipschitz operators through forward evaluations. These methods do not apply to our setup since we have 2 monotone operators, which we activate through their resolvents, and 1 monotone-Lipschitz operator, which we activate through forward evaluations.

The primal-dual method by Combettes and Pesquet [13] and the instances of projective splitting by Johnstone and Eckstein [22, 23] are existing methods that do solve Problem (1). However, these methods do not reduce to DR. We compare the form of these methods in Section 6.

While this paper was under review, there have been exciting developments on splitting methods based on cutting planes (separating hyperplanes): Warped proximal iterations by Bui and Combettes [24] and NOFOB by Giselsson [25] are general frameworks that can solve Problem (1). The methods that arise from these frameworks are different from the methods we present.

3 Preliminaries

In this section, we quickly review known results and set up the notation. The notation and results we discuss are standard, and interested readers can find further information in [26, 27].

Write \mathcal{H} for a real Hilbert space and, respectively, write $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ for its associated scalar product and norm. Write $\text{Id}: \mathcal{H} \rightarrow \mathcal{H}$ for the identity operator. Write $A: \mathcal{H} \rightrightarrows \mathcal{H}$ to denote that A is a set-valued operator. For simplicity, we also write $Ax := A(x)$. When A maps a point to a singleton, we also write $Ax = y$ instead of $Ax = \{y\}$. Write $\text{dom}(A) := \{x \in \mathcal{H} : Ax \neq \emptyset\}$ for the domain of A and $\text{ran}(A) := \{u \in \mathcal{H} : (\exists x \in \mathcal{H}) u \in Ax\}$ for the range of A . Write $\text{gra}(A) := \{(x, u) \in \mathcal{H} \times \mathcal{H} : u \in Ax\}$ for the graph of A . The inverse of A is the set-valued operator defined by $A^{-1}: u \mapsto \{x : u \in Ax\}$. The zero set of A is $\text{zer}(A) := A^{-1}0$. We say that A is monotone if

$$(\forall (x, u), (y, v) \in \text{gra } A) \quad \langle x - y, u - v \rangle \geq 0,$$

and it is maximally monotone if there exists no monotone operator B such that $\text{gra}(B)$ properly contains $\text{gra}(A)$. The resolvent of A is $J_A := (\text{Id} + A)^{-1}$. When A is maximal monotone, J_A is single-valued and $\text{dom } J_A = \mathcal{H}$. A single-valued operator $B: \mathcal{H} \rightarrow \mathcal{H}$ is κ -cocoercive for $\kappa \in]0, \infty[$ if

$$(\forall x, y \in \mathcal{H}) \quad \langle x - y, Bx - By \rangle \geq \kappa \|Bx - By\|^2.$$

A single-valued operator $C: \mathcal{H} \rightarrow \mathcal{H}$ is μ -Lipschitz for $\mu \in]0, \infty[$ if

$$(\forall x, y \in \mathcal{H}) \quad \|Cx - Cy\| \leq \mu \|x - y\|.$$

A single-valued operator $R: \mathcal{H} \rightarrow \mathcal{H}$ is nonexpansive if

$$(\forall x, y \in \mathcal{H}) \quad \|Rx - Ry\| \leq \|x - y\|,$$

i.e., if R is 1-Lipschitz. Let $\theta \in [0, 1]$. A single-valued operator $T: \mathcal{H} \rightarrow \mathcal{H}$ is θ -averaged if $T = (1 - \theta)I + \theta R$ for some nonexpansive operator R . Define the normal cone operator with respect to a nonempty closed convex set $C \subseteq \mathcal{H}$ as

$$N_C(x) = \begin{cases} \emptyset, & \text{if } x \notin C \\ \{y \in \mathcal{H} : \langle y, z - x \rangle \leq 0 \ \forall z \in C\}, & \text{if } x \in C. \end{cases}$$

The Cauchy–Schwartz inequality states $\langle u, v \rangle \leq \|u\|\|v\|$ for any $u, v \in \mathcal{H}$. The Young’s inequality states

$$\langle u, v \rangle \leq \frac{\eta}{2}\|u\|^2 + \frac{1}{2\eta}\|v\|^2,$$

for any $u, v \in \mathcal{H}$ and $\eta > 0$.

Lemma 3.1 *If $C: \mathcal{H} \rightarrow \mathcal{H}$ is μ -Lipschitz continuous, then $\text{Id} - \gamma C$ is one-to-one for $\gamma \in]0, 1/\mu[$.*

Proof Although this result follows immediately from the machinery of scaled relative graphs [28], we provide a proof based on first principles. Let $x, y \in \mathcal{H}$. Then $\|Cx - Cy\| \leq \mu\|x - y\|$ and

$$\|(\text{Id} - \gamma C)x - (\text{Id} - \gamma C)y\| \geq \|x - y\| - \gamma\|Cx - Cy\| \geq (1 - \gamma\mu)\|x - y\|,$$

by Cauchy–Schwartz and μ -Lipschitz continuity. Thus $(\text{Id} - \gamma C)x = (\text{Id} - \gamma C)y$ if and only if $x = y$. \square

A classical result states that J_B is $(1/2)$ -averaged if B is maximal monotone [26, Proposition 23.8]. The following lemma states that J_B is furthermore θ -averaged with $\theta < 1/2$ if B is cocoercive.

Lemma 3.2 *Let $\gamma, \kappa \in]0, \infty[$. If $B: \mathcal{H} \rightarrow \mathcal{H}$ is κ -cocoercive, then $J_{\gamma B}$ is $\frac{1}{2(1+\kappa/\gamma)}$ -averaged and*

$$\|J_{\gamma B}x - J_{\gamma B}y\|^2 \leq \|x - y\|^2 - \left(1 + \frac{2\kappa}{\gamma}\right) \|(\text{Id} - J_{\gamma B})x - (\text{Id} - J_{\gamma B})y\|^2.$$

for any $x, y \in \mathcal{H}$.

Proof Although this result follows immediately from the machinery of scaled relative graphs [28], we provide a proof based on first principles. Let $u = J_{\gamma B}x$ and $v = J_{\gamma B}y$, i.e., $\gamma^{-1}(x - u) = Bu$ and $\gamma^{-1}(y - v) = Bv$. Since B is κ -cocoercive, we have

$$\kappa\|\gamma^{-1}(x - u) - \gamma^{-1}(y - v)\|^2 \leq \langle u - v, \gamma^{-1}(x - u) - \gamma^{-1}(y - v) \rangle.$$

This implies

$$\begin{aligned} (\kappa/\gamma)\|x - u - y + v\|^2 &\leq \langle u - v, x - u - y + v \rangle \\ &= \langle u - v, x - y \rangle - \|u - v\|^2 \\ &= -\frac{1}{2}\|u - v\|^2 + \frac{1}{2}\|x - y\|^2 - \frac{1}{2}\|x - u - y + v\|^2, \end{aligned}$$

which proves the stated inequality. Finally, this inequality is equivalent to $\frac{1}{2(1+\kappa/\gamma)}$ -averagedness of $J_{\gamma B}$ by [26, Proposition 4.35]. \square

4 FBF+DR: Convergent with Further Assumptions

To solve Problem (1), we propose the following iteration

$$(FDRF) \quad \begin{cases} x_{n+1} = J_{\gamma B} z_n \\ y_{n+1} = J_{\gamma A}(2x_{n+1} - z_n - \gamma C x_{n+1}) \\ z_{n+1} = z_n + y_{n+1} - x_{n+1} - \gamma(C y_{n+1} - C x_{n+1}) \end{cases}$$

for $n = 0, 1, \dots$ where $z_0 \in \mathcal{H}$ is a starting point and $\gamma > 0$. We call this method forward-Douglas–Rachford-forward (FDRF) splitting as it combines Tseng’s FBF [2] and Douglas–Rachford [1]; FDRF reduces to FBF when $B = 0$ and to DR when $C = 0$.

