Convergence analysis of the stochastic reflected forward-backward splitting algorithm

Nguyen Van Dung[1] and Bằng Công Vũ

[1] Department of Mathematics, University of Transport and Communications,
3 Cau Giay Street, Hanoi, Vietnam
dungnv@utc.edu.vn; bangcvvn@gmail.com

February 18, 2021

Abstract

We propose and analyze the convergence of a novel stochastic algorithm for solving monotone inclusions that are the sum of a maximal monotone operator and a monotone, Lipschitzian operator. The propose algorithm requires only unbiased estimations of the Lipschitzian operator. We obtain the rate $\mathcal{O}(log(n)/n)$ in expectation for the strongly monotone case, as well as almost sure convergence for the general case. Furthermore, in the context of application to convex-concave saddle point problems, we derive the rate of the primal-dual gap. In particular, we also obtain $\mathcal{O}(1/n)$ rate convergence of the primal-dual gap in the deterministic setting.

Keywords: monotone inclusion, stochastic optimization, stochastic error, monotone operator, operator splitting, reflected method, Lipschitz, composite operator, duality, primal-dual algorithm, ergodic convergence

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

1 Introduction

A wide class of problems in monotone operator theory, variational inequalities, convex optimization, image processing, machine learning, reduces to the problem of solving monotone inclusions involving Lipschitzian operators; see [2, 4, 3, 5, 8, 11, 22, 32, 23, 33, 35] and the references therein. In this paper, we revisit the generic monotone inclusions of finding a zero point of the sum of a maximally monotone operator A and a monotone, μ-Lipschitzian operator B, acting on a real separable Hilbert space $\mathcal{H}$, i.e.,

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

The first splitting method proposed for solving problem was in [33] which is now known as the forward-backward-forward splitting method (FBFSM). Further investigations of this method lead to a new primal-dual splitting method in [4] where B is a linear monotone skew operator in suitable product spaces. A main limitation of the FBFSM is their two calls of B per iteration. This issue was recently resolved in [22] in which the forward reflected backward splitting method (FRBSM) was proposed, namely,

$$\gamma \in]0,+\infty[\,, \quad x_{n+1} = (\mathrm{Id} + \gamma A)^{-1}(x_n - 2\gamma B x_n + \gamma B x_{n-1}). \tag{1.2}$$

An alternative approach to overcome this issue was in [11] where the reflected forward backward splitting method (RFBSM) was proposed:

$$\gamma \in]0,+\infty[\,, \quad x_{n+1} = (\mathrm{Id} + \gamma A)^{-1}(x_n - \gamma B(2x_n - x_{n-1})). \tag{1.3}$$

It is important to stress that the methods FBFSM, RFBSM and RFBSM are limited to the deterministic setting. The stochastic version of FBFSM was investigated in [35] and recently in [14]. Both works [35] and [14] requires two stochastic approximation of B. While, a stochastic version of FRBSM was also considered in [22] for the case when B is a finite sum. However, it remains require to evaluate the operator B.

The objective of this paper is to avoid these above limitations of [35, 14, 22] by considering the stochastic counterpart of (1.3). At each iteration, we use only one unbiased estimation of $B(2x_n - x_{n-1})$ and hence the resulting algorithm shares the same structure as the standard stochastic forward-backward splitting [7, 9, 28]. However, it allows to solve a larger class of problems involving non-cocoercive operators.

In Section 2, we recall the basic notions in convex analysis and monotone operator theory as well as the probability theory, and establish the results which will be used in the proof of the convergence of the proposed method. We present the proposed method and derive the almost sure convergence, convergence in expectation in Section 3. In the last section, we will further apply the proposed algorithm to the convex-concave saddle problem involving the infimal convolutions, and establish the rate of the ergodic convergence of the primal-dual gap.

2 Notation and Background

Let $\mathcal{H}$ be a separable real Hilbert space endowed with the inner product $\langle . \mid . \rangle$ and the associated norm $\|.\|$. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$, and $x \in \mathcal{H}$. We denote the strong convergence and the weak convergence of $(x_n)_{n\in\mathbb{N}}$ to x by $x_n \to x$ and $x_n \rightharpoonup x$, respectively.

Definition 2.1 *Let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a set-valued operator.*

(i) *The domain of A is denoted by* $\mathrm{dom}(A)$ *that is a set of all $x \in \mathcal{H}$ such that $Ax \neq \varnothing$.*

(ii) *The range of A is* $\mathrm{ran}(A) = \big\{u \in \mathcal{H} \mid (\exists x \in \mathcal{H}) u \in Ax\big\}$.

(iii) *The graph of A is* $\mathrm{gra}(A) = \big\{(x,u) \in \mathcal{H} \times \mathcal{H} \mid u \in Ax\big\}$.

(iv) *The inverse of* A *is* $A^{-1}: u \mapsto \big\{x \mid u \in Ax\big\}$.

(v) *The zero set of* A *is* $\operatorname{zer}(A) = A^{-1}0$.

Definition 2.2 *We have the following definitions:*

(i) *We say that* $A: \mathcal{H} \to 2^{\mathcal{H}}$ *is monotone if*

$$\big(\forall (x,u) \in \operatorname{gra} A\big)\big(\forall (y,v) \in \operatorname{gra} A\big) \quad \langle x - y \mid u - v\rangle \geq 0. \tag{2.1}$$

(ii) *We say that* $A: \mathcal{H} \to 2^{\mathcal{H}}$ *is maximally monotone if it is monotone and there exists no monotone operator* B *such that* $\operatorname{gra}(B)$ *properly contains* $\operatorname{gra}(A)$.

(iii) *We say that* $A: \mathcal{H} \to 2^{\mathcal{H}}$ *is* ϕ_A*-uniformly monotone, at* $x \in \operatorname{dom}(A)$, *if there exists an increasing function* $\phi_A\colon\ [0,\infty[\to [0,\infty]$ *that vanishes only at* 0 *such that*

$$\big(\forall u \in Ax\big)\big(\forall (y,v) \in \operatorname{gra} A\big) \quad \langle x - y \mid u - v\rangle \geq \phi_A(\|y - x\|). \tag{2.2}$$

If $\phi_A = \nu_A|\cdot|^2$ *for some* $\nu_A \in\,]0,\infty[$, *then we say that* A *is* ν_A*-strongly monotone.*

(iv) *The resolvent of* A *is*

$$J_A = (\operatorname{Id} + A)^{-1}, \tag{2.3}$$

where Id *denotes the identity operator on* $\mathcal{H}$.

(v) *A single-valued operator* $B: \mathcal{H} \to \mathcal{H}$ *is* 1*-cocoercive or firmly nonexpasive if*

$$(\forall (x,y) \in \mathcal{H}^2)\ \ \langle x - y \mid Bx - By\rangle \geq \|Bx - By\|^2. \tag{2.4}$$

Let $\Gamma_0(\mathcal{H})$ be the class of proper lower semicontinuous convex function from $\mathcal{H}$ to $]-\infty,+\infty]$.

Definition 2.3 *For* $f \in \Gamma_0(\mathcal{H})$*:*

(i) *The proximity operator of* f *is*

$$prox_f : \mathcal{H} \to \mathcal{H} : x \mapsto \underset{y \in \mathcal{H}}{argmin}\big(f(y) + \frac{1}{2}\|x - y\|^2\big). \tag{2.5}$$

(ii) *The conjugate function of* f *is*

$$f^* : a \mapsto \underset{x \in \mathcal{H}}{sup}\big(\,\langle a \mid x\rangle - f(x)\big). \tag{2.6}$$

(iii) *The infimal convolution of the two functions* ℓ *and* g *from* $\mathcal{H}$ *to* $]-\infty,+\infty]$ *is*

$$\ell \,\square\, g\colon x \mapsto \inf_{y \in \mathcal{H}}\,(\ell(y) + g(x - y)). \tag{2.7}$$

Note that $\text{prox}_f = J_{\partial f}$, let $x \in \mathcal{H}$ and set $p = \text{prox}_f x$, we have

$$(\forall y \in \mathcal{H}) \quad f(p) - f(y) \leq \langle y - p \mid p - x \rangle, \tag{2.8}$$

and that

$$(\forall f \in \Gamma_0(\mathcal{H})) \quad (\partial f)^{-1} = \partial f^*. \tag{2.9}$$

Following [25], let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space. A $\mathcal{H}$-valued random variable is a measurable function $X : \Omega \to \mathcal{H}$, 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 $\mathsf{E}[X]$. The conditional expectation of X given a σ-field $\mathcal{A} \subset \mathcal{F}$ is denoted by $\mathsf{E}[X|\mathcal{A}]$. A $\mathcal{H}$-valued random process is a sequence $\{x_n\}$ of $\mathcal{H}$-valued random variables. The abbreviation a.s. stands for 'almost surely'.

Lemma 2.4 ([27, Theorem 1]) *Let $(\mathcal{F}_n)_{n\in\mathbb{N}}$ be an increasing sequence of sub-σ-algebras of $\mathcal{F}$, let $(z_n)_{n\in\mathbb{N}}, (\xi_n)_{n\in\mathbb{N}}, (\zeta_n)_{n\in\mathbb{N}}$ and $(t_n)_{n\in|NN}$ be $[0, +\infty[$-valued random sequences such that, for every $n \in \mathbb{N}$, z_n, ξ_n, ζ_n and t_n are $\mathcal{F}_n$-measurable. Assume moreover that $\sum_{n\in\mathbb{N}} t_n < +\infty$, $\sum_{n\in\mathbb{N}} \zeta_n < +\infty$ a.s. and*

$$(\forall n \in \mathbb{N}) \ \mathsf{E}[z_{n+1}|\mathcal{F}_n] \leq (1 + t_n)z_n + \zeta_n - \xi_n \ a.s.$$

Then z_n converges a.s. and (ξ_n) is summable a.s.

Corollary 2.5 *Let $(\mathcal{F}_n)_{n\in\mathbb{N}}$ be an increasing sequence of sub-σ-algebras of $\mathcal{F}$, let $(x_n)_{n\in\mathbb{N}}$ be $[0, +\infty[$-valued random sequences such that, for every $n \in \mathbb{N}$, x_{n-1} is $\mathcal{F}_n$-measurable and*

$$\sum_{n\in\mathbb{N}} \mathsf{E}[x_n|\mathcal{F}_n] < +\infty \quad a.s \tag{2.10}$$

Then $\sum\limits_{n\in\mathbb{N}} x_n < +\infty$ a.s

Proof. Let us set

$$(\forall n \in \mathbb{N}) \ z_n = \sum_{k=1}^{n-1} x_k.$$

Then, z_n is $\mathcal{F}_n$ measurable. Moreover,

$$\mathsf{E}[z_{n+1}|\mathcal{F}_n] = z_n + \mathsf{E}[x_n|\mathcal{F}_n]$$

Hence, it follows from Lemma 2.4 and (2.10) that $(z_n)_{n\in\mathbb{N}}$ converges a.s. □

The following lemma can be viewed as direct consequence of [7, Proposition 2.3].

Lemma 2.6 *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. Suppose that, for every $x \in C$, $(\|x_{n+1} - x\|)_{n\in\mathbb{N}}$ converges a.s. Suppose that the set of weak sequentially cluster points of $(x_n)_{n\in\mathbb{N}}$ is a subset of C a.s. Then $(x_n)_{n\in\mathbb{N}}$ converges weakly a.s. to a C-valued random vector.*

3 Algorithm and convergences

We propose the following algorithm, for solving (1.1), that requires only the unbiased estimations of the monotone, $\mu-$Lipschitzian operators B.

