A reflected forward-backward splitting method for monotone inclusions involving Lipschitzian operators

Volkan Cevher and Bằng Công Vũ*

Laboratory for Information and Inference Systems, EPFL, Lausanne, Switzerland
volkan.cevher@epfl.ch; bangcvvn@gmail.com

March 24, 2020

Abstract

In this paper, we propose a novel splitting method for finding a zero point of the sum of two monotone operators where one of them is Lipschizian. The weak convergence the method is proved in real Hilbert spaces. Applying the proposed method to composite monotone inclusions involving parallel sums yields a new primal-dual splitting which is different from the existing methods. Connections to existing works are clearly stated. We also provide an application of the proposed method to the image denoising by the total variation.

Keywords: monotone inclusion, monotone operator, operator splitting, cocoercive, forward-backward-forward method, forward-backward algorithm, composite operator, duality, primal-dual algorithm

Mathematics Subject Classifications (2010): 47H05, 49M29, 49M27, 90C25

1 Introduction

The forward-backward-forward splitting method (FBFS) or Tseng's splitting method first appeared in [25]. This method was proposed to find a zero point of the sum of two monotone operators acting on a real Hilbert space $(\mathcal{H}, \langle \cdot \mid \cdot \rangle)$, namely,

$$\text{find } \overline{x} \in \mathcal{H} \text{ such that } 0 \in A\overline{x} + B\overline{x}. \tag{1.1}$$

under the assumption that $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ is a maximally monotone, $B : \mathcal{H} \to \mathcal{H}$ be a monotone and μ-Lipschitzian, i.e.,

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H})\ \|Bx - By\| \leq \mu\|x - y\|, \tag{1.2}$$

and that such a solution exists. The FBFS method operates according to the routine

$$\begin{cases} y_n = x_n - \gamma B x_n \\ z_n = (\mathrm{Id} + \gamma A)^{-1} y_n \\ r_n = z_n - \gamma B z_n \\ x_{n+1} = x_n + r_n - y_n. \end{cases} \tag{1.3}$$

The weak convergence of $(x_n)_{n\in\mathbb{N}}$ to a solution of (1.1) was proved under the condition $0 < \gamma < 1/\mu$. Inexact version of the FBF method was investigated in [5]. Then, variable metric version and the stochastic version the FBF method are in [28] and [26], respectively. One of the most important examples of B is the case when B is a linear skew operator [5] where monotone plus skew model plays a central role in solving primal-dual monotone inclusions and primal-dual convex optimization problems. The main idea of [5] was then developed in [8]. Several developments and extensions of [8] are in [7, 3, 11, 28].

The advantage of this framework is its generality and the main disadvantage of (1.3) is that it requires two calls of B per one iteration. This issue was recently resolved in [17]. Specifically, they propose a forward-reflected-backward splitting method (FRBS) for solving (1.1), namely,

$$\gamma \in \left]0, +\infty\right[, \quad x_{n+1} = (\mathrm{Id} + \gamma A)^{-1}(x_n - 2\gamma B x_n + \gamma B x_{n-1}). \tag{1.4}$$

The weak convergence of the iterates generated by (1.4) is proved under the condition $\gamma \in \left]0, 1/(2\mu)\right[$. If B is linear and A is the normal cone of some non-empty closed convex set K, the FRBS method admits the same structure as the reflected projected gradient methods for variational inequalities [15], namely,

$$\gamma \in \left]0, +\infty\right[, \quad x_{n+1} = (\mathrm{Id} + \gamma N_K)^{-1}(x_n - \gamma B(2x_n - x_{n-1})). \tag{1.5}$$

For any μ-Lipschitzian monotone operator B, the weak convergence of the iterates generated by (1.5) is proved under the condition $\gamma \in \left]0, (\sqrt{2}-1)/\mu\right[$. When N_K is replaced by a subdifferential of some proper lower semicontinuous convex function f, line-search versions (1.5) are proposed in [16].

The objective of this paper is to investigate the convergence of (1.5) for the problem (1.1) for any maximally monotone operator A, i.e., we propose to investigate the convergence of the following reflected forward-backward splitting method (RFBS) for (1.1):

$$\begin{cases} y_n = 2x_n - x_{n-1} \\ x_{n+1} = (\mathrm{Id} + \gamma A)^{-1}(x_n - \gamma B y_n), \end{cases} \tag{1.6}$$

where $\gamma > 0$.

In Section 2, we prove the weak convergence of (1.6) and provide an application to composite monotone inclusions involving the parallel sums and Lipschitzian monotone operators. We compare, in Section 3, the proposed method to several existing methods, for the image denoising by the total variation.

Notations. (See [1]) The scalar products and the associated norms of all Hilbert spaces used in this paper are denoted respectively by $\langle \cdot \mid \cdot \rangle$ and $\|\cdot\|$. We denote by $\mathcal{B}(\mathcal{H}, \mathcal{G})$ the space of all bounded linear operators from $\mathcal{H}$ to $\mathcal{G}$. The symbols $\rightharpoonup$ and $\to$ denote respectively weak and strong convergence. Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a set-valued operator. The domain of A is denoted by $\mathrm{dom}(A)$ that is a set of all $x \in \mathcal{H}$ such that $Ax \neq \varnothing$. The range of A is $\mathrm{ran}(A) = \big\{u \in \mathcal{H} \mid (\exists x \in \mathcal{H}) u \in Ax\big\}$. The graph of A is $\mathrm{gra}(A) = \big\{(x,u) \in \mathcal{H} \times \mathcal{H} \mid u \in Ax\big\}$. The inverse of A is $A^{-1}\colon u \mapsto \big\{x \mid u \in Ax\big\}$. The zero set of A is $\mathrm{zer}(A) = A^{-1}0$. We say that A is monotone if

$$\big(\forall u \in Ax\big)\big(\forall (y,v) \in \mathrm{gra}\, A\big) \quad \langle x - y \mid u - v\rangle \geq 0, \tag{1.7}$$

and it is maximally monotone if there exists no monotone operator B such that $\mathrm{gra}(B)$ properly contains $\mathrm{gra}(A)$. The resolvent of A is

$$J_A = (\mathrm{Id} + A)^{-1}, \tag{1.8}$$

where Id denotes the identity operator on $\mathcal{H}$. A single-valued operator $B\colon \mathcal{H} \to \mathcal{H}$ is β-cocoercive, for some $\beta \in \left]0, +\infty\right[$, if

$$(\forall (x,y) \in \mathcal{H}^2) \ \langle x - y \mid Bx - By\rangle \geq \beta \|Bx - By\|^2. \tag{1.9}$$