We can view FDRF as a fixed-point iteration $z_{n+1} = T z_n$ with

$$T := (\text{Id} - \gamma C) J_{\gamma A}(2J_{\gamma B} - \text{Id} - \gamma C J_{\gamma B}) + \text{Id} - (\text{Id} - \gamma C) J_{\gamma B}.$$

The following result states that T is a fixed-point encoding for Problem (1).

Lemma 4.1 *Assume (A1) and (A2). If $\gamma \in]0, 1/\mu[$, then*

$$\text{zer}(A + B + C) = J_{\gamma B}(\text{Fix}(T)),$$

where $\text{Fix}(T) := \{x \in \mathcal{H} : Tx = x\}$.

Proof Let $x \in \text{zer}(A + B + C)$. Then, there exists $u \in Ax$ and $v \in Bx$ such that $0 = u + v + Cx$. It follows from $v \in Bx$ that $x = J_{\gamma B}z$ where $z = x + \gamma v$. We have $2J_{\gamma B}z - z - \gamma C J_{\gamma B}z = 2x - z - \gamma Cx = x + \gamma u \in (\text{Id} + \gamma A)x$ and $x = J_{\gamma A}(2J_{\gamma B}z - z - \gamma C J_{\gamma B}z)$. Therefore,

$$(\text{Id} - \gamma C) J_{\gamma B}z = (\text{Id} - \gamma C) J_{\gamma A}(2J_{\gamma B}z - z - \gamma C J_{\gamma B}z),$$

which shows that $Tz = z$ and $\text{zer}(A + B + C) \subset J_{\gamma B}(\text{Fix}(T))$. Now, let $z \in \text{Fix}(T)$. By Lemma 3.1, we have $J_{\gamma B}z = J_{\gamma A}(2J_{\gamma B}z - z - \gamma C J_{\gamma B}z)$. Set $x = J_{\gamma B}z$. Then, $z - x \in \gamma Bx$ and $(2x - z - \gamma Cx) - x \in \gamma Ax$. Therefore, $0 \in Ax + Bx + Cx$ and hence $J_{\gamma B}(\text{Fix}(T)) \subset \text{zer}(A + B + C)$. \square

Under further assumptions, FDRF’s $(x_n)_{n \in \mathbb{N}}$ sequence converges weakly to a solution of (1).

Theorem 4.1 *Assume (A1), (A2), and (A3). If furthermore one of the following conditions holds*

- (i) B is κ -cocoercive and $\gamma \in]0, \mu^{-1}/\sqrt{1 + \gamma/(2\kappa)}[$, which is satisfied, for example, if $0 < \gamma < \min\{\kappa, \mu^{-1}\sqrt{2/3}\}$.
- (ii) $B = N_V$ and $C = P_V C_1 P_V$ for some closed vector space V and single-valued operator $C_1 : \mathcal{H} \rightarrow \mathcal{H}$ and $\gamma \in]0, 1/\mu[$,

then $z_n \rightharpoonup z_\star \in \text{Fix}(T)$ and $x_n \rightharpoonup J_{\gamma B}z_\star \in \text{zer}(A + B + C)$ and $y_n \rightharpoonup J_{\gamma B}z_\star$ for (FDRF).

Proof Let $z_\star \in \text{Fix}(T)$ and $x_\star := J_{\gamma B} z_\star$. Set $u_{n+1} := x_{n+1} - z_n + z_\star - x_\star$ and we have

$$\begin{aligned} z_{n+1} - z_\star &= z_n + y_{n+1} - x_{n+1} - \gamma(Cy_{n+1} - Cx_{n+1}) - z_\star \\ &= y_{n+1} - x_\star + \gamma(Cx_{n+1} - Cy_{n+1}) - u_{n+1}. \end{aligned}$$

By

$$\begin{aligned} \|z_{n+1} - z_\star\|^2 &= \|y_{n+1} - x_\star + \gamma(Cx_{n+1} - Cy_{n+1})\|^2 \\ &\quad - 2\langle y_{n+1} - x_\star + \gamma(Cx_{n+1} - Cy_{n+1}), u_{n+1} \rangle + \|u_{n+1}\|^2. \end{aligned} \quad (2)$$

We expand the first term to get

$$\begin{aligned} &\|y_{n+1} - x_\star + \gamma(Cx_{n+1} - Cy_{n+1})\|^2 \\ &= \|y_{n+1} - x_\star\|^2 + 2\gamma\langle y_{n+1} - x_\star, Cx_{n+1} - Cy_{n+1} \rangle + \gamma^2\|Cx_{n+1} - Cy_{n+1}\|^2. \end{aligned} \quad (3)$$

Note

$$\begin{aligned} 2x_{n+1} - z_n - \gamma Cx_{n+1} - y_{n+1} &\in \gamma A y_{n+1}, \\ x_\star - z_\star - \gamma Cx_\star &\in \gamma A x_\star. \end{aligned}$$

Since A and C are monotone, we have

$$\begin{aligned} 0 &\leq \langle y_{n+1} - x_\star, 2x_{n+1} - z_n - \gamma Cx_{n+1} - y_{n+1} - x_\star + z_\star + \gamma Cx_\star \rangle \\ &= \langle y_{n+1} - x_\star, x_{n+1} - y_{n+1} - \gamma Cx_{n+1} + \gamma Cx_\star \rangle + \langle y_{n+1} - x_\star, u_{n+1} \rangle \\ &\leq \langle y_{n+1} - x_\star, x_{n+1} - y_{n+1} + \gamma Cy_{n+1} - \gamma Cx_{n+1} \rangle + \langle y_{n+1} - x_\star, u_{n+1} \rangle, \end{aligned}$$

which implies that

$$\begin{aligned} &2\gamma\langle y_{n+1} - x_\star, Cx_{n+1} - Cy_{n+1} \rangle \\ &\leq 2\langle y_{n+1} - x_\star, x_{n+1} - y_{n+1} \rangle + 2\langle y_{n+1} - x_\star, u_{n+1} \rangle \\ &= \|x_{n+1} - x_\star\|^2 - \|y_{n+1} - x_\star\|^2 - \|x_{n+1} - y_{n+1}\|^2 + 2\langle y_{n+1} - x_\star, u_{n+1} \rangle. \end{aligned} \quad (4)$$