Algorithm 3.1 Let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $]0,+\infty[$. Let x_0, x_{-1} be $\mathcal{H}$-valued, squared integrable random variables. Iterates

$$(\forall n\in\mathbb{N})\quad \left\lfloor \begin{array}{l} y_n = 2x_n - x_{n-1}\\ \text{Finding } r_n : \mathsf{E}[r_n|\mathcal{F}_n] = By_n\\ x_{n+1} = J_{\gamma_n A}(x_n - \gamma_n r_n), \end{array}\right. \tag{3.1}$$

where $\mathcal{F}_n = \sigma(x_0, x_1, \ldots, x_n)$.

Remark 3.2 Here are some remarks.

(i) Algorithm 3.1 is an extension of the reflected forward backward splitting in [11] which itself recovers the projected reflected gradient methods for monotone variational inequalities in [21] as a special case. Further connections to existing works in the deterministic setting can be found in [11] as well as [21].

(ii) In the special case when A is a normal cone operator, $A = N_X$ for some non-empty closed convex set, the iteration (3.1) reduces to the one in [13]. However, as we will see in Remark 3.9, our convergences results are completely different from that of [13].

(iii) The proposed algorithm shares the same structure as the stochastic forward-backward splitting in [7, 9, 28]. The main advantage of (3.1) is the monotonicity and Lipschitzianity of B which is much weaker than cocoercivity assumption in [7, 9, 28].

(iv) Under the current conditions on A and B, an alternative method "Between forward-backward and forward-reflected-backward" for solving Problem 1.1 is presented in [22, Section 6] which remains require to evaluate the operator B as well as its unbiased estimations.

We first prove some lemmas which will be used in the proof of Theorem 3.5 and Theorem 3.8.

Lemma 3.3 *Let $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ be generated by* (3.1). *Suppose that A is ϕ_A-uniformly monotone and B is ϕ_B-uniformly monotone. Let $x\in\operatorname{zer}(A+B)$ and set*

$$(\forall n\in\mathbb{N})\ \epsilon_n = 2\gamma_n\big(\phi_A(\|x_{n+1}-x\|) + \phi_A(\|x_{n+1}-x_n\|) + \phi_B(\|y_n - x\|)\big). \tag{3.2}$$

The following holds.

$$\begin{aligned}&\|x_{n+1}-x\|^2 + \epsilon_n + (3-\frac{\gamma_n}{\gamma_{n-1}})\|x_{n+1}-x_n\|^2 + \frac{\gamma_n}{\gamma_{n-1}}\|x_{n+1}-y_n\|^2 + 2\gamma_n\left\langle r_n - Bx \mid x_{n+1}-x_n\right\rangle\\ &\le \|x_n - x\|^2 + 2\gamma_n\left\langle r_{n-1} - Bx \mid x_n - x_{n-1}\right\rangle + 2\gamma_n\left\langle r_{n-1}-r_n \mid x_{n+1}-y_n\right\rangle + \frac{\gamma_n}{\gamma_{n-1}}\|x_n - y_n\|^2\\ &+ 2\gamma_n\left\langle r_n - By_n \mid x - y_n\right\rangle.\end{aligned} \tag{3.3}$$

Proof. Let $n \in \mathbb{N}$ and $x \in \operatorname{zer}(A+B)$. Set

$$p_{n+1} = \frac{1}{\gamma_n}(x_n - x_{n+1}) - r_n. \tag{3.4}$$

Then, by the definition of the resolvent,

$$p_{n+1} \in Ax_{n+1}. \tag{3.5}$$

Since A is ν_A-uniformly monotone and $-Bx \in Ax$, we obtain

$$\left\langle \frac{x_n - x_{n+1}}{\gamma_n} - r_n + Bx \mid x_{n+1} - x \right\rangle \geq \nu_A \phi_A(\|x_{n+1} - x\|), \tag{3.6}$$

which is equivalent to

$$\langle x_n - x_{n+1} \mid x_{n+1} - x \rangle - \gamma_n \nu_A \phi_A(\|x_{n+1} - x\|) \geq \gamma_n \langle r_n - Bx \mid x_{n+1} - x \rangle. \tag{3.7}$$

Let us estimate the right hand side of (3.7). Using $y_n = 2x_n - x_{n-1}$, we have

$$\begin{aligned}
\langle r_n - Bx \mid x_{n+1} - x \rangle &= \langle r_n - Bx \mid x_{n+1} - y_n \rangle + \langle r_n - By_n \mid y_n - x \rangle + \langle By_n - Bx \mid y_n - x \rangle \\
&= \langle r_n - Bx \mid x_{n+1} - x_n \rangle - \langle r_n - Bx \mid x_n - x_{n-1} \rangle + \langle r_n - By_n \mid y_n - x \rangle \\
&\quad + \langle By_n - Bx \mid y_n - x \rangle \\
&= \langle r_n - Bx \mid x_{n+1} - x_n \rangle - \langle r_{n-1} - Bx \mid x_n - x_{n-1} \rangle + \langle r_{n-1} - r_n \mid x_n - x_{n-1} \rangle \\
&\quad + \langle r_n - By_n \mid y_n - x \rangle + \langle By_n - Bx \mid y_n - x \rangle \\
&= \langle r_n - Bx \mid x_{n+1} - x_n \rangle - \langle r_{n-1} - Bx \mid x_n - x_{n-1} \rangle + \langle r_n - r_{n-1} \mid x_{n+1} - y_n \rangle \\
&\quad + \langle r_{n-1} - r_n \mid x_{n+1} - x_n \rangle + \langle r_n - By_n \mid y_n - x \rangle + \langle By_n - Bx \mid y_n - x \rangle
\end{aligned} \tag{3.8}$$

Using the uniform monotonicity of A again, it follows from (3.4) that

$$\left\langle \frac{x_n - x_{n+1}}{\gamma_n} - r_n - \frac{x_{n-1} - x_n}{\gamma_{n-1}} + r_{n-1} \mid x_{n+1} - x_n \right\rangle \geq \nu_A \phi_A(\|x_{n+1} - x_n\|), \tag{3.9}$$

which is equivalent to

$$\begin{aligned}
\langle r_{n-1} - r_n \mid x_{n+1} - x_n \rangle \geq \nu_A \phi_A(\|x_{n+1} - x_n\|) + \frac{\|x_{n+1} - x_n\|^2}{\gamma_n} \\
+ \left\langle \frac{x_n - y_n}{\gamma_{n-1}} \mid x_{n+1} - x_n \right\rangle.
\end{aligned} \tag{3.10}$$

We have

$$\begin{cases}
2\langle x_n - y_n \mid x_{n+1} - x_n \rangle = \|x_{n+1} - y_n\|^2 - \|x_n - y_n\|^2 - \|x_{n+1} - x_n\|^2 \\
2\langle x_n - x_{n+1} \mid x_{n+1} - x \rangle = \|x_n - x\|^2 - \|x_n - x_{n+1}\|^2 - \|x_{n+1} - x\|^2.
\end{cases} \tag{3.11}$$

Therefore, we derive from (3.7), (3.8), (3.10) and (3.11), and the uniform monotonicity of B that

$$\begin{aligned}
&\|x_n - x\|^2 - \|x_n - x_{n+1}\|^2 - \|x_{n+1} - x\|^2 - 2\gamma_n \nu_A \phi_A(\|x_{n+1} - x\|) \\
&\geq 2\gamma_n \big(\langle r_n - Bx \mid x_{n+1} - x_n \rangle - \langle r_{n-1} - Bx \mid x_n - x_{n-1} \rangle \big) + 2\gamma_n \nu_B \phi_B(\|y_n - x\|) \\
&\quad + 2\gamma_n \langle r_n - r_{n-1} \mid x_{n+1} - y_n \rangle + 2\gamma_n \nu_A \phi_A(\|x_{n+1} - x_n\|) + 2\|x_{n+1} - x_n\|^2 \\
&\quad + \frac{\gamma_n}{\gamma_{n-1}} \big(\|x_{n+1} - y_n\|^2 - \|x_n - y_n\|^2 - \|x_{n+1} - x_n\|^2 \big) + 2\gamma_n \langle r_n - By_n \mid y_n - x \rangle.
\end{aligned} \tag{3.12}$$

Hence,

$$\begin{aligned}&\|x_{n+1}-x\|^2+\epsilon_n+(3-\frac{\gamma_n}{\gamma_{n-1}})\|x_{n+1}-x_n\|^2+\frac{\gamma_n}{\gamma_{n-1}}\|x_{n+1}-y_n\|^2+2\gamma_n\left\langle r_n-Bx\mid x_{n+1}-x_n\right\rangle\\&\leq\|x_n-x\|^2+2\gamma_n\left\langle r_{n-1}-Bx\mid x_n-x_{n-1}\right\rangle+2\gamma_n\left\langle r_{n-1}-r_n\mid x_{n+1}-y_n\right\rangle+\frac{\gamma_n}{\gamma_{n-1}}\|x_n-y_n\|^2\\&+2\gamma_n\left\langle r_n-By_n\mid x-y_n\right\rangle,\end{aligned}\tag{3.13}$$

which proves (3.3). □

We also have the following lemma where (3.14) was used in [21] as well as in [11].

Lemma 3.4 *For every $n\in\mathbb{N}$, we have following estimations*

$$2\left\langle By_{n-1}-By_n\mid x_{n+1}-y_n\right\rangle\leq\mu(1+\sqrt{2})\|y_n-x_n\|^2+\mu\|x_n-y_{n-1}\|^2+\mu\sqrt{2}\|y_n-x_{n+1}\|^2,\tag{3.14}$$

and

$$\begin{aligned}\mathbf{T}_n&=\frac{1}{\gamma_n}\|x_n-x\|^2+\mu\|x_n-y_{n-1}\|^2+(\frac{1}{\gamma_{n-1}}+\mu(1+\sqrt{2}))\|x_n-x_{n-1}\|^2+2\alpha_{n-1}\\&\geq\frac{1}{2\gamma_n}\|x_n-x\|^2,\end{aligned}\tag{3.15}$$

where $\alpha_n=\left\langle By_n-Bx\mid x_{n+1}-x_n\right\rangle$.