The parallel sum of two operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ is $A \Box B = (A^{-1} + B^{-1})^{-1}$. The class of all lower semicontinuous convex functions $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ such that $\mathrm{dom}\, f = \big\{x \in \mathcal{H} \mid f(x) < +\infty\big\} \neq \varnothing$ is denoted by $\Gamma_0(\mathcal{H})$. Now, let $f \in \Gamma_0(\mathcal{H})$. The subdifferential of $f \in \Gamma_0(\mathcal{H})$ is the maximally monotone operator

$$\partial f\colon \mathcal{H} \to 2^{\mathcal{H}}\colon x \mapsto \big\{u \in \mathcal{H} \mid (\forall y \in \mathcal{H}) \quad \langle y - x \mid u\rangle + f(x) \leq f(y)\big\} \tag{1.10}$$

Moreover, the proximity operator of f is

$$\mathrm{prox}_f = J_{\partial f}\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto \underset{y \in \mathcal{H}}{\mathrm{argmin}}\, f(y) + \frac{1}{2}\|x - y\|^2. \tag{1.11}$$

Various closed-form expressions of the proximity operators are in [1, Chapter 24].

2 Weak convergence

We first prove an auxiliary result which will be used to prove the weak convergence of the sequence generated by the reflected forward-backward splitting.

Lemma 2.1 *Let $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ be generated by* (1.6). *Set*

$$(\forall n \in \mathbb{N}) \ p_{n+1} = x_n - \gamma B y_n - x_{n+1}. \tag{2.1}$$

Suppose that B is β-cocoercive. Then, for every $x \in \mathrm{zer}(A + B)$,

$$\begin{aligned} &\|x_{n+1} - x\|^2 + \|x_{n+1} - y_n\|^2 + \|x_{n+1} - x_n\|^2 + \|p_{n+1} + \gamma Bx\|^2 + \gamma(2\beta - \gamma)\|By_n - Bx\|^2 \\ &\leq \|x_n - x\|^2 + \|p_n + \gamma Bx\|^2 - \gamma^2\|By_{n-1} - Bx\|^2 + 2\gamma \left\langle By_{n-1} - By_n \mid x_{n+1} - y_n\right\rangle. \end{aligned} \tag{2.2}$$

If B is monotone, we remain have (2.2) *with $\beta = 0$.*

Proof. Let $x \in \operatorname{zer}(A+B)$ and $n \in \mathbb{N}$. By the definition of the resolvent, $p_{n+1} \in \gamma A x_{n+1}$. We have

$$\langle x_n + \gamma B y_{n-1} - x_{n-1} \mid x_{n+1} - y_n \rangle = -\langle p_n \mid x_{n+1} - y_n \rangle . \tag{2.3}$$

Let us recall that

$$y_n = 2x_n - x_{n-1} \text{ which is equivalent to } x_n - x_{n-1} = y_n - x_n. \tag{2.4}$$

Then it follows that

$$\langle y_n - x_n \mid x_{n+1} - y_n \rangle = -\langle p_n + \gamma B y_{n-1} \mid x_{n+1} - y_n \rangle , \tag{2.5}$$

and

$$\langle x_{n+1} - x_n \mid x - x_{n+1} \rangle = -\langle p_{n+1} + \gamma B y_n \mid x - x_{n+1} \rangle . \tag{2.6}$$

We have

$$\begin{cases} 2\langle y_n - x_n \mid x_{n+1} - y_n \rangle = \|x_n - x_{n+1}\|^2 - \|x_n - y_n\|^2 - \|x_{n+1} - y_n\|^2 \\ 2\langle x_{n+1} - x_n \mid x - x_{n+1} \rangle = \|x_n - x\|^2 - \|x_n - x_{n+1}\|^2 - \|x_{n+1} - x\|^2. \end{cases} \tag{2.7}$$

In turn,

$$\|x_{n+1} - x\|^2 + \|x_n - y_n\|^2 + \|x_{n+1} - y_n\|^2 = \|x_n - x\|^2 + 2\Gamma_n + 2\gamma \langle B y_{n-1} - B y_n \mid x_{n+1} - y_n \rangle , \tag{2.8}$$

where

$$\begin{aligned} \Gamma_n &= \langle p_n + \gamma B y_n \mid x_{n+1} - y_n \rangle + \langle p_{n+1} + \gamma B y_n \mid x - x_{n+1} \rangle \\ &= \langle p_n \mid x_{n+1} - y_n \rangle + \langle p_{n+1} \mid x - x_{n+1} \rangle + \gamma \langle B y_n \mid x - y_n \rangle . \end{aligned} \tag{2.9}$$

Since γA is monotone and $-\gamma B x \in \gamma A x$, we obtain

$$\begin{aligned} \langle p_{n+1} \mid x - x_{n+1} \rangle &\le \langle p_{n+1} \mid x - x_{n+1} \rangle + \langle -\gamma B x - p_{n+1} \mid x - x_{n+1} \rangle \\ &= \gamma \langle B x \mid x_{n+1} - x \rangle . \end{aligned} \tag{2.10}$$

Since B is β-cocoercive, we also have

$$\gamma \langle B y_n \mid x - y_n \rangle \le \gamma \langle B x \mid x - y_n \rangle - \gamma\beta \|B y_n - B x\|^2. \tag{2.11}$$

Adding (2.10) and (2.11), and using (2.4), the monotonicity of γA, we get

$$\begin{aligned} \Gamma_n &\le \langle p_n \mid x_{n+1} - y_n \rangle + \gamma \langle B x \mid x_{n+1} - y_n \rangle - \gamma\beta \|B y_n - B x\|^2 \\ &= \langle p_n + \gamma B x \mid x_{n+1} - x_n \rangle - \langle p_n + \gamma B x \mid x_n - x_{n-1} \rangle - \gamma\beta \|B y_n - B x\|^2 \\ &\le \langle p_{n+1} + \gamma B x \mid x_{n+1} - x_n \rangle - \langle p_n + \gamma B x \mid x_n - x_{n-1} \rangle - \gamma\beta \|B y_n - B x\|^2. \end{aligned} \tag{2.12}$$