Combining (3) and (4), we get

$$\begin{aligned} &\|y_{n+1} - x_\star + \gamma(Cx_{n+1} - Cy_{n+1})\|^2 \\ &\leq \|x_{n+1} - x_\star\|^2 - \|x_{n+1} - y_{n+1}\|^2 + \gamma^2\|Cx_{n+1} - Cy_{n+1}\|^2 \\ &\quad + 2\langle y_{n+1} - x_\star, u_{n+1} \rangle. \end{aligned}$$

Applying this bound to (2), we get

$$\begin{aligned} \|z_{n+1} - z_\star\|^2 &\leq \|x_{n+1} - x_\star\|^2 - \|x_{n+1} - y_{n+1}\|^2 + \gamma^2\|Cx_{n+1} - Cy_{n+1}\|^2 \\ &\quad - 2\gamma\langle Cx_{n+1} - Cy_{n+1}, u_{n+1} \rangle + \|u_{n+1}\|^2. \end{aligned} \quad (5)$$

(i) We consider the case where B is κ -cocoercive. From (5), we get

$$\begin{aligned}
& \|z_{n+1} - z_\star\|^2 \\
& \leq \|z_n - z_\star\|^2 - \|x_{n+1} - y_{n+1}\|^2 + \gamma^2 \|Cx_{n+1} - Cy_{n+1}\|^2 \\
& \quad - 2\gamma \langle Cx_{n+1} - Cy_{n+1}, u_{n+1} \rangle - \frac{2\kappa}{\gamma} \|u_{n+1}\|^2 \\
& \leq \|z_n - z_\star\|^2 - \|x_{n+1} - y_{n+1}\|^2 + \gamma^2 \left(1 + \frac{\gamma}{2\kappa(1-\varepsilon')}\right) \|Cx_{n+1} - Cy_{n+1}\|^2 \\
& \quad - \frac{2\kappa\varepsilon'}{\gamma} \|u_{n+1}\|^2 \\
& \leq \|z_n - z_\star\|^2 - \|x_{n+1} - y_{n+1}\|^2 + \gamma^2 \left(1 + \frac{\gamma}{2\kappa(1-\varepsilon')}\right) \mu^2 \|x_{n+1} - y_{n+1}\|^2 \\
& \quad - \frac{2\kappa\varepsilon'}{\gamma} \|u_{n+1}\|^2 \\
& = \|z_n - z_\star\|^2 - \varepsilon \|x_{n+1} - y_{n+1}\|^2 - \frac{2\kappa\varepsilon'}{\gamma} \|u_{n+1}\|^2, \tag{6}
\end{aligned}$$

where $0 < \varepsilon' < 1$. The first inequality follows from Lemma 3.2, the second inequality follows from Young's inequality, the third inequality follows from μ -Lipschitz continuity of C , and the final equality follows from the definition $\varepsilon := 1 - \gamma^2 \left(1 + \frac{\gamma}{2\kappa(1-\varepsilon')}\right) \mu^2$. We choose $\varepsilon' > 0$ small enough so that $\varepsilon > 0$.

(ii) If $B = N_V$ and $C = P_V C_1 P_V$, then

$$\begin{aligned}
& \langle Cx_{n+1} - Cy_{n+1}, u_{n+1} \rangle \\
& = \langle C_1 P_V x_{n+1} - C_1 P_V y_{n+1}, P_V(x_{n+1} - z_n) + P_V(z_\star - x_\star) \rangle = 0.
\end{aligned}$$

Hence, (5) becomes,

$$\begin{aligned}
& \|z_{n+1} - z_\star\|^2 \\
& = \|x_{n+1} - x_\star\|^2 - \|x_{n+1} - y_{n+1}\|^2 + \gamma^2 \|Cx_{n+1} - Cy_{n+1}\|^2 + \|u_{n+1}\|^2 \\
& \leq \|z_n - z_\star\|^2 - \|x_{n+1} - y_{n+1}\|^2 + \gamma^2 \|Cx_{n+1} - Cy_{n+1}\|^2 \\
& \leq \|z_n - z_\star\|^2 - \|x_{n+1} - y_{n+1}\|^2 + \gamma^2 \mu^2 \|x_{n+1} - y_{n+1}\|^2 \\
& = \|z_n - z_\star\|^2 - \varepsilon \|x_{n+1} - y_{n+1}\|^2,
\end{aligned}$$

where the first inequality follows from $\|x_{n+1} - x_\star\|^2 \leq \|z_n - z_\star\|^2 - \|u_{n+1}\|^2$, which follows from (1/2)-averagedness of P_V , the second inequality follows from μ -Lipschitz continuity of C , and the final equality follows from the definition $\varepsilon := 1 - \gamma^2 \mu^2 > 0$.

In cases (i) and (ii) both, we have

$$(\forall z_\star \in \text{Fix}(T)) \quad \|z_{n+1} - z_\star\|^2 \leq \|z_n - z_\star\|^2 - \varepsilon \|x_{n+1} - y_{n+1}\|^2$$

with $\varepsilon > 0$, which shows that $(z_n)_{n \in \mathbb{N}}$ is Fejér monotone with respect to $\text{Fix}(T)$ and

$$\sum_{n \in \mathbb{N}} \|x_{n+1} - y_{n+1}\|^2 < +\infty,$$

which implies $x_{n+1} - y_{n+1} \rightarrow 0$. Let us prove that every weak cluster point of $(z_n)_{n \in \mathbb{N}}$ is in $\text{Fix}(T)$. Let \bar{z} be a weak cluster point of $(z_n)_{n \in \mathbb{N}}$, i.e., there exists a subsequence $(z_{k_n})_{n \in \mathbb{N}}$ such that $z_{k_n} \rightharpoonup \bar{z}$. Consider two cases:

(i) We consider the case where B is κ -cocoercive. From the second negative term in (6), we get

$$\begin{aligned} \sum_{n \in \mathbb{N}} \|u_n\|^2 < +\infty &\implies u_n \rightarrow 0 \implies x_{n+1} - z_n \rightarrow x_\star - z_\star \\ &\implies Bx_{n+1} \rightarrow Bx_\star = BJ_{\gamma B}z_\star \end{aligned}$$

where the last implication follows from $x_{n+1} = J_{\gamma B}z_n$, $z_n - x_{n+1} = \gamma Bx_{n+1}$, $x_\star = J_{\gamma B}z_\star$, and $z_\star - x_\star = \gamma Bx_\star$. Since $z_{k_n} \rightharpoonup \bar{z}$, we have $x_{1+k_n} \rightharpoonup \bar{x} = \bar{z} - \gamma Bx_\star$ and $y_{1+k_n} \rightharpoonup \bar{z} - \gamma Bx_\star$. Since $x_{1+k_n} \rightharpoonup \bar{x}$ and $Bx_{1+k_n} \rightarrow Bx_\star$ and $\text{gra}(B)$ is closed under $\mathcal{H}^{\text{weak}} \times \mathcal{H}^{\text{strong}}$ [26, Proposition 20.38], we get $Bx_\star = B\bar{x}$. Hence, $\bar{x} = J_{\gamma B}\bar{z}$. By definition of the FDRF iteration, we have

$$\underbrace{x_{1+k_n} - z_{k_n}}_{\rightarrow -\gamma B\bar{x}} + \underbrace{x_{1+k_n} - y_{1+k_n}}_{\rightarrow 0} + \underbrace{\gamma C y_{1+k_n} - \gamma C x_{1+k_n}}_{\rightarrow 0} \in \underbrace{\gamma A y_{1+k_n}}_{\rightarrow \bar{x}} + \underbrace{\gamma C y_{1+k_n}}_{\rightarrow \bar{x}}.$$