Proof. Let $n\in\mathbb{N}$. We have

$$\begin{aligned}2\left\langle By_{n-1}-By_n\mid x_{n+1}-y_n\right\rangle&\leq2\|x_{n+1}-y_n\|\|By_{n-1}-By_n\|\\&\leq2\mu\|x_{n+1}-y_n\|\|y_{n-1}-y_n\|\\&\leq\frac{\mu}{\sqrt{2}}\|y_n-y_{n-1}\|^2+\mu\sqrt{2}\|x_{n+1}-y_n\|^2\\&=\frac{\mu}{\sqrt{2}}\|y_n-x_n+x_n-y_{n-1}\|^2+\mu\sqrt{2}\|x_{n+1}-y_n\|^2\\&\leq\frac{\mu}{\sqrt{2}}\big((1+\frac{1}{\sqrt{2}-1})\|y_n-x_n\|^2+(1+\sqrt{2}-1)\|x_n-y_{n-1}\|^2\big)\\&\quad+\mu\sqrt{2}\|x_{n+1}-y_n\|^2\\&=\mu(1+\sqrt{2})\|x_n-y_n\|^2+\mu\|x_n-y_{n-1}\|^2+\mu\sqrt{2}\|x_{n+1}-y_n\|^2.\end{aligned}$$

Since $\alpha_{n-1}=\left\langle By_{n-1}-Bx\mid x_n-x_{n-1}\right\rangle$, we obtain

$$\begin{aligned}2|\alpha_{n-1}|&\leq2\mu\|y_{n-1}-x\|\|x_n-x_{n-1}\|\\&\leq2\mu(\|x_n-y_{n-1}\|+\|x_n-x\|)\|x_n-x_{n-1}\|\\&\leq\mu(\|x_n-y_{n-1}\|^2+\|x_n-x\|^2+2\|x_n-x_{n-1}\|^2).\end{aligned}\tag{3.16}$$

Therefore, we derive from (3.16) and the definition of $\mathbf{T}_n$ that

$$\begin{aligned}\mathbf{T}_n&\geq\frac{1}{2\gamma_n}\|x_n-x\|^2+(\frac{1}{2\gamma_n}-\mu)\|x_n-x\|^2+(\frac{1}{\gamma_{n-1}}+\mu(-1+\sqrt{2}))\|x_n-x_{n-1}\|^2\\&\geq\frac{1}{2\gamma_n}\|x_n-x\|^2,\end{aligned}$$

which proves (3.15). □

Theorem 3.5 *The following hold.*

(i) *Let* $(\gamma_n)_{n\in\mathbb{N}}$ *be a nondecreasing sequence in* $\left]0, \frac{\sqrt{2}-1}{\mu}\right[$, *satisfies*

$$\tau = \inf_{n\in\mathbb{N}}(\frac{2}{\gamma_n} - \frac{1}{\gamma_{n-1}} - \mu(1+\sqrt{2})) > 0$$

In the setting of Algorithm 3.1, assume that the following condition are satisfied for $\mathcal{F}_n = \sigma((x_k)_{0\le k\le n})$

$$\sum_{n\in\mathbb{N}} \mathsf{E}[\|r_n - By_n\|^2|\mathcal{F}_n] < +\infty \quad a.s \tag{3.17}$$

Then (x_n) *converges weakly to a random varibale* $\overline{x}\colon \Omega \to \operatorname{zer}(A+B)$ *a.s.*

(ii) *Suppose that* $\operatorname{dom}(A)$ *is bounded, A or B is uniformly monotone. Let* $(\gamma_n)_{n\in\mathbb{N}}$ *be a monotone decreasing sequence in* $\left]0, \frac{\sqrt{2}-1}{\mu}\right[$ *such that*

$$(\gamma_n)_{n\in\mathbb{N}} \in \ell_2(\mathbb{N})\backslash\ell_1(\mathbb{N}) \quad and \quad \sum_{n\in\mathbb{N}} \gamma_n^2 \mathsf{E}[\|r_n - By_n\|^2|\mathcal{F}_n] < \infty \ a.s. \tag{3.18}$$

Then $(x_n)_{n\in\mathbb{N}}$ *converges strongly a unique solution* $\overline{x}$.

Proof. (i): Let $n \in \mathbb{N}$. Applying Lemma 3.3 with $\nu_A = \nu_B = 0$, we have,

$$\begin{aligned}&\frac{1}{\gamma_n}\|x_{n+1}-x\|^2 + \frac{1}{\gamma_{n-1}}\|x_{n+1}-y_n\|^2 + (\frac{3}{\gamma_n} - \frac{1}{\gamma_{n-1}})\|x_n - x_{n+1}\|^2 + 2\delta_n \\ &\le \frac{1}{\gamma_n}\|x_n - x\|^2 + \frac{1}{\gamma_{n-1}}\|x_n - y_n\|^2 + 2\delta_{n-1} + 2\langle r_{n-1} - r_n \mid x_{n+1} - y_n\rangle + 2\beta_n\end{aligned} \tag{3.19}$$

where

$$\begin{cases}\delta_n = \langle r_n - Bx \mid x_{n+1} - x_n\rangle \\ \beta_n = \langle r_n - By_n \mid x - y_n\rangle\end{cases}$$

Let χ be in $\left]0, \frac{\tau}{2}\right[$, it follows from the Cauchy Schwarz's inequality and (3.14) that

$$\begin{aligned}2\langle r_{n-1} - r_n \mid x_{n+1} - y_n\rangle &= 2\langle r_{n-1} - By_{n-1} + By_{n-1} - By_n + By_n - r_n \mid x_{n+1} - y_n\rangle \\ &\le \frac{\|r_{n-1} - By_{n-1}\|^2}{\chi} + \chi\|x_{n+1} - y_n\|^2 + \frac{\|r_n - By_n\|^2}{\chi} + \chi\|x_{n+1} - y_n\|^2 \\ &+ \mu(1+\sqrt{2})\|y_n - x_n\|^2 + \mu\|x_n - y_{n-1}\|^2 + \mu\sqrt{2}\|y_n - x_{n+1}\|^2\end{aligned} \tag{3.20}$$

Hence, we derive from (3.19) and (3.20) that

$$\begin{aligned}&\frac{1}{\gamma_n}\|x_{n+1}-x\|^2 + (\frac{1}{\gamma_{n-1}} - \mu\sqrt{2} - 2\chi)\|x_{n+1}-y_n\|^2 + (\frac{3}{\gamma_n} - \frac{1}{\gamma_{n-1}})\|x_n - x_{n+1}\|^2 + 2\delta_n \\ &\le \frac{1}{\gamma_n}\|x_n - x\|^2 + (\frac{1}{\gamma_{n-1}} + \mu(1+\sqrt{2}))\|x_n - y_n\|^2 + \mu\|x_n - y_{n-1}\|^2 + 2\delta_{n-1} + 2\beta_n \\ &+ \frac{\|r_{n-1} - By_{n-1}\|^2}{\chi} + \frac{\|r_n - By_n\|^2}{\chi}.\end{aligned} \tag{3.21}$$

In turn, using $\gamma_n \le \gamma_{n+1}$ and $x_n - y_n = x_{n-1} - x_n$

$$
\begin{aligned}
&\frac{1}{\gamma_{n+1}}\|x_{n+1}-x\|^2 + \mu\|x_{n+1}-y_n\|^2 + \big(\frac{3}{\gamma_n} - \frac{1}{\gamma_{n-1}}\big)\|x_n - x_{n+1}\|^2 + 2\delta_n \\
&\le \frac{1}{\gamma_n}\|x_n - x\|^2 + \mu\|x_n - y_{n-1}\|^2 + \big(\frac{3}{\gamma_{n-1}} - \frac{1}{\gamma_{n-2}}\big)\|x_n - x_{n-1}\|^2 + 2\delta_{n-1} + 2\beta_n \\
&- \big(\frac{1}{\gamma_{n-1}} - \mu(1+\sqrt{2}) - 2\chi\big)\|x_{n+1} - y_n\|^2 - \big(\frac{2}{\gamma_{n-1}} - \frac{1}{\gamma_{n-2}} - \mu(1+\sqrt{2})\big)\|x_n - x_{n-1}\|^2 \\
&\qquad + \frac{\|r_{n-1} - By_{n-1}\|^2 + \|r_n - By_n\|^2}{\chi}.
\end{aligned} \tag{3.22}
$$

Let us set

$$
\theta_n = \frac{1}{\gamma_n}\|x_n - x\|^2 + \mu\|x_n - y_{n-1}\|^2 + \big(\frac{3}{\gamma_{n-1}} - \frac{1}{\gamma_{n-2}}\big)\|x_n - x_{n-1}\|^2 + 2\delta_{n-1} + \frac{\|r_{n-1} - By_{n-1}\|^2}{\chi}. \tag{3.23}
$$

We have

$$
\begin{aligned}
2|\delta_{n-1}| &= 2\left|\left\langle r_{n-1} - By_{n-1} \mid x_n - x_{n-1}\right\rangle + 2\left\langle By_{n-1} - Bx \mid x_n - x_{n-1}\right\rangle\right| \\
&\le \frac{\|r_{n-1} - By_{n-1}\|^2}{\chi} + \chi\|x_n - x_{n-1}\|^2 + 2\mu\|y_{n-1} - x\|\|x_n - x_{n-1}\| \qquad (3.24)\\
&\le \frac{\|r_{n-1} - By_{n-1}\|^2}{\chi} + \chi\|x_n - x_{n-1}\|^2 + 2\mu\big(\|x_n - y_{n-1}\| + \|x_n - x\|\big)\|x_n - x_{n-1}\| \\
&\le \frac{\|r_{n-1} - By_{n-1}\|^2}{\chi} + \chi\|x_n - x_{n-1}\|^2 + \mu\big(\|x_n - y_{n-1}\|^2 + \|x_n - x\|^2 + 2\|x_n - x_{n-1}\|^2\big)
\end{aligned}
$$

$$
\Rightarrow \theta_n \ge \big(\frac{1}{\gamma_n} - \mu\big)\|x_n - x\|^2 + \big(\frac{3}{\gamma_{n-1}} - \frac{1}{\gamma_{n-2}} - \chi - 2\mu\big)\|x_n - x_{n-1}\|^2 \ge \mu\|x_n - x\|^2 \ge 0 \tag{3.25}
$$

Moreover, it follows from (3.1) that

$$
\mathsf{E}[\beta_n|\mathcal{F}_n] = 0. \tag{3.26}
$$

Therefore, by taking the conditional expectation both sides of (3.22) with respect to $\mathcal{F}_n$, we obtain

$$
\begin{aligned}
\mathsf{E}[\theta_{n+1}|\mathcal{F}_n] &\le \theta_n - \big(\frac{1}{\gamma_{n-1}} - \mu(\sqrt{2}+1) - 2\chi\big)\mathsf{E}[\|x_{n+1} - y_n\|^2|\mathcal{F}_n] \\
&\quad - \big(\frac{2}{\gamma_{n-1}} - \frac{1}{\gamma_{n-2}} - \mu(1+\sqrt{2})\big)\|x_n - x_{n-1}\|^2 + 2\frac{\mathsf{E}[\|r_n - By_n\|^2|\mathcal{F}_n]}{\chi}.
\end{aligned} \tag{3.27}
$$

It follows from our conditions on step sizes $(\gamma_n)_{n\in\mathbb{N}}$ that

$$
\frac{1}{\gamma_{n-1}} - \mu(\sqrt{2}+1) - 2\chi > 0 \text{ and } \frac{2}{\gamma_{n-1}} - \frac{1}{\gamma_{n-2}} - \mu(1+\sqrt{2}) - \chi > 0, \tag{3.28}
$$

Now, in view of Lemma 2.4, we get

$$
\theta_n \to \bar{\theta} \;\text{ and }\; \|x_n - x_{n-1}\| \to 0 \;\; a.s. \tag{3.29}
$$

From (3.17) and Corollary 2.5, we have

$$\sum_{n\in\mathbb{N}} \|r_{n-1} - By_{n-1}\|^2 < +\infty \ \Rightarrow \|r_{n-1} - By_{n-1}\| \to 0 \ \ a.s \tag{3.30}$$

Since $(\theta_n)_{n\in\mathbb{N}}$ converges, it is bounded and therefore, using (3.25), it follows that $(\|x_n - x\|)_{n\in\mathbb{N}}$ and $(x_n)_{n\in\mathbb{N}}$ are bounded. Hence $(y_n)_{n\in\mathbb{N}}$ is also bounded. In turn, from (3.24), we derive

$$\delta_{n-1} \to 0 \ \ a.s \tag{3.31}$$

Moreover,

$$\|x_n - y_n\| = \|x_n - x_{n-1}\| \to 0, \text{ and } \|x_n - y_{n-1}\| \leq \|x_n - x_{n-1}\| + \|x_{n-1} - y_{n-1}\| \to 0. \tag{3.32}$$

