STOCHASTIC FORWARD-DOUGLAS-RACHFORD SPLITTING FOR MONOTONE INCLUSIONS*

VOLKAN CEVHER†, BẰNG CÔNG VŨ†, AND ALP YURTSEVER†

Abstract. We propose a stochastic Forward-Douglas-Rachford Splitting framework for finding a zero point of the sum of three maximally monotone operators in real separable Hilbert space, where one of the operators is cocoercive. We characterize the rate of convergence in expectation in the case of strongly monotone operators. We provide guidance on step-size sequences that achieve this rate, even if the strong convexity parameter is unknown.

Key words. monotone inclusion, monotone operator, operator splitting, cocoercive, forward-backward algorithm, composite operator, duality, primal-dual algorithm

AMS subject classifications. 47H05, 49M29, 49M27, 90C25

1. Introduction. Forward-backward and Douglas-Rachford splitting methods are two fundamental algorithms for solving monotone inclusion problems in real Hilbert spaces [5]. These methods have also been successfully applied for solving various convex optimization problems from a vast set of applications; see [2, 8, 12, 35, 16, 17, 15] and the references therein.

Recently, a new splitting method appeared in the literature, unifying these two fundamental methods into a generalized forward-backward splitting method [30]. However, it is shown in [9] that the generalized forward-backward splitting is in fact a special instance of a more general splitting method, namely the forward-Douglas-Rachford. Towards this goal, [18] extends the forward-Douglas-Rachford to the three-operator splitting method.

These key results so far remained within the so-called deterministic setting. However, solving monotone inclusions in the stochastic setting is of great interest [13, 14, 29, 32, 33, 34, 37]. In parallel to their deterministic counterparts, stochastic forward-backward splitting methods are proposed in [13, 14, 32]. A stochastic version of the Douglas-Rachford splitting can also be found in [13]. Based on the product space reformulation technique and the work in [16], additional stochastic primal-dual methods are introduced in [14, 29, 34] for solving monotone inclusions involving cocoercive operators.

The objective of this work is to extend these recent deterministic methods involving three operators [18, 9, 30] to the stochastic setting. To this end, we focus on the stochastic version of the three-operator splitting method in [18] for solving the following general problem:

Problem 1. Let β be a strictly positive number, $(\boldsymbol{\mathcal{H}}, \langle \cdot \mid \cdot \rangle)$ be a real separable Hilbert space, $\boldsymbol{A}\colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}}$ and $\boldsymbol{B}\colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}}$ be maximally monotone operators, $\boldsymbol{U}$ be self adjoint and positive definite, and let $\boldsymbol{Q}\colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}$ satisfy

$$(\forall \boldsymbol{x} \in \boldsymbol{\mathcal{H}})(\forall \boldsymbol{y} \in \boldsymbol{\mathcal{H}}) \quad \langle \boldsymbol{x} - \boldsymbol{y} \mid \boldsymbol{Q}\boldsymbol{x} - \boldsymbol{Q}\boldsymbol{y} \rangle \geq \beta \langle \boldsymbol{Q}\boldsymbol{x} - \boldsymbol{Q}\boldsymbol{y} \mid \boldsymbol{U}(\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{Q}\boldsymbol{y}) \rangle .$$

*Submitted to the editors 2 September 2016.

Funding: This work is funded in part by the European Commission under Grant ERC Future Proof, SNF 200021-146750, and SNF CRSII2-147633.

†École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland (volkan.cehver@epfl.ch, bang.vu@epfl.ch, alp.yurtsever@epfl.ch).

Let $\boldsymbol{\mathcal{P}}$ be the set of all points $\boldsymbol{x}$ in $\boldsymbol{\mathcal{H}}$ such that

$$0 \in \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{x} + \boldsymbol{Q}\boldsymbol{x}. \tag{1}$$

Our aim is to find a random vector that is $\boldsymbol{\mathcal{P}}$-valued almost surely.

Problem 1 covers a wide class of primal monotone inclusions, primal-dual monotone inclusions in product space, convex optimization, stochastic optimization, split feasibility, and variational inequalities [4, 6, 5, 12, 17, 23, 36].

The paper is organized as follows. Section 2 introduces the notation and recalls the necessary background in monotone operator theory. Then, section 3 presents the main algorithm and proves its weak almost sure convergence. Finally, section 4 establishes the algorithm's rate of convergence in expectation.

2. Notation, Background and Preliminary results. Throughout the paper, $\mathcal{H}$ is a real separable Hilbert space. We denote by $\langle \cdot \mid \cdot \rangle$ and $\|\cdot\|$ the scalar product and its associated norm in $\mathcal{H}$. The symbols $\rightharpoonup$ and $\to$ denote weak and strong convergence respectively. We denote by $\ell^1_+(\mathbb{N})$ the set of summable sequences in $[0, +\infty[$, and by $\mathcal{B}(\mathcal{H})$ the space of linear operators from $\mathcal{H}$ into itself.

Let $U \in \mathcal{B}(\mathcal{H})$ be a self-adjoint and positive definite operator. Then, we define

$$\langle x \mid y \rangle_U = \langle Ux \mid y \rangle, \quad \text{and} \quad \|x\|_U = \sqrt{\langle Ux \mid x \rangle}.$$

The set of all fixed points of $T\colon \mathcal{H} \to \mathcal{H}$ is

$$\operatorname{Fix}(T) = \{x \in \mathcal{H} \mid Tx = x\}.$$

Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a set-valued operator. The domain and the graph of A are

$$\operatorname{dom} A = \{x \in \mathcal{H} \mid Ax \neq \varnothing\}, \quad \text{and} \quad \operatorname{gra} A = \{(x,u) \in \mathcal{H} \times \mathcal{H} \mid u \in Ax\}.$$

The set of zeros and the range of A are

$$\operatorname{zer} A = \{x \in \mathcal{H} \mid 0 \in Ax\}, \quad \text{and} \quad \operatorname{ran} A = \{u \in \mathcal{H} \mid (\exists x \in \mathcal{H}) u \in Ax\}.$$

The inverse of A is

$$A^{-1}\colon \mathcal{H} \to 2^{\mathcal{H}}\colon u \mapsto \{x \in \mathcal{H} \mid u \in Ax\}.$$

We denote by Id the identity operator of $\mathcal{H}$, then the resolvent of A is defined as

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

The parallel sum of $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ is defined as

$$A \,\square\, B = (A^{-1} + B^{-1})^{-1}.$$

A is a monotone operator if

$$\langle x - y \mid u - v \rangle \geq 0, \qquad \forall (x,u) \in \operatorname{gra} A, \ \ \forall (y,v) \in \operatorname{gra} A.$$

Moreover, it is maximally monotone if there exists no monotone operator $\widetilde{A}\colon \mathcal{H} \to \mathcal{H}$ such that $\operatorname{gra} A \subset \operatorname{gra} \widetilde{A} \neq \operatorname{gra} A$.

Let $\Gamma_0(\mathcal{H})$ be the class of proper lower semicontinuous convex functions from $\mathcal{H}$ to $]-\infty,+\infty]$. For any self-adjoint strongly positive operator $U\in\mathcal{B}(\mathcal{H})$ and $f\in\Gamma_0(\mathcal{H})$, we define the proximal operator as

$$\operatorname{prox}_f^U\colon \mathcal{H}\to\mathcal{H}\colon x\mapsto \underset{y\in\mathcal{H}}{\operatorname{argmin}}\left(f(y)+\frac{1}{2}\|x-y\|_U^2\right).$$

We denote prox_f^I by prox_f for notational simplicity. Note that $\operatorname{prox}_f^U = J_{U^{-1}\partial f}$.

The conjugate function of f is

$$f^*\colon a\mapsto \sup_{x\in\mathcal{H}}\big(\langle a\mid x\rangle - f(x)\big).$$

Note that $\forall f\in\Gamma_0(\mathcal{H})$ and $\forall x\in\operatorname{dom}\partial f$,

$$y\in\partial f(x)\Leftrightarrow x\in\partial f^*(y),$$

or equivalently, $(\partial f)^{-1}=\partial f^*$.

The infimal convolution of the two functions f and g from $\mathcal{H}$ to $]-\infty,+\infty]$ is defined as

$$f\,\square\, g\colon x\mapsto \inf_{y\in\mathcal{H}}\big(f(y)+g(x-y)\big).$$

The strong relative interior of a subset C of $\mathcal{H}$ is the set of points $x\in C$ such that the cone generated by $-x+C$ is a closed vector subspace of $\mathcal{H}$. We refer to [5] for an account of the main results of convex analysis, monotone operator theory, and the theory of nonexpansive operators in the context of Hilbert spaces.

We will use a family of functions $(\varphi_c)_{c\in\mathbb{R}}$, where φ_c for ant $c\in\mathbb{R}$ is defined as

$$\varphi_c\colon\ ]0,+\infty[\,\to\mathbb{R}\colon t\mapsto\begin{cases}(t^c-1)/c & \text{if } c\neq 0;\\ \log t & \text{if } c=0.\end{cases}$$

Let $(\boldsymbol{\Omega},\boldsymbol{\mathcal{F}},\mathsf{P})$ be a probability space. A $\mathcal{H}$-valued random variable is a measurable function $x\colon\boldsymbol{\Omega}\to\mathcal{H}$, where $\mathcal{H}$ is endowed with the Borel σ-algebra. We denote by $\sigma(x)$ the σ-field generated by x. The expectation of a random variable x is denoted by $\mathbf{E}[x]$. The conditional expectation of x given a σ-field $\mathcal{A}\subset\mathcal{F}$ is denoted by $\mathbf{E}[x|\mathcal{A}]$. Given a random variable $y\colon\boldsymbol{\Omega}\to\mathcal{H}$, the conditional expectation of x given y is denoted by $\mathbf{E}[x|y]$. Throughout the text and inside the algorithms, we use the letter $\boldsymbol{r}$ to denote an unbiased estimate. See [24] for more details on probability theory in Hilbert spaces. A $\mathcal{H}$-valued random process is a sequence $(x_n)_{n\in\mathbb{N}}$ of $\mathcal{H}$-valued random variables. The following lemma is a special case of [14, Proposition 2.3].

Lemma 1. *Let C be a non-empty closed subset of $\mathcal{H}$ and let $(x_n)_{n\in\mathbb{N}}$ be a $\mathcal{H}$-valued random process. For every $n\in\mathbb{N}$, set $\mathcal{F}_n=\sigma(x_0,\ldots,x_n)$. Suppose that, for every $x\in C$, there exist $[0,+\infty[$-valued random sequences $(\xi_n(x))_{n\in\mathbb{N}}$, $(\zeta_n(x))_{n\in\mathbb{N}}$ and $(t_n(x))_{n\in\mathbb{N}}$ such that, for every $n\in\mathbb{N}$, $\xi_n(x)$, $\zeta_n(x)$ and $t_n(x)$ are $\mathcal{F}_n$-measurable, $(\zeta_n(x))_{n\in\mathbb{N}}$ and $(t_n(x))_{n\in\mathbb{N}}$ are summable almost surely, and*

$$(\forall n\in\mathbb{N})\quad \mathbf{E}[\|x_{n+1}-x\|^2|\mathcal{F}_n]\le(1+t_n(x))\|x_n-x\|^2+\zeta_n(x)-\xi_n(x)\ \text{almost surely.}$$

Then, the followings hold:

(i) *$(x_n)_{n\in\mathbb{N}}$ is bounded almost surely.*

(ii) *There exists $\widetilde{\Omega}\subset\Omega$ such that $\mathsf{P}(\widetilde{\Omega})=1$ and, $(\|x_n(\omega)-x\|)_{n\in\mathbb{N}}$ converges for every $\omega\in\widetilde{\Omega}$ and $x\in C$.*

(iii) *Suppose that the set of weak cluster points of* $(x_n)_{n\in\mathbb{N}}$ *is a subset of* C *almost surely. Then* $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a* C*-valued random vector almost surely.*

LEMMA 2. [12, Lemma 3.7] *Let* $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ *be maximally monotone, let* $U \in \mathcal{B}(\mathcal{H})$ *be self-adjoint and strongly positive, and let* $\mathcal{G}$ *be the real Hilbert space obtained by endowing* $\mathcal{H}$ *with the scalar product* $(x,y) \mapsto \langle x \mid y\rangle_{U^{-1}}$. *Then, the followings hold:*
(i) $UA\colon \mathcal{G} \to 2^{\mathcal{G}}$ *is maximally monotone.*
(ii) $J_{UA}\colon \mathcal{G} \to \mathcal{G}$ *is firmly nonexpansive.*

LEMMA 3. [18, Lemma 3.2] *Let* $\boldsymbol{U}$ *be self-adjoint and positive definite, and let* γ *be a strictly positive number. Set* $T_2 = J_{\gamma\boldsymbol{UB}}$ *and* $T_1 = J_{\gamma\boldsymbol{UA}}$, *and* $\mathbf{T} = \mathrm{Id} - T_2 + T_1 \circ (2T_2 - \mathrm{Id} - \gamma\boldsymbol{UQ} \circ T_2)$. *Then,* $\mathrm{Fix}(\mathbf{T}) \neq \varnothing$ *whenever* $\mathrm{zer}(\boldsymbol{A}+\boldsymbol{B}+\boldsymbol{Q}) \neq \varnothing$. *Furthermore,*

$$\mathrm{zer}(\boldsymbol{A}+\boldsymbol{B}+\boldsymbol{Q}) = J_{\gamma\boldsymbol{UB}}(\mathrm{Fix}(\mathbf{T})).$$

3. Weak almost sure convergence.

In this section, we propose a stochastic forward-Douglas-Rachford splitting method for solving problem 1, and we provide our main theorem which guarantees the weak almost sure convergence of the proposed algorithm under mild assumptions. Then, we present some important special cases and variations of the proposed algorithm, and we characterize their convergence guarantees.

3.1. Stochastic Forward-Douglas-Rachford splitting method.

Let γ and $(\lambda_n)_{n\in\mathbb{N}}$ be strictly positive numbers, and let $(\boldsymbol{r}_n)_{n\in\mathbb{N}}$ and $\overline{\boldsymbol{x}}_0$ be squared integrable $\boldsymbol{\mathcal{H}}$-valued random vectors. Then, stochastic forward-Douglas-Rachford splitting method (SFDR) applies to the problem 1 as described in Algorithm 1.

Algorithm 1 Stochastic Forward-Douglas-Rachford splitting method (SFDR)

for $n = 0, 1, 2, \ldots$ **do**
  $\boldsymbol{x}_n = J_{\gamma\boldsymbol{UB}}\overline{\boldsymbol{x}}_n$
  $\overline{\boldsymbol{x}}_{n+1} = \overline{\boldsymbol{x}}_n + \lambda_n\big(J_{\gamma\boldsymbol{UA}}(2\boldsymbol{x}_n - \overline{\boldsymbol{x}}_n - \gamma\boldsymbol{U}\boldsymbol{r}_n) - \boldsymbol{x}_n\big)$
end for

Next theorem provides the weak almost sure convergence guarantee of SFDR.

THEOREM 4. *Let* $(\varepsilon,\alpha) \in \left]0,1\right[^2$ *satisfy* $\varepsilon \le \lambda_n \le 2-\varepsilon-\alpha$ *and* $\varepsilon \le \gamma \le \alpha(2\beta-\varepsilon)$, *and suppose that*

$$\mathrm{zer}(\boldsymbol{A}+\boldsymbol{B}+\boldsymbol{Q}) \neq \varnothing. \tag{2}$$

Also assume that the following conditions are satisfied with $\boldsymbol{\mathcal{F}}_n = \sigma(\overline{\boldsymbol{x}}_0,\ldots,\overline{\boldsymbol{x}}_n)$,
(i) $\mathbf{E}[\boldsymbol{r}_n|\boldsymbol{\mathcal{F}}_n] = \boldsymbol{Q}\boldsymbol{x}_n$ *almost surely.*
(ii) $\sum_{n\in\mathbb{N}} \mathbf{E}[\|\boldsymbol{r}_n - \boldsymbol{Q}\boldsymbol{x}_n\|^2|\boldsymbol{\mathcal{F}}_n] < +\infty$ *almost surely.*

Then, the followings hold for some random vectors $\boldsymbol{x}$ *and* $\overline{\boldsymbol{y}}$, *which are* $\boldsymbol{\mathcal{P}}$*-valued and* $\mathrm{Fix}(T)$*-valued respectively.*
(i) $\overline{\boldsymbol{x}}_n \rightharpoonup \overline{\boldsymbol{y}}$ *almost surely.*
(ii) $\boldsymbol{x}_n \rightharpoonup \boldsymbol{x}$ *almost surely.*
(iii) $\boldsymbol{x}_n \to \boldsymbol{x}$ *almost surely, if one of the following conditions is satisfied for some* $\widetilde{\boldsymbol{\Omega}} \subset \boldsymbol{\Omega}$ *with* $\mathsf{P}(\widetilde{\boldsymbol{\Omega}}) = 1$.
  (a) $\boldsymbol{Q}$ *is demiregular at* $\boldsymbol{x}(\omega)$, *for every* $\omega \in \widetilde{\boldsymbol{\Omega}}$.
  (b) $\boldsymbol{A}$ *is uniformly monotone at* $\boldsymbol{x}(\omega)$, *for every* $\omega \in \widetilde{\boldsymbol{\Omega}}$.
  (c) $\boldsymbol{B}$ *is uniformly monotone* $\boldsymbol{x}(\omega)$, *for every* $\omega \in \widetilde{\boldsymbol{\Omega}}$.