Since $A + C$ is maximal monotone (A and C are maximal monotone with $\text{dom } C = \mathcal{H}$ [26, Corollary 25.5]) $\text{gra}(A + C)$ is closed under $\mathcal{H}^{\text{weak}} \times \mathcal{H}^{\text{strong}}$ [26, Proposition 20.38] and we get

$$-\gamma B\bar{x} \in \gamma A\bar{x} + \gamma C\bar{x},$$

which shows that $\bar{x} \in \text{zer}(A + B + C)$. Furthermore, $\bar{x} - \bar{z} \in \gamma A\bar{x} + \gamma C\bar{x}$ and hence $\bar{x} = J_{\gamma A}(2\bar{x} - \bar{z} - \gamma C\bar{x}) = J_{\gamma B}\bar{z}$. Therefore,

$$(\text{Id} - \gamma C)J_{\gamma A}(2\bar{x} - \bar{z} - \gamma C\bar{x}) + \bar{z} - (\text{Id} - \gamma C)J_{\gamma B}\bar{z} = \bar{z},$$

or equivalently $T\bar{z} = \bar{z}$. Hence, $z_n \rightharpoonup z_\star$ and $x_n \rightharpoonup J_{\gamma B}z_\star$. Since we have $x_{n+1} - y_{n+1} \rightarrow 0$, we conclude $y_n \rightharpoonup J_{\gamma B}z_\star$.

(ii) We consider the case where $B = N_V$. Then, $J_{\gamma B} = P_V$ is weakly continuous and hence $x_{1+k_n} \rightharpoonup \bar{x} = P_V\bar{z}$. Then, we have

$$\begin{aligned} p_{k_n} &:= \underbrace{x_{1+k_n} - z_{k_n}}_{\rightarrow \bar{x} - \bar{z}} + \underbrace{x_{1+k_n} - y_{1+k_n}}_{\rightarrow 0} + \underbrace{\gamma C y_{1+k_n} - \gamma C x_{1+k_n}}_{\rightarrow 0} \\ &\in \underbrace{\gamma A y_{1+k_n}}_{\rightarrow \bar{x}} + \underbrace{\gamma C y_{1+k_n}}_{\rightarrow \bar{x}}. \end{aligned}$$

Since $A + C$ is maximal monotone, $x_{1+k_n} = P_V z_{k_n}$, $x_{1+k_n} - y_{1+k_n} \rightarrow 0$, we have

$$p_{k_n} \rightharpoonup \bar{x} - \bar{z}, \quad P_{V^\perp} y_{1+k_n} \rightarrow 0, \quad \text{and} \quad P_V p_{k_n} \rightarrow 0.$$

From [26, Example 26.7] we have $\bar{x} \in \text{zer}(A + C + N_V)$ and $\bar{x} - \bar{z} \in (\gamma A + \gamma C)\bar{x}$. Hence $\bar{x} = J_{\gamma A}(2\bar{x} - \bar{z} - \gamma C\bar{x}) = J_{\gamma B}\bar{z}$. Therefore,

$$(\text{Id} - \gamma C)J_{\gamma A}(2\bar{x} - \bar{z} - \gamma C\bar{x}) + \bar{z} - (\text{Id} - \gamma C)J_{\gamma B}\bar{z} = \bar{z},$$

or equivalently $T\bar{z} = \bar{z}$. Hence $z_n \rightharpoonup \bar{z}$ and $x_n \rightharpoonup J_{\gamma B}\bar{z}$. \square

Under condition (i), B is single-valued and one can alternatively use FBF [2] or FBHF [9] by utilizing forward evaluations of B rather than the resolvent $J_{\gamma B}$. However, many cocoercive operators B require similar computational costs for evaluating B and J_B , and, in such cases, it may be advantageous to use J_B instead of the forward evaluation B [29].

Under condition (ii), the operator $B = N_V$ enforces a linear equality constraint. Consider

$$\text{find } x \in \mathcal{H} \quad \text{such that} \quad 0 \in \sum_{i=1}^m A_i x + C_i x,$$

where A_1, \dots, A_m are maximal monotone and C_1, \dots, C_m are monotone and Lipschitz. The equivalent formulation

$$\text{find } \mathbf{x} \in \mathcal{H}^m \quad \text{such that} \quad 0 \in N_V(\mathbf{x}) + \sum_{i=1}^m (A_i x_i + C_i x_i),$$

where $\mathbf{x} = (x_1, \dots, x_m)$ and $V = \{\mathbf{x} \in \mathcal{H}^m : x_1 = \dots = x_m\}$ is the consensus set, is an important instance of case (ii). (This problem class is the motivation for Raguet et al.'s forward-Douglas-Rachford [5,6].) When $B = N_V$ and V is the consensus set, FDRF reduces to Banert's relaxed forward backward, presented in the thesis [10]. Finally, Briceño-Arias's forward-partial inverse-forward [11] is also applicable under this setup. Briceño-Arias's method is different from our FDRF, but it can also be considered a "forward-Douglas-Rachford-forward splitting" as it reduces to DRS and FBF as special cases.

FDRF resembles forward-Douglas-Rachford (FDR) splitting [5,6,7,8] but is different due to the correction term $\gamma(Cy_{n+1} - Cx_{n+1})$. For convergence, FDR requires C to be cocoercive or (with a slight modification) B to be strongly monotone [8, Theorems 1.1 and 1.2]. In contrast, Theorem 4.1 states that FDRF converges when B is cocoercive.

However, FDRF does not converge in full generality. The following result establishes that Assumptions (A1), (A2), and (A3) are not sufficient to ensure that FDRF converges.

Theorem 4.2 *Given any $\gamma > 0$, there exist operators A , B , and C satisfying Assumptions (A1), (A2), and (A3) such that the FDRF iterates $(z_n)_{n \in \mathbb{N}}$ and $(x_n)_{n \in \mathbb{N}}$ diverge.*

Proof Let $\mathcal{H} = \mathbb{R}^2$ and let A , B , and C satisfy

$$J_{\gamma A}(x, y) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \quad B(x, y) = \begin{bmatrix} 0 & \gamma^{-1} \cot(\omega/2) \\ -\gamma^{-1} \cot(\omega/2) & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix},$$

$$C(x, y) = \begin{bmatrix} 0 & \mu \\ -\mu & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix},$$

where \cot denotes cotangent and $\omega > 0$ is small. Then, A , B , and C are maximally monotone and $\{0\} = \text{zer}(A + B + C)$. With direct calculations, we

get

$$T(x, y) = \begin{bmatrix} \frac{1}{2}(1 + \cos(\omega) + \gamma\mu \sin(\omega)) & \frac{1}{2}(\gamma\mu - \gamma\mu \cos(\omega) + \sin(\omega)) \\ \frac{1}{2}(-\gamma\mu + \gamma\mu \cos(\omega) - \sin(\omega)) & \frac{1}{2}(1 + \cos(\omega) + \gamma\mu \sin(\omega)) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