Therefore, we derive from (3.23), (3.29), (3.30), (3.31), (3.32) and Lemma 2.4 that

$$(\|x_n - x\|)_{n\in\mathbb{N}} \text{ converges a.s.} \tag{3.33}$$

Let x^* be a weak cluster point of $(x_n)_{n\in\mathbb{N}}$. Then, there exists a subsequence $(x_{n_k})_{k\in\mathbb{N}}$ which converges weakly to x^* a.s. By (3.32), $y_{n_k} \rightharpoonup x^*$ a.s. Let us next set

$$z_n = (I + \gamma_n A)^{-1}(x_n - \gamma_n By_n). \tag{3.34}$$

Then, since $J_{\gamma_n A}$ is nonexpansive, we have

$$\|x_{n+1} - z_n\| \leq \gamma_n \|By_n - r_n\| \to 0 \text{ a.s.} \tag{3.35}$$

It follows from $x_{n_k} \rightharpoonup x^*$ that $x_{n_k+1} \rightharpoonup x^*$ and hence from (3.35) that $z_{n_k}(\omega) \rightharpoonup x^*(\omega)$. Since $z_{n_k} = (I + \gamma_{n_k} A)^{-1}(x_{n_k} - \gamma_{n_k} By_{n_k})$, we have

$$\frac{x_{n_k} - z_{n_k}}{\gamma_{n_k}} - By_{n_k} + Bz_{n_k} \in (A + B)z_{n_k}, \tag{3.36}$$

From (3.32) and (3.35), we have

$$\lim_{k\to\infty} \|x_{n_k} - z_{n_k}\| = \lim_{k\to\infty} \|y_{n_k} - z_{n_k}\| = 0 \tag{3.37}$$

Since B is μ-Lipschitz and $(\gamma_n)_{n\in\mathbb{N}}$ is bounded away from 0, it follows that

$$\frac{x_{n_k} - z_{n_k}}{\gamma_{n_k}} - By_{n_k} + Bz_{n_k} \to 0 \quad a.s. \tag{3.38}$$

Using [2, Corollary 25.5], the sum $A + B$ is maximally monotone and hence, its graph is closed in $\mathcal{H}^{weak} \times \mathcal{H}^{strong}$ [2, Proposition 20.38]. Therefore, $0 \in (A + B)x^*$ a.s., that is $x^* \in \operatorname{zer}(A + B)$ a.s. By Lemma 2.6, the sequence $(x_n)_{n\in\mathbb{N}}$ converges weakly to $\bar{x} \in \operatorname{zer}(A+B)$ and the proof is complete

(ii) It follows from Lemma 3.3 and (3.14) that

$$\begin{aligned}
&\|x_{n+1} - x\|^2 + \frac{\gamma_n}{\gamma_{n-1}}\|x_{n+1} - y_n\|^2 + \big(3 - \frac{\gamma_n}{\gamma_{n-1}}\big)\|x_n - x_{n+1}\|^2 + 2\gamma_n\alpha_n \\
&\quad + 2\gamma_n \langle r_n - By_n \mid x_{n+1} - x_n\rangle + \epsilon_n \\
&\leq \|x_n - x\|^2 + 2\gamma_n \langle r_{n-1} - By_{n-1} + By_{n-1} - By_n + By_n - r_n \mid x_{n+1} - y_n\rangle + \frac{\gamma_n}{\gamma_{n-1}}\|x_n - y_n\|^2 \\
&\quad + 2\gamma_n\alpha_{n-1} + 2\gamma_n \langle r_{n-1} - By_{n-1} \mid x_n - x_{n-1}\rangle + 2\gamma_n\beta_n \\
&\leq \|x_n - x\|^2 + \gamma_n\big(\mu(1+\sqrt{2})\|y_n - x_n\|^2 + \mu\|x_n - y_{n-1}\|^2 + \mu\sqrt{2}\|y_n - x_{n+1}\|^2\big) + \frac{\gamma_n}{\gamma_{n-1}}\|x_n - y_n\|^2 \\
&\quad + 2\gamma_n\alpha_{n-1} + 2\gamma_n \langle r_{n-1} - By_{n-1} \mid x_n - x_{n-1}\rangle + 2\gamma_n\beta_n + 2\gamma_n \langle r_{n-1} - By_{n-1} + By_n - r_n \mid x_{n+1} - y_n\rangle ,
\end{aligned} \tag{3.39}$$

For any $\eta \in \left]0, \frac{1-\gamma_0\mu(1+\sqrt{2})}{10}\right[$, using the Cauchy Schwarz's inequality, we have

$$\begin{cases} 2\gamma_n \langle r_{n-1} - By_{n-1} \mid x_{n+1} - y_n \rangle & \leq \frac{\gamma_n^2}{\eta}\|r_{n-1} - By_{n-1}\|^2 + \eta\|x_{n+1} - y_n\|^2 \\ 2\gamma_n \langle r_n - By_n \mid x_{n+1} - y_n \rangle & \leq \frac{\gamma_n^2}{\eta}\|r_n - By_n\|^2 + \eta\|x_{n+1} - y_n\|^2 \\ 2\gamma_n \langle r_{n-1} - By_{n-1} \mid x_n - x_{n-1} \rangle & \leq \frac{\gamma_n^2}{\eta}\|r_{n-1} - By_{n-1}\|^2 + \eta\|x_n - x_{n-1}\|^2 \\ 2\gamma_n \langle r_n - By_n \mid x_{n+1} - x_n \rangle & \leq \frac{\gamma_n^2}{\eta}\|r_n - By_n\|^2 + \eta\|x_{n+1} - x_n\|^2 \end{cases} \tag{3.40}$$

We have

$$\|x_{n+1} - y_n\|^2 \leq 2(\|x_{n+1} - x_n\|^2 + \|x_n - y_n\|^2) \tag{3.41}$$

Therefore, we derive from (3.39), (3.40), (3.41), the monotonic decreasing of $(\gamma_n)_{n\in\mathbb{N}}$ and $y_n - x_n = x_n - x_{n-1}$ that

$$\begin{aligned} &\|x_{n+1} - x\|^2 + (\frac{\gamma_n}{\gamma_{n-1}} - \gamma_n\mu\sqrt{2})\|x_{n+1} - y_n\|^2 + 2\|x_{n+1} - x_n\|^2 + 2\gamma_n\alpha_n + \epsilon_n \\ &\leq \|x_n - x\|^2 + \gamma_n\mu\|x_n - y_{n-1}\|^2 + (1 + \gamma_0\mu(1+\sqrt{2}))\|x_n - x_{n-1}\|^2 + 2\gamma_{n-1}\alpha_{n-1} \\ &- 2(\gamma_{n-1} - \gamma_n)\alpha_{n-1} + 2\gamma_n\beta_n + 2\frac{\gamma_{n-1}^2}{\eta}\|r_{n-1} - By_{n-1}\|^2 + 2\frac{\gamma_n^2}{\eta}\|r_n - By_n\|^2 \\ &+ 5\eta\big(\|x_{n+1} - x_n\|^2 + \|x_n - y_n\|^2\big) \end{aligned} \tag{3.42}$$

Since dom(A) is bounded, there exists $M > 0$ such that $(\forall n \in \mathbb{N})$ $|\alpha_n| \leq M$, and hence (3.42) implies that

$$\begin{aligned} &\|x_{n+1} - x\|^2 + \gamma_n\mu\|x_{n+1} - y_n\|^2 + (2 - 5\eta)\|x_{n+1} - x_n\|^2 + 2\gamma_n\alpha_n \\ &\leq \|x_n - x\|^2 + \gamma_{n-1}\mu\|x_n - y_{n-1}\|^2 + (2 - 5\eta)\|x_n - x_{n-1}\|^2 + 2\gamma_{n-1}\alpha_{n-1} \\ &- (\frac{\gamma_n}{\gamma_{n-1}} - \gamma_n\mu(\sqrt{2}+1))\|x_{n+1} - y_n\|^2 - (1 - \gamma_0\mu(1+\sqrt{2}) - 10\eta)\|x_n - x_{n-1}\|^2 \\ &+ 2(\gamma_{n-1} - \gamma_n)M + 2\frac{\gamma_{n-1}^2}{\eta}\|r_{n-1} - By_{n-1}\|^2 + 2\frac{\gamma_n^2}{\eta}\|r_n - By_n\|^2 - \epsilon_n + 2\gamma_n\beta_n. \end{aligned} \tag{3.43}$$

Let us set

$$p_n = \|x_n - x\|^2 + \gamma_{n-1}\mu\|x_n - y_{n-1}\|^2 + (2 - 5\eta)\|x_n - x_{n-1}\|^2 + 2\gamma_{n-1}\alpha_{n-1} + 2\frac{\gamma_{n-1}^2}{\eta}\|r_{n-1} - By_{n-1}\|^2. \tag{3.44}$$

Then, by taking the conditional expectation with respect to $\mathcal{F}_n$ both sides of (3.43) and using $\mathsf{E}[r_n|\mathcal{F}_n] = By_n$, we get

$$\begin{aligned} \mathsf{E}[p_{n+1}|\mathcal{F}_n] &\leq p_n - (\frac{\gamma_n}{\gamma_{n-1}} - \gamma_n\mu(\sqrt{2}+1))\mathsf{E}[\|x_{n+1} - y_n\|^2|\mathcal{F}_n] - (1 - \gamma_0\mu(1+\sqrt{2}) - 10\eta)\|x_n - x_{n-1}\|^2 \\ &+ 2(\gamma_{n-1} - \gamma_n)M + 4\frac{\gamma_n^2}{\eta}\mathsf{E}[\|r_n - By_n\|^2|\mathcal{F}_n] - \mathsf{E}[\epsilon_n|\mathcal{F}_n] \end{aligned} \tag{3.45}$$

Note that,

$$\begin{cases} \frac{\gamma_n}{\gamma_{n-1}} - \gamma_n\mu(\sqrt{2}+1) > 0, \\ 1 - \gamma_0\mu(1+\sqrt{2}) - 10\eta > 0, \\ \sum_{n\in\mathbb{N}}(\gamma_{n-1} - \gamma_n)M = \gamma_0 M \end{cases} \tag{3.46}$$

Similar to (3.15), we have p_n is a nonnegative sequence. In turn, Lemma 2.4 and (3.45) give,

$$p_n \to \bar{p}, \quad \|x_n - x_{n-1}\| \to 0 \;\text{ and } \sum_{n\in\mathbb{N}} \mathsf{E}[\epsilon_n|\mathcal{F}_n] < +\infty \quad a.s. \tag{3.47}$$