Proof. Let T_1 and T_2 be defined as in Lemma 3. Set $P = \operatorname{Id} - T_2$, $R = 2T_2 - \operatorname{Id} - \gamma \boldsymbol{U}\boldsymbol{Q} \circ T_2$, and $Y = \operatorname{Id} - \gamma \boldsymbol{U}\boldsymbol{Q} \circ T_2$. Then we get $\mathbf{T} = P + T_1 \circ R$. For all $n \in \mathbb{N}$, set $R_n = 2T_2 - \operatorname{Id} - \gamma \boldsymbol{U}\boldsymbol{r}_n$, $Y_n = \operatorname{Id} - \gamma \boldsymbol{U}\boldsymbol{r}_n$, and $\mathbf{T}_n = P + T_1 \circ R_n$. Then, $2P + R_n = Y_n$. Hence, Algorithm 1 yields

$$(\forall n \in \mathbb{N}) \quad \overline{\boldsymbol{x}}_{n+1} = \overline{\boldsymbol{x}}_n + \lambda_n (\mathbf{T}_n \overline{\boldsymbol{x}}_n - \overline{\boldsymbol{x}}_n).$$

Let $\overline{\boldsymbol{x}} \in \operatorname{Fix}(\mathbf{T})$. Then, upon setting $\boldsymbol{V} = \boldsymbol{U}^{-1}$, using [5, Corollay 2.3], we obtain

$$(\forall n \in \mathbb{N}) \quad \|\overline{\boldsymbol{x}}_{n+1} - \overline{\boldsymbol{x}}\|_{\boldsymbol{V}}^2 = (1-\lambda_n)\|\overline{\boldsymbol{x}}_n - \overline{\boldsymbol{x}}\|_{\boldsymbol{V}}^2 + \lambda_n \|\mathbf{T}_n \overline{\boldsymbol{x}}_n - \overline{\boldsymbol{x}}\|_{\boldsymbol{V}}^2 - \lambda_n(1-\lambda_n)\|\mathbf{T}_n \overline{\boldsymbol{x}}_n - \overline{\boldsymbol{x}}_n\|_{\boldsymbol{V}}^2. \tag{3}$$

Moreover, upon setting

$$(\forall n \in \mathbb{N}) \quad \xi_n = \left\langle -T_1 \circ R_n \overline{\boldsymbol{x}}_n + T_1 \circ R \overline{\boldsymbol{x}} \mid \gamma \boldsymbol{r}_n - \gamma \boldsymbol{Q} \circ T_2 \overline{\boldsymbol{x}} \right\rangle,$$

and using the firm expansiveness of T_1 and P with respective to the scalar product $\langle \cdot \mid \cdot \rangle_{\boldsymbol{V}}$, we have

$$\begin{aligned}
\|\mathbf{T}_n \overline{\boldsymbol{x}}_n - \overline{\boldsymbol{x}}\|_{\boldsymbol{V}}^2 &= \|\mathbf{T}_n \overline{\boldsymbol{x}}_n - \mathbf{T}\overline{\boldsymbol{x}}\|_{\boldsymbol{V}}^2 \\
&= \|P\overline{\boldsymbol{x}}_n - P\overline{\boldsymbol{x}}\|_{\boldsymbol{V}}^2 + 2\left\langle P\overline{\boldsymbol{x}}_n - P\overline{\boldsymbol{x}} \mid T_1 \circ R_n \overline{\boldsymbol{x}}_n - T_1 \circ R\overline{\boldsymbol{x}} \right\rangle_{\boldsymbol{V}} \\
&\quad + \|T_1 \circ R_n \overline{\boldsymbol{x}}_n - T_1 \circ R\overline{\boldsymbol{x}}\|_{\boldsymbol{V}}^2 \\
&\le \left\langle P\overline{\boldsymbol{x}}_n - P\overline{\boldsymbol{x}} \mid \overline{\boldsymbol{x}}_n - \overline{\boldsymbol{x}} \right\rangle_{\boldsymbol{V}} + \left\langle T_1 \circ R_n \overline{\boldsymbol{x}}_n - T_1 \circ R\overline{\boldsymbol{x}} \mid R_n \overline{\boldsymbol{x}}_n - R\overline{\boldsymbol{x}} \right\rangle_{\boldsymbol{V}} \\
&\quad + 2\left\langle P\overline{\boldsymbol{x}}_n - P\overline{\boldsymbol{x}} \mid T_1 \circ R_n \overline{\boldsymbol{x}}_n - T_1 \circ R\overline{\boldsymbol{x}} \right\rangle_{\boldsymbol{V}} \\
&\le \left\langle P\overline{\boldsymbol{x}}_n - P\overline{\boldsymbol{x}} \mid \overline{\boldsymbol{x}}_n - \overline{\boldsymbol{x}} \right\rangle_{\boldsymbol{V}} + \left\langle T_1 \circ R_n \overline{\boldsymbol{x}}_n - T_1 \circ R\overline{\boldsymbol{x}} \mid Y_n \overline{\boldsymbol{x}}_n - Y\overline{\boldsymbol{x}} \right\rangle_{\boldsymbol{V}} \\
&= \left\langle \mathbf{T}_n \overline{\boldsymbol{x}}_n - \mathbf{T}\overline{\boldsymbol{x}} \mid \overline{\boldsymbol{x}}_n - \overline{\boldsymbol{x}} \right\rangle_{\boldsymbol{V}} + \xi_n.
\end{aligned} \tag{4}$$

Since

$$2\left\langle \mathbf{T}_n \overline{\boldsymbol{x}}_n - \mathbf{T}\overline{\boldsymbol{x}} \mid \overline{\boldsymbol{x}}_n - \overline{\boldsymbol{x}} \right\rangle_{\boldsymbol{V}} = \|\mathbf{T}_n \overline{\boldsymbol{x}}_n - \overline{\boldsymbol{x}}\|_{\boldsymbol{V}}^2 + \|\overline{\boldsymbol{x}}_n - \overline{\boldsymbol{x}}\|_{\boldsymbol{V}}^2 - \|(\operatorname{Id} - \mathbf{T}_n)\overline{\boldsymbol{x}}_n\|_{\boldsymbol{V}}^2,$$

we derive from (4) that

$$\|\mathbf{T}_n \overline{\boldsymbol{x}}_n - \overline{\boldsymbol{x}}\|_{\boldsymbol{V}}^2 \le \|\overline{\boldsymbol{x}}_n - \overline{\boldsymbol{x}}\|_{\boldsymbol{V}}^2 - \|(\operatorname{Id} - \mathbf{T}_n)\overline{\boldsymbol{x}}_n\|_{\boldsymbol{V}}^2 + 2\xi_n. \tag{5}$$

Now, let us estimate ξ_n. We have $-T_1 \circ R_n = \operatorname{Id} - \mathbf{T}_n - T_2$, and $-T_1 \circ R = \operatorname{Id} - \mathbf{T} - T_2$. Hence, it follows that

$$\begin{aligned}
\xi_n &= \left\langle (\operatorname{Id} - \mathbf{T}_n)\overline{\boldsymbol{x}}_n - (\operatorname{Id} - \mathbf{T})\overline{\boldsymbol{x}} \mid \gamma \boldsymbol{r}_n - \gamma \boldsymbol{Q} \circ T_2 \overline{\boldsymbol{x}} \right\rangle - \gamma \left\langle T_2 \overline{\boldsymbol{x}}_n - T_2 \overline{\boldsymbol{x}} \mid \boldsymbol{r}_n - \boldsymbol{Q} \circ T_2 \overline{\boldsymbol{x}} \right\rangle \\
&= \gamma \left\langle (\operatorname{Id} - \mathbf{T}_n)\overline{\boldsymbol{x}}_n \mid \boldsymbol{Q} \circ T_2 \overline{\boldsymbol{x}}_n - \boldsymbol{Q} \circ T_2 \overline{\boldsymbol{x}} \right\rangle + \gamma \left\langle (\operatorname{Id} - \mathbf{T}_n)\overline{\boldsymbol{x}}_n \mid \boldsymbol{r}_n - \boldsymbol{Q} \circ T_2 \overline{\boldsymbol{x}}_n \right\rangle \\
&\quad - \gamma \left\langle T_2 \overline{\boldsymbol{x}}_n - T_2 \overline{\boldsymbol{x}} \mid \boldsymbol{r}_n - \boldsymbol{Q} \circ T_2 \overline{\boldsymbol{x}} \right\rangle \\
&= \gamma \left\langle (\operatorname{Id} - \mathbf{T}_n)\overline{\boldsymbol{x}}_n \mid \boldsymbol{Q} \circ T_2 \overline{\boldsymbol{x}}_n - \boldsymbol{Q} \circ T_2 \overline{\boldsymbol{x}} \right\rangle + \gamma \left\langle (\mathbf{T} - \mathbf{T}_n)\overline{\boldsymbol{x}}_n \mid \boldsymbol{r}_n - \boldsymbol{Q} \circ T_2 \overline{\boldsymbol{x}}_n \right\rangle \\
&\quad + \gamma \left\langle (\operatorname{Id} - \mathbf{T})\overline{\boldsymbol{x}}_n \mid \boldsymbol{r}_n - \boldsymbol{Q} \circ T_2 \overline{\boldsymbol{x}}_n \right\rangle - \gamma \left\langle T_2 \overline{\boldsymbol{x}}_n - T_2 \overline{\boldsymbol{x}} \mid \boldsymbol{r}_n - \boldsymbol{Q} \circ T_2 \overline{\boldsymbol{x}} \right\rangle.
\end{aligned}$$

Then, we can obtain the following estimate:

$$\begin{aligned}
2\gamma \left\langle (\operatorname{Id} - \mathbf{T}_n)\overline{\boldsymbol{x}}_n \mid \boldsymbol{Q} \circ T_2 \overline{\boldsymbol{x}}_n - \boldsymbol{Q} \circ T_2 \overline{\boldsymbol{x}} \right\rangle &\le \alpha \|(\operatorname{Id} - \mathbf{T}_n)\overline{\boldsymbol{x}}_n\|_{\boldsymbol{V}}^2 \\
&\quad + (\gamma^2/\alpha)\|\boldsymbol{U}\boldsymbol{Q} \circ T_2 \overline{\boldsymbol{x}}_n - \boldsymbol{U}\boldsymbol{Q} \circ T_2 \overline{\boldsymbol{x}}\|_{\boldsymbol{V}}^2,
\end{aligned}$$

and

$$\begin{aligned}\langle(\mathbf{T}-\mathbf{T}_n)\overline{\boldsymbol{x}}_n \mid \boldsymbol{r}_n-\boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}_n\rangle &\leq \|T_1\circ R\overline{\boldsymbol{x}}_n - T_1\circ R_n\overline{\boldsymbol{x}}_n\|_{\boldsymbol{V}}\|\boldsymbol{U}\boldsymbol{r}_n-\boldsymbol{U}\boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}_n\|_{\boldsymbol{V}}\\ &\leq \|R_1\overline{\boldsymbol{x}}_n - R_n\overline{\boldsymbol{x}}_n\|_{\boldsymbol{V}}\|\boldsymbol{U}\boldsymbol{r}_n-\boldsymbol{U}\boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}_n\|_{\boldsymbol{V}}\\ &\leq \gamma\|\boldsymbol{U}\boldsymbol{r}_n-\boldsymbol{U}\boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}_n\|_{\boldsymbol{V}}^2\\ &\leq \gamma\|\boldsymbol{r}_n-\boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}_n\|_{\boldsymbol{U}}^2\\ &\leq \gamma\|\boldsymbol{U}\|\|\boldsymbol{r}_n-\boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}_n\|^2.\end{aligned}$$

Therefore, (5) becomes

$$\begin{aligned}\|\mathbf{T}_n\overline{\boldsymbol{x}}_n-\overline{\boldsymbol{x}}\|_{\boldsymbol{V}}^2 &\leq \|\overline{\boldsymbol{x}}_n-\overline{\boldsymbol{x}}\|_{\boldsymbol{V}}^2-(1-\alpha)\|(\mathrm{Id}-\mathbf{T}_n)\overline{\boldsymbol{x}}_n\|_{\boldsymbol{V}}^2+\zeta_n\\ &\quad+(\gamma^2/\alpha)\|\boldsymbol{U}\boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}_n-\boldsymbol{U}\boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}\|_{\boldsymbol{V}}^2+2\gamma^2\|\boldsymbol{U}\|\|\boldsymbol{r}_n-\boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}_n\|^2,\end{aligned}\tag{6}$$

where we set

$$\zeta_n = 2\gamma\left\langle(\mathrm{Id}-\mathbf{T})\overline{\boldsymbol{x}}_n \mid \boldsymbol{r}_n-\boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}_n\right\rangle - 2\gamma\left\langle T_2\overline{\boldsymbol{x}}_n-T_2\overline{\boldsymbol{x}} \mid \boldsymbol{r}_n-\boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}\right\rangle.$$

Since $\mathbf{T}$ and T_2 are continuous, $\overline{\boldsymbol{x}}_n$ is $\boldsymbol{\mathcal{F}}_n$-measurable, and $(\mathrm{Id}-\mathbf{T})\overline{\boldsymbol{x}}_n$ and $T_2\overline{\boldsymbol{x}}_n-T_2\overline{\boldsymbol{x}}$ are $\boldsymbol{\mathcal{F}}_n$-measurable, we have

$$\begin{aligned}\mathbf{E}[\zeta_n|\boldsymbol{\mathcal{F}}_n] &= 2\gamma\mathbf{E}[\langle(\mathrm{Id}-T)\overline{\boldsymbol{x}}_n \mid \boldsymbol{r}_n-\boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}_n\rangle|\boldsymbol{\mathcal{F}}_n]\\ &\qquad - 2\gamma\mathbf{E}[\langle T_2\overline{\boldsymbol{x}}_n-T_2\overline{\boldsymbol{x}} \mid \boldsymbol{r}_n-\boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}\rangle|\boldsymbol{\mathcal{F}}_n]\\ &= 2\gamma\left\langle(\mathrm{Id}-\mathbf{T})\overline{\boldsymbol{x}}_n \mid \mathbf{E}[\boldsymbol{r}_n-\boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}_n|\boldsymbol{\mathcal{F}}_n]\right\rangle\\ &\qquad - 2\gamma\left\langle T_2\overline{\boldsymbol{x}}_n-T_2\overline{\boldsymbol{x}} \mid \mathbf{E}[\boldsymbol{r}_n-\boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}|\boldsymbol{\mathcal{F}}_n]\right\rangle\\ &= -2\gamma\left\langle T_2\overline{\boldsymbol{x}}_n-T_2\overline{\boldsymbol{x}} \mid \boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}_n-\boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}\right\rangle\\ &\leq -2\beta\gamma\|\boldsymbol{U}\boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}_n-\boldsymbol{U}\boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}\|_{\boldsymbol{V}}^2.\end{aligned}$$

Taking conditional expectation of both sides with respect to $\boldsymbol{\mathcal{F}}_n$ in (6), we obtain,

$$\begin{aligned}\mathbf{E}[\|\mathbf{T}_n\overline{\boldsymbol{x}}_n-\overline{\boldsymbol{x}}\|_{\boldsymbol{V}}^2|\boldsymbol{\mathcal{F}}_n] &\leq \|\overline{\boldsymbol{x}}_n-\overline{\boldsymbol{x}}\|_{\boldsymbol{V}}^2-(1-\alpha)\mathbf{E}[\|(\mathrm{Id}-\mathbf{T}_n)\overline{\boldsymbol{x}}_n\|_{\boldsymbol{V}}^2|\boldsymbol{\mathcal{F}}_n]\\ &\quad-\gamma(2\beta-\gamma/\alpha)\|\boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}_n-\boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}\|_{\boldsymbol{U}}^2\\ &\quad+2\gamma^2\mathbf{E}[\|\boldsymbol{U}\|\|\boldsymbol{r}_n-\boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}_n\|^2|\boldsymbol{\mathcal{F}}_n].\end{aligned}$$

Next, we set

$$(\forall n\in\mathbb{N})\quad \tau_{1,n}=\lambda_n(1-\lambda_n)+\lambda_n(1-\alpha)\quad\text{and}\quad \tau_{2,n}=\lambda_n\gamma(2\beta-\gamma/\alpha).$$

Then, $(\forall n\in\mathbb{N})\,\tau_{1,n}\geq\varepsilon^2$ and $\tau_{2,n}\geq\varepsilon^3$. Now inserting (6) to (3), we obtain, for every $n\in\mathbb{N}$,

$$\begin{aligned}&\mathbf{E}[\|\overline{\boldsymbol{x}}_{n+1}-\overline{\boldsymbol{x}}\|_{\boldsymbol{V}}^2|\boldsymbol{\mathcal{F}}_n] \leq \|\overline{\boldsymbol{x}}_n-\overline{\boldsymbol{x}}\|_{\boldsymbol{V}}^2-\tau_{1,n}\mathbf{E}[\|(\mathrm{Id}-\mathbf{T}_n)\overline{\boldsymbol{x}}_n\|_{\boldsymbol{V}}^2|\boldsymbol{\mathcal{F}}_n]\\ &\quad-\tau_{2,n}\|\boldsymbol{Q}\circ T_{2,n}\overline{\boldsymbol{x}}_n-\boldsymbol{Q}\circ T_{2,n}\overline{\boldsymbol{x}}\|_{\boldsymbol{V}}^2+2\gamma^2\mathbf{E}[\|\boldsymbol{U}\|\|\boldsymbol{r}_n-\boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}_n\|^2|\boldsymbol{\mathcal{F}}_n]\\ &\leq \|\overline{\boldsymbol{x}}_n-\overline{\boldsymbol{x}}\|_{\boldsymbol{V}}^2-\varepsilon^3\|\boldsymbol{Q}\circ T_{2,n}\overline{\boldsymbol{x}}_n-\boldsymbol{Q}\circ T_{2,n}\overline{\boldsymbol{x}}\|_{\boldsymbol{U}}^2\\ &\quad-\varepsilon^2\mathbf{E}[\|(\mathrm{Id}-\mathbf{T}_n)\overline{\boldsymbol{x}}_n\|_{\boldsymbol{V}}^2|\boldsymbol{\mathcal{F}}_n]\|+2\gamma^2\mathbf{E}[\|\boldsymbol{U}\|\|\boldsymbol{r}_n-\boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}_n\|^2|\boldsymbol{\mathcal{F}}_n].\end{aligned}$$