The 2×2 matrix defining T has eigenvalues

$$|\lambda_1|^2 = |\lambda_2|^2 = (\cos(\omega/2) + \gamma\mu \sin(\omega/2))^2 = 1 + \gamma\mu\omega + \mathcal{O}(\omega^2).$$

So $|\lambda_1|^2 = |\lambda_2|^2 > 1$ for small enough ω . Therefore, FDRF with $z_0 \neq 0$ diverges in the sense that $\|z_n\| \rightarrow \infty$ and $\|x_n\| \rightarrow \infty$. \square

In splitting methods, step size requirements often depend on the assumptions, rather than on the specific operators. Theorem 4.2 rules out the possibility of proving a result like “Assuming (A1), (A2), and (A3), FDRF converges for $\gamma \in]0, \gamma_{\max}(\mu)[$ ”, where $\gamma_{\max}(\mu)$ is some function that depends μ . However, Theorem 4.2 does not rule out the possibility that one can examine the specific operators A , B , and C (to gain more information beyond the Lipschitz parameter of C) and then select $\gamma > 0$ to obtain convergence.

5 FRB+DR: Convergent in General

To solve Problem (1), we propose the following iteration

$$(FRDR) \quad \begin{cases} x_{n+1} = J_{\gamma B}(x_n - \gamma u_n - \gamma(2Cx_n - Cx_{n-1})) \\ y_{n+1} = J_{\beta A}(2x_{n+1} - x_n + \beta u_n) \\ u_{n+1} = u_n + \frac{1}{\beta}(2x_{n+1} - x_n - y_{n+1}) \end{cases}$$

for $n = 0, 1, \dots$, where $x_0, x_{-1}, u_0 \in \mathcal{H}$ are starting points and $\gamma > 0$, $\beta > 0$. We call this method forward-reflected-Douglas–Rachford (FRDR) splitting as it combines Malitsky and Tam’s FRB [12] and Douglas–Rachford [1]. Note FRDR evaluates operator C only once per iteration, since the evaluation of Cx_{n-1} from the previous iteration can be reused. In contrast, FDRF evaluates C twice per iteration.

FRDR reduces to FRB when $A = 0$ and to DR when $C = 0$ and $\beta = \gamma$. When $A = 0$, we have $J_{\beta A} = \text{Id}$, $u_n = 0$, and the iteration is independent of β . FRB converges when $\gamma < 1/(2\mu)$ [12, Theorem 2.5], which is consistent with the parameter range of Theorem 5.1 with $\beta \rightarrow \infty$. When $C = 0$ and $\beta = \gamma$, one recovers DR with $z_n = x_n - \gamma u_n$.

Without any further assumptions, the $(x_n)_{n \in \mathbb{N}}$ sequence of FRDR converges weakly to a solution of (1).

Theorem 5.1 *Assume (A1), (A2), (A3), $0 < \beta$, and $0 < \gamma < \beta/(1 + 2\mu\beta)$. Then $x_n \rightharpoonup x_\star \in \text{zer}(A + B + C)$ for (FRDR).*

Proof Consider the Hilbert space $\mathcal{H} \times \mathcal{H}$ equipped with an alternative inner product and norm

$$\begin{aligned} \langle (x, u), (y, v) \rangle_{\mathcal{H} \times \mathcal{H}} &:= (1/\gamma) \langle x, y \rangle - \langle x, v \rangle - \langle y, u \rangle + \beta \langle u, v \rangle, \\ \|(x, u)\|_{\mathcal{H} \times \mathcal{H}}^2 &:= (1/\gamma) \|x\|^2 - 2 \langle x, u \rangle + \beta \|u\|^2. \end{aligned}$$

Since $\gamma < \beta$, the inner product and norm are valid. Let $x_\star \in \text{zer}(A + B + C)$, $u_\star \in Ax_\star$, and $-u_\star \in (B + C)x_\star$.

Define

$$\begin{aligned} \tilde{A}y_{n+1} &:= u_n + \frac{1}{\beta}(2x_{n+1} - x_n - y_{n+1}), \\ \tilde{B}x_{n+1} &:= \frac{1}{\gamma}(x_n - x_{n+1}) - u_n - 2Cx_n + Cx_{n-1}, \end{aligned}$$

$\tilde{A}x_\star := u_\star$, and $\tilde{B}x_\star := -u_\star - Cx_\star$ so that $\tilde{A}y_{n+1} \in Ay_{n+1}$, $\tilde{B}x_{n+1} \in Bx_{n+1}$, $\tilde{A}x_\star \in Ax_\star$, and $\tilde{B}x_\star \in Bx_\star$. Define

$$\begin{aligned} V_n &:= \|(x_n, u_n) - (x_\star, u_\star)\|_{\mathcal{H} \times \mathcal{H}}^2 + \frac{1}{2} \|(x_n, u_n) - (x_{n-1}, u_{n-1})\|_{\mathcal{H} \times \mathcal{H}}^2 \\ &\quad + 2 \langle Cx_n - Cx_{n-1}, x_\star - x_n \rangle \\ S_n &:= \frac{1}{2} \|(x_n, u_n) - (x_{n+1}, u_{n+1})\|_{\mathcal{H} \times \mathcal{H}}^2 + \frac{1}{2} \|(x_n, u_n) - (x_{n-1}, u_{n-1})\|_{\mathcal{H} \times \mathcal{H}}^2. \end{aligned}$$

We have

$$\begin{aligned} &\|(x_{n+1}, u_{n+1}) - (x_\star, u_\star)\|_{\mathcal{H} \times \mathcal{H}}^2 \\ &= \|(x_n, u_n) - (x_\star, u_\star)\|_{\mathcal{H} \times \mathcal{H}}^2 - \|(x_{n+1}, u_{n+1}) - (x_n, u_n)\|_{\mathcal{H} \times \mathcal{H}}^2 \\ &\quad - 2 \left\langle \tilde{B}x_{n+1} - \tilde{B}x_\star, x_{n+1} - x_\star \right\rangle - 2 \left\langle \tilde{A}y_{n+1} - \tilde{A}x_\star, y_{n+1} - x_\star \right\rangle \\ &\quad + 2 \langle Cx_n - Cx_{n-1}, x_n - x_{n+1} \rangle - 2 \langle Cx_\star - Cx_n, x_\star - x_{n+1} \rangle \\ &\quad + 2 \langle Cx_n - Cx_{n-1}, x_\star - x_n \rangle \\ &\stackrel{(a)}{\leq} \|(x_n, u_n) - (x_\star, u_\star)\|_{\mathcal{H} \times \mathcal{H}}^2 - \|(x_{n+1}, u_{n+1}) - (x_n, u_n)\|_{\mathcal{H} \times \mathcal{H}}^2 \\ &\quad + 2 \langle Cx_n - Cx_{n-1}, x_n - x_{n+1} \rangle - 2 \langle Cx_\star - Cx_n, x_\star - x_{n+1} \rangle \\ &\quad + 2 \langle Cx_n - Cx_{n-1}, x_\star - x_n \rangle \\ &\stackrel{(b)}{\leq} \|(x_n, u_n) - (x_\star, u_\star)\|_{\mathcal{H} \times \mathcal{H}}^2 - \|(x_{n+1}, u_{n+1}) - (x_n, u_n)\|_{\mathcal{H} \times \mathcal{H}}^2 \\ &\quad + L \|x_n - x_{n-1}\|^2 + L \|x_{n+1} - x_n\|^2 - 2 \langle Cx_{n+1} - Cx_n, x_\star - x_{n+1} \rangle \\ &\quad + 2 \langle Cx_n - Cx_{n-1}, x_\star - x_n \rangle. \end{aligned}$$