Using the same argument as the proof of (i),

$$\lim_{n\to\infty} \|x_n - x\|^2 = \bar{p}. \tag{3.48}$$

Now, let us consider the case where A is ϕ_A-uniformly monotone. We then derive from (3.47) that

$$\sum_{n\in\mathbb{N}} \gamma_n \mathsf{E}[\phi_A(\|x_{n+1} - x\|)|\mathcal{F}_n] < +\infty, \tag{3.49}$$

hence Corollary 2.5 impies that

$$\sum_{n\in\mathbb{N}} \gamma_n \phi_A(\|x_{n+1} - x\|) < +\infty. \tag{3.50}$$

Since $\sum_{n\in\mathbb{N}} \gamma_n = \infty$, it follows from (3.50) that $\underline{\lim}\phi_A(\|x_{n+1} - x\|) = 0$. Thus, there exists a subsequence $(k_n)_{n\in\mathbb{N}}$ such that $\phi_A(\|x_{k_n} - x\|) \to 0$ and hence $\|x_{k_n} - x\| \to 0$. Therefore, by (3.48), we obtain $x_n \to x$. We next consider that case when B is ϕ_B-uniformly monotone. Since $y_n = 2x_n - x_{n-1}$, by the triangle inequality,

$$\|x_n - x\| - \|x_n - x_{n-1}\| \le \|y_n - x\| \le \|x_n - x\| + \|x_{n-1} - x_n\|, \tag{3.51}$$

and by (3.47), we obtain $\lim_{n\to\infty} \|y_n - x\| = \lim_{n\to\infty} \|x_n - x\|$. Hence, by using the same argument as the case A is uniformly monotone, we obtain $y_n \to x$ and hence $x_n \to x$. The proof of the theorem is complete. □

Remark 3.6 For $0 < \gamma < \dfrac{\sqrt{2}-1}{\mu}$, $\dfrac{1}{2-\gamma\mu(1+\sqrt{2})} < c < 1$. Then for every $(\gamma_n)_{n\in\mathbb{N}} \subset [c\gamma, \gamma]^{\mathbb{N}}$, we have $\tau = \inf_{n\in\mathbb{N}} (\dfrac{2}{\gamma_n} - \dfrac{1}{\gamma_{n-1}} - \mu(1+\sqrt{2})) > 0$

Corollary 3.7 *Let $\gamma \in \left]0, (\sqrt{2}-1)/\mu\right[$. Let x_0, x_{-1} be $\mathcal{H}$-valued, squared integrable random variables.*

$$(\forall n \in \mathbb{N}) \quad \left\lfloor \begin{array}{l} y_n = 2x_n - x_{n-1} \\ \mathsf{E}[r_n|\mathcal{F}_n] = By_n \\ x_{n+1} = J_{\gamma A}(x_n - \gamma r_n). \end{array} \right. \tag{3.52}$$

Suppose that

$$\sum_{n\in\mathbb{N}} \mathsf{E}[\|r_n - By_n\|^2|\mathcal{F}_n] < +\infty \;\; a.s. \tag{3.53}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a random variable $\overline{x}\colon \Omega \to \operatorname{zer}(A+B)$ a.s.

Theorem 3.8 *Suppose that A is ν-strongly monotone. Define*

$$\forall n \in \mathbb{N}) \quad \gamma_n = \frac{1}{2\nu(n+1)}. \tag{3.54}$$

Suppose that there exists a constant c such that

$$(\forall n \in \mathbb{N})\ \mathsf{E}[\|r_n - By_n\|^2|\mathcal{F}_n] \leq c. \tag{3.55}$$

Then

$$(\forall n > n_0)\mathsf{E}\left[\|x_n - \overline{x}\|^2\right] = \mathcal{O}(\log(n+1)/(n+1)), \tag{3.56}$$

where n_0 is the smallest integer such that $n_0 > 4\nu^{-1}\mu(1+\sqrt{2})$.

Proof. Let $n \in \mathbb{N}$. It follows from (3.54) that

$$1 + 2\nu\gamma_n = \frac{2\nu(n+2)}{2\nu(n+1)} = \frac{\gamma_n}{\gamma_{n+1}}. \tag{3.57}$$

Set

$$\begin{cases} \rho_{1,n} & = \langle r_n - By_n \mid x - y_n\rangle + \langle r_{n-1} - By_{n-1} \mid x_n - x_{n-1}\rangle - \langle r_n - By_n \mid x_{n+1} - x_n\rangle, \\ \rho_{2,n} & = \langle r_{n-1} - By_{n-1} - r_n + By_n \mid x_{n+1} - y_n\rangle, \\ \rho_n & = \rho_{1,n} + \rho_{2,n}. \end{cases} \tag{3.58}$$

Hence, by applying Lemma 3.3 with $\phi_B = 0$ and $\phi_A = \nu|\cdot|^2$, we obtain

$$\begin{aligned} &(1+2\nu_A\gamma_n)\|x_{n+1} - x\|^2 + (3 + 2\nu\gamma_n - \frac{\gamma_n}{\gamma_{n-1}})\|x_{n+1} - x_n\|^2 + \frac{\gamma_n}{\gamma_{n-1}}\|x_{n+1} - y_n\|^2 + 2\gamma_n\alpha_n \\ &\leq \|x_n - x\|^2 + 2\gamma_n\alpha_{n-1} + 2\gamma_n\langle By_{n-1} - By_n \mid x_{n+1} - y_n\rangle + \frac{\gamma_n}{\gamma_{n-1}}\|x_n - y_n\|^2 + 2\gamma_n\rho_n. \end{aligned} \tag{3.59}$$

We derive from Lemma 3.4 and (3.59) that

$$\begin{aligned} &\frac{1}{\gamma_{n+1}}\|x_{n+1} - x\|^2 + (\frac{1}{\gamma_{n-1}} - \mu\sqrt{2})\|x_{n+1} - y_n\|^2 + \big(\frac{3}{\gamma_n} + 2\nu - \frac{1}{\gamma_{n-1}}\big)\|x_n - x_{n+1}\|^2 + 2\alpha_n \\ &\leq \frac{1}{\gamma_n}\|x_n - x\|^2 + (\frac{1}{\gamma_{n-1}} + \mu(1+\sqrt{2}))\|x_n - x_{n-1}\|^2 + \mu\|x_n - y_{n-1}\|^2 + 2\alpha_{n-1} + 2\rho_n. \end{aligned} \tag{3.60}$$

Now, using the definition of $\mathbf{T}_n$, we can rewrite (3.60) as

$$\begin{aligned} \mathbf{T}_{n+1} \leq \mathbf{T}_n + 2\rho_n &- (\frac{1}{\gamma_{n-1}} - \mu(\sqrt{2}+1))\|x_{n+1} - y_n\|^2 \\ &- \big(\frac{2}{\gamma_n} + 2\nu - \frac{1}{\gamma_{n-1}} - \mu(\sqrt{2}+1)\big)\|x_n - x_{n+1}\|^2 \end{aligned} \tag{3.61}$$

Let us rewrite $\rho_{2,n}$ as

$$\begin{aligned} \rho_{2,n} = &\langle r_{n-1} - By_{n-1} \mid x_{n+1} - x_n\rangle - \langle r_{n-1} - By_{n-1} \mid x_n - x_{n-1}\rangle \\ &- \langle r_n - By_n \mid x_{n+1} - x_n\rangle + \langle r_n - By_n \mid x_n - x_{n-1}\rangle, \end{aligned} \tag{3.62}$$

which implies that

$$\rho_n = \langle r_n - By_n \mid x - x_n\rangle + \langle r_{n-1} - By_{n-1} \mid x_{n+1} - x_n\rangle - 2\langle r_n - By_n \mid x_{n+1} - x_n\rangle. \tag{3.63}$$

Taking the conditional expectation with respect to $\mathcal{F}_n$, we obtain

$$\mathsf{E}[\mathbf{T}_{n+1}|\mathcal{F}_n] \leq \mathbf{T}_n + 2\mathsf{E}[\rho_n|\mathcal{F}_n] - (\frac{1}{\gamma_{n-1}} - \mu(\sqrt{2}+1))\mathsf{E}[\|x_{n+1} - y_n\|^2|\mathcal{F}_n]$$
$$- \big(\frac{2}{\gamma_n} + 2\nu - \frac{1}{\gamma_{n-1}} - \mu(\sqrt{2}+1)\big)\mathsf{E}[\|x_n - x_{n+1}\|^2|\mathcal{F}_n]. \tag{3.64}$$

By the definition of ρ_n in (3.63), we have

$$\begin{aligned}
2\mathsf{E}[\rho_n|\mathcal{F}_n] &= 2\mathsf{E}[\langle r_{n-1} - By_{n-1} \mid x_{n+1} - x_n\rangle\,|\mathcal{F}_n] - 4\mathsf{E}[\langle r_n - By_n \mid x_{n+1} - x_n\rangle\,|\mathcal{F}_n] \\
&\leq 2\gamma_{n-1}\mathsf{E}[\|r_{n-1} - By_{n-1}\|^2|\mathcal{F}_n] + \frac{1}{2\gamma_{n-1}}\mathsf{E}[\|x_{n+1} - x_n\|^2|\mathcal{F}_n] \\
&\quad + 16\gamma_n\mathsf{E}[\|r_n - By_n\|^2|\mathcal{F}_n] + \frac{1}{4\gamma_n}\mathsf{E}[\|x_{n+1} - x_n\|^2|\mathcal{F}_n]
\end{aligned} \tag{3.65}$$

In turn, it follows from (3.64) that

$$\mathsf{E}[\mathbf{T}_{n+1}|\mathcal{F}_n] \leq \mathbf{T}_n - (\frac{1}{\gamma_{n-1}} - \mu(\sqrt{2}+1))\mathsf{E}[\|x_{n+1} - y_n\|^2|\mathcal{F}_n]$$
$$- \big(\frac{1}{4\gamma_n} - \mu(\sqrt{2}+1)\big)\mathsf{E}[\|x_n - x_{n+1}\|^2|\mathcal{F}_n] + 18\gamma_{n-1}c. \tag{3.66}$$

Note that for $n > n_0$, $\frac{1}{4\gamma_n} - \mu(1+\sqrt{2}) \geq 0$, and hence taking expectation both the sides of (3.66), we obtain

$$(\forall n > n_0)\ \mathsf{E}\left[\mathbf{T}_{n+1}\right] \leq \mathsf{E}\left[\mathbf{T}_{n_0}\right] + c\sum_{k=n_0}^{n}\gamma_k, \tag{3.67}$$

which proves the desired result by invoking Lemma 3.4. □

Remark 3.9 We have some comparisons to existing work.

(i) Under the standard condition (3.18), we obtain the strong almost sure convergence of the iterates, when A or B is uniformly monotone, as in the context of the stochastic forward-backward splitting [28]. In the general case, to ensure the weak almost sure convergence, we not only need the step-size bounded away from 0 but also the summable condition in (3.17). These conditions were used in [7, 9, 29, 30].

(ii) In the case when A is a normal cone in Euclidean spaces and the weak sharpness of B is satisfied, as it was shown in [13, Proposition1], the strong almost sure convergence of $(x_n)_{n\in\mathbb{N}}$ is obtained under the condition (3.18). Without imposing additional conditions on B such as weak sharpness [13], uniform monotonicity [28], the problem of proving the almost sure convergence of the iterates under the condition (3.18) is still open.

(iii) When A is strongly monotone, we obtained the rate $\mathcal{O}(\log(n+1)/(n+1))$ which is slower than the rate $\mathcal{O}(1/(n+1))$ of the stochastic forward-backward splitting [28] and their extensions in [12, 34] . The main reason is the monotonicity and Lipschitzianity of B is weaker than the cocoercivity of B as in [28, 29].