Let us set

$$T_n = \|x_n - x\|^2 - 2\langle p_n + \gamma B x \mid x_n - x_{n-1} \rangle . \tag{2.13}$$

It follows from (2.12) and (2.8) that

$$T_{n+1} + \|x_n - y_n\|^2 + \|x_{n+1} - y_n\|^2 + 2\gamma\beta \|B y_n - B x\|^2 \le T_n + 2\gamma \langle B y_{n-1} - B y_n \mid x_{n+1} - y_n \rangle . \tag{2.14}$$

By (2.1), we have $p_{n+1} + x_{n+1} - x_n = -\gamma By_n$, and hence,

$$\begin{aligned} -2\langle p_{n+1} + \gamma Bx \mid x_{n+1} - x_n\rangle &= \|x_{n+1} - x_n\|^2 + \|p_{n+1} + \gamma Bx\|^2 \\ &\quad - \|p_{n+1} + \gamma Bx + x_{n+1} - x_n\|^2 \\ &= \|x_{n+1} - x_n\|^2 + \|p_{n+1} + \gamma Bx\|^2 - \gamma^2\|By_n - Bx\|^2, \end{aligned} \tag{2.15}$$

which implies that

$$T_{n+1} = \|x_{n+1} - x\|^2 + \|x_{n+1} - x_n\|^2 + \|p_{n+1} + \gamma Bx\|^2 - \gamma^2\|By_n - Bx\|^2. \tag{2.16}$$

Therefore, using (2.4) again, (2.14) becomes

$$\begin{aligned} &\|x_{n+1} - x\|^2 + \|x_{n+1} - y_n\|^2 + \|x_{n+1} - x_n\|^2 + \|p_{n+1} + \gamma Bx\|^2 + \gamma(2\beta - \gamma)\|By_n - Bx\|^2 \\ &\leq \|x_n - x\|^2 + \|p_n + \gamma Bx\|^2 - \gamma^2\|By_{n-1} - Bx\|^2 + 2\gamma\langle By_{n-1} - By_n \mid x_{n+1} - y_n\rangle, \end{aligned} \tag{2.17}$$

which proves the desired result. In the case when B is monotone, setting $\beta = 0$, we obtain the second conclusion. □

The main result of this paper is stated in the following theorem where we prove the weak convergence of the sequence generated by the reflected forward-backward splitting to a zero point of $A + B$.

Theorem 2.2 *Let $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ be generated by* (1.6). *Suppose that $\gamma \in \left]0, (\sqrt{2} - 1)/\mu\right[$, then $x_n \rightharpoonup \overline{x} \in \operatorname{zer}(A + B)$ and $y_n \rightharpoonup \overline{x} \in \operatorname{zer}(A + B)$.*

Proof. Since B is monotone, by Lemma 2.1, we have

$$\begin{aligned} &\|x_{n+1} - x\|^2 + \|x_{n+1} - y_n\|^2 + \|x_{n+1} - x_n\|^2 + \|p_{n+1} + \gamma Bx\|^2 - \gamma^2\|By_n - Bx\|^2 \\ &\leq \|x_n - x\|^2 + \|p_n + \gamma Bx\|^2 - \gamma^2\|By_{n-1} - Bx\|^2 + 2\gamma\langle By_{n-1} - By_n \mid x_{n+1} - y_n\rangle. \end{aligned} \tag{2.18}$$

Since B is μ-Lipschitz continuous, for any $\tau_1 > 0$ and $\tau_2 > 0$, we obtain

$$\begin{aligned} 2\langle By_n - By_{n-1} \mid y_n - x_{n+1}\rangle &\leq 2\mu\|y_n - y_{n-1}\|\|y_n - x_{n+1}\| \\ &\leq \frac{\mu}{\tau_1}\|y_n - y_{n-1}\|^2 + \mu\tau_1\|y_n - x_{n+1}\|^2 \\ &\leq \frac{\mu}{\tau_1}\big((1 + \frac{1}{\tau_2})\|y_n - x_n\|^2 + (1 + \tau_2)\|x_n - y_{n-1}\|^2\big) + \mu\tau_1\|y_n - x_{n+1}\|^2 \\ &= \alpha_1\|y_n - x_n\|^2 + \alpha_2\|x_n - y_{n-1}\|^2 + \mu\tau_1\|y_n - x_{n+1}\|^2, \end{aligned} \tag{2.19}$$

where we set

$$\alpha_1 = \frac{\mu}{\tau_1}(1 + \frac{1}{\tau_2}) \text{ and } \alpha_2 = \frac{\mu}{\tau_1}(1 + \tau_2). \tag{2.20}$$

Therefore, we derive from (2.18) that

$$\begin{aligned} &\|x_{n+1} - x\|^2 + \gamma\alpha_2\|x_{n+1} - y_n\|^2 + \|x_{n+1} - x_n\|^2 + \|p_{n+1} + \gamma Bx\|^2 - \gamma^2\|By_n - Bx\|^2 \\ &\leq \|x_n - x\|^2 + \|x_{n-1} - x_n\|^2 + \|p_n + \gamma Bx\|^2 - \gamma^2\|By_{n-1} - Bx\|^2 \\ &\quad - (1 - \gamma\alpha_1)\|x_n - y_n\|^2 + \gamma\alpha_2\|x_n - y_{n-1}\|^2 - (1 - \gamma(\alpha_2 + \mu\tau_1))\|x_{n+1} - y_n\|^2. \end{aligned} \tag{2.21}$$

Therefore, we need the following condition on γ,

$$0 < \gamma < \frac{1}{\alpha_1} = \frac{\tau_1\tau_2}{\mu(1+\tau_2)} \text{ and } 0 < \gamma < \frac{1}{\alpha_2+\mu\tau_1} = \frac{\tau_1}{\mu(1+\tau_2+\tau_1^2)}. \tag{2.22}$$