Inequality (a) follows from monotonicity of A and B . Inequality (b) follows from

$$\begin{aligned} 2 \langle Cx_n - Cx_{n-1}, x_n - x_{n+1} \rangle &\leq \frac{1}{\mu} \|Cx_n - Cx_{n-1}\|^2 + \mu \|x_n - x_{n+1}\|^2 \\ &\leq \mu (\|x_n - x_{n+1}\|^2 + \|x_n - x_{n-1}\|^2), \end{aligned}$$

which follows Young's inequality and Lipschitz continuity of C , and

$$-2 \langle Cx_\star - Cx_n, x_\star - x_{n+1} \rangle \leq -2 \langle Cx_{n+1} - Cx_n, x_\star - x_{n+1} \rangle$$

which follows from monotonicity of C . Reorganizing, we get

$$\begin{aligned} V_{n+1} &\leq V_n - \frac{1}{2} (\|(x_{n+1}, u_{n+1}) - (x_n, u_n)\|_{\mathcal{H} \times \mathcal{H}}^2 + \|(x_n, u_n) - (x_{n-1}, u_{n-1})\|_{\mathcal{H} \times \mathcal{H}}^2) \\ &\quad + L (\|x_n - x_{n-1}\|^2 + \|x_{n+1} - x_n\|^2) \end{aligned}$$

we now add

$$\begin{aligned} \frac{\beta\gamma\mu}{\beta - \gamma} \left(\frac{1}{\beta} \|x_n - x_{n+1}\|^2 - 2 \langle x_n - x_{n+1}, u_n - u_{n+1} \rangle + \beta \|u_n - u_{n+1}\|^2 \right) &\geq 0 \\ \frac{\beta\gamma\mu}{\beta - \gamma} \left(\frac{1}{\beta} \|x_n - x_{n-1}\|^2 - 2 \langle x_n - x_{n-1}, u_n - u_{n-1} \rangle + \beta \|u_n - u_{n-1}\|^2 \right) &\geq 0 \end{aligned}$$

to the right-hand-side (nonnegativity follows from Young's inequality) to get

$$V_{n+1} \leq V_n - \frac{\beta - \gamma - 2\mu\gamma\beta}{2(\beta - \gamma)} S_n. \quad (7)$$

Using the telescoping sum argument with (7), we get

$$V_0 - \frac{\beta - \gamma - 2\mu\gamma\beta}{2(\beta - \gamma)} \sum_{i=0}^n S_i \geq V_n$$

Next, we have

$$\begin{aligned} V_n &\stackrel{(a)}{\geq} \frac{1}{\gamma} \|x_n - x_\star\|^2 - 2 \langle x_n - x_\star, u_n - u_\star \rangle + \beta \|u_n - u_\star\|^2 \\ &\quad + \frac{1}{2\gamma} \|x_n - x_{n-1}\|^2 - \langle x_n - x_{n-1}, u_n - u_{n-1} \rangle + \frac{\beta}{2} \|u_n - u_{n-1}\|^2 \\ &\quad - \mu (\|x_n - x_{n-1}\|^2 + \|x_\star - x_n\|^2) \\ &\stackrel{(b)}{\geq} \frac{1}{2} \|(x_n, u_n) - (x_\star, u_\star)\|_{\mathcal{H} \times \mathcal{H}}^2 \\ &\quad + \left(\frac{1}{2\gamma} - \frac{1}{2\beta} - \mu \right) \|x_n - x_\star\|^2 + \left(\frac{1}{2\gamma} - \frac{1}{2\beta} - \mu \right) \|x_n - x_{n-1}\|^2 \\ &\stackrel{(c)}{\geq} \frac{1}{2} \|(x_n, u_n) - (x_\star, u_\star)\|_{\mathcal{H} \times \mathcal{H}}^2. \end{aligned}$$

Inequality (a) follows from

$$\begin{aligned} 2 \langle Cx_n - Cx_{n-1}, x_\star - x_n \rangle &\leq \frac{1}{\mu} \|Cx_n - Cx_{n-1}\|^2 + \mu \|x_\star - x_n\|^2 \\ &\leq \mu (\|x_n - x_{n-1}\|^2 + \|x_\star - x_n\|^2), \end{aligned}$$

which follows Young's inequality and Lipschitz continuity of C , inequality (b) follows from

$$\begin{aligned} \frac{\beta}{2}\|u_n - u_\star\|^2 - \langle x_n - x_\star, u_n - u_\star \rangle &\geq -\frac{1}{2\beta}\|x_n - x_\star\|^2 \\ \frac{\beta}{2}\|u_n - u_{n-1}\|^2 - \langle x_n - x_{n-1}, u_n - u_{n-1} \rangle &\geq -\frac{1}{2\beta}\|x_n - x_{n-1}\|^2, \end{aligned}$$

Young's inequality, and inequality (c) follows from $\gamma < \beta/(1 + 2\mu\beta)$. Putting these together, we have

$$V_0 - \frac{\beta - \gamma - 2\mu\gamma\beta}{2(\beta - \gamma)} \sum_{i=0}^n S_i \geq \frac{1}{2} \|(x_n, u_n) - (x_\star, u_\star)\|_{\mathcal{H} \times \mathcal{H}}^2.$$

This implies that the sequence $((x_n, u_n))_{n \in \mathbb{N}}$ is bounded and $S_n \rightarrow 0$. Since

$$S_n \geq \frac{1}{2} \|(x_{n+1}, u_{n+1}) - (x_n, u_n)\|_{\mathcal{H} \times \mathcal{H}}^2,$$

$S_n \rightarrow 0$ implies $x_{n+1} - x_n \rightarrow 0$ and $u_{n+1} - u_n \rightarrow 0$. Since

$$u_{n+1} - u_n = (1/\beta)(2x_{n+1} - x_n - y_{n+1})$$

we also have $x_{n+1} - y_{n+1} \rightarrow 0$.