Hence, the sequence $(\overline{\boldsymbol{x}}_n)_{n\in\mathbb{N}}$ is a stochastic Fejér monotone with respect to the target set $\mathrm{Fix}(\mathbf{T})$. Therefore, the following estimate holds

$$(\forall n\in\mathbb{N})\quad \begin{cases}(\overline{\boldsymbol{x}}_n)_{n\in\mathbb{N}}\text{ is bounded almost surely;}\\ \mathbf{E}[\|(\mathrm{Id}-\mathbf{T}_n)\overline{\boldsymbol{x}}_n|^2\boldsymbol{\mathcal{F}}_n]=\lambda_n^{-1}\mathbf{E}[\|(\overline{\boldsymbol{x}}_{n+1}-\overline{\boldsymbol{x}}_n)|^2\boldsymbol{\mathcal{F}}_n]\to 0;\\ \boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}_n-\boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}\to 0.\end{cases}\tag{7}$$

(i): Let $\overline{\boldsymbol{y}}$ be a weakly cluster point of $(\overline{\boldsymbol{x}}_n)_{n\in\mathbb{N}}$, i.e there exists a subsequence $(\overline{\boldsymbol{x}}_{k_n})_{n\in\mathbb{N}}$ such that $\overline{\boldsymbol{x}}_{k_n} \rightharpoonup \overline{\boldsymbol{y}}$ almost surely. Moreover, we also have

$$\begin{aligned}\|\overline{\boldsymbol{x}}_n - \mathbf{T}\overline{\boldsymbol{x}}_n\|^2_{\boldsymbol{V}} &= \mathbf{E}[\|\overline{\boldsymbol{x}}_n - \mathbf{T}\overline{\boldsymbol{x}}_n\|^2_{\boldsymbol{V}}|\boldsymbol{\mathcal{F}}_n]\\ &\leq 2\mathbf{E}[\|\overline{\boldsymbol{x}}_n - \mathbf{T}_n\overline{\boldsymbol{x}}_n\|^2_{\boldsymbol{V}}|\boldsymbol{\mathcal{F}}_n] + 2\mathbf{E}[\|\mathbf{T}\overline{\boldsymbol{x}}_n - \mathbf{T}_n\overline{\boldsymbol{x}}_n\|^2_{\boldsymbol{V}}|\boldsymbol{\mathcal{F}}_n]\\ &\to 0,\end{aligned}$$

which implies that $\overline{\boldsymbol{x}}_n - \mathbf{T}\overline{\boldsymbol{x}}_n \to 0$ almost surely. To sum up, we have

$$\overline{\boldsymbol{x}}_{k_n} \rightharpoonup \overline{\boldsymbol{y}} \quad \text{and} \quad \overline{\boldsymbol{x}}_{k_n} - \mathbf{T}\overline{\boldsymbol{x}}_{k_n} \to 0 \quad \text{almost surely.}$$

Therefore, by demi-closed principle, $\overline{\boldsymbol{y}} \in \operatorname{Fix}(\mathbf{T})$ almost surely. Now, using Lemma 1, $(\overline{\boldsymbol{x}}_n)_{n\in\mathbb{N}}$ converges weaky to a random vector $\overline{\boldsymbol{y}}$, $\operatorname{Fix}(\mathbf{T})$-valued almost surely.

(ii) Note that T_2 is nonexpansive and $(\overline{\boldsymbol{x}}_n)_{n\in\mathbb{N}}$ is bounded almost surely, and $(\forall n \in \mathbb{N})$ $\boldsymbol{x}_n = T_2\overline{\boldsymbol{x}}_n$, and hence $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ is also bounded almost surely. Let $\boldsymbol{z}$ be a weak cluster point of $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$, i.e., there exists a subsequence $(\boldsymbol{x}_{k_n})_{n\in\mathbb{N}}$ that converges weakly to $\boldsymbol{z}$ almost surely. Since $\operatorname{gra}(\boldsymbol{Q})$ is weak-to-strong sequentially closed and since $\boldsymbol{Q}\boldsymbol{x}_n \to \boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}}$, we get $\boldsymbol{Q}\circ T_2\overline{\boldsymbol{x}} = \boldsymbol{Q}\boldsymbol{z}$ and $\boldsymbol{Q}\boldsymbol{x}_{k_n} \to \boldsymbol{Q}\boldsymbol{z}$. Set $\boldsymbol{a}_n = T_1(2\boldsymbol{x}_n - \overline{\boldsymbol{x}}_n - \gamma\boldsymbol{U}\boldsymbol{Q}\boldsymbol{x}_n)$. Then, we derive from (7) that $\boldsymbol{a}_n - \boldsymbol{x}_n \to 0$ almost surely. We have

$$\begin{cases}(\boldsymbol{x}_{k_n}, \boldsymbol{Q}\boldsymbol{x}_{k_n}) \in \operatorname{gra}(\boldsymbol{Q}), \quad \boldsymbol{x}_{k_n} \rightharpoonup \boldsymbol{z}, \quad \boldsymbol{Q}\boldsymbol{x}_{k_n} \to \boldsymbol{Q}\boldsymbol{z};\\ (\boldsymbol{x}_{k_n}, \gamma^{-1}\boldsymbol{V}(\overline{\boldsymbol{x}}_{k_n} - \boldsymbol{x}_{k,n})) \in \operatorname{gra}(\boldsymbol{B}), \quad \gamma^{-1}\boldsymbol{V}(\overline{\boldsymbol{x}}_{k_n} - \boldsymbol{x}_{k,n}) \rightharpoonup \gamma^{-1}\boldsymbol{V}(\overline{\boldsymbol{y}} - \boldsymbol{z});\\ (\boldsymbol{a}_{k_n}, \gamma^{-1}\boldsymbol{V}(\boldsymbol{x}_{k_n} - \boldsymbol{a}_{k_n} + \boldsymbol{x}_{k_n} - \overline{\boldsymbol{x}}_{k_n}) - \boldsymbol{Q}\boldsymbol{x}_{k_n}) \in \operatorname{gra}(\boldsymbol{A});\\ \gamma^{-1}\boldsymbol{V}(\boldsymbol{x}_{k_n} - \boldsymbol{a}_{k_n} + \boldsymbol{x}_{k_n} - \overline{\boldsymbol{x}}_{k_n}) - \boldsymbol{Q}\boldsymbol{x}_{k_n} \rightharpoonup \gamma^{-1}\boldsymbol{V}(\boldsymbol{z} - \overline{\boldsymbol{y}}) - \boldsymbol{Q}\boldsymbol{z}.\end{cases}$$

Therefore, by [5, Proposition 25.5], $\boldsymbol{z} \in \operatorname{zer}(\boldsymbol{A} + \boldsymbol{B} + \boldsymbol{Q})$ and $\gamma^{-1}\boldsymbol{V}(\overline{\boldsymbol{y}} - \boldsymbol{z}) \in \boldsymbol{B}\boldsymbol{z} \Rightarrow \boldsymbol{z} = J_{\gamma\boldsymbol{U}\boldsymbol{B}}\overline{\boldsymbol{y}}$. Since every subsequence of the bounded sequence $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ converges weakly to $J_{\gamma\boldsymbol{U}\boldsymbol{B}}\overline{\boldsymbol{y}}$, we conclude that $\boldsymbol{x}_n \rightharpoonup J_{\gamma\boldsymbol{U}\boldsymbol{B}}\overline{\boldsymbol{y}}$ almost surely.

(iii): Let $\boldsymbol{\Omega}_1$ be the set of all $\omega \in \boldsymbol{\Omega}$ such that $\boldsymbol{x}_n(\omega) \rightharpoonup \boldsymbol{x}(\omega)$ and $\boldsymbol{Q}\boldsymbol{x}_n(\omega) \to \boldsymbol{Q}\boldsymbol{x}(\omega)$, and $\boldsymbol{a}_n(\omega) - \boldsymbol{x}_n(\omega) \to 0$. Then, $\mathsf{P}(\boldsymbol{\Omega}_1) = 1$ and hence $\mathsf{P}(\boldsymbol{\Omega}_1 \cap \widetilde{\boldsymbol{\Omega}}) = 1$.

(iii)(a): For every $\omega \in \boldsymbol{\Omega}_1 \cap \boldsymbol{\Omega}$. We have $\boldsymbol{x}_n(\omega) \rightharpoonup \boldsymbol{x}(\omega)$ and $\boldsymbol{Q}\boldsymbol{x}_n \to \boldsymbol{Q}\boldsymbol{x}(\omega)$. Since $\boldsymbol{Q}$ is demiregular at $\boldsymbol{x}(\omega)$, by definition, we obtain $\boldsymbol{x}_n(\omega) \to \boldsymbol{x}(\omega)$.

(iii)(b): As in the proof of (ii), we have

$$\begin{cases}(\boldsymbol{a}_n, \gamma^{-1}\boldsymbol{V}(\boldsymbol{x}_n - \boldsymbol{a}_n + \boldsymbol{x}_n - \overline{\boldsymbol{x}}_n) - \boldsymbol{Q}\boldsymbol{x}_n) \in \operatorname{gra}(\boldsymbol{A})\\ \gamma^{-1}\boldsymbol{V}(\boldsymbol{x} - \overline{\boldsymbol{y}}) - \boldsymbol{Q}\boldsymbol{x} \in \boldsymbol{A}\boldsymbol{x}.\end{cases} \tag{8}$$

Since $\boldsymbol{A}$ is uniformly monotone at $\boldsymbol{x}(\omega)$, there exists an increasing function, vanishing only at 0, $\phi\colon [0, +\infty[\to [0, +\infty]$ such that

$$\begin{aligned}&\phi(\|(\boldsymbol{a}_n - \boldsymbol{x})(\omega)\|)\\ &\leq \big\langle(\boldsymbol{a}_n - \boldsymbol{x})(\omega) \,|\, \big(\gamma^{-1}\boldsymbol{V}(\boldsymbol{x}_n - \boldsymbol{a}_n + \boldsymbol{x}_n - \overline{\boldsymbol{x}}_n) - \boldsymbol{Q}\boldsymbol{x}_n - \gamma^{-1}\boldsymbol{V}(\boldsymbol{x} - \overline{\boldsymbol{y}}) + \boldsymbol{Q}\boldsymbol{x}\big)(\omega)\big\rangle\\ &\qquad = \langle(\boldsymbol{a}_n - \boldsymbol{x})(\omega) \mid (\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{Q}\boldsymbol{x}_n)(\omega)\rangle\\ &\qquad\quad + \big\langle(\boldsymbol{a}_n - \boldsymbol{x})(\omega) \mid \gamma^{-1}\boldsymbol{V}(\boldsymbol{x}_n - \boldsymbol{a}_n + \boldsymbol{x}_n - \overline{\boldsymbol{x}}_n - \boldsymbol{x} + \overline{\boldsymbol{y}})(\omega)\big\rangle\\ &\qquad = t_{1,n} + t_{2,n},\end{aligned} \tag{9}$$

where we set

$$\begin{cases} t_{1,n} = \langle (\boldsymbol{a}_n - \boldsymbol{x})(\omega) \mid (\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{Q}\boldsymbol{x}_n)(\omega)\rangle \\ t_{2,n} = \big\langle (\boldsymbol{a}_n - \boldsymbol{x})(\omega) \mid \gamma^{-1}\boldsymbol{V}(\boldsymbol{x}_n - \boldsymbol{a}_n + \boldsymbol{x}_{_n} - \overline{\boldsymbol{x}}_n - \boldsymbol{x} + \overline{\boldsymbol{y}})(\omega)\big\rangle \\ t_{3,n} = \big\langle (\boldsymbol{a}_n - \boldsymbol{x})(\omega) \mid \gamma^{-1}\boldsymbol{V}(\boldsymbol{x}_n - \boldsymbol{a}_n)(\omega)\big\rangle \\ t_{4,n} = \big\langle (\boldsymbol{a}_n - \boldsymbol{x}_n)(\omega) \mid \gamma^{-1}\boldsymbol{V}(\boldsymbol{x}_{_n} - \overline{\boldsymbol{x}}_n - \boldsymbol{x} + \overline{\boldsymbol{y}})(\omega)\big\rangle . \end{cases}$$

Let us estimate each term on the right hand side of (9). Since $\boldsymbol{a}_n(\omega) - \boldsymbol{x}(\omega)$ converges weakly to 0, its bounded and since $\boldsymbol{Q}\boldsymbol{x}_n(\omega) \to \boldsymbol{Q}\boldsymbol{x}(\omega)$, we have

$$t_{1,n} = \langle (\boldsymbol{a}_n - \boldsymbol{x})(\omega) \mid (\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{Q}\boldsymbol{x}_n)(\omega)\rangle \leq \|(\boldsymbol{a}_n - \boldsymbol{x})(\omega)\|\, \|\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{Q}\boldsymbol{x}_n)(\omega)\| \to 0.$$

We next consider the second term of (9). We have

$$\begin{aligned} t_{2,n} &= t_{3,n} + \big\langle (\boldsymbol{a}_n - \boldsymbol{x})(\omega) \mid \gamma^{-1}\boldsymbol{V}(\boldsymbol{x}_{_n} - \overline{\boldsymbol{x}}_n - \boldsymbol{x} + \overline{\boldsymbol{y}})(\omega)\big\rangle \\ &= t_{3,n} + t_{4,n} + \big\langle (\boldsymbol{x}_n - \boldsymbol{x})(\omega) \mid \gamma^{-1}\boldsymbol{V}(\boldsymbol{x}_{_n} - \overline{\boldsymbol{x}}_n - \boldsymbol{x} + \overline{\boldsymbol{y}})(\omega)\big\rangle \\ &\leq t_{3,n} + t_{4,n} \\ &\to 0, \end{aligned} \tag{10}$$

where the last inequality follows from $\big\langle (\boldsymbol{x}_n - \boldsymbol{x})(\omega) \mid \gamma^{-1}\boldsymbol{V}(\boldsymbol{x}_{_n} - \overline{\boldsymbol{x}}_n - \boldsymbol{x} + \overline{\boldsymbol{y}})(\omega)\big\rangle \leq 0$ because of the monotonicity of $\boldsymbol{B}$. Therefore, $\phi(\|(\boldsymbol{a}_n - \boldsymbol{x})(\omega)\| \to 0$ and hence $(\boldsymbol{a}_n - \boldsymbol{x})(\omega) \to 0$. Therefore, $\boldsymbol{x}_n \to \boldsymbol{x}$ almost surely.

(iii)(c): Using the strong monotonicity of $\boldsymbol{B}$ and (10), there exists an increasing function $\psi\colon [0,+\infty[\to [0,+\infty]$ vanishing only at 0 such that

$$\begin{aligned} \psi(\|(\boldsymbol{x}_n - \boldsymbol{x})(\omega)\|) &\leq \big\langle (\boldsymbol{x}_n - \boldsymbol{x})(\omega) \mid \gamma^{-1}\boldsymbol{V}(\overline{\boldsymbol{x}}_n - \boldsymbol{x}_n + \boldsymbol{x} - \overline{\boldsymbol{y}})(\omega)\big\rangle \\ &\leq 2|t_{3,n}| + 2|t_{4,n}| \\ &\to 0, \end{aligned}$$

which implies that $(\boldsymbol{x}_n - \boldsymbol{x})(\omega) \to 0$ and hence $\boldsymbol{x}_n - \boldsymbol{x} \to 0$ almost surely. ∎

Corollary 5. *Let $f \in \Gamma_0(\boldsymbol{\mathcal{H}})$, $g \in \Gamma_0(\boldsymbol{\mathcal{H}})$, and let $h\colon \boldsymbol{\mathcal{H}} \to]-\infty,+\infty]$ be a convex differentiable function with a β^{-1}-Lipschitz gradient. Let $\boldsymbol{\mathcal{P}}$ be the set of all solutions $\boldsymbol{x}$ to*

$$\underset{\boldsymbol{x}\in\boldsymbol{\mathcal{H}}}{\text{minimize}}\; f(\boldsymbol{x}) + g(\boldsymbol{x}) + h(\boldsymbol{x}), \tag{11}$$

under the condition

$$0 \in \operatorname{zer}(\partial f + \partial g + \nabla h). \tag{12}$$

Let $(\varepsilon,\alpha) \in]0,1[^2$, let γ be a strictly positive number, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence of strictly positive numbers such that $\varepsilon \leq \lambda_n \leq 2 - \varepsilon - \alpha$ and $\varepsilon \leq \gamma \leq \alpha(2\beta\|\boldsymbol{U}\|^{-1} - \varepsilon)$. Let $(\boldsymbol{r}_n)_{n\in\mathbb{N}}$ be a sequence of squared integrable $\boldsymbol{\mathcal{H}}$-valued random vectors. Let $\overline{\boldsymbol{x}}_0$ be a squared integrable $\boldsymbol{\mathcal{H}}$-valued random vectors.

Algorithm 2 SFDR for sum of three functions

for $n = 0, 1, 2, \ldots$ **do**
  $\boldsymbol{x}_n = \operatorname{prox}^{\boldsymbol{U}^{-1}}_{\gamma f} \overline{\boldsymbol{x}}_n$
  $\overline{\boldsymbol{x}}_{n+1} = \overline{\boldsymbol{x}}_n + \lambda_n\big(\operatorname{prox}^{\boldsymbol{U}^{-1}}_{\gamma g}(2\boldsymbol{x}_n - \overline{\boldsymbol{x}}_n - \gamma\boldsymbol{U}\boldsymbol{r}_n) - \boldsymbol{x}_n\big)$
end for

Suppose that the following conditions are satisfied with $\boldsymbol{\mathcal{F}}_n = \sigma(\overline{\boldsymbol{x}}_0, \ldots, \overline{\boldsymbol{x}}_n)$,

(i) $\mathbf{E}[\boldsymbol{r}_n|\boldsymbol{\mathcal{F}}_n] = \nabla h(\boldsymbol{x}_n)$ *almost surely.*
(ii) $\sum_{n\in\mathbb{N}} \mathbf{E}[\|\boldsymbol{r}_n - \boldsymbol{Q}\boldsymbol{x}_n\|^2|\boldsymbol{\mathcal{F}}_n] < +\infty$ *almost surely.*

Then, the following holds for some random vectors $\boldsymbol{x}$, $\boldsymbol{\mathcal{P}}$*-valued, and* $\overline{\boldsymbol{y}}$, $\mathrm{Fix}(\mathbf{T})$*-valued, where* $\mathbf{T}$ *is defined in Lemma 3 with* $T_2 = \mathrm{prox}_{\gamma f}^{\boldsymbol{U}^{-1}}$ *and* $T_1 = \mathrm{prox}_{\gamma g}^{\boldsymbol{U}^{-1}}$,

(i) $\overline{\boldsymbol{x}}_n \rightharpoonup \overline{\boldsymbol{y}}$ *almost surely.*
(ii) $\boldsymbol{x}_n \rightharpoonup \boldsymbol{x}$ *almost surely.*
(iii) $\boldsymbol{x}_n \to \boldsymbol{x}$ *almost surely, if one of the following conditions is satisfied for some* $\widetilde{\boldsymbol{\Omega}} \subset \boldsymbol{\Omega}$ *with* $\mathsf{P}(\widetilde{\boldsymbol{\Omega}}) = 1$.
   (a) *If* h *is uniformly convex at* $\boldsymbol{x}(\omega)$*, for every* $\omega \in \widetilde{\boldsymbol{\Omega}}$
   (b) *If* f *is uniformly convex at* $\boldsymbol{x}(\omega)$*, for every* $\omega \in \widetilde{\boldsymbol{\Omega}}$.
   (c) *If* g *is uniformly convex* $\boldsymbol{x}(\omega)$*, for every* $\omega \in \widetilde{\boldsymbol{\Omega}}$.