The optimal bound of γ happens when

$$\frac{\tau_1\tau_2}{\mu(1+\tau_2)} = \frac{\tau_1}{\mu(1+\tau_2+\tau_1^2)}, \tag{2.23}$$

which implies that

$$\tau_2\tau_1^2 = 1-\tau_2^2. \tag{2.24}$$

Hence,

$$0 < \gamma^2 < \frac{\tau_2\tau_2\tau_1^2}{\mu^2(1+\tau_2)^2} = \frac{\tau_2(1-\tau_2)}{\mu^2(1+\tau_2)} = \frac{(1+\tau_2-1)(2-(1+\tau_2))}{\mu^2(1+\tau_2)}. \tag{2.25}$$

We express the last term as

$$\frac{(1+\tau_2-1)(2-(1+\tau_2))}{\mu^2(1+\tau_2)} = \frac{1}{\mu^2}(3-1-\tau_2-\frac{2}{1+\tau_2}), \tag{2.26}$$

which attains the maximum when $1+\tau_2 = \sqrt{2}$ and hence $\tau_1 = \sqrt{2}$. It follows that $\gamma^2 < \mu^{-2}(3-2\sqrt{2})$. Thus

$$0 < \gamma < \frac{\sqrt{2}-1}{\mu}. \tag{2.27}$$

We next set

$$E_n = \|x_n - x\|^2 + \|x_{n-1}-x_n\|^2 + \|p_n+\gamma Bx\|^2 + \gamma\alpha_2\|x_n - y_{n-1}\|^2 - \gamma^2\|By_{n-1}-Bx\|^2. \tag{2.28}$$

Then, we can rewrite (2.21) as

$$E_{n+1} \le E_n - (1-\gamma\mu(1+\sqrt{2}))\|x_n-y_n\|^2 - (1-\gamma\mu(1+\sqrt{2}))\|x_{n+1}-y_n\|^2. \tag{2.29}$$

We have

$$\begin{aligned}\gamma^2\|By_{n-1}-Bx\|^2 &\le 2\gamma^2\|By_{n-1}-Bx_n\|^2 + 2\gamma^2\|Bx_n - Bx\|^2\\ &\le 2\gamma^2\mu^2\|y_{n-1}-x_n\|^2 + 2\gamma^2\mu^2\|x_n-x\|^2.\end{aligned} \tag{2.30}$$

Then, since $\gamma < (\sqrt{2}-1)/\mu$, we have $\gamma < 2/\mu$ and hence there exists $\epsilon > 0$ such that

$$\begin{aligned}E_n &\ge \gamma\mu(1-2\gamma\mu)\|x_n-y_{n-1}\|^2 + (1-2\gamma^2\mu^2)\|x_n-x\|^2\\ &\ge \epsilon\big(\|y_{n-1}-x_n\|^2 + \|x_n-x\|^2\big)\\ &\ge 0.\end{aligned} \tag{2.31}$$

In turn, we derive from (2.29) that

$$\begin{cases} E_n \to \overline{\zeta} \in \mathbb{R}\\ \sum_{n\in\mathbb{N}}\|x_n-y_n\|^2 < +\infty\\ \sum_{n\in\mathbb{N}}\|x_{n+1}-y_n\|^2 < +\infty.\end{cases} \tag{2.32}$$

Since $(E_n)_{n\in\mathbb{N}}$ converges, it is bounded and therefore, it follows from (2.31) that $(\|x_n - x\|)_{n\in\mathbb{N}}$ and $(x_n)_{n\in\mathbb{N}}$ are bounded. Hence, $(y_n)_{n\in\mathbb{N}}$ and $(By_n - Bx)_{n\in\mathbb{N}}$ are also bounded. Since $x_n - y_n \to 0$, we obtain $\langle y_n - x_n \mid By_{n-1} - Bx\rangle \to 0$. We have

$$\begin{aligned}\|p_n + \gamma Bx\|^2 - \gamma^2\|By_{n-1} - Bx\|^2 &= \|x_{n-1} - x_n\|^2 - 2\gamma \langle x_{n-1} - x_n \mid By_{n-1} - Bx\rangle \\ &= \|y_n - x_n\|^2 + 2\gamma \langle y_n - x_n \mid By_{n-1} - Bx\rangle \\ &\to 0. \end{aligned} \tag{2.33}$$

Therefore, $\|x_n - x\| \to \overline{\zeta}$. Let $\overline{x}$ be a weak cluster point of $(x_n)_{n\in\mathbb{N}}$, then there exists $x_{k_n} \rightharpoonup \overline{x}$. Note that $x_{k_n} - x_{k_n+1} - \gamma By_{k_n} + \gamma Bx_{k_n+1} \to 0$ and

$$x_{k_n} - x_{k_n+1} - \gamma By_{k_n} + \gamma Bx_{k_n+1} = p_{k_n+1} + \gamma Bx_{k_n+1} \in \gamma(A+B)x_{k_n+1}. \tag{2.34}$$

Since $A + B$ is maximally monotone, its graph is closed in $\mathcal{H}^{strong} \times \mathcal{H}^{weak}$. Therefore, it follows from (2.34) that $\overline{x} \in \operatorname{zer}(A + B)$. Using Opial's result [18, 19], we obtain $x_n \rightharpoonup \overline{x} \in \operatorname{zer}(A + B)$. □

Proposition 2.3 *If B is β-cocoercive for some $\beta \in\,]0,\infty[$, then the conclusion of Theorem 2.2 remains valid for $\gamma < \beta/2$.*

Proof. By Lemma 2.1, we get

$$\begin{aligned}&\|x_{n+1} - x\|^2 + \|x_{n+1} - y_n\|^2 + \|x_{n+1} - x_n\|^2 + \|p_{n+1} + \gamma Bx\|^2 + \gamma(2\beta - \gamma)\|By_n - Bx\|^2 \\ &\le \|x_n - x\|^2 + \|p_n + \gamma Bx\|^2 - \gamma^2\|By_{n-1} - Bx\|^2 + 2\gamma \langle By_{n-1} - By_n \mid x_{n+1} - y_n\rangle . \end{aligned} \tag{2.35}$$