(iv) In the case when A is a normal cone to a nonempty closed convex set X in Euclidean spaces, the work in [13] obtained the rate $1/\sqrt{n}$ of the gap function defined by $X \ni x \mapsto \sup_{y\in X} \langle By \mid x-y\rangle$. This rate of convergence was firstly established in [19] for solving variational inequalities with stochastic mirror-prox algorithm. Therefore, they differ from our results in the present paper.

We provide an generic special case which was widely studied in the stochastic optimization; see [1, 31, 15, 16, 18, 20] for instances.

Corollary 3.10 *Let $f \in \Gamma_0(\mathcal{H})$ and let $h\colon \mathcal{H} \to \mathbb{R}$ be a convex differentiable function, with μ-Lipschitz continuous gradient, given by an expectation form $h(x) = \mathsf{E}_\xi[H(x,\xi)]$. In the expectation, ξ is a random vector whose probability distribution is supported on a set $\Omega_P \subset \mathbb{R}^m$, and $H\colon \mathcal{H}\times\Omega_p \to \mathbb{R}$ is convex function with respect to the variable x. The problem is to*

$$\underset{x\in\mathcal{H}}{\text{minimize}}\ f(x) + h(x), \tag{3.68}$$

under the following assumptions:

(i) $\operatorname{zer}(\partial f + \nabla h) \neq \varnothing$.

(ii) *It is possible to obtain independent and identically distributed (i.i.d.) samples $(\xi_n)_{n\in\mathbb{N}}$ of ξ.*

(iii) *Given $(x,\xi) \in \mathcal{H}\times\Omega_P$, one can find a point $\nabla H(x,\xi)$ such that $\mathsf{E}[\nabla H(x,\xi)] = \nabla h(x)$.*

Let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $]0,+\infty[$. Let x_0, x_{-1} be in $\mathcal{H}$.

$$(\forall n\in\mathbb{N})\quad \left\lfloor \begin{array}{l} y_n = 2x_n - x_{n-1} \\ x_{n+1} = \operatorname{prox}_{\gamma_n f}(x_n - \gamma_n \nabla H(y_n,\xi_n)). \end{array}\right. \tag{3.69}$$

Then, the following hold.

(i) *If f is ν-strongly monotone, for some $\nu \in]0,+\infty[$, and there exists a constant c such that*

$$\mathsf{E}[\|\nabla H(y_n,\xi_n) - \nabla h(y_n)\|^2 | \xi_0,\ldots,\xi_{n-1}] \leq c. \tag{3.70}$$

Then, for the learning rate $(\forall n\in\mathbb{N})\ \gamma_n = \frac{1}{2\nu(n+1)}$. We obtain

$$(\forall n > n_0)\mathsf{E}\left[\|x_n - \overline{x}\|^2\right] = \mathcal{O}(\log(n+1)/(n+1)), \tag{3.71}$$

where n_0 is the smallest integer such that $n_0 > 2\nu^{-1}\mu(1+\sqrt{2})$, and $\overline{x}$ is the unique solution to (3.68).

(ii) *If f is not strongly monotone, let $(\gamma_n)_{n\in\mathbb{N}}$ be a non-decreasing sequence in $\left]0, \frac{\sqrt{2}-1}{\mu}\right[$, satisfies $\tau = \underset{n\in\mathbb{N}}{\inf}\left(\frac{2}{\gamma_n} - \frac{1}{\gamma_{n-1}} - \mu(1+\sqrt{2})\right) > 0$ and*

$$\sum_{n\in\mathbb{N}} \mathsf{E}[\nabla H(y_n,\xi_n) - \nabla h(y_n)\|^2 | \xi_0,\ldots,\xi_{n-1}] < +\infty \quad a.s \tag{3.72}$$

Then (x_n) converges weakly to a random variable $\overline{x}\colon \Omega \to \operatorname{zer}(\partial f + \nabla h)$ a.s.

Proof. The conclusions are followed from Theorem 3.5 & 3.8 where

$$A = \partial f, B = \nabla h, \text{ and } (\forall n \in \mathbb{N})\ r_n = \nabla H(y_n, \xi_n). \tag{3.73}$$

□

Remark 3.11 The algorithm (3.69) as well as the convergence results appear to be new. Algorithm (3.69) is different from the standard stochastic proximal gradient [1, 31, 15, 16] only the evaluation of the stochastic gradients at the reflections $(y_n)_{n\in\mathbb{N}}$.

4 Ergodic convergences

In this section, we focus on the class of primal-dual problem which was firstly investigated in [8]. This typical structured primal-dual framework covers a widely class of convex optimization problems and it has found many applications to image processing, machine learning [8, 10, 26, 6, 24]. We further exploit the duality nature of this framework to obtain a new stochastic primal-dual splitting method and focus on the ergodic convergence of the primal-dual gap.

Problem 4.1 Let $f \in \Gamma_0(\mathcal{H})$, $g \in \Gamma_0(\mathcal{G})$ and let $h\colon \mathcal{H} \to \mathbb{R}$ be a convex differentiable function, with μ_h-Lipschitz continuous gradient, given by an expectation form $h(x) = \mathsf{E}_\xi[H(x,\xi)]$. In the expectation, ξ is a random vector whose probability distribution P is supported on a set $\Omega_p \subset \mathbb{R}^m$, and $H\colon \mathcal{H} \times \Omega \to \mathbb{R}$ is convex function with respect to the variable x. Let $\ell \in \Gamma_0(\mathcal{G})$ be a convex differentiable function with μ_ℓ-Lipschitz continuous gradient, and given by an expectation form $\ell(v) = \mathsf{E}_\xi[L(v,\xi)]$. In the expectation, ζ is a random vector whose probability distribution is supported on a set $\Omega_D \subset \mathbb{R}^d$, and $L\colon \mathcal{G} \times \Omega_D \to \mathbb{R}$ is convex function with respect to the variable v. Let $K\colon \mathcal{H} \to \mathcal{G}$ be a bounded linear operator. The primal problem is to

$$\underset{x\in\mathcal{H}}{\text{minimize}}\ h(x) + (\ell^*\square g)(Kx) + f(x), \tag{4.1}$$

and the dual problem is to

$$\underset{v\in\mathcal{G}}{\text{minimize}}\ (h+f)^*(-K^*v) + g^*(v) + \ell(v), \tag{4.2}$$

under the following assumptions:

(i) There exists a point $(x^\star, v^\star) \in \mathcal{H} \times \mathcal{G}$ such that the primal-dual gap function defined by

$$\begin{aligned} G :& \mathcal{H} \times \mathcal{G} \to \mathbb{R} \cup \{-\infty, +\infty\} \\ & (x,v) \mapsto h(x) + f(x) + \langle Kx \mid v\rangle - g^*(v) - \ell(v) \end{aligned} \tag{4.3}$$

verifies the following condition:

$$\big(\forall x \in \mathcal{H}\big)\big(\big(\forall v \in \mathcal{G}\big)\ G(x^\star, v) \le G(x^\star, v^\star) \le G(x, v^\star), \tag{4.4}$$

(ii) It is possible to obtain independent and identically distributed (i.i.d.) samples $(\xi_n, \zeta_n)_{n\in\mathbb{N}}$ of (ξ, ζ).

(iii) Given $(x, v, \xi, \zeta) \in \mathcal{H} \times \mathcal{G} \times \Omega_P \times \Omega_D$, one can find a point $(\nabla H(x, \xi), \nabla L(v, \xi))$ such that

$$\mathsf{E}_{(\xi,\zeta)}[(\nabla H(x, \xi), \nabla L(v, \zeta))] = (\nabla h(x), \nabla \ell(v)). \tag{4.5}$$

Using the standard technique as in [8], we derive from (3.69) the following stochastic primal-dual splitting method, Algorithm 4.2, for solving Problem 4.1. The weak almost sure convergence and the convergence in expectation of the resulting algorithm can be derived easily from Corollary 3.10 and hence we omit them here.

Algorithm 4.2 Let $(x_0, x_{-1}) \in \mathcal{H}^2$ and $(v_0, v_{-1}) \in \mathcal{G}^2$. Let $(\gamma_n)_{n\in\mathbb{N}}$ be a non-negative sequence. Iterates

$$\begin{array}{l} \text{For } n = 0, 1 \ldots, \\ \left\lfloor \begin{array}{l} y_n = 2x_n - x_{n-1} \\ u_n = 2v_n - v_{n-1} \\ x_{n+1} = \operatorname{prox}_{\gamma_n f}(x_n - \gamma_n \nabla H(y_n, \xi_n) - \gamma_n K^* u_n) \\ v_{n+1} = \operatorname{prox}_{\gamma_n g^*}(v_n - \gamma_n \nabla L(u_n, \zeta_n) + \gamma_n K y_n) \end{array} \right. \end{array} \tag{4.6}$$

Theorem 4.3 *Let $x_0 = x_{-1}$, $v_0 = v_{-1}$. Set $\mu = 2\max\{\mu_h, \ \mu_\ell\} + \|K\|$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a decreasing sequence in $\left]0, \frac{1}{2\mu}\right[$ such that*

$$e_0 = \sum_{n\in\mathbb{N}} \gamma_n^2 \mathsf{E}\left[\|\nabla H(y_n, \xi_n) - \nabla h(y_n)\|^2 + \|\nabla L(u_n, \zeta_n) - \nabla \ell(u_n)\|^2\right] < \infty. \tag{4.7}$$

For every $N \in \mathbb{N}$, define

$$\hat{x}_N = \Big(\sum_{n=0}^{N} \gamma_n x_{n+1}\Big) / \Big(\sum_{n=0}^{N} \gamma_n\Big) \text{ and } \hat{v}_N = \Big(\sum_{n=0}^{N} \gamma_n v_{n+1}\Big) / \Big(\sum_{n=0}^{N} \gamma_n\Big). \tag{4.8}$$

Assume that $\operatorname{dom} f$ *and* $\operatorname{dom} g^*$ *are bounded. Then the following holds:*

$$\mathsf{E}[G(\hat{x}_N, v) - G(x, \hat{v}_N)] \leq \Big(\frac{1}{2}\|(x_0, v_0) - (x, v)\|^2 + \gamma_0 c(x, v) + e_0\Big) \Big/ \Big(\sum_{k=0}^{N} \gamma_k\Big)^{-1}. \tag{4.9}$$

where

$$c(x, v) = \|K\| \sup_{n\in\mathbb{N}} \{\mathsf{E}\left[|\langle x_{n+1} - x \mid v_{n+1} - v_n\rangle|\right] + \mathsf{E}\left[|\langle x_{n+1} - x_n \mid v_{n+1} - v\rangle|\right]\} < \infty. \tag{4.10}$$

Proof. We first note that (4.10) holds because of the boundedness of $\operatorname{dom} f$ and $\operatorname{dom} g^*$. Since ℓ is a convex, differentiable function with μ_ℓ-Lipschitz continuous gradient, using the descent lemma,

$$\ell(u) \leq \ell(q) + \langle \nabla \ell(q) \mid u - q\rangle + \frac{\mu_\ell}{2}\|u - q\|^2. \tag{4.11}$$

Since ℓ is convex, $\ell(q) \leq \ell(w) + \langle \nabla \ell(q) \mid q - w\rangle$. Adding this inequality to (4.11), we obtain

$$\ell(u) \leq \ell(w) + \langle \nabla \ell(q) \mid u - w\rangle + \frac{\mu_\ell}{2}\|u - q\|^2. \tag{4.12}$$