Proof. The conclusions (i)(ii) follow from Theorem 4 with $\boldsymbol{A} = \partial f$, $\boldsymbol{B} = \partial g$, $\boldsymbol{Q} = \nabla h$, and [5, Theorem 16.2], and [5, Propostion 16.5]. The last assertion follows from the fact that if f is uniformly convex at a point in the domain of ∂f, then ∂f is uniformly monotone at that point; and hence, it is demiregular [2, Proposition 2.4(v)]. ∎

Corollary 6. *Let* m *and* β *be, respectively, strictly positive integer and real. For every* $i \in \{1,\ldots,m\}$*, let* $(\mathcal{H}_i, \langle\cdot \mid \cdot\rangle)$ *be real Hilbert space,* $A_i\colon \mathcal{H}_i \to 2^{\mathcal{H}_i}$ *and* $B_i\colon \mathcal{H}_i \to 2^{\mathcal{H}_i}$ *be maximally monotone, and let* $Q_i\colon \mathcal{H}_1 \times \ldots \times \mathcal{H}_m \to \mathcal{H}_i$ *be such that, for every* $\boldsymbol{x} = (x_1,\ldots,x_m)$ *and every* $\boldsymbol{y} = (y_1,\ldots,y_m)$,

$$\sum_{i=1}^{m} \langle Q_i\boldsymbol{x} - Q_i\boldsymbol{y} \mid x_i - y_i\rangle \geq \beta \sum_{i=1}^{m} \|Q_i\boldsymbol{x} - Q_i\boldsymbol{y}\|_{U_i}^2, \tag{13}$$

for some self adjoint positive definite operator U_i *on* $\mathcal{H}_i$*. Suppose that the set* $\boldsymbol{\mathcal{X}}$ *of all point* $\boldsymbol{x} = (x_1,\ldots,x_m)$ *to the following coupled system of inclusion*

$$\begin{cases} 0 & \in A_1x_1 + B_1x_1 + Q_1\boldsymbol{x} \\ \vdots & \\ 0 & \in A_mx_m + B_mx_m + Q_m\boldsymbol{x} \end{cases}$$

is non-empty. For every $i \in \{1,\ldots,m\}$*, let* U_i *be a self adjoint, positive definite operator on* $\mathcal{H}_i$*,* $\overline{x}_{i,0}$ *and* $(r_{i,n})_{n\in\mathbb{N}}$ *be, respectively, a vector and a random process,* $\mathcal{H}_i$*-valued, squared integrable. Let* $(\varepsilon,\alpha) \in \left]0,1\right[^2$*, let* γ *be a strictly positive number, let* $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence of strictly positive numbers such that* $\varepsilon \leq \lambda_n \leq 2 - \varepsilon - \alpha$ *and* $\varepsilon \leq \gamma \leq \alpha(2\beta - \varepsilon)$.

Algorithm 3 SFDR for multivariate monotone inclusions

for $n = 0,1,2,\ldots$ **do**
  $x_{i,n} = J_{\gamma U_i B_i}\overline{x}_{i,n}$
  $\overline{x}_{i,n+1} = \overline{x}_{i,n} + \lambda_n\big(J_{\gamma U_i A_i}(2x_{i,n} - \overline{x}_{i,n} - \gamma U_i r_{i,n}) - \overline{x}_{i,n}\big)$
end for

Suppose that the following conditions are satisfied $\forall i \in \{1,\ldots,m\}$ *with* $\boldsymbol{\mathcal{F}}_n = \sigma(\overline{x}_{i,0},\ldots,\overline{x}_{i,n})_{1\leq i\leq m}$,

(i) $\mathbf{E}[r_{i,n}|\boldsymbol{\mathcal{F}}_n] = Q_i(x_{1,n},\ldots,x_{m,n})$ *almost surely,*
(ii) $\sum_{n\in\mathbb{N}} \mathbf{E}[\|r_{i,n} - Q_i(x_{1,n},\ldots,x_{m,n})\|^2|\boldsymbol{\mathcal{F}}_n] < +\infty$ *almost surely.*

Then, the following hold for some random vectors $\boldsymbol{x}$, $\boldsymbol{\mathcal{X}}$*-valued almost surely.,*

(i) $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ *converges weakly to* $\boldsymbol{x}$ *almost surely.*
(ii) *Suppose that, for every* $\omega \in \widetilde{\boldsymbol{\Omega}} \subset \boldsymbol{\Omega}$ *with* $\mathsf{P}(\widetilde{\boldsymbol{\Omega}}) = 1$, *and*
  (a) $(Q_1, \ldots, Q_m)$ *is demiregular at* $(x_1(\omega), \ldots, x_m(\omega))$, *then, for every* $j \in \{1, \ldots, m\}, x_{j,n} \to x_j$ *almost surely.*
  (b) A_j *or* B_j *is uniformly motone at* $x_j(\omega)$, *for some* $j \in \{1, \ldots, m\}$, *then* $x_{j,n} \to \overline{x}_j$ *almost surely.*

Proof. Define $\boldsymbol{\mathcal{H}} = \mathcal{H}_1 \oplus \ldots \oplus \mathcal{H}_m$ with the scalar product and the norm

$$\langle \boldsymbol{x} \mid \boldsymbol{y} \rangle = \sum_{i=1}^m \langle x_i \mid y_i \rangle \quad \text{and } \|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x} \mid \boldsymbol{x} \rangle}, \tag{14}$$

where $\boldsymbol{x} = (x_1, \ldots, x_m)$ and $\boldsymbol{y} = (y_1, \ldots, y_m)$ denote the generic elements in $\boldsymbol{\mathcal{H}}$. We next define

$$\begin{cases} \boldsymbol{U} \colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}} \colon \boldsymbol{x} \mapsto (U_1 x_1, \ldots, U_m x_m);\\ \boldsymbol{A} \colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}} \colon \boldsymbol{x} \mapsto \bigtimes_{i=1}^m A_i x_i;\\ \boldsymbol{B} \colon \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}} \colon \boldsymbol{x} \mapsto \bigtimes_{i=1}^m B_i x_i;\\ \boldsymbol{Q} \colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}} \colon \boldsymbol{x} \mapsto (Q_1 \boldsymbol{x}, \ldots, Q_m \boldsymbol{x}). \end{cases}$$

Then, $\boldsymbol{U}$ is self-adjoint, positive definite on $\boldsymbol{\mathcal{H}}$, $\boldsymbol{A}$ and $\boldsymbol{B}$ are maximally monotone on $\boldsymbol{\mathcal{H}}$ [5, Proposition 20.23]. Therefore, in view of Lemma 3, $\boldsymbol{UA}$ and $\boldsymbol{UB}$ is maximally monotone with respect to $\langle \cdot \mid \cdot \rangle_{\boldsymbol{U}^{-1}}$. By [5, Propostion 23.16], we also have

$$(\forall \boldsymbol{x} \in \boldsymbol{\mathcal{H}})(\forall \gamma \in \left]0, +\infty\right[) \quad \begin{cases} J_{\gamma \boldsymbol{UA}} \boldsymbol{x} = (J_{\gamma U_i A_i} x_i)_{1\leq i\leq m}\\ J_{\gamma \boldsymbol{UB}} \boldsymbol{x} = (J_{\gamma U_i B_i} x_i)_{1\leq i\leq m}. \end{cases}$$

Moreover, in view of (13) and (14), $\boldsymbol{Q}$ is a β-cocoercive operator. Upon setting,

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \boldsymbol{x}_n = (x_{1,n}, \ldots, x_{m,n})\\ \overline{\boldsymbol{x}}_{1,n} = (\overline{x}_{1,n}, \ldots, \overline{x}_{m,n})\\ \boldsymbol{r}_n = (r_{1,n}, \ldots, r_{m,n}), \end{cases}$$

Algorithm 5 reduces to a special case of Algorithm 1. Moreover, every specific conditions on operators as well as the stepsize γ, the relaxation parameter $(\lambda_n)_{n\in\mathbb{N}}$ are satisfied.

(i): This assertion follows from Theorem 4 (i).

(iii)(a): Suppose that $\boldsymbol{Q}$ is demiregular at $\boldsymbol{x}(\omega)$. By Theorem 4(iii)(a), $\boldsymbol{x}_n \to \boldsymbol{x}$ almost surely, this is equivalent to $(\forall i \in 1, \ldots, m)\ x_{i,n} \to x_i(\omega)$.

(iii)(b) Now, suppose that A_j is uniformly monotone at $x_j(\omega)$, for every $\omega \in \widetilde{\boldsymbol{\Omega}}$. Let $\boldsymbol{\Omega}_2$ be the set of all $\omega \in \boldsymbol{\Omega}$ such that (8) holds. Set $\boldsymbol{\Omega}^* = \boldsymbol{\Omega}_1 \cap \boldsymbol{\Omega}_2 \cap \widetilde{\boldsymbol{\Omega}}$. Then $\mathsf{P}(\boldsymbol{\Omega}^*) = 1$. Fix $\omega \in \boldsymbol{\Omega}$. We can rewrite (8), with $\boldsymbol{a}_n = (a_{i,n})_{1\leq i\leq m}$ and $\boldsymbol{x} = (x_1, \ldots, x_m)$, and, $\overline{\boldsymbol{y}} = (\overline{y}_1, \ldots, \overline{y}_m)$, as

$$\begin{cases} (a_{i,n}, \gamma^{-1} U_i^{-1}(x_{i,n} - a_{i,n} + x_{i,n} - \overline{x}_{i,n}) - C x_{i,n}) \in \operatorname{gra}(A_i),\\ \gamma^{-1} U_i^{-1}(x_i - \overline{y}_i) - C x_i \in A_i x_i^*. \end{cases}$$

Since A_j is uniformly monotone at $x_j(\omega)$, there exists an increasing function, vanishing only at 0, $\phi_j\colon [0,+\infty[\to [0,+\infty]$ such that

$$\begin{aligned}\phi_j(\|(a_{j,n}-x_j)(\omega)\|) &\leq \big\langle (a_{j,n}-x_j)(\omega) | \big(\gamma^{-1}U_j^{-1}(x_{j,n}-a_{j,n}+x_{j,n}-\overline{x}_{j,n})(\omega)\big\rangle \\ &\quad + \big\langle (a_{j,n}-x_j)(\omega) | -Q_j\boldsymbol{x}_n - \gamma^{-1}U_j^{-1}(x_j-\overline{y}_j)+Q_j\boldsymbol{x}\big)(\omega)\big\rangle.\end{aligned}$$

Now using the monotonicity of A_i with $i \neq j$, we get

$$\begin{aligned}0 &\leq \big\langle (a_{i,n}-x_i)(\omega) \mid \big(\gamma^{-1}U_i^{-1}(x_{i,n}-a_{i,n}+x_{i,n}-\overline{x}_n)\big\rangle \\ &\quad + \big\langle (a_{i,n}-x_i)(\omega) \mid -Q_i x_{i,n} - \gamma^{-1}U_i^{-1}(x_i-\overline{y}_i)+\boldsymbol{Q}\boldsymbol{x}\big)(\omega)\big\rangle .\end{aligned}$$

By adding the last two inequalities, we arrive at

$$\begin{aligned}\phi_j(\|(a_{j,n}-x_j)(\omega)\|) &\leq \big\langle (\boldsymbol{a}_n-\boldsymbol{x})(\omega) \mid \big(\gamma^{-1}\boldsymbol{V}(\boldsymbol{x}_n-\boldsymbol{a}_n+\boldsymbol{x}_{_n}-\overline{\boldsymbol{x}}_n)\big\rangle \\ &\quad + \big\langle (\boldsymbol{a}_n-\boldsymbol{x})(\omega) \mid -\boldsymbol{Q}\boldsymbol{x}_n - \gamma^{-1}\boldsymbol{V}(\boldsymbol{x}-\overline{\boldsymbol{y}})+\boldsymbol{Q}\boldsymbol{x}\big)(\omega)\big\rangle \\ &= \langle (\boldsymbol{a}_n-\boldsymbol{x})(\omega) \mid (\boldsymbol{Q}\boldsymbol{x}-\boldsymbol{Q}\boldsymbol{x}_n)(\omega)\rangle \\ &\quad + \big\langle (\boldsymbol{a}_n-\boldsymbol{x})(\omega) \mid \gamma^{-1}\boldsymbol{V}(\boldsymbol{x}_n-\boldsymbol{a}_n+\boldsymbol{x}_{_n}-\overline{\boldsymbol{x}}_n-\boldsymbol{x}+\overline{\boldsymbol{y}})(\omega)\big\rangle \\ &= t_{1,n}+t_{2,n} \\ &\to 0.\end{aligned}$$

Therefore, $a_{j,n}(\omega) \to x_j(\omega)$ and hence $x_{j,n} \to x_j$ almost surely.

In the case when B_j is uniformly monotone at $x_j(\omega)$, by using the same manner as above, there exists an increasing function $\phi_j\colon [0,+\infty[\to [0,+\infty]$ such that

$$\begin{aligned}\psi(\|(x_{j,n}-x_j)(\omega)\|) &\leq \big\langle (\boldsymbol{x}_n-\boldsymbol{x})(\omega) \mid \gamma^{-1}\boldsymbol{V}(\overline{\boldsymbol{x}}_n-\boldsymbol{x}_n+\boldsymbol{x}-\overline{\boldsymbol{y}})(\omega)\big\rangle \\ &\leq 2|t_{3,n}|+2|t_{4,n}| \\ &\to 0,\end{aligned}$$

which implies that $x_{j,n}(\omega) \to x_j(\omega)$ and hence $x_{j,n} \to x_j$ almost surely. ☐

Remark 7. Here are some comments and connections to existing work.

(i) Our results in this section appears to be new, there is no stochastic primal method for finding a zero point of the sum of three operators in the literature. Some existing stochastic primal method are devoted to find a zero of the sum of two operators as in [14, 33, 34, 37].

(ii) In the deterministic setting, where we take $(\forall n \in \mathbb{N})$ $\boldsymbol{r}_n = \boldsymbol{B}\boldsymbol{x}_n$ and $\boldsymbol{U} = \mathrm{Id}$. The Algorithm 1 reduces to [18, Algorithm 1] where their convergence results can be found in [18, Theorem 3.1]. Further connections to two operator splitting methods in [9, 25, 26] can be found.

(iii) When $\boldsymbol{B}$ is a normal cone to closed vector subspace $\boldsymbol{\mathcal{W}}$ and $\boldsymbol{U} = \mathrm{Id}$, we obtain a stochastic version of algorithm in [9, Eq.(3.8)]. Furthermore, when $\boldsymbol{\mathcal{H}}$ is a product space and $\boldsymbol{\mathcal{W}}$ is its diagonal subspace, we obtain a stochastic version of the algorithm in [30].

(iv) In the special case when $\boldsymbol{B}$ is zero and $\boldsymbol{U} = \mathrm{Id}$, then $(\forall n \in \mathbb{N})$ $\boldsymbol{x}_n = \overline{\boldsymbol{x}}_n$, and hence Algorithm 1 reduces to the stochastic forward-backward splitting which is recently investigated in [13, 14, 33].

(v) In the case when $(B_i)_{1\leq i\leq m}$ are zero, (14) was firstly solved by the algorithm in [2] and then in [34] in the stochastic setting.

(vi) In the case of minimization problems, almost sure convergence of the stochastic projected gradient method were also investigated in [4, 6, 27], and of the stochastic proximal gradient method [1, 32].

(vii) We present an application of our framework to the minimization of sum of three convex functions in [38], and provide some numerical examples on Markowitz portfolio optimization and support vector machine classification.

3.2. Composite monotone inclusions involving cocoercive operators. By using the reformulation product space techniques, our result can be applied to solving a wide class of composite monotone inclusions involving cocoercive operators as in [35, 36]. For simple, we provide an application to the following generic problem.

PROBLEM 2. Let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert spaces and $L \in \mathcal{B}(\mathcal{H},\mathcal{G})$, let U and V be self adjoint positive definite operators on H and $\mathcal{G}$, respectively. Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{G} \to 2^{\mathcal{G}}$ be maximally monotone, let $C\colon \mathcal{H} \to \mathcal{H}$ be such that

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \langle x - y \mid Cx - Cy\rangle \geq \mu\|Cx - Cy\|_U^2,$$

for some strictly positive number μ, and let $D\colon \mathcal{G} \to 2^{\mathcal{G}}$ be maximally monotone such that

$$(\forall v \in \mathcal{G})(\forall w \in \mathcal{G}) \quad \big\langle v - w \mid D^{-1}v - D^{-1}w\big\rangle \geq \nu\|D^{-1}v - D^{-1}w\|_V^2,$$

for some strictly positive number ν. Let $z \in \mathcal{H}$ and $r \in \mathcal{G}$. The primal problem is to

$$\text{find } \overline{x} \in \mathcal{H} \text{ such that} \quad z \in A\overline{x} + L^*(B\Box D)(L\overline{x}) + C\overline{x}, \tag{15}$$

and the dual problem is to

$$\text{find } \overline{v}^* \in \mathcal{G} \text{ such that} \quad -r \in -L(A+C)^{-1}(-L^*\overline{v}^*) + B^{-1}\overline{v}^* + D^{-1}\overline{v}^*, \tag{16}$$

We denote by $\mathcal{P}$ and $\mathcal{D}$ the set of solutions to (15) and (16), respectively.