Let us estimate the term $q_n = 2\gamma \langle By_{n-1} - By_n \mid x_{n+1} - y_n\rangle$. Let $\epsilon \in\,]0,\infty[$ be such that $0 < \gamma < (1-\epsilon)\beta/2$. We have

$$\begin{aligned} q_n &= 2\gamma \langle By_{n-1} - Bx \mid x_{n+1} - y_n\rangle + 2\gamma \langle Bx - By_n \mid x_{n+1} - y_n\rangle \\ &\le \frac{2\gamma^2}{1-\epsilon}\|By_{n-1} - Bx\|^2 + \frac{2\gamma^2}{1-\epsilon}\|By_n - Bx\|^2 + (1-\epsilon)\|x_{n+1} - y_n\|^2, \end{aligned} \tag{2.36}$$

and thus, we derive from (2.35) that

$$\begin{aligned}&\|x_{n+1} - x\|^2 + \|x_{n+1} - x_n\|^2 + \|p_{n+1} + \gamma Bx\|^2 + \gamma^2\frac{1+\epsilon}{1-\epsilon}\|By_n - Bx\|^2 + \epsilon\|x_{n+1} - y_n\|^2 \\ &\le \|x_n - x\|^2 + \|p_n + \gamma Bx\|^2 + \gamma^2\frac{1+\epsilon}{1-\epsilon}\|By_{n-1} - Bx\|^2 - \gamma(2\beta - \frac{4\gamma}{1-\epsilon})\|By_n - Bx\|^2. \end{aligned} \tag{2.37}$$

Since $\gamma < (1-\epsilon)\beta/2$, we obtain

$$\begin{cases} \|x_n - x\|^2 + \|p_n + \gamma Bx\|^2 + \gamma^2\frac{1+\epsilon}{1-\epsilon}\|By_{n-1} - Bx\|^2 \to \overline{\xi} \in \mathbb{R}, \\ \sum_{n\in\mathbb{N}} \|By_n - Bx\|^2 < \infty, \\ \sum_{n\in\mathbb{N}} \|x_n - x_{n+1}\|^2 < +\infty, \\ \sum_{n\in\mathbb{N}} \|y_n - x_{n+1}\|^2 < +\infty. \end{cases} \tag{2.38}$$

Since $p_{n+1} + \gamma Bx = \gamma(Bx - By_n) + x_n - x_{n+1} \to 0$, it follows that $\|x_n - x\|^2 \to \overline{\xi} \in \mathbb{R}$ and hence $(x_n)_{n\in\mathbb{N}}$ is bounded. Let $\overline{x}$ be a weak cluster point of $(x_n)_{n\in\mathbb{N}}$, then there exists a subsequence $(x_{k_n})_{n\in\mathbb{N}}$ of $(x_n)_{n\in\mathbb{N}}$ such that $x_{k_n} \rightharpoonup \overline{x}$. Note that $By_{k_n} \to Bx$ and $x_n - y_n = x_{n-1} - x_n \to 0$.

Since B is maximally monotone, its graph is closed in $\mathcal{H}^{strong} \times \mathcal{H}^{weak}$, we obtain $Bx = B\overline{x}$ and thus $By_{k_n} \to B\overline{x}$. Since A is maximally monotone, its graph is closed in $\mathcal{H}^{strong} \times \mathcal{H}^{weak}$, passing limit from

$$\frac{x_{k_n} - x_{k_n+1}}{\gamma} - By_{k_n} = p_{k_n+1}/\gamma \in Ax_{k_n+1}, \tag{2.39}$$

we obtain $\overline{x} \in \operatorname{zer}(A+B)$. Therefore, using Opial's result [18], we obtain $x_n \rightharpoonup \overline{x} \in \operatorname{zer}(A+B)$. □

Remark 2.4 Here are some remarks.

(i) A special case of Theorem 2.2 is in [15] when $A = N_S$, the normal cone operator to a closed convex set $S \subset \mathcal{H}$. The line-search versions for the case where $A = \partial f$ for some $f \in \Gamma_0(\mathbb{R}^d)$ are in [16]. The connection to the existing work concerning with solving variational inequalities can be found in [15, 16]. For the compactness, we do not cite all of them here.

(ii) In the case, B is β-cocoercive, $\mu = 1/\beta$ and the range of the step size is relaxed from $\left]0, (\sqrt{2}-1)/\mu\right[$ to $]0, 0.5\beta[$ which is relatively small in comparison to the standard forward-backward splitting [1].

(iii) In the case, B is linear, (1.6) is exactly the same as the one in [17] where the convergence is proved under the condition $\gamma \in]0, 0.5/\mu[$. The computational cost of (1.6) and [17] are much cheaper than that of FBFS in [25].

Corollary 2.5 *Let $B\colon \mathcal{H} \to \mathcal{H}$ be a monotone and μ_0-Lipschitzian, $\mu_0 \in]-\infty, +\infty[$, and $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone. Let m be a strictly positive integer and let $(\mathcal{G}_i)_{1\leq i\leq m}$ be real Hilbert spaces. For every $i \in \{1,\ldots,m\}$, let $A_i\colon \mathcal{G}_i \to 2^{\mathcal{G}_i}$ be a maximally monotone, and let $B_i\colon \mathcal{G}_i \to 2^{\mathcal{G}_i}$ be a maximally monotone such that B_i^{-1} is μ_i-Lipschitzian operator for some $\mu_i \in]-\infty, +\infty[$, let $L_i\colon \mathcal{H} \to \mathcal{G}_i$ be a bounded linear operator such that $0 \neq \sum_{i=1}^m \|L_i\|^2$. Suppose that*

$$0 \in \operatorname{ran}\big(A + \sum_{i=1}^m L_i^*(A_i \square B_i)L_i + B\big). \tag{2.40}$$

The primal inclusion is to find $\overline{x}$ such that

$$0 \in A\overline{x} + \sum_{i=1}^m L_i^*(A_i \square B_i)L_i\overline{x} + B\overline{x}, \tag{2.41}$$

and the dual inclusion is to find $(\overline{v}_i)_{1\leq i\leq m} \in (\mathcal{G}_i)_{1\leq i\leq m}$ such that

$$(\forall i \in \{1,\ldots,m\})\ 0 \in L_i(A+B)^{-1}(-\sum_{i=1}^m L_i^*\overline{v}_i) + A_i^{-1}\overline{v}_i + B_i^{-1}\overline{v}_i. \tag{2.42}$$

Set

$$\mu = \max\{\mu_0, \ldots, \mu_m\} + \sqrt{\sum_{i=1}^m \|L_i\|^2}. \tag{2.43}$$