Now consider a weakly convergent subsequence $((x_{k_n}, u_{k_n}))_{n \in \mathbb{N}}$ such that $(x_{k_n}, u_{k_n}) \rightharpoonup (\bar{x}, \bar{u})$. Note that x_{n+1} and y_{n+1} are defined by the inclusion

$$\begin{bmatrix} \frac{1}{\gamma}(x_n - x_{n+1}) + 2Cx_n - Cx_{n-1} - Cx_{n+1} \\ \frac{1}{\beta}(2x_{n+1} - y_{n+1} - x_n) \end{bmatrix} \in \begin{bmatrix} (B + C)x_{n+1} + u_n \\ Ay_{n+1} - u_n \end{bmatrix}.$$

The right-hand side is a maximal monotone operator on $\mathcal{H} \times \mathcal{H}$ (equipped with the usual inner product) [26, Propositions 20.23, 20.38] and the left-hand side strongly converges to 0 since C is continuous. Since $\text{gra}(B + C)$ is closed under $\mathcal{H}^{\text{weak}} \times \mathcal{H}^{\text{strong}}$ [26, Proposition 20.38], we have

$$-\bar{u} \in (B + C)\bar{x}, \quad \bar{u} \in A\bar{x}.$$

Adding these we also get $\bar{x} \in \text{zer}(A + B + C)$. Finally, since $(V_n)_{n \in \mathbb{N}}$ is a monotonically decreasing nonnegative sequence, it has a limit. Since C is continuous, $(x_n)_{n \in \mathbb{N}}$ and $(u_n)_{n \in \mathbb{N}}$ are bounded sequences, and $x_n - x_{n-1} \rightarrow 0$ and $u_n - u_{n-1} \rightarrow 0$, we have

$$\lim_{n \rightarrow \infty} V_n = \lim_{n \rightarrow \infty} \|(x_n, u_n) - (x_\star, u_\star)\|_{\mathcal{H} \times \mathcal{H}}^2.$$

By plugging in $(x_\star, u_\star) = (\bar{x}, \bar{u})$, we conclude that the entire sequence weakly converges to (\bar{x}, \bar{u}) . \square

The proof of Theorem 5.1 closely follows Malitsky and Tam's analysis [12, Lemma 2.4]. In fact, FRDR can be thought of as an instance FRB on a primal-dual system with an auxiliary metric. Naively translating Malitsky and Tam's convergence analysis via a change of coordinates leads to a step size requirement of $\mu < (\beta - \gamma)/(1 + \gamma\beta + \sqrt{(1 + \gamma\beta)^2 - 4\gamma(\beta - \gamma)})$ and $0 < \gamma < \beta$. With a direct analysis, we obtain the better (and simpler) requirement of $\mu < (\beta - \gamma)/(2\beta\gamma)$.

The discovery of this proof was computer-assisted in the sense that we used the performance estimation problem (PEP) [30, 31, 32] and a computer algebra system (CAS). We briefly describe the strategy here.

The proof of Theorem 5.1 crucially relies on finding the Lyapunov function V_n and showing $V_{n+1} \leq V_n - S_n$. For a fixed numerical value of β , γ , and μ , roughly speaking, the PEP allows us to pose a semidefinite program (SDP) which solution indicates whether a candidate Lyapunov function is nonincreasing. (A proof establishing a Lyapunov function is nonincreasing is a nonnegative combination of known inequalities, and the SDP automates the search.) We used the SDP to quickly experiment with many candidate Lyapunov functions and identified that the V_n used in the proof of Theorem 5.1 is a nonincreasing quantity.

To get a general proof, we numerically solved the SDP for many values of β , γ , and μ and deduced the general symbolic solution. (The general proof is equivalent to a general solution of the SDP that symbolically depends on β , γ , and μ .) We relied on a CAS, Mathematica, to work through the symbolic calculations. In deducing the symbolic form of the proof, we utilized the observed structure of the solution. For example, the optimal SDP matrices were rank deficient, so we set the determinant of the corresponding symbolic matrix to 0 and eliminated a degree of freedom.

Finally, we translated the symbolic calculations into a traditional proof that is verifiable by humans without the aid of computer software. This step involved some further simplifications, such as replacing the identity

$$2 \langle u, v \rangle = \eta \|u\|^2 + \frac{1}{\eta} \|v\|^2 - \|\sqrt{\eta}u - (1/\sqrt{\eta})v\|^2$$

with Young's inequality

$$2 \langle u, v \rangle \leq \eta \|u\|^2 + \frac{1}{\eta} \|v\|^2.$$

6 Comparison with Other Methods

We now quickly examine other existing methods applied to Problem (1) to see how they differ from FDRF and FRDR. We leave the comparison of these methods, in terms of their computational effectiveness, as a direction of future work. Note that Problem (1) can be reformulated into the primal dual system

$$\text{find } x, u \in \mathcal{H} \quad \text{such that} \quad \begin{bmatrix} 0 \\ 0 \end{bmatrix} \in \begin{bmatrix} Bx \\ A^{-1}u \end{bmatrix} + \begin{bmatrix} C & \text{Id} \\ -\text{Id} & 0 \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix}. \quad (8)$$

Combettes–Pesquet. The method of [13] can be thought of as FBF applied to the primal-dual system (8):

$$\begin{cases} \bar{x}_{n+1} = J_{\gamma B}(x_n - \gamma(Cx_n + u_n)) \\ y_{n+1} = J_{\gamma^{-1}A}(x_n + \gamma^{-1}u_n) \\ x_{n+1} = \bar{x}_{n+1} - \gamma(C\bar{x}_{n+1} - Cx_n) - \gamma^2(x_n - y_{n+1}) \\ u_{n+1} = u_n + \gamma(\bar{x}_{n+1} - y_{n+1}). \end{cases}$$

This method solves Problem (1) with an appropriate choice of $\gamma > 0$. This method does not reduce to DR when $C = 0$.

Malitsky–Tam. We can plainly applying FRB [12] to (8):

$$\begin{cases} x_{n+1} = J_{\gamma B}(x_n - \gamma(2Cx_n - Cx_{n-1} + 2u_n - u_{n-1})) \\ y_{n+1} = J_{\gamma^{-1}A}(2x_n - x_{n-1} + \gamma^{-1}u_n) \\ u_{n+1} = u_n + \gamma(2x_n - x_{n-1} - y_{n+1}). \end{cases}$$

This method solves Problem (1) with an appropriate choice of $\gamma > 0$. This method does not reduce to DR when $C = 0$.

Briceño-Arias. When $B = N_V$ and $V \subset \mathcal{H}$ is a closed vector space, i.e., in case (ii) of Theorem 4.1, forward-partial inverse-forward [11] applies:

$$\begin{cases} x_{n+1} = J_{\gamma A}(z_n - \gamma J_{\gamma B} C J_{\gamma B} z_n) \\ y_{n+1} = J_{\gamma B}(2x_{n+1} - z_n + \gamma J_{\gamma B} C J_{\gamma B} z_n) - x_{n+1} + z_n - \gamma J_{\gamma B} C J_{\gamma B} z_n \\ z_{n+1} = y_{n+1} - \gamma(J_{\gamma B} C J_{\gamma B} y_{n+1} - J_{\gamma B} C J_{\gamma B} z_{n+1}). \end{cases}$$

This method reduces to DRS when $C = 0$ and to FBF when $B = 0$. However, this method does not apply in the general setup when B is not a normal cone operator.