Let γ and $(\lambda_n)_{n\in\mathbb{N}}$ be strictly positive numbers. Let $(r_{1,n})_{n\in\mathbb{N}}$ and $(r_{2,n})_{n\in\mathbb{N}}$ be, respectively, sequences of square integrable $\mathcal{H}$-valued and $\mathcal{G}$-valued random vectors. Let $\overline{x}_0$ and $\overline{v}_0$ be square integrable $\mathcal{H}$-valued and $\mathcal{G}$-valued random vectors.

Algorithm 4 Primal-dual SFDR

for $n = 0, 1, 2, \ldots$ **do**
  $x_n = \overline{x}_n - \gamma U L^* v_n$
  $v_n = \overline{v}_n + \gamma V L x_n$
  $\overline{x}_{n+1} = \overline{x}_n + \lambda_n\big(J_{\gamma UA}(2x_n - \overline{x}_n - \gamma U(r_{1,n} - z)) - x_n\big)$
  $\overline{v}_{n+1} = \overline{v}_n + \lambda_n\big(J_{\gamma VB^{-1}}(2v_n - \overline{v}_n - \gamma V(r_{2,n} + r)) - v_n\big)$
end for

COROLLARY 8. *Under the same conditions of Problem 2. Set $\beta = \min\{\mu, \nu\}$ and suppose that*

$$z \in \operatorname{ran}(A + L^*(B \,\Box\, D)(L \cdot - r) + C).$$

Let $(\varepsilon, \alpha) \in \,]0,1[^2$. Suppose that $\varepsilon \leq \lambda_n \leq 2 - \varepsilon - \alpha$ and $\varepsilon \leq \gamma \leq \alpha(2\beta - \varepsilon)$. Let $(\overline{x}_n, \overline{v}_n)_{n\in\mathbb{N}}$ be sequences generated by Algorithm 4. Suppose that the following conditions are satisfied with $\boldsymbol{\mathcal{F}}_n = \sigma(\overline{x}_0, \overline{v}_0, \ldots, \overline{x}_n, \overline{v}_n)$,

(i) $(\forall n \in \mathbb{N}) \quad \mathbf{E}[r_{1,n}|\boldsymbol{\mathcal{F}}_n] = Cx_n$ *and* $\mathbf{E}[r_{2,n}|\boldsymbol{\mathcal{F}}_n] = D^{-1}v_n$ *almost surely.*

(ii) $\sum_{n\in\mathbb{N}} \mathbf{E}[\|r_{1,n} - Cx_n\|^2|\boldsymbol{\mathcal{F}}_n] + \mathbf{E}[\|r_{2,n} - D^{-1}v_n\|^2|\boldsymbol{\mathcal{F}}_n] < +\infty$ *almost surely.*

Then, the following holds for some random vector (x^*, v^*), $\mathcal{P}\times\mathcal{D}$-valued almost surely.

(i) $(x_n)_{n\in\mathbb{N}}$ *converges weakly to* x^* *almost surely.*
(ii) $(v_n)_{n\in\mathbb{N}}$ *converges weakly to* v^* *almost surely.*
(iii) *Suppose that, for every* $\omega \in \widetilde{\mathbf{\Omega}} \subset \mathbf{\Omega}$ *with* $\mathsf{P}(\widetilde{\mathbf{\Omega}}) = 1$,
  (a) C *is demiregular or* A *is uniformly monotone at* $x^*(\omega)$*. Then,* $x_n \to x^*$ *almost surely.*
  (b) D^{-1} *is demiregular or* B^{-1} *is uniformly convex* $v^*(\omega)$*. Then,* $v_n \to v^*$ *almost surely.*

Proof. Define $\boldsymbol{U}\colon (x,v) \mapsto (Ux, Vv)$ and $\boldsymbol{S}\colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}\colon (x,v) \mapsto (L^*v, -Lx)$. Then $\boldsymbol{U}$ is a self-adjoint, positive definite, $\boldsymbol{S}$ is monotone and skew, hence it is maximally monotone [5, Example 20.30]. Set $(\forall n \in \mathbb{N})$ $\boldsymbol{x}_n = (x_n, v_n), \overline{\boldsymbol{x}}_n = (\overline{x}_n, \overline{v}_n)$. We have

$$\begin{cases} x_n = \overline{x}_n - \gamma U L^* v_n \\ v_n = \overline{v}_n - \gamma V L x_n \end{cases} \Leftrightarrow \begin{cases} \overline{x}_n - x_n = \gamma U L^* v_n \\ \overline{v}_n - v_n = -\gamma V L x_n \end{cases} \Leftrightarrow \overline{\boldsymbol{x}}_n - \boldsymbol{x}_n \in \gamma \boldsymbol{U}\boldsymbol{S}\boldsymbol{x}_n,$$

Therefore, for any $\gamma \in [0, \infty[$,

$$(\forall n \in \mathbb{N}) \quad \boldsymbol{x}_n = J_{\gamma \boldsymbol{U}\boldsymbol{S}} \overline{\boldsymbol{x}}_n.$$

Define $\boldsymbol{A}\colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}\colon (x,v) \mapsto (-z + Ax) \times (r + B^{-1}v)$ and $\boldsymbol{Q}\colon \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}\colon (x,v) \mapsto (Cx, D^{-1}v)$. Then, for every $\boldsymbol{x} = (x,v)$ and $\boldsymbol{y} = (y,w)$ in $\boldsymbol{\mathcal{H}}$, we have

$$\begin{aligned} \langle \boldsymbol{x} - \boldsymbol{y} \mid \boldsymbol{Q}\boldsymbol{x} - \boldsymbol{Q}\boldsymbol{y} \rangle &= \langle x - y \mid Cx - Cy \rangle + \left\langle v - w \mid D^{-1}v - D^{-1}w \right\rangle \\ &\geq \mu \|Cx - Cy\|_U^2 + \nu \|D^{-1}v - D^{-1}w\|_V^2 \\ &\geq \beta \|\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{Q}\boldsymbol{y}\|_{\boldsymbol{U}}^2. \end{aligned}$$

Moreover, it follows from [5, Proposition 20.23 and 23.16] that $\boldsymbol{A}$ is maximally monotone with the resolvent below

$$(\forall \gamma \in]0, +\infty[)(\forall \boldsymbol{x} \in \boldsymbol{\mathcal{H}}) \ J_{\gamma \boldsymbol{U}\boldsymbol{A}} \boldsymbol{x} = (J_{\gamma U A} x, J_{\gamma V B^{-1}} v).$$

Under the condition (2), we obtain $Z = \operatorname{zer}(\boldsymbol{A} + \boldsymbol{S} + \boldsymbol{Q}) \neq \varnothing$. Furthermore, as shown in [11] that

$$(x,v) \in \operatorname{zer}(\boldsymbol{A} + \boldsymbol{S} + \boldsymbol{Q}) \Longrightarrow x \in \mathcal{P} \text{ and } v \in \mathcal{D}.$$

Upon setting $(\forall n \in \mathbb{N})$ $\boldsymbol{r}_n = (r_{1,n}, r_{2,n})$, the sequence $(\overline{x}_n, \overline{v}_n)_{n\in\mathbb{N}}$ defined by Algorithm 4 satisfies

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} \boldsymbol{x}_n = J_{\gamma \boldsymbol{U}\boldsymbol{S}} \overline{\boldsymbol{x}}_n \\ \overline{\boldsymbol{x}}_{n+1} = \overline{\boldsymbol{x}}_n + \lambda_n \big(J_{\gamma \boldsymbol{U}\boldsymbol{A}}(2\boldsymbol{x}_n - \overline{\boldsymbol{x}}_n - \gamma \boldsymbol{U}\boldsymbol{r}_n) - \boldsymbol{x}_n \big). \end{array} \right. \tag{17}$$

(17) is a special case of Algorithm 1 with $\boldsymbol{B} = \boldsymbol{S}$. Let us check the conditions to ensure the convergence of the iteration generated by (17). First, the specific conditions on the operator are satisfied as shown above. Second, our assumptions on the relaxation parameter $(\lambda_n)_{n\in\mathbb{N}}$ and the step size γ are exact the same as in Theorem 4. Third, our conditions on the stochastic gradient $(r_{1,n})_{n\in\mathbb{N}}$ and $(r_{2,n})_{n\in\mathbb{N}}$ show that the specific conditions on $(\boldsymbol{r}_n)_{n\in\mathbb{N}}$ in Theorem 4 are also satisfied. Hence, the conclusions (i) and (ii) follow from Theorem 4.

(iii)(a) Suppose that C is demiregular or A is uniformly monotone at $x^*(\omega)$, for every $\omega \in \widetilde{\mathbf{\Omega}}$. Let $\mathbf{\Omega}_2$ be the set of all $\omega \in \mathbf{\Omega}$ such that (8) holds. Set $\mathbf{\Omega}^* = \mathbf{\Omega}_1 \cap \mathbf{\Omega}_2 \cap \widetilde{\mathbf{\Omega}}$.

Then $\mathsf{P}(\boldsymbol{\Omega}^*) = 1$. For every $\omega \in \boldsymbol{\Omega}^*$, we have $\boldsymbol{Q}\boldsymbol{x}_n(\omega) \to \boldsymbol{Q}\boldsymbol{x}(\omega)$, we have $Cx_n(\omega) \to Cx^*(\omega)$. Since $x_n(\omega) \rightharpoonup x(\omega)$, it follows from the definition of demiregular operator that $x_n(\omega) \to x(\omega)$. Next, suppose that A is uniformly monotone at $x^*(\omega)$. We can rewrite (8), with $\boldsymbol{a}_n = (a_n, b_n)$ and $\boldsymbol{x} = (x^*, v^*)$, and, $\overline{\boldsymbol{y}} = (\overline{y}_1, \overline{y}_2)$, as

$$\begin{cases} (a_n, \gamma^{-1}U^{-1}(x_n - a_n + x_{_n} - \overline{x}_n) - Cx_n) \in \mathrm{gra}(A) \\ \gamma^{-1}U^{-1}(x^* - \overline{y}_1) - Cx^* \in Ax^* \\ (b_n, \gamma^{-1}V^{-1}(v_n - b_n + v_{_n} - \overline{v}_n) - D^{-1}v_n) \in \mathrm{gra}(B^{-1}) \\ \gamma^{-1}V^{-1}(v^* - \overline{y}_2) - D^{-1}v^* \in B^{-1}v^*. \end{cases}$$

Since A is uniformly monotone at $x^*(\omega)$, there exists an increasing function, vanishing only at 0, $\phi\colon\ [0, +\infty[\, \to [0, +\infty]$ such that

$$\begin{aligned} \phi(\|(a_n - x^*)(\omega)\|) \leq & \left\langle (a_n - x)(\omega) \mid \big(\gamma^{-1}U^{-1}(x_n - a_n + x_n - \overline{x}_n)\right\rangle \\ & + \left\langle (a_n - x)(\omega) \mid -Cx_n - \gamma^{-1}U^{-1}(x^* - \overline{y}_1) + Cx^*\big)(\omega)\right\rangle. \end{aligned}$$

Now using the monotonicity of B^{-1}, we get

$$\begin{aligned} 0 \leq & \left\langle (a_n - x)(\omega) \mid \big(\gamma^{-1}U^{-1}(x_n - a_n + x_n - \overline{x}_n)\right\rangle \\ & + \left\langle (a_n - x)(\omega) \mid -Cx_n - \gamma^{-1}U^{-1}(x^* - \overline{y}_1) + Cx^*\big)(\omega)\right\rangle. \end{aligned}$$

By adding the last two inequalities, we arrive at

$$\begin{aligned} \phi(\|(a_n - x)(\omega)\|) &\leq \left\langle (\boldsymbol{a}_n - \boldsymbol{x})(\omega) \mid \big(\gamma^{-1}\boldsymbol{V}(\boldsymbol{x}_n - \boldsymbol{a}_n + \boldsymbol{x}_{_n} - \overline{\boldsymbol{x}}_n)\right\rangle \\ &\quad + \left\langle (\boldsymbol{a}_n - \boldsymbol{x})(\omega) \mid -\boldsymbol{Q}\boldsymbol{x}_n - \gamma^{-1}\boldsymbol{V}(\boldsymbol{x} - \overline{\boldsymbol{y}}) + \boldsymbol{Q}\boldsymbol{x}\big)(\omega)\right\rangle \\ &= \langle (\boldsymbol{a}_n - \boldsymbol{x})(\omega) \mid (\boldsymbol{Q}\boldsymbol{x} - \boldsymbol{Q}\boldsymbol{x}_n)(\omega)\rangle \\ &\quad + \left\langle (\boldsymbol{a}_n - \boldsymbol{x})(\omega) \mid \gamma^{-1}\boldsymbol{V}(\boldsymbol{x}_n - \boldsymbol{a}_n + \boldsymbol{x}_{_n} - \overline{\boldsymbol{x}}_n - \boldsymbol{x} + \overline{\boldsymbol{y}})(\omega)\right\rangle \\ &= t_{1,n} + t_{2,n} \\ &\to 0. \end{aligned}$$

□

Therefore, $a_n(\omega) \to x^*(\omega)$ and hence $x_n \to x^*$ almost surely.

(iii)(b): Using the same argument as the proof of (iii)(b).

A direct consequence of above result is an application to minimization problems.

COROLLARY 9. *Let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert space, and $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Let $f \in \Gamma_0(\mathcal{H})$ and $g \in \Gamma_0(\mathcal{G})$, and let $h\colon \mathcal{H} \to \mathbb{R}$ be a convex differentiable function with μ-Lipschitz continuous gradient, for some $\mu \in \,]0, +\infty[$, and let ℓ be ν-strongly convex, for some $\nu \in \,]0, +\infty[$. Let $z \in \mathcal{H}$ and $r \in \mathcal{G}$. Denote the solution set to the primal problem as $\mathcal{P}_1$:*

$$\underset{x \in \mathcal{H}}{\mathit{minimize}}\, (f(x) - \langle z \mid x\rangle) + (\ell \,\square\, g)(Lx - r) + h(x),$$

and we denote the solution set to the dual problem as $\mathcal{D}_1$:

$$\underset{v \in \mathcal{G}}{\mathit{minimize}}\, (f^* \,\square\, h^*)(z - L^*v) + g^*(v) + \ell^*(v) + \langle v \mid r\rangle.$$

Suppose that

$$z \in \mathrm{ran}(\partial f + L^*(\partial g \,\square\, \partial \ell)(L \cdot - r) + \nabla h).$$

Let γ and $(\lambda_n)_{n\in\mathbb{N}}$ be sequences of strictly positive numbers, let U and V be respectively self adjoint operators on $\mathcal{H}$ and $\mathcal{G}$. Let $(r_{1,n})_{n\in\mathbb{N}}$ and $(r_{1,n})_{n\in\mathbb{N}}$ be, respectively, sequences of square integrable $\mathcal{H}$-valued and $\mathcal{G}$-valued random vectors. Let $\overline{x}_0$ and $\overline{v}_0$ be square integrable $\mathcal{H}$-valued and $\mathcal{G}$-valued random vectors.

Algorithm 5 Primal-dual SFDR for composite minimization problem

for $n = 0, 1, 2, \ldots$ **do**
$x_n = \overline{x}_n - \gamma U L^* v_n$
$v_n = \overline{v}_n + \gamma V L x_n$
$\overline{x}_{n+1} = \overline{x}_n + \lambda_n\big(\operatorname{prox}_{\gamma f}^{U^{-1}}(2x_n - \overline{x}_n - \gamma U(r_{1,n} - z)) - x_n\big)$
$\overline{v}_{n+1} = \overline{v}_n + \lambda_n\big(\operatorname{prox}_{\gamma g^*}^{V^{-1}}(2v_n - \overline{v}_n - \gamma V(r_{2,n} + r)) - v_n\big)$
end for

Let $(\varepsilon,\alpha) \in \,]0,1[^2$ and $\beta = \min\{\mu,\nu\}$. Suppose that $\varepsilon \le \lambda_n \le 2-\varepsilon-\alpha$ and $\varepsilon \le \gamma \le \alpha(2\beta\max\{\|U\|,\|V\|\}^{-1}-\varepsilon)$. Let $(\overline{x}_n,\overline{v}_n)_{n\in\mathbb{N}}$ be sequences generated by Algorithm 5. Suppose that the following conditions are satisfied with $\boldsymbol{\mathcal{F}}_n = \sigma(\overline{x}_0,\overline{v}_0,\ldots,\overline{x}_n,\overline{v}_n)$,

(i) $(\forall n\in\mathbb{N})\quad \mathbf{E}[r_{1,n}|\boldsymbol{\mathcal{F}}_n] = \nabla h(x_n)$ *and* $\mathbf{E}[r_{2,n}|\boldsymbol{\mathcal{F}}_n] = \nabla\ell^*(v_n)$ *almost surely.*

(ii) $\sum_{n\in\mathbb{N}}\mathbf{E}[\|r_{1,n}-\nabla h(x_n)\|^2|\boldsymbol{\mathcal{F}}_n] + \mathbf{E}[\|r_{2,n}-\nabla\ell^*(v_n)\|^2|\boldsymbol{\mathcal{F}}_n] < +\infty$ *almost surely.*

Then, the following holds for some random vector (x^*,v^*), $\mathcal{P}_1\times\mathcal{D}_1$-valued almost surely.

(i) $(x_n)_{n\in\mathbb{N}}$ *converges weakly to* x^*.

(ii) $(v_n)_{n\in\mathbb{N}}$ *converges weakly to* v^*.