Let $\gamma \in \left]0, (\sqrt{2}-1)/\mu\right[$, $(x_0, x_{-1}) \in \mathcal{H}^2$ *and, for every* $i \in \{1, \ldots, m\}$, *let* $(v_{i,0}, v_{i,-1}) \in \mathcal{G}_i^2$. *Iterate, for every* $n \in \mathbb{N}$,

$$\begin{cases} x_{n+1} & = J_{\gamma A}(x_n - \gamma B(2x_n - x_{n-1}) - \gamma \sum_{i=1}^m L_i^*(2v_{i,n} - v_{i,n-1})) \\ \text{For} & i = 1, \ldots, m \\ v_{i,n+1} & = J_{\gamma A_i^{-1}}(v_{i,n} - \gamma B_i^{-1}(2v_{i,n} - v_{i,n-1}) + \gamma L_i(2x_n - x_{n-1})). \end{cases} \tag{2.44}$$

Then $x_n \rightharpoonup \overline{x}$ *solves* (2.41) *and* $(v_{1,n}, \ldots, v_{m,n}) \rightharpoonup (\overline{v}_1, \ldots, \overline{v}_m)$ *solves* (2.42).

Proof. We use the technique in [8]. Let $\mathcal{K} = \mathcal{H} \oplus \mathcal{G}_1 \oplus \ldots \oplus \mathcal{G}_m$ be the Hilbert direct sum of the Hilbert spaces $\mathcal{H}$ and $(\mathcal{G}_i)_{1\leq i\leq m}$, where the scalar product and the the associated norm of $\mathcal{K}$ are respectively defined as, for any $(x, \boldsymbol{v}) = (x, v_1, \ldots, v_m) \in \mathcal{K}$ and $(y, \boldsymbol{w}) = (x, w_1, \ldots, w_m) \in \mathcal{K}$,

$$\langle\langle \cdot \mid \cdot \rangle\rangle \colon \big((x, \boldsymbol{v}), (y, \boldsymbol{w})\big) \mapsto \langle x \mid y \rangle + \sum_{i=1}^m \langle v_i \mid w_i \rangle\,, \tag{2.45}$$

and

$$|||\colon|||\colon (x, \boldsymbol{v}) \mapsto \sqrt{\|x\|^2 + \sum_{i=1}^m \|v_i\|^2}. \tag{2.46}$$

Let us define

$$\begin{cases} \boldsymbol{B}\colon \mathcal{K} \to \mathcal{K}\colon (x, v_1, \ldots, v_m) \mapsto (Bx + \sum_{i=1}^m L_i^* v_i, -L_1 x + B_1^{-1} v_1, \ldots, -L_m x + B_m^{-1} v_m) \\ \boldsymbol{A}\colon \mathcal{K} \to 2^{\mathcal{K}}\colon (x, v_1, \ldots, v_m) \mapsto Ax \times A_1^{-1} v_1 \times \ldots, \times A_m^{-1} v_m. \end{cases} \tag{2.47}$$

It is shown in [8, Eq. (3.12)] and [8, Eq. (3.13)] that under the condition (2.40), $\mathrm{zer}(\boldsymbol{A} + \boldsymbol{B}) \neq \varnothing$. Furthermore, [8, Eq. (3.21)] and [8, Eq. (3.22)] yield

$$(\overline{x}, \overline{v}_1, \ldots, \overline{v}_m) \in \mathrm{zer}(\boldsymbol{A} + \boldsymbol{B}) \Rightarrow \overline{x} \text{ solves (2.41) and } (\overline{v}_1, \ldots, \overline{v}_m) \text{ solves (2.42)}. \tag{2.48}$$

It is show in [8] that $\boldsymbol{B}$ is monotone and μ-Lipschitzian and using [1, Proposition 20.23], $\boldsymbol{A}$ is also a maximally monotone operator. Furthermore, it follows from [1, Proposition 23.16] that

$$(\forall \boldsymbol{x} = (x, v_1, \ldots, v_m) \in \mathcal{K})(\forall \gamma \in]0, +\infty[)\ J_{\gamma \boldsymbol{A}} \boldsymbol{x} = \big(J_{\gamma A} x, J_{\gamma A_1^{-1}} v_1, \ldots J_{\gamma A_m^{-1}} v_m\big), \tag{2.49}$$

For every $n \in \mathbb{N}$, set $\boldsymbol{x}_n = (x_n, v_{1,n}, \ldots, v_{m,n})$. Then the proposed algorithm can be rewritten in the space $\mathcal{K}$ as follows

$$\boldsymbol{x}_{n+1} = J_{\gamma \boldsymbol{A}}(\boldsymbol{x}_n - \gamma \boldsymbol{B}(2\boldsymbol{x}_n - \boldsymbol{x}_{n-1})). \tag{2.50}$$

In view of Theorem 2.2(ii), $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ converges weakly to $\overline{\boldsymbol{x}} = (\overline{x}, \overline{v}_1, \ldots, \overline{v}_m)$ in $\mathrm{zer}(\boldsymbol{A} + \boldsymbol{B})$. By (2.48), it follows that $x_n \rightharpoonup \overline{x}$ solves (2.41) and $(v_{1,n}, \ldots, v_{m,n}) \rightharpoonup (\overline{v}_1, \ldots, \overline{v}_m)$ solves (2.42). □

Remark 2.6 Here are some remarks:

(i) The structured primal-dual monotone inclusions (2.41)-(2.42) is firstly introduced in [8] and then [2, 24, 28, 11, 26]. Various special can be found in the literature [8, 20, 23].

(ii) The iteration (2.44) is different from the one in [8] and (2.44) requires only one call of $B, (B_i)_{1\leq i\leq m}, (L_i)_{1\leq i\leq m}$ per itearation.

(iii) Even when $B, (B_i)_{1\leq i\leq m}$ are restricted to be cocoercive, (2.44) is different from the one in [27].

(iv) Using the same idea as in [8], concretes applications to minimization problem involving the parallel sums are straightforward and we omit them here.