Johnstone–Eckstein. The method of [22, 23] is based on the notion of projective splitting and bears little resemblance the other methods. The method is very flexible, and there are multiple instances applicable to Problem (1). The following instance follows the presentation of [22]:

$$\begin{cases} x_{n+1}^A = J_{\gamma A}(z_n + \gamma w_n^A) \\ x_{n+1}^B = J_{\gamma B}(z_n + \gamma w_n^B) \\ x_{n+1}^C = z_n - \gamma(Cz_n - w_n^C) \\ z_{n+1} = z_n - \frac{\alpha_n}{\gamma^2} (3z_n - x_{n+1}^A - x_{n+1}^B - \gamma C x_{n+1}^C + \gamma(w_n^A + w_n^B + w_n^C)) \\ w_{n+1}^A = w_n^A - \alpha_n(x_{n+1}^A - x_n^C) \\ w_{n+1}^B = w_n^B - \alpha_n(x_{n+1}^B - x_n^C) \\ w_{n+1}^C = -w_{n+1}^A - w_{n+1}^B. \end{cases}$$

The scalar parameter α_n is computed each iteration with a formula given in [22]. This method does not reduce to DR or FBF.

7 Conclusions

In this paper, we considered the monotone inclusion problem with a sum of 3 operators, in which 2 are monotone and 1 is monotone-Lipschitz, and studied combinations of methods type “forward-Douglas-Rachford-forward”. We presented FDRF, a combination of DR and FBF, and showed that it converges with further assumptions, but not generally. We then presented FRDR, a combination of DR and FRB, and showed that it converges in general. Moreover, FRDR has a lower computational cost per iteration since it evaluates the monotone-Lipschitz operator only once per iteration. Therefore, we conclude FRDR to be the better forward-Douglas-Rachford-forward method.

Acknowledgements Ernest Ryu was partially supported by AFOSR MURI FA9550-18-1-0502, NSF Grant DMS-1720237, and ONR Grant N000141712162. Bàng Công Vũ’s research work was funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under Grant No. 102.01-2017.05.

References

1. Lions, P.L., and Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. *SIAM J. Num. Anal.* **16**(6), 964–979 (1979)
2. Tseng, P.: A modified forward-backward splitting method for maximal monotone mappings. *SIAM J. Control Optim.* **38**(2), 431–446 (2000)
3. Bruck, R.E.: On the weak convergence of an ergodic iteration for the solution of variational inequalities for monotone operators in Hilbert space. *J. Math. Anal. Appl.* **61**(1), 159–164 (1977)
4. Passty, G.B.: Ergodic convergence to a zero of the sum of monotone operators in Hilbert space. *J. Math. Anal. Appl.* **72**(2), 383–390 (1979)
5. Raguet, H., Fadili, J., and Peyré, G.: Generalized forward-backward splitting. *SIAM J. Imaging Sci.* **6**(3), 1199–1226 (2013)
6. Raguet, H.: A note on the forward-Douglas–Rachford splitting for monotone inclusion and convex optimization. *Optim. Lett.*, **13**(4), 717–740 (2018)
7. Briceño-Arias, L.M.: Forward–Douglas–Rachford splitting and forward–partial inverse method for solving monotone inclusions. *Optimization.* **64**(5), 1239–1261 (2015)
8. Davis, D. and Yin, W.: A Three-Operator Splitting Scheme and its Optimization Applications. *Set-Valued Var. Anal.* **25**(4), 829–858 (2017)
9. Briceño-Arias, L.M., and Davis, D.: Forward-Backward-Half Forward Algorithm for Solving Monotone Inclusions. *SIAM J. Optim.* **28**(4), 2839–2871 (2018)
10. Banert, S., *A relaxed forward-backward splitting algorithm for inclusions of sums of monotone operators*. Masters thesis, Technische Universität Chemnitz, (2012)
11. Briceño-Arias, L.M.: Forward-Partial Inverse-Forward Splitting for Solving Monotone Inclusions. *J. Optim. Theory Appl.* **166**(2), 391–413 (2015)
12. Malitsky, Y., and Tam, M.K.: A Forward-Backward Splitting Method for Monotone Inclusions Without Cocoercivity. *arXiv:1808.04162*. (2018)
13. Combettes, P.L., and Pesquet, J.-C.: Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators. *Set-Valued Var. Anal.* **20**(2), 307–330 (2012)
14. Combettes, P.L.: Systems of structured monotone inclusions: duality, algorithms, and applications. *SIAM J. Optim.* **23**(4), 2420–2447 (2013)
15. Rockafellar, R.T., Monotone operators associated with saddle-functions and minimax problems. In F. E. Browder, editor, *Nonlinear Functional Analysis, Part 1, volume 18 of Proceedings of Symposia in Pure Mathematics.* **18**, 241–250 (1970)

16. Boţ, R.I., and Hendrich, C.: A Douglas–Rachford type primal-dual method for solving inclusions with mixtures of composite and parallel-sum type monotone operators. *SIAM J. Optim.* **23**(4), 2541–2565 (2013)
17. Latafat, P., and Patrinos, P.: Asymmetric forward-backward-adjoint splitting for solving monotone inclusions involving three operators. *Comput. Optim. Appl.*, **68**(1), 57–93 (2017)
18. Ryu, E.K., Uniqueness of DRS as the 2 Operator Resolvent-Splitting and Impossibility of 3 Operator Resolvent-Splitting. *Math. Program.* (2019)
19. Condat, L.: A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. *J. Optim. Theory Appl.* **158**(2), 460–479 (2013)
20. Vũ, B.C.: A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Adv. Comput. Math.* **38**(3), 667–681 (2013)
21. Yan, M.: A new primaldual algorithm for minimizing the sum of three functions with a linear operator. *J. Sci. Comput.* **76**(3), 1698–1717 (2018)
22. Johnstone, P.R., Eckstein, J.: Projective Splitting with Forward Steps: Asynchronous and Block-Iterative Operator Splitting. *arXiv:1803.07043*. (2018)
23. Johnstone, P.R., Eckstein, J.: Projective Splitting with Forward Steps only Requires Continuity. *arXiv:1809.07180*. (2018)
24. Bùi, M. N. and Combettes, P. L.: Warped proximal iterations for monotone inclusions. *arXiv:1908.07077* (2019)
25. Giselsson, P.: Nonlinear forward-backward splitting with projection correction. *arXiv:1908.07449* (2019)
26. Bauschke, H.H., Combettes, P.L.: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, New York, 2nd ed., (2017)
27. Ryu, E.K. and Boyd, S. *Primer on Monotone Operator Methods*. *Appl. Comput. Math.* **15**(1), 3–43 (2016)
28. Ryu, E.K., Hannah, R., and Yin, W.: Scaled Relative Graph: Nonexpansive operators via 2D Euclidean Geometry. *arXiv:1902.09788* (2019)
29. Combettes, P.L. and Glaudin, L.E.: Proximal Activation of Smooth Functions in Splitting Algorithms for Convex Minimization. *SIAM J. Imaging Sci.* (2018)
30. Drori, Y. and Teboulle, M. Performance of first-order methods for smooth convex minimization: a novel approach. *Math. Program.* **145**(1–2), 451–482 (2014)
31. Taylor, A.B., Hendrickx, and J.M., and Glineur, F.: Smooth strongly convex interpolation and exact worst-case performance of first-order methods. *Math. Program.* **161**(1–2), 307–345 (2017)
32. Ryu, E.K., Taylor, A.B., Bergeling, C., and Giselsson, P.: Operator Splitting Performance Estimation: Tight contraction factors and optimal parameter selection. *arXiv:1812.00146* (2018)