In particular, applying (4.12) with $u = v_{n+1}$, $w = v$ and $q = u_n$, we get

$$\ell(v_{n+1}) \leq \ell(v) + \langle \nabla \ell(u_n) \mid v_{n+1} - v \rangle + \frac{\mu_\ell}{2} \|v_{n+1} - u_n\|^2. \tag{4.13}$$

Moreover, it follows from (4.6) that

$$-\left(v_{n+1} - v_n + \gamma_n \nabla L(u_n, \zeta_n) - \gamma_n K y_n\right) \in \gamma_n \partial g^*(v_{n+1}), \tag{4.14}$$

and hence, using the convexity of g^*,

$$g^*(v) - g^*(v_{n+1}) \geq \frac{1}{\gamma_n} \left\langle v_{n+1} - v \mid v_{n+1} - v_n + \gamma_n \nabla L(u_n, \zeta_n) - \gamma_n K y_n \right\rangle. \tag{4.15}$$

Therefore, we derive from (4.13), (4.15) and (4.3) that

$$\begin{aligned} G(x_{n+1}, v) - G(x_{n+1}, v_{n+1}) &= \langle K x_{n+1} \mid v - v_{n+1} \rangle - g^*(v) + g^*(v_{n+1}) - \ell(v) + \ell(v_{n+1}) \\ &\leq \langle K x_{n+1} \mid v - v_{n+1} \rangle + \frac{1}{\gamma_n} \left\langle v - v_{n+1} \mid v_{n+1} - v_n + \gamma_n \nabla L(u_n, \zeta_n) - \gamma_n K y_n \right\rangle \\ &\qquad + \langle \nabla \ell(u_n) \mid v_{n+1} - v \rangle + \frac{\mu_\ell}{2} \|v_{n+1} - u_n\|^2 \\ &= \langle K(x_{n+1} - y_n) \mid v - v_{n+1} \rangle + \frac{1}{\gamma_n} \langle v - v_{n+1} \mid v_{n+1} - v_n \rangle + \frac{\mu_\ell}{2} \|v_{n+1} - u_n\|^2 \\ &\qquad + \langle \nabla \ell(u_n) - \nabla L(u_n, \zeta_n) \mid v_{n+1} - v \rangle. \end{aligned} \tag{4.16}$$

By the same way, since h is convex differentiable with μ_h-Lipschitz gradient, we have

$$h(x_{n+1}) - h(x) \leq \langle \nabla h(y_n) \mid x_{n+1} - x \rangle + \frac{\mu_h}{2} \|x_{n+1} - y_n\|^2. \tag{4.17}$$

Moreover, it follows from (4.6) that

$$-\left(x_{n+1} - x_n + \gamma_n \nabla H(y_n, \xi_n) + \gamma_n K^*(u_n)\right) \in \gamma_n \partial f(x_{n+1}), \tag{4.18}$$

and hence, by the convexity of f,

$$f(x_{n+1}) - f(x) \leq \frac{1}{\gamma_n} \left\langle x - x_{n+1} \mid x_{n+1} - x_n + \gamma_n \nabla H(y_n, \xi_n) + \gamma_n K^*(u_n) \right\rangle. \tag{4.19}$$

In turn, using the definition of G as in (4.4), we have

$$\begin{aligned} G(x_{n+1}, v_{n+1}) - G(x, v_{n+1}) &= h(x_{n+1}) - h(x) + \langle K(x_{n+1} - x) \mid v_{n+1} \rangle + f(x_{n+1}) - f(x) \\ &\leq \langle \nabla h(y_n) \mid x_{n+1} - x \rangle + \frac{\mu_h}{2} \|x_{n+1} - y_n\|^2 + \langle K(x_{n+1} - x) \mid v_{n+1} \rangle \\ &\quad + \frac{1}{\gamma_n} \left\langle x - x_{n+1} \mid x_{n+1} - x_n + \gamma_n \nabla H(y_n, \xi_n) + \gamma_n K^*(u_n) \right\rangle \\ &= \langle K(x_{n+1} - x) \mid v_{n+1} - u_n \rangle + \frac{1}{\gamma_n} \langle x - x_{n+1} \mid x_{n+1} - x_n \rangle \\ &\quad + \frac{\mu_h}{2} \|x_{n+1} - y_n\|^2 + \langle \nabla h(y_n) - \nabla H(y_n, \xi_n) \mid x_{n+1} - x \rangle. \end{aligned} \tag{4.20}$$

Let us set

$$\begin{cases} \bar{x}_{n+1} &= \operatorname{prox}_{\gamma_n f}(x_n - \gamma_n \nabla h(y_n) - \gamma_n K^*(u_n)), \\ \bar{v}_{n+1} &= \operatorname{prox}_{\gamma_n g^*}(v_n - \gamma_n \nabla \ell(u_n) + \gamma_n K(y_n)). \end{cases} \tag{4.21}$$

Then, using the Cauchy Schwarz's inequality and the nonexpansiveness of $\operatorname{prox}_{\gamma_n f}$, we obtain

$$\begin{aligned}
&\langle \nabla h(y_n) - \nabla H(y_n,\xi_n) \mid x_{n+1} - x\rangle \\
&\quad= \langle \nabla h(y_n) - \nabla H(y_n,\xi_n) \mid x_{n+1} - \bar{x}_{n+1}\rangle + \langle \nabla h(y_n) - \nabla H(y_n,\xi_n) \mid \bar{x}_{n+1} - x\rangle \\
&\quad\leq \|\nabla H(y_n,\xi_n) - \nabla h(y_n)\|\|x_{n+1} - \bar{x}_{n+1}\| + \langle \nabla h(y_n) - \nabla H(y_n,\xi_n) \mid \bar{x}_{n+1} - x\rangle \\
&\quad\leq \gamma_n\|\nabla H(y_n,\xi_n) - \nabla h(y_n)\|^2 + \langle \nabla h(y_n) - \nabla H(y_n,\xi_n) \mid \bar{x}_{n+1} - x\rangle . \qquad (4.22)
\end{aligned}$$

By the same way,

$$\begin{aligned}
&\langle \nabla \ell(u_n) - \nabla L(u_n,\zeta_n) \mid v_{n+1} - v\rangle \\
&\quad\leq \gamma_n\|\nabla L(u_n,\zeta_n) - \nabla \ell(u_n)\|^2 + \langle \nabla \ell(u_n) - \nabla L(u_n,\zeta_n) \mid \bar{v}_{n+1} - v\rangle . \qquad (4.23)
\end{aligned}$$

It follows from (4.16), (4.20) and (4.22), (4.23) that

$$\begin{aligned}
&G(x_{n+1},v) - G(x,v_{n+1}) \\
&\quad\leq \frac{1}{\gamma_n}\big(\langle v - v_{n+1} \mid v_{n+1} - v_n\rangle + \langle x_n - x_{n+1} \mid x_{n+1} - x\rangle \big) + \frac{\mu_h}{2}\|x_{n+1} - y_n\|^2 \\
&\qquad+ \langle K(x_{n+1} - y_n) \mid v - v_{n+1}\rangle + \langle K(x_{n+1} - x) \mid v_{n+1} - u_n\rangle + \frac{\mu_\ell}{2}\|v_{n+1} - u_n\|^2 \\
&\qquad+ \langle \nabla h(y_n) - \nabla H(y_n,\xi_n) \mid \bar{x}_{n+1} - x\rangle + \langle \nabla \ell(u_n) - \nabla L(u_n,\zeta_n) \mid \bar{v}_{n+1} - v\rangle \\
&\qquad\qquad+ \gamma_n\|\nabla H(y_n,\xi_n) - \nabla h(y_n)\|^2 + \gamma_n\|\nabla L(u_n,\zeta_n) - \nabla \ell(u_n)\|^2, \qquad (4.24)
\end{aligned}$$

which is equivalent to

$$\begin{aligned}
&\gamma_n\big(G(x_{n+1},v) - G(x,v_{n+1})\big) \\
&\quad\leq \big(\langle v - v_{n+1} \mid v_{n+1} - v_n\rangle + \langle x_n - x_{n+1} \mid x_{n+1} - x\rangle \big) + \frac{\mu_h\gamma_n}{2}\|x_{n+1} - y_n\|^2 \\
&\qquad+ \gamma_n\big(\langle K(x_{n+1} - y_n) \mid v - v_{n+1}\rangle + \langle K(x_{n+1} - x) \mid v_{n+1} - u_n\rangle \big) + \frac{\mu_\ell\gamma_n}{2}\|v_{n+1} - u_n\|^2 \\
&\qquad+ \gamma_n^2\big(\|\nabla H(y_n,\xi_n) - \nabla h(y_n)\|^2 + \|\nabla L(u_n,\zeta_n) - \nabla \ell(u_n)\|^2\big) \\
&\qquad+ \gamma_n\big(\langle \nabla h(y_n) - \nabla H(y_n,\xi_n) \mid \bar{x}_{n+1} - x\rangle + \langle \nabla \ell(u_n) - \nabla L(u_n,\zeta_n) \mid \bar{v}_{n+1} - v\rangle \big). \qquad (4.25)
\end{aligned}$$

For simple, set $\mu_0 = \max\{\mu_h, \mu_\ell\}$ and let us define some notations in the space $\mathcal{H} \times \mathcal{G}$ where the scalar product and the associated norm are defined in the normal manner,

$$\begin{cases}
\mathsf{x} &= (x,v), \quad \mathsf{x}_n = (x_n,v_n), \quad \mathsf{y}_n = (y_n,u_n),\ \overline{\mathbf{x}}_n = (\overline{x}_n,\overline{v}_n), \\
\mathsf{r}_n &= (\nabla H(y_n,\xi_n), \nabla L(u_n,\zeta_n)), \\
\mathsf{R}_n &= (\nabla h(y_n), \nabla \ell(u_n)),
\end{cases} \qquad (4.26)$$

and

$$S\colon \mathcal{H}\times\mathcal{G} \to \mathcal{H}\times\mathcal{G}\colon (x,v) \mapsto (K^*v, -Kx). \qquad (4.27)$$

Then, one has $\|S\| = \|K\|$ and

$$\begin{aligned}
&\langle K(x_{n+1} - y_n) \mid v - v_{n+1}\rangle + \langle K(x_{n+1} - x) \mid v_{n+1} - u_n\rangle \\
&\quad= \langle S(\mathsf{x}_{n+1} - \mathsf{x}_n) \mid \mathsf{x}_{n+1} - \mathsf{x}\rangle - \langle S(\mathsf{x}_n - \mathsf{x}_{n-1}) \mid \mathsf{x}_n - \mathsf{x}\rangle - \langle S(\mathsf{x}_n - \mathsf{x}_{n-1}) \mid \mathsf{x}_{n+1} - \mathsf{x}_n\rangle \\
&\quad\leq d_{n+1} - d_n + \frac{\|K\|}{2}\big(\|\mathsf{x}_n - \mathsf{x}_{n-1})\|^2 + \|\mathsf{x}_{n+1} - \mathsf{x}_n\|^2\big), \qquad (4.28)
\end{aligned}$$

where we set $d_n = \langle S(\mathsf{x}_n - \mathsf{x}_{n-1}) \mid \mathsf{x}_n - \mathsf{x}\rangle$. Moreover, we also have

$$\begin{aligned}\langle v - v_{n+1} \mid v_{n+1} - v_n\rangle + \langle x_n - x_{n+1} \mid x_{n+1} - x\rangle \\ &= \langle \mathsf{x}_{n+1} - \mathsf{x}_n \mid \mathsf{x} - \mathsf{x}_{n+1}\rangle \\ &= \frac{1}{2}\|\mathsf{x}_n - \mathsf{x}\|^2 - \frac{1}{2}\|\mathsf{x}_n - \mathsf{x}_{n+1}\|^2 - \frac{1}{2}\|\mathsf{x}_{n+1} - \mathsf{x}\|^2. \end{aligned} \tag{4.29}$$