3 Numerical experiment

We provide an application of the proposed method to the image denoising by the total variation which was investigated in [9, Problem 4.16]. Let $z_0 \in \mathbb{R}^{N\times N}$ be the ideal image and z be its corrupted observation of the form

$$z = z_0 + w, \tag{3.1}$$

where w is the realization of a noise process. The goal is to recover z_0 from z. Let us recall some notations in [9, Section 4] and the references therein. The discrete gradient operator is defined by

$$\nabla\colon \mathbb{R}^{N\times N} \to \mathbb{R}^{N\times N} \oplus \mathbb{R}^{N\times N} \colon (x_{k,l})_{1\leq k,l\leq N} \mapsto \big(\eta^{(1)}_{k,l}, \eta^{(2)}_{k,l}\big)_{1\leq k,l\leq N}, \tag{3.2}$$

where

$$(\forall (k,l) \in \{1,\ldots,N\}^2) \quad \begin{cases} \eta^{(1)}_{k,l} &= x_{k+1,l} - x_{k,l}, \quad \text{if } k < N;\\ \eta^{(1)}_{N,l} &= 0;\\ \eta^{(2)}_{k,l} &= x_{k,l+1} - x_{k,l}, \quad \text{if } l < N;\\ \eta^{(2)}_{k,N} &= 0. \end{cases} \tag{3.3}$$

Then the discrete total variation function is defined by

$$\mathrm{tv}\colon \mathbb{R}^{N\times N} \to \mathbb{R}\colon x \mapsto \|\nabla x\|_{2,1} = \sum_{1\leq k,l\leq N} \sqrt{|\eta^{(1)}_{k,l}|^2 + |\eta^{(2)}_{k,l}|^2}. \tag{3.4}$$

Now let X be a nonempty closed convex set that model prior knowledge on the ideal image z_0. To recover z_0 from z, we focus on solving the following problem

$$\underset{x\in X}{\text{minimize}}\; \xi\, \mathrm{tv}(x) + \frac{1}{2}\|x - z\|^2, \tag{3.5}$$

where ξ is a strictly positive number, together with its dual problem

$$\underset{v\in Y}{\text{minimize}}\; \widetilde{\sigma}_X(z + \xi\, \mathrm{div}\, v), \tag{3.6}$$

where $\mathrm{div} = -\nabla^*$ [9, Eq. (4.45)], $\widetilde{\sigma}_X$ is the Moreau envelope of the support function σ_X [9, Eq. (2.13)], and Y is given by

$$Y = \Big\{\big(\eta^{(1)}_{k,l}, \eta^{(2)}_{k,l}\big)_{1\leq k,l\leq N} \in \mathbb{R}^{N\times N} \oplus \mathbb{R}^{N\times N} \mid \max_{1\leq k,l\leq N} \sqrt{|\eta^{(1)}_{k,l}|^2 + |\eta^{(2)}_{k,l}|^2} \leq 1\Big\}. \tag{3.7}$$

It is shown in [9] that the set of solutions to (3.6) is nonempty. Moreover, if $\overline{v}$ is a solution to (3.6), then $x^* = P_X(z + \xi\, \mathrm{div}\, \overline{v})$ is the solution to (3.5). Furthermore, the function $v \mapsto \widetilde{\sigma}_X(z + \xi\, \mathrm{div}\, v)$ is the differentiable convex function with 8ξ-Lipschitz continuous gradient. Therefore, the dual problem can be solved by various methods in the literature. In this section, we will compare the following methods with $\xi = 0.1$ and $X = \big\{x \in \mathbb{R}^{N\times N} \mid (\forall (k,l) \in \{1,\ldots,N\}^2)\ 0 \leq x_{k,l} \leq 1\big\}$.

(i) Tseng's splitting method (FBFS) with $\gamma = 0.99/(8\xi)$.

(ii) Forward-reflected-backward splitting method (FRBS) with $\gamma = 0.49/(8\xi)$.

(iii) Reflected forward-backward splitting method (RFBS) with $\gamma = 0.49/(8\xi)$.

(iv) Dual forward-backward splitting (DFBS) with $\gamma = 1.99/(8\xi)$ [9].

We use the parrotgray image with $N = 256$ and $w = \mathcal{N}(0, 0.1)$. The optimal solution x^* is found by the primal-dual method in [27]. The results, which are implemented by Octave 4.1, are presented in Figures 1, 2 and 3.

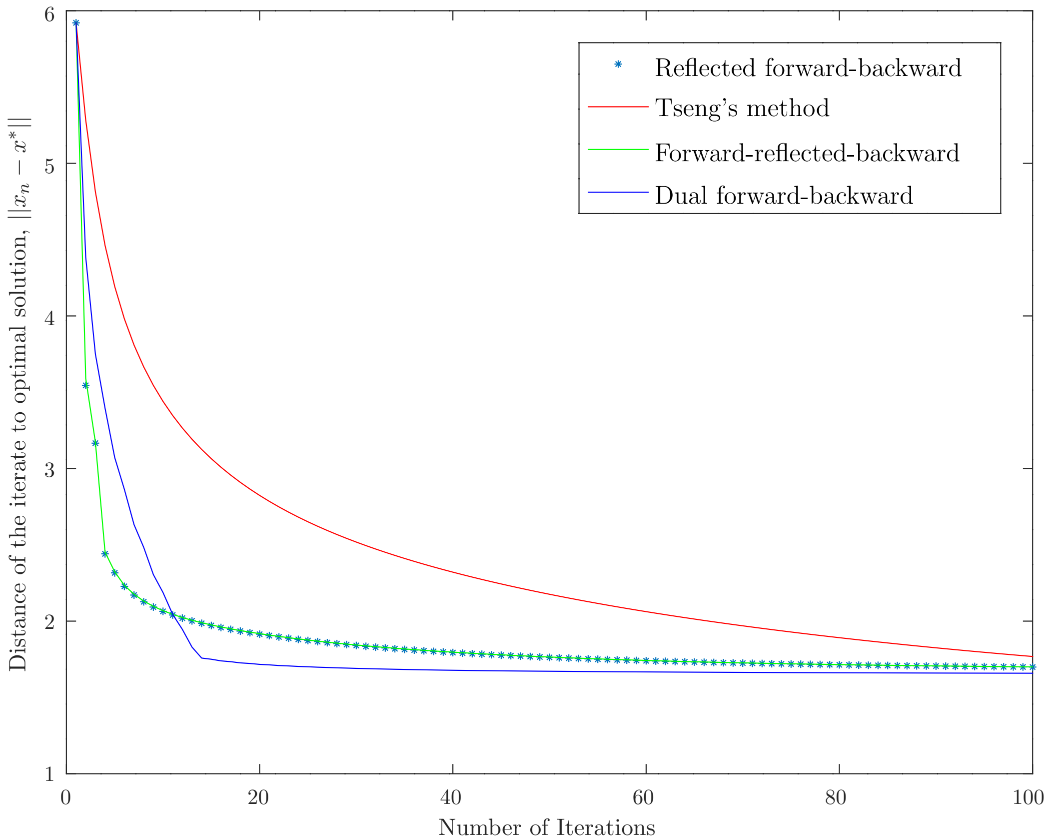

Figure 1: Convergence results for the parrotgray image.