(iii) *Suppose that, for every* $\omega\in\widetilde{\boldsymbol{\Omega}}\subset\boldsymbol{\Omega}$ *with* $\mathsf{P}(\widetilde{\boldsymbol{\Omega}}) = 1$,
 (a) h *or* f *is uniformly convex at* $x^*(\omega)$. *Then,* $x_n\to x^*$ *almost surely.*
 (b) ℓ^* *or* g^* *is uniformly convex* $v^*(\omega)$. *Then,* $v_n\to v^*$ *almost surely.*

Proof. Using the same argument as in the proof of [35, Corollary 4.2], the conclusions follow from Corollary 8. ∎

Remark 10. Here are some remarks and comments.

(i) The stochastic primal-dual splitting algorithms in this section appear to be new even in the deterministic setting. In the deterministic setting when $C = 0$ and $D^{-1} = 0$, by taking $(\forall n\in\mathbb{N})$ $r_{1,n} = 0$ and $r_{2,n} = 0$, and $U = \mathrm{Id}$, $V = \mathrm{Id}$, algorithm (4) reduces to a error-free version of [10, Eq. (2.22)]. A variant of this algorithm can be found in [8].

(ii) Our approach follows the reformulation technique in [11] and [35], we reformulate the primal-dual inclusions to the form of (1) and then apply the Algorithm 1 directly.

(iii) Our conditions on the stochastic estimation $(r_{1,n})_{n\in\mathbb{N}}$ and $(r_{2,n})_{n\in\mathbb{N}}$ are used in [33], and they differ from that of [14, 29, 37].

(iv) The algorithms in this section require the inverse of $(\mathrm{Id}+\gamma^2 UL^*VL)$ which is quite simple in some specific applications in [30, 7, 8].

4. Convergence rate. We provide in this section, the convergence rate in expectations of norm squared error of a variant version of Algorithm 1, for the case where either $\boldsymbol{Q}$ or $\boldsymbol{B}$ is strongly monotone. In this case, the problem has a unique solution, say $\boldsymbol{x}^*$. The results obtained share the same convergence rates as existing working in [33] for strongly monotone inclusions.

Let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence of strictly positive number, let $(\boldsymbol{r}_n)_{n\in\mathbb{N}}$ be a sequence of squared integrable $\boldsymbol{\mathcal{H}}$-valued random vectors. Let $\overline{\boldsymbol{x}}_{A,0}$ be a squared integrable

$\mathcal{H}$-valued random vectors, $\overline{x}_{B,0} = J_{\gamma_0 UB}\overline{x}_{A,0}$ and $u_{B,0} = (U\gamma_0)^{-1}(\mathrm{Id} - J_{\gamma_0 UB})\overline{x}_{A,0}$.

Algorithm 6 SFDR for strongly monotone inclusions

for $n = 0, 1, 2, \ldots$ **do**
  $\overline{x}_{B,n+1} = J_{\gamma_n UB}(\overline{x}_{A,n} + \gamma_n u_{B,n})$
  $u_{B,n+1} = \gamma_n^{-1}(\overline{x}_{A,n} - \overline{x}_{B,n+1}) + u_{B,n}$
  $\overline{x}_{A,n+1} = J_{\gamma_{n+1} UA}(\overline{x}_{B,n+1} - \gamma_{n+1} u_{B,n+1} - \gamma_{n+1} U r_{n+1})$
end for

THEOREM 11. *Set $V = U^{-1}$ and suppose that there exist $\mu_Q \in \,]0,+\infty[$ and $\mu_B \in [0,+\infty[$, and $\beta \in [0,+\infty[$, respectively, such that*

$$(\forall (x,y) \in \mathcal{H}^2)(w \in Bx)(v \in By) \quad \langle x - y \mid w - v\rangle \geq \mu_B \|x - y\|_V^2$$

and

$$\langle x - y \mid Qx - Qy\rangle \geq \mu_Q \|x - y\|_V^2. \tag{18}$$

Furthemore, suppose that $(\forall n \in \mathbb{N})\ 0 \leq \gamma_n < \min\{2(1-\eta)\beta, (2\eta\mu_Q)^{-1}\}$, for some $\eta \in \,]0,1[$, and the following conditions are satisfied, for $(\forall n \in \mathbb{N})\mathcal{F}_n = \sigma(\overline{x}_{A,k})_{0\leq k\leq n}$,
(i) $(\forall n \in \mathbb{N}) \quad \mathbf{E}[r_{n+1}|\mathcal{F}_n] = Q\overline{x}_{B,n+1}$ *almost surely.*
(ii) $(\exists c \in [0,+\infty[)(\exists t \in \mathbb{R})(\forall n \in \mathbb{N})\ \sum_{k=0}^n \mathbf{E}[\|r_k - Q\overline{x}_{B,k}\|^2] \leq cn^t$.

Then, the following holds
(i) *For every $n \in \mathbb{N}$,*

$$\begin{aligned}(1 + 2\gamma_n\mu_B)\mathbf{E}[\|\overline{x}_{B,n+1} - x^*\|_V^2|\mathcal{F}_n] + \gamma_n^2\mathbf{E}[\|u_{B,n+1} - x^*\|_V^2|\mathcal{F}_n]\\ \leq (1 - 2\gamma_n\mu_Q\eta)\|\overline{x}_{B,n} - x^*\|_V^2 + \gamma_n^2\|u_{B,n} - x^*\|_V^2 + 2\gamma_n^2\mathbf{E}[\|r_n - Q\overline{x}_{B,n}\|^2|\mathcal{F}_n]\end{aligned} \tag{19}$$

(ii) *For every $n \in \mathbb{N}$, define*

$$\gamma_{n+1} = \big(1 + 2\gamma_k\mu_B\big)^{-1}\Big(-\gamma_n^2\mu_Q\eta + \sqrt{(\gamma_n^2\mu_Q\eta)^2 + (1 + 2\gamma_k\mu_B)\gamma_n^2}\Big),$$

Then, $\mathbf{E}[\|\overline{x}_{B,n} - x^*\|^2] = O(1/n^2) + O(1/n^{2-t})$.
(iii) *Let $\alpha \in \,]0,1]$ and $(\tau_0, c) \in \,]0,+\infty[^2$. Suppose that $(2\mathbf{E}[\|r_k - Q\overline{x}_{B,k}\|^2])_{n\in\mathbb{N}}$ and $(\mathbf{E}[\|u_{B,n} - x^*\|_V^2])_{n\in\mathbb{N}}$ are uniformly bounded by $\tau_0(2\mu_Q c\eta)^2$ (and hence the condition (ii) is satisfied with $t = 1$). Let n_0 be a strictly positive integer be such that $c_0 = 2c\mu_Q\eta \leq n_0^\alpha$ and $c \leq \min\{2(1-\eta)\beta, (2\eta\mu_Q)^{-1}\}n_0^\alpha$. Set $t_0 = 1 - 2^{\alpha-1} \geq 0$ and $(\forall n \in \mathbb{N})\ s_n = \mathbf{E}[\|\overline{x}_{B,n} - x^*\|_V^2]$, and $(\forall n \geq 2n_0)\ \gamma_n = c.n^{-\alpha}$ for some $\alpha \in \,]0,1[$. Then, for every $n \geq 2n_0$,*

$$\begin{aligned}&s_{n+1} \leq\\ &\begin{cases}\Big(\tau_0 c_0^2\varphi_{1-2\alpha}(n) + s_{n_0}\exp\Big(\frac{c_0 n_0^{1-\alpha}}{1-\alpha}\Big)\Big)\exp\Big(\frac{-c_0 t_0 (n+1)^{1-\alpha}}{1-\alpha}\Big) + \frac{\tau_0 2^\alpha c_0}{(n-2)^\alpha} & \text{if } \alpha \in \,]0,1[,\\ s_{n_0}\big(\frac{n_0}{n+1}\big)^{c_0} + \frac{\tau_0 c_0^2}{(n+1)^{c_0}}(1 + \frac{1}{n_0})^{c_0}\varphi_{c_0-1}(n) & \text{if } \alpha = 1.\end{cases}\\ &= \begin{cases}\mathcal{O}(1/n^\alpha) & \text{if } 0 < \alpha \leq 1\\ \mathcal{O}(1/n) & \alpha = 1 \text{ and } c_0 \geq 1.\end{cases}\end{aligned}$$

Proof. (i): We have

$$(\forall n \in \mathbb{N}) \qquad \begin{cases} \boldsymbol{u}_{A,n} = \gamma_n^{-1}(\overline{\boldsymbol{x}}_{B,n} - \overline{\boldsymbol{x}}_{A,n}) - (\boldsymbol{u}_{B,n} + \boldsymbol{U}\boldsymbol{r}_n) \in \boldsymbol{U}\boldsymbol{A}\overline{\boldsymbol{x}}_{A,n} \\ \boldsymbol{u}_{B,n} \in \boldsymbol{U}\boldsymbol{B}\overline{\boldsymbol{x}}_{B,n} \\ \gamma_n(\boldsymbol{u}_{B,n+1} - \boldsymbol{u}_{B,n}) = \overline{\boldsymbol{x}}_{A,n} - \overline{\boldsymbol{x}}_{B,n+1} \\ \gamma_n(\boldsymbol{u}_{A,n} + \boldsymbol{u}_{B,n} + \boldsymbol{U}\boldsymbol{r}_n) = \overline{\boldsymbol{x}}_{B,n} - \overline{\boldsymbol{x}}_{A,n} \\ \gamma_n(\boldsymbol{u}_{B,n+1} + \boldsymbol{u}_{A,n} + \boldsymbol{U}\boldsymbol{r}_n) = \overline{\boldsymbol{x}}_{B,n} - \overline{\boldsymbol{x}}_{B,n+1}. \end{cases}$$

Note that $\boldsymbol{V} = \boldsymbol{U}^{-1}$ and, for every $n \in \mathbb{N}$, set

$$\begin{cases} \chi_n = 2\gamma_n\big(\langle \overline{\boldsymbol{x}}_{A,n} - \boldsymbol{x}^* \mid \boldsymbol{u}_{A,n} + \boldsymbol{U}\boldsymbol{r}_n\rangle_{\boldsymbol{V}} + \langle \overline{\boldsymbol{x}}_{B,n+1} - \boldsymbol{x}^* \mid \boldsymbol{u}_{B,n+1}\rangle_{\boldsymbol{V}}\big) \\ \chi_{1,n} = 2\langle \overline{\boldsymbol{x}}_{B,n+1} - \boldsymbol{x}^* \mid \overline{\boldsymbol{x}}_{B,n} - \overline{\boldsymbol{x}}_{B,n+1}\rangle_{\boldsymbol{V}} \\ \chi_{2,n} = 2\langle \overline{\boldsymbol{x}}_{A,n} - \overline{\boldsymbol{x}}_{B,n+1} \mid \overline{\boldsymbol{x}}_{B,n} - \overline{\boldsymbol{x}}_{A,n}\rangle_{\boldsymbol{V}} \\ \chi_{3,n} = 2\gamma_n\langle \overline{\boldsymbol{x}}_{B,n+1} - \overline{\boldsymbol{x}}_{A,n} \mid \boldsymbol{u}_{B,n} - \boldsymbol{u}_B^*\rangle_{\boldsymbol{V}} \\ \qquad = 2\gamma_n^2\langle \boldsymbol{u}_{B,n} - \boldsymbol{u}_{B,n+1} \mid \boldsymbol{u}_{B,n} - \boldsymbol{u}_B^*\rangle_{\boldsymbol{V}} \\ \chi_{4,n} = 2\gamma_n\langle \overline{\boldsymbol{x}}_{B,n+1} - \overline{\boldsymbol{x}}_{A,n} \mid \boldsymbol{u}_B^*\rangle_{\boldsymbol{V}}. \end{cases}$$

Then, simple calculations show that

$$(20) \qquad \begin{cases} \chi_{2,n} = \|\overline{\boldsymbol{x}}_{B,n} - \overline{\boldsymbol{x}}_{B,n+1}\|_{\boldsymbol{V}}^2 - \|\overline{\boldsymbol{x}}_{A,n} - \overline{\boldsymbol{x}}_{B,n+1}\|_{\boldsymbol{V}}^2 - \|\overline{\boldsymbol{x}}_{B,n} - \overline{\boldsymbol{x}}_{A,n}\|_{\boldsymbol{V}}^2 \\ \chi_{1,n} = \|\overline{\boldsymbol{x}}_{B,n} - \boldsymbol{x}^*\|_{\boldsymbol{V}}^2 - \|\overline{\boldsymbol{x}}_{B,n+1} - \boldsymbol{x}^*\|_{\boldsymbol{V}}^2 - \|\overline{\boldsymbol{x}}_{B,n} - \overline{\boldsymbol{x}}_{B,n+1}\|_{\boldsymbol{V}}^2 \\ \chi_{3,n} = \gamma_n^2\big(\|\boldsymbol{u}_{B,n+1} - \boldsymbol{u}_{B,n}\|_{\boldsymbol{V}}^2 + \|\boldsymbol{u}_{B,n} - \boldsymbol{u}_B^*\|_{\boldsymbol{V}}^2 - \|\boldsymbol{u}_{B,n+1} - \boldsymbol{u}_B^*\|_{\boldsymbol{V}}^2\big) \\ \qquad = \|\overline{\boldsymbol{x}}_{A,n} - \overline{\boldsymbol{x}}_{B,n+1}\|_{\boldsymbol{V}}^2 + \gamma_n^2\big(\|\boldsymbol{u}_{B,n} - \boldsymbol{u}_B^*\|_{\boldsymbol{V}}^2 - \|\boldsymbol{u}_{B,n+1} - \boldsymbol{u}_B^*\|_{\boldsymbol{V}}^2\big). \end{cases}$$

Furthermore, for every $n \in \mathbb{N}$, we can express χ_n as follows.

$$\begin{aligned} \chi_n &= 2\gamma_n\big(\langle \overline{\boldsymbol{x}}_{A,n} - \overline{\boldsymbol{x}}_{B,n+1} \mid \boldsymbol{u}_{A,n} + \boldsymbol{U}\boldsymbol{r}_n\rangle_{\boldsymbol{V}} \\ &\quad + \langle \overline{\boldsymbol{x}}_{B,n+1} - \boldsymbol{x}^* \mid \boldsymbol{u}_{B,n+1} + \boldsymbol{u}_{A,n} + \boldsymbol{U}\boldsymbol{r}_n\rangle_{\boldsymbol{V}}\big) \\ &= 2\gamma_n\langle \overline{\boldsymbol{x}}_{A,n} - \overline{\boldsymbol{x}}_{B,n+1} \mid \boldsymbol{u}_{A,n} + \boldsymbol{U}\boldsymbol{r}_n\rangle_{\boldsymbol{V}} + \chi_{1,n} \\ &= 2\gamma_n\big(\langle \overline{\boldsymbol{x}}_{A,n} - \overline{\boldsymbol{x}}_{B,n+1} \mid \boldsymbol{u}_{A,n} + \boldsymbol{U}\boldsymbol{r}_n + \boldsymbol{u}_{B,n}\rangle_{\boldsymbol{V}} \\ &\quad - \langle \overline{\boldsymbol{x}}_{A,n} - \overline{\boldsymbol{x}}_{B,n+1} \mid \boldsymbol{u}_{B,n}\rangle_{\boldsymbol{V}}\big) + \chi_{1,n} \\ &= \chi_{2,n} + \chi_{1,n} + 2\gamma_n\langle \overline{\boldsymbol{x}}_{B,n+1} - \overline{\boldsymbol{x}}_{A,n} \mid \boldsymbol{u}_{B,n}\rangle_{\boldsymbol{V}} \\ &= \chi_{2,n} + \chi_{1,n} + 2\gamma_n\langle \overline{\boldsymbol{x}}_{B,n+1} - \overline{\boldsymbol{x}}_{A,n} \mid \boldsymbol{u}_{B,n} - \boldsymbol{u}_B^*\rangle_{\boldsymbol{V}} \\ &\quad + 2\gamma_n\langle \overline{\boldsymbol{x}}_{B,n+1} - \overline{\boldsymbol{x}}_{A,n} \mid \boldsymbol{u}_B^*\rangle_{\boldsymbol{V}} \\ &= \chi_{2,n} + \chi_{1,n} + \chi_{3,n} + \chi_{4,n}. \end{aligned}$$

Now, summing the equalities in (20), we obtain,

$$\begin{aligned} \chi_n = \|\overline{\boldsymbol{x}}_{B,n} - \boldsymbol{x}^*\|_{\boldsymbol{V}}^2 - \|\overline{\boldsymbol{x}}_{B,n+1} - \boldsymbol{x}^*\|_{\boldsymbol{V}}^2 - \|\overline{\boldsymbol{x}}_{B,n} - \overline{\boldsymbol{x}}_{A,n}\|_{\boldsymbol{V}}^2 \\ + \gamma_n^2\big(\|\boldsymbol{u}_{B,n} - \boldsymbol{u}_B^*\|_{\boldsymbol{V}}^2 - \|\boldsymbol{u}_{B,n+1} - \boldsymbol{u}_B^*\|_{\boldsymbol{V}}^2\big) + \chi_{4,n}. \end{aligned}$$