Furthermore, using the triangle inequality, we obtain

$$\frac{\mu_h\gamma_n}{2}\|x_{n+1} - y_n\|^2 + \frac{\mu_\ell\gamma_n}{2}\|v_{n+1} - u_n\|^2 \leq \gamma_n\mu_0\big(\|\mathsf{x}_n - \mathsf{x}_{n+1}\|^2 + \|\mathsf{x}_n - \mathsf{x}_{n-1}\|^2\big). \tag{4.30}$$

Finally, we can rewrite the two last terms in (4.25) as

$$\begin{aligned}&\gamma_n^2\big(\|\nabla H(y_n,\xi_n) - \nabla h(y_n)\|^2 + \|\nabla L(u_n,\zeta_n) - \nabla \ell(u_n)\|^2\big)\\ &\quad + \gamma_n\big(\langle \nabla h(y_n) - \nabla H(y_n,\xi_n) \mid \bar{x}_{n+1} - x\rangle + \langle \nabla \ell(u_n) - \nabla L(u_n,\zeta_n) \mid \bar{v}_{n+1} - v\rangle \big)\\ &= \gamma_n^2\|\mathsf{r}_n - \mathsf{R}_n\|^2 + \gamma_n \langle \mathsf{r}_n - \mathsf{R}_n \mid \mathsf{x} - \bar{\mathsf{x}}_{n+1}\rangle \end{aligned} \tag{4.31}$$

Therefore, inserting (4.28), (4.29), (4.30) and (4.31) into (4.25) and rearranging, we get

$$\begin{aligned}\gamma_n\big(G(x_{n+1},v)-G(x,v_{n+1})\big) \leq \frac{1}{2}\|\mathsf{x}_n - \mathsf{x}\|^2 - \frac{1}{2}\|\mathsf{x}_{n+1} - \mathsf{x}\|^2 + \gamma_n d_{n+1} - \gamma_n d_n\\ - (\frac{1}{2} - \gamma_n\mu_0 - \frac{\gamma_n\|K\|}{2})\|\mathsf{x}_n - \mathsf{x}_{n+1}\|^2 + (\gamma_n\mu_0 + \frac{\gamma_n\|K\|}{2})\|\mathsf{x}_n - \mathsf{x}_{n-1}\|^2\\ + \gamma_n^2\|\mathsf{r}_n - \mathsf{R}_n\|^2 + \gamma_n \langle \mathsf{r}_n - \mathsf{R}_n \mid \mathsf{x} - \bar{\mathsf{x}}_{n+1}\rangle. \end{aligned} \tag{4.32}$$

Let us set

$$b_n = \frac{1}{2}\|\mathsf{x}_n - \mathsf{x}\|^2 + (\gamma_n\mu_0 + \frac{\gamma_n\|K\|}{2})\|\mathsf{x}_n - \mathsf{x}_{n-1}\|^2 - \gamma_n d_n. \tag{4.33}$$

We have

$$\begin{aligned}&|\gamma_n d_n| \leq \gamma_n\|K\|\|\mathsf{x}_n - \mathsf{x}\|\|\mathsf{x}_\mathsf{n} - \mathsf{x}_{n-1}\| \leq \frac{\gamma_n\|K\|}{2}\big(\|\mathsf{x}_n - \mathsf{x}\|^2 + \|\mathsf{x}_\mathsf{n} - \mathsf{x}_{n-1}\|^2\big)\\ &\Rightarrow b_n \geq 0 \ \forall n \in \mathbb{N}\end{aligned}$$

Then, we can rewrite (4.32) as

$$\begin{aligned}\gamma_n\big(G(x_{n+1},v)-G(x,v_{n+1})\big) \leq b_n - b_{n+1} - (\frac{1}{2} - 2\gamma_n\mu_0 - \frac{2\gamma_n\|K\|}{2})\|\mathsf{x}_n - \mathsf{x}_{n+1}\|^2\\ + (\gamma_n - \gamma_{n+1})d_{n+1} + \gamma_n^2\|\mathsf{r}_n - \mathsf{R}_n\|^2 + \gamma_n \langle \mathsf{r}_n - \mathsf{R}_n \mid \mathsf{x} - \bar{\mathsf{x}}_{n+1}\rangle. \end{aligned} \tag{4.34}$$

Now, using our assumption, since $\bar{\mathsf{x}}_{n+1}$ is independent of (ξ_n,ζ_n), we have

$$\mathsf{E}\left[\langle \mathsf{r}_n - \mathsf{R}_n \mid \mathsf{x} - \bar{\mathsf{x}}_{n+1}\rangle \,|(\xi_0,\zeta_0),\ldots(\xi_{n-1},\zeta_{n-1})\right] = 0. \tag{4.35}$$

Moreover, the condition on the learning rate gives

$$\frac{1}{2} - 2\gamma_n\mu_0 - \gamma_n\|K\| \geq 0 \text{ and } \gamma_n - \gamma_{n+1} \geq 0. \tag{4.36}$$

Therefore, taking expectation both sides of (4.34), we obtain

$$\mathsf{E}[\gamma_n\big(G(x_{n+1},v)-G(x,v_{n+1})\big)] \leq \mathsf{E}\left[b_n\right] - \mathsf{E}\left[b_{n+1}\right] + (\gamma_n - \gamma_{n+1})c(x,v) + \gamma_n^2\mathsf{E}\left[\|\mathsf{r}_n - \mathsf{R}_n\|^2\right]. \tag{4.37}$$

Now, for any $N \in \mathbb{N}$, summing (4.34) from $n = 0$ to $n = N$ and invoking the convexity-concavity of G, we arrive at the desired result. □

Remark 4.4 Here are some remarks.

(i) To the best of our knowledge, this is first work establishing the rate convergence of the primal-dual gap for structure convex optimization involving infimal convolutions.

(ii) The results presented in this Section are new even in the deterministic setting. In this case, by setting $\gamma_n \equiv \gamma$, our results share the same rate convergence $\mathcal{O}(1/N)$ of the primal-dual gap as in [17]. While in the stochastic setting, our results share the same rate convergence of the primal-dual gap as in [30] under the same conditions on $(\gamma_n)_{n\in\mathbb{N}}$ and variances as in (4.7). However, the work in [30] are limited to the case ℓ is a constant function.

References

[1] Y. F. Atchade, G. Fort and E. Moulines, On perturbed proximal gradient algorithms, *J. Mach. Learn. Res.*, Vol. 18, pp. 310–342, 2017.

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

[3] M.N. Bùi and P. L . Combettes, Multivariate monotone inclusions in saddle form, https://arxiv.org/pdf/2002.06135.pdf.

[4] 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.

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

[6] A. Chambolle and T. Pock, An introduction to continuous optimization for imaging, *Acta Numer.*, Vol. 25, pp. 161-319, 2016.

[7] P.L. Combettes and J.-C. Pesquet, Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping, *SIAM J. Optim.*, Vol. 25, pp. 1221-1248, 2015.

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

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

[10] P.L. Combettes and J.-C. Pesquet, Proximal splitting methods in signal processing. Fixed-point algorithms for inverse problems in science and engineering,*Optim. Appl.*, Vol.49, pp. 185-212, Springer, New York, 2011.

[11] V. Cevher and B. C. Vũ, A reflected forward-backward splitting method for monotone inclusions involving Lipschitzian operators, *Set-Valued Var. Anal.,* 2020.

[12] V. Cevher, B. C. Vũ and A. Yurtsever, Stochastic Forward Douglas-Rachford Splitting Method for Monotone Inclusions. In: Giselsson P., Rantzer A. (eds) Large-Scale and Distributed Optimization, *Lecture Notes in Mathematics*, Vol 2227, Springer, Cham 2018.

[13] S. Cui and U. V. Shanbhag, On the analysis of reflected gradient and splitting methods for monotone stochastic variational inequality problems, *IEEE 55th Conference on Decision and Control (CDC)*, ARIA Resort & Casino December 12-14, 2016, Las Vegas, USA, 2016.

[14] S. Cui and U. V. Shanbhag, Variance-Reduced Proximal and Splitting Schemes for Monotone Stochastic Generalized Equations https://arxiv.org/abs/2008.11348

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

[16] A. Defazio, F.Bach and S. Lacoste-Julien, SAGA:afast increment algradient method with support for non-strongly convex composite objectives, *Adv. Neural Inf. Process. Syst.*, Vol. 27, pp. 1646–1654,2014.

[17] Y. Drori, S. Sabach, and M. Teboulle, A simple algorithm for a class of nonsmooth convex-concave saddle-point problems, *Oper. Res. Lett.*, Vol. 43, pp. 209-214, 2015.

[18] 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.

[19] A. Juditsky, A. Nemirovski, C. Tauvel, Solving variational inequalities with stochastic mirror-prox algorithm, *Stoch. Syst.* Vol. 1, pp. 17–58, 2011.

[20] J.T. Kwok, C. Hu and W. Pan, Accelerated gradient methods for stochastic optimization and online learning, *Adv. Neural Inf. Process. Syst.*, Vol. 22, pp.781-789, 2009.

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

[22] Y. Malitsky and M. K.Tam, A Forward-Backward Splitting Method for Monotone Inclusions Without Cocoercivity, *SIAM J. Optim.*, Vol. 30, pp. 1451-1472, 2020.

[23] 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.

[24] H. Ouyang, N. He, L. Tran, and A. Gray, Stochastic alternating direction method of multipliers, *In Proceedings of the 30th International Conference on Machine Learning*, Atlanta, GA, USA, 2013.

[25] M. Ledoux, M. Talagrand, Probability in Banach spaces: isoperimetry and processes, Springer, New York, 1991

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

[27] H. Robbins and D. Siegmund, A convergence theorem for non negative almost supermartingales and some applications. In: Rustagi JS, editor. *Optimizing methods in statistic*, New York (NY): Academic Press; pp. 233-257, 1971.

[28] L. Rosasco, S. Villa and, B. C. Vũ Stochastic forward-backward splitting for monotone inclusions, *J. Optim. Theory Appl.*, Vol. 169, pp. 388-406, 2016.

[29] L. Rosasco, S. Villa, B. C. Vũ, A stochastic inertial forward-backward splitting algorithm for multivariate monotone inclusions, *Optimization*, Vol. 65, pp. 1293-1314, 2016.

[30] L. Rosasco, S. Villa, B. C. Vũ, A First-order stochastic primal-dual algorithm with correction step, *Numer. Funct. Anal. Optim.*, Vol. 38, pp.602-626, 2017.

[31] L. Rosasco, S. Villa, B. C. Vũ, Convergence of Stochastic Proximal Gradient Algorithm, *Applied Mathematics & Optimization*, 2019. https://link.springer.com/article/10.1007/s00245-019-09617-7

[32] E. K. Ryu and B. C. Vũ, Finding the Forward-Douglas-Rachford-Forward Method, *J. Optim. Theory Appl.*, Vol. 184, pp. 858–876, 2020.

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

[34] A. Yurtsever, B. C. Vũ, V. Cevher, Stochastic three-composite convex minimization, *Advances in Neural Information Processing Systems*, pp. 4329-4337, 2016.

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