Acknowledgments. We thank the referees for their suggestions and correction which helped to improve the first version of the manuscript. The work of B. Cong Vu and Volkan Cevher were supported by European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no 725594 - time-data).

Figure 2: Original parrotgray image (left) and noisy image (right) with $w = \mathcal{N}(0, 0.1)$.

References

[1] H. H. Bauschke, P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, New York, 2nd ed., 2017.

[2] R. I. Boţ and C. Hendrich, *A Douglas–Rachford type primal-dual method for solving inclusions with mixtures of composite and parallel-sum type monotone operators*, SIAM J. Optim., 23 (2013), pp. 2541–2565.

[3] R. I. Boţ and C. Hendrich, *Convergence Analysis for a Primal-Dual Monotone + Skew Splitting Algorithm with Applications to Total Variation Minimization*, J. Math. Imaging Vis., 49 (2014), pp. 551–568.

[4] L. M. Briceño-Arias, *Forward–Douglas–Rachford splitting and forward–partial inverse method for solving monotone inclusions*, Optimization, 64 (2015), pp. 1239–1261.

[5] L. M. Briceño-Arias and P. L. Combettes, *A monotone+skew splitting model for composite monotone inclusions in duality*, SIAM J. Optim., 21 (2011), pp. 1230–1250.

Figure 3: Images densoising: RFBS (top left), FRBS (top right), FBFS (bottom left), DFBS (bottom right).

[6] L. M. Briceño-Arias and D. Davis, *Forward-Backward-Half Forward Algorithm for Solving Monotone Inclusions*, SIAM J. Optim., 28 (2018), pp. 2839–2871.

[7] P. L. Combettes, *Systems of structured monotone inclusions: duality, algorithms, and applications*, *SIAM J. Optim.*, 23 (2013), pp. 2420–2447.

[8] P. L. Combettes and J.-C. Pesquet, *Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators*, *Set-Valued Var. Anal.*, 20 (2012), pp. 307–330.

[9] P. L. Combettes, Đ. Dũng and B. C. Vũ, *Dualization of Signal Recovery Problems*, *Set-Valued Var. Anal.*, (18) (2010), pp. 373-404.

[10] D. Davis and W. Yin, *A Three-Operator Splitting Scheme and its Optimization Applications*, Set-Valued Var. Anal., 25 (2017), pp. 829–858.

[11] Đ. Dũng and B. C. Vũ, *A splitting algorithm for system of composite monotone inclusions*, Vietnam J. Maths., 43 (2015), pp. 323-341.

[12] N. Komodakis and J.-C.Pesquet, *Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems*, IEEE Signal processing magazine, 32 (2015), pp. 31-54.

[13] P. Latafat and P. Patrinos, *Asymmetric forward-backward-adjoint splitting for solving monotone inclusions involving three operators*, Comput. Optim. Appl., 68 (2017), pp. 57–93.

[14] P. L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Num. Anal., 16 (1979), pp. 964–979.

[15] Y. Malitsky, *Projected reflected gradient methods for monotone variational inequalities*, SIAM J. Control Optim., 25 (2015), pp. 502–520.

[16] Y. Malitsky, *Proximal extrapolated gradient methods for variational inequalities*, Optim. Methods Softw., 33 (2018), pp. 140–164.

[17] Y. Malitsky, and M. K.Tam, *A Forward-Backward Splitting Method for Monotone Inclusions Without Cocoercivity*, *arXiv preprint.* (2018).

[18] Z. Opial, *Weak convergence of the sequence of successive approximations for nonexpansive mappings*, Bull. Amer. Math. Soc., 73 (1967), pp. 591-597.

[19] G. B. Passty, *Ergodic convergence to a zero of the sum of monotone operators in Hilbert space*, J. Math. Anal. Appl., 72 (1979), pp. 383–390.

[20] M. Q. Pham, L. Duval, C. Chaux and J.-C. Pesquet, *A primal-dual proximal algorithm for sparse template-based adaptive filtering: Application to seismic multiple removal*, IEEE Trans. Signal Process., 62 (2014), pp. 4256–4269.

[21] H. Raguet, *A note on the forward-Douglas-Rachford splitting for monotone inclusion and convex optimization*, Optim. Lett. (2018), https://doi.org/10.1007/s11590-018-1272-8

[22] R. Raguet, J. Fadili and G. Peyré, *Generalized forward-backward splitting*, SIAM J. Imaging Sci., 6 (2013), pp. 1199–1226.

[23] A. Repetti, E. Chouzenoux and J.-C. Pesquet, *A penalized weighted least squares approach for restoring data corrupted with signal-dependent noise*, In Proceedings of the 20th European Signal Processing (SIPCO 2012), 1553-1557, Bucharest, Romania, august 27-31, (2012).

[24] E. K. Ryu and B. C. Vũ, *Finding the Forward-Douglas-Rachford-Forward Method*, 2019.

[25] P. Tseng, *A modified forward-backward splitting method for maximal monotone mappings*, SIAM J. Control Optim., 38 (2000), pp. 431–446.

[26] B. C. Vũ, Almost sure convergence of the forward-backward-forward splitting algorithm, *Optim. Lett.*, 10 (2016), pp. 781–803.

[27] B. C. Vũ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators*, Adv. Comput. Math., 38 (2013), pp. 667–681.

[28] B. C. Vũ, *A variable metric extension of the forward–backward–forward algorithm for monotone operators*, Numer. Funct. Anal. Optim., 34 (2013), pp. 1050–1065.