It follows from $\boldsymbol{u}_A^* \in \boldsymbol{U}\boldsymbol{A}\boldsymbol{x}^*$ and the monotonicity of $\boldsymbol{U}\boldsymbol{A}$ that

$$\langle \overline{\boldsymbol{x}}_{A,n} - \boldsymbol{x}^* \mid \boldsymbol{u}_{A,n} - \boldsymbol{u}_A^*\rangle_{\boldsymbol{V}} \geq 0$$

and hence

$$\begin{aligned}
\chi_n &= 2\gamma_n\big(\langle \overline{\boldsymbol{x}}_{A,n} - \boldsymbol{x}^* \mid \boldsymbol{u}_{A,n} - \boldsymbol{u}_A^* \rangle_{\boldsymbol{V}} + \langle \overline{\boldsymbol{x}}_{A,n} - \boldsymbol{x}^* \mid \boldsymbol{u}_A^* + \boldsymbol{U}\boldsymbol{r}_n \rangle_{\boldsymbol{V}} \\
&\quad + \langle \overline{\boldsymbol{x}}_{B,n+1} - \boldsymbol{x}^* \mid \boldsymbol{u}_{B,n+1} \rangle_{\boldsymbol{V}} \big) \\
&\geq 2\gamma_n\big(\langle \overline{\boldsymbol{x}}_{A,n} - \boldsymbol{x}^* \mid \boldsymbol{u}_A^* + \boldsymbol{U}\boldsymbol{r}_n \rangle_{\boldsymbol{V}} + \langle \overline{\boldsymbol{x}}_{B,n+1} - \boldsymbol{x}^* \mid \boldsymbol{u}_{B,n+1} \rangle_{\boldsymbol{V}} \big) \\
&= 2\gamma_n\big(\langle \overline{\boldsymbol{x}}_{A,n} - \boldsymbol{x}^* \mid \boldsymbol{u}_A^* + \boldsymbol{U}\boldsymbol{r}_n \rangle_{\boldsymbol{V}} + \langle \overline{\boldsymbol{x}}_{B,n+1} - \boldsymbol{x}^* \mid \boldsymbol{u}_{B,n+1} - \boldsymbol{u}_B^* \rangle_{\boldsymbol{V}} \\
&\quad + \langle \overline{\boldsymbol{x}}_{B,n+1} - \boldsymbol{x}^* \mid \boldsymbol{u}_B^* \rangle_{\boldsymbol{V}} \big) \\
&\geq 2\gamma_n\big(\langle \overline{\boldsymbol{x}}_{A,n} - \boldsymbol{x}^* \mid \boldsymbol{u}_A^* + \boldsymbol{U}\boldsymbol{r}_n \rangle_{\boldsymbol{V}} + \mu_B \|\overline{\boldsymbol{x}}_{B,n+1} - \boldsymbol{x}^*\|_{\boldsymbol{V}}^2 \\
&\quad + \langle \overline{\boldsymbol{x}}_{B,n+1} - \boldsymbol{x}^* \mid \boldsymbol{u}_B^* \rangle_{\boldsymbol{V}} \big),
\end{aligned}$$

where the last inequality follows from the assumption that $\boldsymbol{UB}$ is μ_B-strongly monotone. Set

$$(\forall n \in \mathbb{N}) \quad \boldsymbol{x}_{A,n}^e = J_{\gamma_n \boldsymbol{UA}}((\overline{\boldsymbol{x}}_{B,n} - \gamma_n \boldsymbol{U}\boldsymbol{u}_{B,n} - \gamma_n \boldsymbol{UQ}\overline{\boldsymbol{x}}_{B,n})).$$

Then using the firm non-expansiveness of $J_{\gamma_n \boldsymbol{UA}}$ with respect to the norm $\|\cdot\|_{\boldsymbol{V}}$, we get

$$(\forall n \in \mathbb{N}) \quad \|\boldsymbol{x}_{A,n}^e - \overline{\boldsymbol{x}}_{A,n}\|_{\boldsymbol{V}} \leq \gamma_n \|\boldsymbol{U}(\boldsymbol{Q}\overline{\boldsymbol{x}}_{B,n} - \boldsymbol{r}_n)\|_{\boldsymbol{U}^{-1}} = \gamma_n \|\boldsymbol{Q}\overline{\boldsymbol{x}}_{B,n} - \boldsymbol{r}_n\|_{\boldsymbol{U}}.$$

Let us set

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \chi_{5,n} = \left\langle \overline{\boldsymbol{x}}_{A,n} - \boldsymbol{x}_{A,n}^e \mid \boldsymbol{r}_n - \boldsymbol{Q}\overline{\boldsymbol{x}}_{B,n} \right\rangle \\ \chi_{6,n} = \left\langle \boldsymbol{x}_{A,n}^e - \boldsymbol{x}^* \mid \boldsymbol{r}_n - \boldsymbol{Q}\overline{\boldsymbol{x}}_{B,n} \right\rangle \\ \chi_{7,n} = \chi_{5,n} + \chi_{6,n} = \langle \overline{\boldsymbol{x}}_{A,n} - \boldsymbol{x}^* \mid \boldsymbol{r}_n - \boldsymbol{Q}\overline{\boldsymbol{x}}_{B,n} \rangle . \end{cases}$$

Then, we have

$$\begin{aligned}
(\forall n \in \mathbb{N}) \quad \chi_{5,n} &= \left\langle \overline{\boldsymbol{x}}_{A,n} - \boldsymbol{x}_{A,n}^e \mid \boldsymbol{U}\boldsymbol{r}_n - \boldsymbol{UQ}\overline{\boldsymbol{x}}_{B,n} \right\rangle_{\boldsymbol{V}} \\
&\leq \|\overline{\boldsymbol{x}}_{A,n} - \boldsymbol{x}_{A,n}^e\|_{\boldsymbol{V}} \|\boldsymbol{U}\boldsymbol{r}_n - \boldsymbol{UQ}\overline{\boldsymbol{x}}_{B,n}\|_{\boldsymbol{V}} \\
&\leq \gamma_n \|\boldsymbol{r}_n - \boldsymbol{Q}\overline{\boldsymbol{x}}_{B,n}\|_{\boldsymbol{U}}^2,
\end{aligned}$$

and since $\boldsymbol{x}_{A,n}^e$ is $\boldsymbol{\mathcal{F}}_n$-measurable, we obtain

$$(\forall n \in \mathbb{N}) \quad \mathbf{E}[\chi_{6,n} | \boldsymbol{\mathcal{F}}_n] = \left\langle \boldsymbol{x}_{A,n}^e - \boldsymbol{x}^* \mid \mathbf{E}[\boldsymbol{r}_n - \boldsymbol{Q}\overline{\boldsymbol{x}}_{B,n} | \boldsymbol{\mathcal{F}}_n] \right\rangle = 0.$$

Furthermore, for every $\eta \in]0,1[$, since $\boldsymbol{UQ}$ is β-cocoercive and μ_Q-strongly monotone, we have

$$2\left\langle \overline{\boldsymbol{x}}_{A,n}-\boldsymbol{x}^* \mid \boldsymbol{U}\boldsymbol{r}_n\right\rangle_{\boldsymbol{V}} = 2\left\langle \overline{\boldsymbol{x}}_{A,n}-\boldsymbol{x}^* \mid \boldsymbol{U}\boldsymbol{Q}\overline{\boldsymbol{x}}_{B,n}\right\rangle_{\boldsymbol{V}} + 2\left\langle \overline{\boldsymbol{x}}_{A,n}-\boldsymbol{x}^* \mid \boldsymbol{U}\boldsymbol{r}_n-\boldsymbol{U}\boldsymbol{Q}\overline{\boldsymbol{x}}_{B,n}\right\rangle_{\boldsymbol{V}}$$

(21)

$$\begin{aligned}
&= 2\left\langle \overline{\boldsymbol{x}}_{A,n}-\overline{\boldsymbol{x}}_{B,n} \mid \boldsymbol{U}\boldsymbol{Q}\overline{\boldsymbol{x}}_{B,n}-\boldsymbol{U}\boldsymbol{Q}\boldsymbol{x}^*\right\rangle_{\boldsymbol{V}} + 2\left\langle \overline{\boldsymbol{x}}_{A,n}-\boldsymbol{x}^* \mid \boldsymbol{Q}\boldsymbol{x}^*\right\rangle \\
&\quad + 2\left\langle \overline{\boldsymbol{x}}_{B,n}-\boldsymbol{x}^* \mid \boldsymbol{U}\boldsymbol{Q}\overline{\boldsymbol{x}}_{B,n}-\boldsymbol{U}\boldsymbol{Q}\boldsymbol{x}^*\right\rangle_{\boldsymbol{V}} + 2\chi_{7,n} \\
&\geq \frac{-1}{2\beta(1-\eta)}\|\overline{\boldsymbol{x}}_{A,n}-\overline{\boldsymbol{x}}_{B,n}\|_{\boldsymbol{V}}^2 - 2\beta(1-\eta)\|\boldsymbol{Q}\overline{\boldsymbol{x}}_{B,n}-\boldsymbol{Q}\boldsymbol{x}^*\|_{\boldsymbol{U}}^2 + 2\eta\mu_Q\|\overline{\boldsymbol{x}}_{B,n}-\boldsymbol{x}^*\|_{\boldsymbol{V}}^2 \\
&\quad + 2\beta(1-\eta)\|\boldsymbol{Q}\overline{\boldsymbol{x}}_{B,n}-\boldsymbol{Q}\boldsymbol{x}^*\|_{\boldsymbol{U}}^2 + 2\left\langle \overline{\boldsymbol{x}}_{A,n}-\boldsymbol{x}^* \mid \boldsymbol{Q}\boldsymbol{x}^*\right\rangle + 2\chi_{7,n} \\
&\geq \frac{-1}{2\beta(1-\eta)}\|\overline{\boldsymbol{x}}_{A,n}-\overline{\boldsymbol{x}}_{B,n}\|_{\boldsymbol{V}}^2 + 2\eta\mu_Q\|\overline{\boldsymbol{x}}_{B,n}-\boldsymbol{x}^*\|_{\boldsymbol{V}}^2 + 2\chi_{7,n}
\end{aligned}$$

(22)

$$+ 2\left\langle \overline{\boldsymbol{x}}_{A,n}-\boldsymbol{x}^* \mid \boldsymbol{Q}\boldsymbol{x}^*\right\rangle.$$

Now, inserting (22) into (6), we arrive at

$$\chi_n \geq 2\gamma_n\big(\left\langle \overline{\boldsymbol{x}}_{A,n}-\boldsymbol{x}^* \mid \boldsymbol{u}_A^*+\boldsymbol{U}\boldsymbol{Q}\boldsymbol{x}^*\right\rangle_{\boldsymbol{V}} + \left\langle \overline{\boldsymbol{x}}_{B,n+1}-\boldsymbol{x}^* \mid \boldsymbol{u}_B^*\right\rangle_{\boldsymbol{V}}\big) + 2\gamma_n\chi_{7,n}$$

(23)

$$\frac{-\gamma_n}{2\beta(1-\eta)}\|\overline{\boldsymbol{x}}_{A,n}-\overline{\boldsymbol{x}}_{B,n}\|_{\boldsymbol{V}}^2 + 2\eta\mu_Q\gamma_n\|\overline{\boldsymbol{x}}_{B,n}-\boldsymbol{x}^*\|^2 + 2\mu_B\gamma_n\|\overline{\boldsymbol{x}}_{B,n+1}-\boldsymbol{x}^*\|_{\boldsymbol{V}}^2,$$

since it follows that

$$\begin{aligned}
&2\gamma_n\big(\left\langle \overline{\boldsymbol{x}}_{A,n}-\boldsymbol{x}^* \mid \boldsymbol{u}_A^*+\boldsymbol{U}\boldsymbol{Q}\boldsymbol{x}^*\right\rangle_{\boldsymbol{V}} + \left\langle \overline{\boldsymbol{x}}_{B,n+1}-\boldsymbol{x}^* \mid \boldsymbol{u}_B^*\right\rangle_{\boldsymbol{V}}\big) - \chi_{4,n} \\
&\qquad = 2\gamma_n(\left\langle \overline{\boldsymbol{x}}_{A,n}-\boldsymbol{x}^* \mid \boldsymbol{u}_A^*+\boldsymbol{u}_{\boldsymbol{B}}^*+\boldsymbol{U}\boldsymbol{Q}\boldsymbol{x}^*\right\rangle_{\boldsymbol{V}} \\
&\qquad = 0.
\end{aligned}$$

We derive from (23) and (6) that

$$\begin{aligned}
&(1+2\gamma_k\mu_B)\|\overline{\boldsymbol{x}}_{B,n+1}-\boldsymbol{x}^*\|_{\boldsymbol{V}}^2 + (1-\frac{\gamma_n}{2(1-\eta)\beta})\|\overline{\boldsymbol{x}}_{A,n}-\overline{\boldsymbol{x}}_{B,n}\|_{\boldsymbol{V}}^2 \\
&\qquad + \gamma_n^2\|\boldsymbol{u}_{B,n+1}-\boldsymbol{x}^*\|_{\boldsymbol{V}}^2 \\
&\qquad \leq (1-2\gamma_n\mu_Q\eta)\|\overline{\boldsymbol{x}}_{B,n}-\boldsymbol{x}^*\|_{\boldsymbol{V}}^2 + \gamma_n^2\|\boldsymbol{u}_{B,n}-\boldsymbol{x}^*\|_{\boldsymbol{V}}^2 - 2\gamma_n\chi_{7,n}.
\end{aligned}$$

By our assumptions of $(\gamma_n)_{n\in\mathbb{N}}$, we have $(1-\frac{\gamma_n}{2(1-\eta)\beta}) \geq 0$,

$$(1+2\gamma_k\mu_B)\|\overline{\boldsymbol{x}}_{B,n+1}-\boldsymbol{x}^*\|_{\boldsymbol{V}}^2 + \gamma_n^2\|\boldsymbol{u}_{B,n+1}-\boldsymbol{x}^*\|_{\boldsymbol{V}}^2 \leq (1-2\gamma_n\mu_Q\eta)\|\overline{\boldsymbol{x}}_{B,n}-\boldsymbol{x}^*\|_{\boldsymbol{V}}^2 + \gamma_n^2\|\boldsymbol{u}_{B,n}-\boldsymbol{x}^*\|_{\boldsymbol{V}}^2 - 2\gamma_n\chi_{7,n}.$$

Now, taking the conditonal expectation with respect to $\boldsymbol{\mathcal{F}}_n$, we obtain

$$\begin{aligned}
&(1+2\gamma_n\mu_B)\mathbf{E}[\|\overline{\boldsymbol{x}}_{B,n+1}-\boldsymbol{x}^*\|_{\boldsymbol{V}}^2|\boldsymbol{\mathcal{F}}_n] + \gamma_n^2\mathbf{E}[\|\boldsymbol{u}_{B,n+1}-\boldsymbol{x}^*\|_{\boldsymbol{V}}^2|\boldsymbol{\mathcal{F}}_n] \\
&\leq (1-2\gamma_n\mu_Q\eta)\|\overline{\boldsymbol{x}}_{B,n}-\boldsymbol{x}^*\|_{\boldsymbol{V}}^2 + \gamma_n^2\|\boldsymbol{u}_{B,n}-\boldsymbol{x}^*\|_{\boldsymbol{V}}^2 - 2\gamma_n\mathbf{E}[\chi_{7,n}|\boldsymbol{\mathcal{F}}_n] \\
&= (1-2\gamma_n\mu_Q\eta)\|\overline{\boldsymbol{x}}_{B,n}-\boldsymbol{x}^*\|_{\boldsymbol{V}}^2 + \gamma_n^2\|\boldsymbol{u}_{B,n}-\boldsymbol{x}^*\|_{\boldsymbol{V}}^2 - 2\gamma_n\mathbf{E}[\chi_{5,n}|\boldsymbol{\mathcal{F}}_n] \\
&\leq (1-2\gamma_n\mu_Q\eta)\|\overline{\boldsymbol{x}}_{B,n}-\boldsymbol{x}^*\|_{\boldsymbol{V}}^2 + \gamma_n^2\|\boldsymbol{u}_{B,n}-\boldsymbol{x}^*\|_{\boldsymbol{V}}^2 + 2\gamma_n^2\mathbf{E}[\|\boldsymbol{r}_n-\boldsymbol{Q}\overline{\boldsymbol{x}}_{B,n}\|^2|\boldsymbol{\mathcal{F}}_n],
\end{aligned}$$

which proves (19).

(ii) As indicated in the proof of [18], we have

$$(\forall n \in \mathbb{N}) \quad \gamma_n^{-2}(1+2\gamma_n\mu_B) = \gamma_{n+1}^{-2}(1-2\gamma_{n+1}\mu_Q\eta).$$

and

$$\lim_{n\to\infty}(n+1)\gamma_n = (\eta\mu_Q+\mu_B)^{-1}. \tag{24}$$

Therefore, by dividing both sides of (19) by γ_n^2, and taking the expectations, we obtain

$$\begin{aligned}&\gamma_{n+1}^{-2}(1-2\gamma_{n+1}\mu_Q\eta)\mathbf{E}[\|\overline{\boldsymbol{x}}_{B,n+1}-\boldsymbol{x}^*\|^2_{\boldsymbol{V}}]] + \mathbf{E}[\|\boldsymbol{u}_{B,n+1}-\boldsymbol{x}^*\|^2_{\boldsymbol{V}}]\\ &\le \gamma_n^{-2}(1-2\gamma_n\mu_Q\eta)\mathbf{E}[\|\overline{\boldsymbol{x}}_{B,n}-\boldsymbol{x}^*\|^2_{\boldsymbol{V}}] + \mathbf{E}[\|\boldsymbol{u}_{B,n}-\boldsymbol{x}^*\|^2_{\boldsymbol{V}}] + 2\mathbf{E}[\|\boldsymbol{r}_n-\boldsymbol{Q}\overline{\boldsymbol{x}}_{B,n}\|^2].\end{aligned}$$

Now, summing this inequality from $n=0$ to $n=N$, we get

$$\begin{aligned}\gamma_{N+1}^{-2}(1-2\gamma_{N+1}\mu_Q\eta)\mathbf{E}[\|\overline{\boldsymbol{x}}_{B,N+1}-\boldsymbol{x}^*\|^2_{\boldsymbol{V}}]] &\le \gamma_0^{-2}(1-2\gamma_0\mu_Q\eta)\mathbf{E}\|\overline{\boldsymbol{x}}_{B,0}-\boldsymbol{x}^*\|^2_{\boldsymbol{V}}\\ &+ \mathbf{E}[\|\boldsymbol{u}_{B,0}-\boldsymbol{x}^*\|^2] + \sum_{k=0}^{N}\mathbf{E}[\|\boldsymbol{r}_n-\boldsymbol{Q}\overline{\boldsymbol{x}}_{B,n}\|^2].\end{aligned}$$

In view of (24), (ii) follows from (6).

(iii): Set $\theta_n = 2\gamma_n\mu_Q\eta = c_0 n^{-\alpha}$. Hence

$$s_{n+1} \le (1-\theta_n)s_n + \tau_0\theta_n^2.$$

Hence, the result follows from [32, Lemma 4.4] ◻

COROLLARY 12. *Consider the problem* (11) *under the same conditions on* f, h, g *as well as* (12) *as in Corollary* 5*. Furthermore, assume that* h *is* μ_h*-strongly convex, for some* $\mu_h \in]0,+\infty[$*, and* g *is* μ_g*-strongly convex, for some* $\mu_g \in [0,+\infty[$*. Let* $(\gamma_n)_{n\in\mathbb{N}}$ *be such that* $(\forall n \in \mathbb{N})\ 0 \le \gamma_n < \min\{2(1-\eta)\beta, (2\eta\mu_h)^{-1}\}$*, for some* $\eta \in]0,1[$*, and let* $(\boldsymbol{r}_n)_{n\in\mathbb{N}}$ *be a sequence of squared integrable* $\boldsymbol{\mathcal{H}}$*-valued random vectors. Let* $\overline{\boldsymbol{x}}_{f,0}$ *be a squared integrable* $\boldsymbol{\mathcal{H}}$*-valued random vectors,* $\overline{\boldsymbol{x}}_{g,0} = \operatorname{prox}_{\gamma_0 g}\overline{\boldsymbol{x}}_{f,0}$ *and* $\boldsymbol{u}_{g,0} = (\gamma_0)^{-1}(\mathrm{Id}-\operatorname{prox}_{\gamma_0 g})\overline{\boldsymbol{x}}_{f,0}$.

Algorithm 7 SFDR for strongly convex functions

for $n = 0,1,2,\ldots$ **do**
  $\overline{\boldsymbol{x}}_{g,n+1} = \operatorname{prox}_{\gamma_n g}(\overline{\boldsymbol{x}}_{f,n} + \gamma_n\boldsymbol{u}_{g,n})$
  $\boldsymbol{u}_{g,n+1} = \gamma_n^{-1}(\overline{\boldsymbol{x}}_{f,n} - \overline{\boldsymbol{x}}_{g,n+1}) + \boldsymbol{u}_{g,n}$
  $\overline{\boldsymbol{x}}_{f,n+1} = \operatorname{prox}_{\gamma_{n+1} f}(\overline{\boldsymbol{x}}_{g,n+1} - \gamma_{n+1}\boldsymbol{u}_{g,n+1} - \gamma_{n+1}\boldsymbol{r}_{n+1})$
end for

Suppose that the following conditions are satisfied, for every $n \in \mathbb{N}$, $\boldsymbol{\mathcal{F}}_n = \sigma(\overline{\boldsymbol{x}}_{f,k})_{0\le k\le n}$,

(i) $(\forall n \in \mathbb{N})\quad \mathbf{E}[\boldsymbol{r}_{n+1}|\boldsymbol{\mathcal{F}}_n] = \nabla h(\overline{\boldsymbol{x}}_{g,n+1})$ *almost surely.*

(ii) $(\exists c \in [0,+\infty[)(\exists t \in \mathbb{R})(\forall n \in \mathbb{N})\ \sum_{k=0}^{n}\mathbf{E}[\|r_k - \nabla h(\overline{\boldsymbol{x}}_{g,k})\|^2] \le cn^t$.

Then, the following holds

(i) *For every* $n \in \mathbb{N}$,

$$\begin{aligned}&(1+2\gamma_n\mu_g)\mathbf{E}[\|\overline{\boldsymbol{x}}_{g,n+1}-\boldsymbol{x}^*\|^2|\boldsymbol{\mathcal{F}}_n] + \gamma_n^2\mathbf{E}[\|\boldsymbol{u}_{g,n+1}-\boldsymbol{x}^*\|^2|\boldsymbol{\mathcal{F}}_n]\\ &\le (1-2\gamma_n\mu_h\eta))\|\overline{\boldsymbol{x}}_{g,n}-\boldsymbol{x}^*\|^2 + \gamma_n^2\|\boldsymbol{u}_{g,n}-\boldsymbol{x}^*\|^2 + 2\gamma_n^2\mathbf{E}[\|\boldsymbol{r}_n - \nabla h(\overline{\boldsymbol{x}}_{g,n})\|^2|\boldsymbol{\mathcal{F}}_n]\end{aligned}$$

(ii) *For every* $n \in \mathbb{N}$, *define*

$$\gamma_{n+1} = \frac{-\gamma_n^2\mu_h\eta + \sqrt{(\gamma_n^2\mu_h\eta)^2 + (1+2\gamma_n\mu_g)\gamma_n^2}}{1+2\gamma_n\mu_g},$$

Then, $\mathbf{E}[\|\overline{\boldsymbol{x}}_{g,n} - \boldsymbol{x}^*\|^2] = O(1/n^2) + O(1/n^{2-t})$.

(iii) *Let* $\alpha \in]0,1]$ *and* $(\tau_0, c) \in]0,+\infty[^2$. *Suppose that* $(2\mathbf{E}[\|r_k - \nabla h(\overline{\boldsymbol{x}}_{g,k})\|^2])_{n\in\mathbb{N}}$ *and* $(\mathbf{E}[\|\boldsymbol{u}_{g,n} - \boldsymbol{x}^*\|^2_{\boldsymbol{V}}])_{n\in\mathbb{N}}$ *are uniformly bounded by* $\tau_0(2\mu_Q c\eta)^2$ *(and hence the condition (ii) is satisfied with* $t=1$*). Let* n_0 *be a strictly positive integer be such that* $c_0 = 2c\mu_Q\eta \leq n_0^\alpha$ *and* $c \leq \min\{2(1-\eta)\beta, (2\eta\mu_Q)^{-1}\}n_0^\alpha$. *Set* $(\forall n \in \mathbb{N})$ $s_n = \mathbf{E}[\|\overline{\boldsymbol{x}}_{g,n} - \boldsymbol{x}^*\|^2_{\boldsymbol{V}}]$, *and* $(\forall n \geq 2n_0)$ $\gamma_n = cn^{-\alpha}$ *for some* $\alpha \in]0,1[$. *Then, for every* $n \geq 2n_0$,

$$s_{n+1} \leq \begin{cases} \mathcal{O}(1/n^\alpha) & \text{if } 0 < \alpha < 1 \\ \mathcal{O}(1/n) & \alpha = 1 \text{ and } c_0 \geq 1. \end{cases}$$

Remark 13. In the case when we know that $(\forall n \in \mathbb{N})$ $\overline{\boldsymbol{x}}_{B,n} \in \mathcal{M} \ni \boldsymbol{x}^*$, the condition (18) can be replaced by

$$(\forall (\boldsymbol{x},\boldsymbol{y}) \in \mathcal{M}) \quad \langle \boldsymbol{x} - \boldsymbol{y} \mid \boldsymbol{Q}\boldsymbol{x} - \boldsymbol{Q}\boldsymbol{y}\rangle \geq \mu_Q\|\boldsymbol{x} - \boldsymbol{y}\|^2_{\boldsymbol{V}}.$$

Remark 14. In the case where $\boldsymbol{Q}$ is monotone, and $\boldsymbol{UQ}$ is Lipschitzian with a Lipschitz constant β_0 with respectively the norm $\|\cdot\|_{\boldsymbol{V}}$, and $\mu_B > 0$ (i.e. $\boldsymbol{UB}$ is indeed strongly monotone), and

$$(\forall n \in \mathbb{N}) \quad \gamma_{n+1} = \gamma_n(1 + 2\gamma_n(\mu_B - \gamma_n\beta_0^2/2))^{-1/2},$$

under the same conditions on the stochastic estimate $(\boldsymbol{r}_n)_{n\in\mathbb{N}}$ as in Theorem 11, we also have $\mathbf{E}[\|\overline{\boldsymbol{x}}_{B,n} - \boldsymbol{x}^*\|^2_{\boldsymbol{V}}] = O(1/n^2) + O(1/n^{2-t})$.

Proof. Using (21), since $\boldsymbol{UQ}$ is monotone and Lipschitzian with respect to the norm $\|\cdot\|_{\boldsymbol{V}}$, we have

$$\begin{aligned}
2\left\langle \overline{\boldsymbol{x}}_{A,n} - \boldsymbol{x}^* \mid \boldsymbol{U}\boldsymbol{r}_n\right\rangle_{\boldsymbol{V}} &= 2\left\langle \overline{\boldsymbol{x}}_{A,n} - \overline{\boldsymbol{x}}_{B,n} \mid \boldsymbol{UQ}\overline{\boldsymbol{x}}_{B,n} - \boldsymbol{UQ}\boldsymbol{x}^*\right\rangle_{\boldsymbol{V}} \\
&\quad + 2\left\langle \overline{\boldsymbol{x}}_{A,n} - \boldsymbol{x}^* \mid \boldsymbol{Q}\boldsymbol{x}^*\right\rangle + 2\left\langle \overline{\boldsymbol{x}}_{B,n} - \boldsymbol{x}^* \mid \boldsymbol{UQ}\overline{\boldsymbol{x}}_{B,n} - \boldsymbol{UQ}\boldsymbol{x}^*\right\rangle_{\boldsymbol{V}} + 2\chi_{7,n} \\
&\geq 2\left\langle \overline{\boldsymbol{x}}_{A,n} - \overline{\boldsymbol{x}}_{B,n} \mid \boldsymbol{UQ}\overline{\boldsymbol{x}}_{B,n} - \boldsymbol{UQ}\boldsymbol{x}^*\right\rangle_{\boldsymbol{V}} + 2\left\langle \overline{\boldsymbol{x}}_{A,n} - \boldsymbol{x}^* \mid \boldsymbol{Q}\boldsymbol{x}^*\right\rangle \\
&\quad + 2\chi_{7,n} \\
&\geq \frac{-1}{\gamma_n}\|\overline{\boldsymbol{x}}_{A,n} - \overline{\boldsymbol{x}}_{B,n}\|^2_{\boldsymbol{V}} - \gamma_n\|\boldsymbol{Q}\overline{\boldsymbol{x}}_{B,n} - \boldsymbol{Q}\boldsymbol{x}^*\|^2_{\boldsymbol{U}} \\
&\quad + 2\chi_{7,n} + 2\left\langle \overline{\boldsymbol{x}}_{A,n} - \boldsymbol{x}^* \mid \boldsymbol{Q}\boldsymbol{x}^*\right\rangle \\
&\geq \frac{-1}{\gamma_n}\|\overline{\boldsymbol{x}}_{A,n} - \overline{\boldsymbol{x}}_{B,n}\|^2_{\boldsymbol{V}} - \gamma_n\beta_0^2\|\overline{\boldsymbol{x}}_{B,n} - \boldsymbol{x}^*\|^2_{\boldsymbol{V}} \\
&\quad + 2\chi_{7,n} + 2\left\langle \overline{\boldsymbol{x}}_{A,n} - \boldsymbol{x}^* \mid \boldsymbol{Q}\boldsymbol{x}^*\right\rangle.
\end{aligned}$$

Now, using the same argument as in the proof of Theorem 11, we derive that $\mathbf{E}[\|\overline{\boldsymbol{x}}_{B,n} - \boldsymbol{x}^*\|^2_{\boldsymbol{V}}] = O(1/n^2) + O(1/n^{2-t})$. ∎

Remark 15. We have the following remarks:

(i) The rate of convergence in expectation for solving strongly monotone inclusions is also investigated in [33] for the case of the stochastic forward-backward splitting. The best rate obtained in [33] is $O(1/n)$ which is the same as convergence rate here for $t = 1$.

(ii) The rate $O(1/n)$ for variational inequalities involving Lipschitzian and monotone operator is also obtained in [28].

(iii) In the case of minimization, the further connections to existing work in [1, 3, 19, 20, 21, 23, 31] can be found in [32].

REFERENCES

[1] Y. F. Atchade, G. Fort and E. Moulines, On stochastic proximal gradient algorithms, arXiv:1402.2365, 2014.

[2] H. Attouch, L. M. Briceño-Arias, and P. L. Combettes, A parallel splitting method for coupled monotone inclusions, *SIAM J. Control Optim.*, vol. 48, pp. 3246–3270, 2010.

[3] F. Bach and E. Moulines, Non-asymptotic analysis of stochastic approximation algorithms for machine learning, *Advances in Neural Information Processing Systems*, pp. 451-459, 2011.

[4] K. Barty and J.-S. Roy and C. Strugarek, Hilbert-valued perturbed subgradient algorithms, *Math. Oper. Res.*,vol. 32, pp. 551–562, 2007.

[5] H. H. Bauschke, P. L. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, New York (2011).

[6] A. Bennar and J.-M. Monnez, Almost sure convergence of a stochastic approximation process in a convex set, *Int. J. Appl. Math.*, vol. 20, pp. 713-722, 2007.

[7] P. Bianchi, W. Hachem and F. Iutzeler, A stochastic coordinate descent primal-dual algorithm and applications to large-scale composite optimization, arXiv:1407.0898v3, 2015.

[8] K. Bredies and H. Sun, Preconditioned Douglas–Rachford splitting methods for convex-concave saddle-point problems, *SIAM J. Numer. Anal.*, vol. 53, pp. 421-444, 2015.

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

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

[11] P. L. Combettes, 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.*, vol. 20, pp. 307–330, 2012.

[12] P. L. Combettes and B. C. Vũ, Variable metric forward-backward splitting with applications to monotone inclusions in duality, *Optimization*, vol. 63, pp. 1289-1318, 2013.

[13] P. L. Combettes and J.-C. Pesquet, Stochastic approximations and perturbations in forward-backward splitting for monotone operators, *Pure and Applied Functional Analysis*, vol. 1, pp. 13-37, 2016.

[14] P. L. Combettes and J.-C. Pesquet, Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping, preprint, 2014. arXiv:1404.7536v4, 2015.

[15] P. Combettes and J.-C. Pesquet, A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery, IEEE *J. Sel. Topics Signal Process.*, vol. 1, pp. 564-574, 2007.

[16] P. L Combettes, L. Condat, J.-C. Pesquet, and B. C. Vũ, A forward-backward view of some primal-dual optimization methods in image recovery, *In Proc. Int. Conf. Image Process.*, Paris, France, 27-30 Oct. 2014

[17] P. L. Combettes and V. R. Wajs, Signal recovery by proximal forward-backward splitting, *Multiscale Model. Simul.* vol. 4, pp. 1168–1200, 2005.

[18] D. Davis and W. Yin, A three-Operator splitting scheme and its optimization applications, arXiv:1504.01032v1, 2015.

[19] J. Duchi and Y. Singer, Efficient online and batch learning using forward backward splitting, *J. Mach. Learn. Res.*, vol. 10, pp. 2899–2934, 2009.

[20] S. Ghadimi and G. Lan, Optimal stochastic approximation algorithms for strongly convex stochastic composite optimization i: A generic algorithmic framework, *SIAM J. Optim.*, vol. 22, pp. 1469–1492, 2012.

[21] E. Hazan and S. Kale, Beyond the regret minimization barrier: an optimal algorithm for stochastic strongly-convex optimization, *Journal of Machine Learning Research*, vol. 15, pp. 2489-2512, 2014.

[22] J. T. Kwok, C. Hu and W. Pan, Accelerated gradient methods for stochastic optimization and online learning, *Advances in Neural Information Processing Systems*, vol. 22, pp. 781–789, 2009.

[23] G. Lan, An optimal method for stochastic composite optimization, *Math. Program.* vol. 133, pp. 365–397, 2012.

[24] M. Ledoux and M. Talagrand, *Probability in Banach Spaces: Isoperimetry and Processes.* Springer, New York 1991.

[25] P.-L. Lions, B. Mercier, Splitting algorithms for the sum of two nonlinear operators, *SIAM J. Numer. Anal.*, vol. 16, pp. 964–979, 1979.
[26] B. Mercier, Topics in finite element solution of elliptic problems (Lectures on Mathematics, no. 63). Tata Institute of Fundamental Research, Bombay, 1979.
[27] J.-M. Monnez, Almost sure convergence of stochastic gradient processes with matrix step sizes, *Statist. Probab. Lett.*, vol. 76, pp. 531-536, 2006.
[28] A. Nemirovski, Prox-method with rate of convergence $O(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems, *SIAM J. Optim.*, vol. 15, pp. 229-251, 2004.
[29] J.-C. Pesquet and A. Repetti, A class of randomized primal-dual algorithms for distributed optimization, *J. Nonlinear Convex Anal.,*, vol. 16, pp. 2453-2490, 2015.
[30] H. Raguet, J. Fadili, and G. Peyré, Generalized forward-backward splitting, *SIAM J. Imaging Sci.*, vol. 6, pp. 1199–1226, 2013.
[31] A. Rakhlin and O. Shamir and K. Sridaran, Making gradient descent optimal for strongly convex stochastic optimization, *Proceedings of the 29th International Conference on Machine Learning*, 2012.
[32] L. Rosasco, S. Villa, and B. C. Vũ, Convergence of stochastic proximal gradient algorithm, arXiv:1403.5074v3, 2014.
[33] L. Rosasco, S. Villa, and B. C. Vũ, Stochastic forward-backward splitting method for solving monotone inclusions, *J. Optim. Theory Appl.*, vol. 169, pp. 388–406, 2016.
[34] L. Rosasco, S. Villa, and B. C. Vũ, A stochastic inertial forward-backward splitting algorithm for multivariate monotone inclusions, *Optimization*, vol. 65, pp. 1293–1314, 2016.
[35] B. C. Vũ, A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Adv. Comput. Math.*, vol. 38, pp. 667–681, 2013.
[36] B. C. Vũ, A splitting algorithm for coupled system of primal-dual monotone inclusions, *J. Optim. Theory Appl.*, vol. 164, pp. 993–1025, 2015.
[37] B. C. Vũ, Almost sure convergence of the forward-backward-forward splitting algorithm, *Optim. Lett.*, vol.10, pp. 781–803, 2016.
[38] A. Yurtsever, B. C. Vũ and V. Cevher, Stochastic three-composite convex minimization, *Advances in Neural Information Processing Systems*, Barcelona, 2016.