STOCHASTIC APPROXIMATIONS AND PERTURBATIONS IN FORWARD-BACKWARD SPLITTING FOR MONOTONE OPERATORS

PATRICK L. COMBETTES AND JEAN-CHRISTOPHE PESQUET

Abstract. We investigate the asymptotic behavior of a stochastic version of the forward-backward splitting algorithm for finding a zero of the sum of a maximally monotone set-valued operator and a cocoercive operator in Hilbert spaces. In our general setting, only stochastic approximations of the cocoercive operator are available and stochastic perturbations in the evaluation of the resolvents of the set-valued operator are possible. In addition, relaxations and not necessarily vanishing proximal parameters are allowed. Weak and strong almost sure convergence properties of the iterates are established under mild conditions on the underlying stochastic processes. Leveraging these results, we also establish the almost sure convergence of the iterates of a stochastic variant of a primal-dual proximal splitting method for composite minimization problems.

1. Introduction

Throughout the paper, H is a separable real Hilbert space with scalar product $\langle\cdot\mid\cdot\rangle$, associated norm $\|\cdot\|$, and Borel σ-algebra $\mathcal{B}$.

A large array of problems arising in Hilbertian nonlinear analysis are captured by the following simple formulation.

Problem 1.1. Let $\mathsf{A}\colon\mathsf{H}\to 2^{\mathsf{H}}$ be a set-valued maximally monotone operator, let $\vartheta\in\,]0,+\infty[$, and let $\mathsf{B}\colon\mathsf{H}\to\mathsf{H}$ be a ϑ-cocoercive operator, i.e.,

$$(\forall \mathsf{x}\in\mathsf{H})(\forall \mathsf{y}\in\mathsf{H})\quad \langle \mathsf{x}-\mathsf{y}\mid \mathsf{B}\mathsf{x}-\mathsf{B}\mathsf{y}\rangle \geqslant \vartheta\|\mathsf{B}\mathsf{x}-\mathsf{B}\mathsf{y}\|^2, \tag{1.1}$$

such that

$$\mathsf{F}=\big\{\mathsf{z}\in\mathsf{H}\;\big|\;0\in\mathsf{A}\mathsf{z}+\mathsf{B}\mathsf{z}\big\}\neq\varnothing. \tag{1.2}$$

The problem is to find a point in F.

Instances of Problem 1.1 are found in areas such as evolution inclusions [2], Nash equilibria [7], image recovery [8, 10, 16], inverse problems [9, 13], signal processing [21], statistics [25], machine learning [26], variational inequalities

2010 *Mathematics Subject Classification.* 46N10, 90C15, 65K10, 90C48.

Key words and phrases. convex optimization, forward-backward algorithm, monotone operators, primal-dual algorithm, proximal gradient method, stochastic approximation.

This work was supported by the CNRS MASTODONS project under grant 2013MesureHD and by the CNRS Imag'in project under grant 2015OPTIMISME.

[31, 52], mechanics [40, 41], structure design [50], and optimization [4, 38, 51]. For instance, an important specialization of Problem 1.1 in the context of convex optimization is the following [4, Section 27.3].

Problem 1.2. Let $\mathsf{f}\colon \mathsf{H} \to \left]-\infty,+\infty\right]$ be a proper lower semicontinuous convex function, let $\vartheta \in \left]0,+\infty\right[$, and let $\mathsf{g}\colon \mathsf{H} \to \mathbb{R}$ be a differentiable convex function such that ∇g is ϑ^{-1}-Lipschitz continuous on H. The problem is to

$$\underset{\mathsf{x}\in\mathsf{H}}{\text{minimize}}\;\; \mathsf{f}(\mathsf{x}) + \mathsf{g}(\mathsf{x}), \tag{1.3}$$

under the assumption that $\mathsf{F} = \text{Argmin}(\mathsf{f}+\mathsf{g}) \neq \varnothing$.

A standard method to solve Problem 1.1 is the forward-backward algorithm [14, 38, 52], which constructs a sequence $(\mathsf{x}_n)_{n\in\mathbb{N}}$ in H by iterating

$$(\forall n \in \mathbb{N})\quad \mathsf{x}_{n+1} = \mathsf{J}_{\gamma_n\mathsf{A}}(\mathsf{x}_n - \gamma_n\mathsf{B}\mathsf{x}_n), \quad \text{where} \quad 0 < \gamma_n < 2\vartheta. \tag{1.4}$$

Recent theoretical advances on deterministic versions of this algorithm can be found in [6, 11, 20, 22]. Let us also stress that a major motivation for studying the forward-backward algorithm is that it can be applied not only to Problem 1.1 *per se*, but also to systems of coupled monotone inclusions via product space reformulations [2], to strongly monotone composite inclusions problems via duality arguments [16, 20], and to primal-dual composite problems via renorming in the primal-dual space [20, 53]. Thus, new developments on (1.4) lead to new algorithms for solving these problems as well.

Our paper addresses the following stochastic version of (1.4) in which, at each iteration n, due to uncertainties on the underlying mathematical model, $\mathsf{B}x_n$ is not known exactly and is available only through some stochastic approximation u_n. In addition, a_n stands for a stochastic perturbation modeling the approximate implementation of the resolvent operator $\mathsf{J}_{\gamma_n\mathsf{A}}$. Let $(\Omega, \mathcal{F}, \mathsf{P})$ be the underlying probability space. An H-valued random variable is a measurable map $x\colon (\Omega,\mathcal{F}) \to (\mathsf{H},\mathcal{B})$ and, for every $p \in \left[1,+\infty\right[$, $L^p(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ denotes the space of equivalence classes of H-valued random variable x such that $\int_\Omega \|x\|^p d\mathsf{P} < +\infty$.

Algorithm 1.3. *Consider the setting of Problem 1.1. Let x_0, $(u_n)_{n\in\mathbb{N}}$, and $(a_n)_{n\in\mathbb{N}}$ be random variables in $L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $\left]0,1\right]$, and let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $\left]0,2\vartheta\right[$. Set*

$$(\forall n \in \mathbb{N})\quad x_{n+1} = x_n + \lambda_n\big(\mathsf{J}_{\gamma_n\mathsf{A}}(x_n - \gamma_n u_n) + a_n - x_n\big). \tag{1.5}$$

The first instances of the stochastic iteration (1.5) can be traced back to [44] in the context of the gradient method, i.e., when $\mathsf{A} = 0$ and B is the gradient of a convex function. Stochastic approximations in the gradient method were then investigated in the Russian literature of the late 1960s and early 1970s [27, 28, 29, 33, 42, 49]. Stochastic gradient methods have also been used extensively in adaptive signal processing, in control, and in machine learning, e.g., [3, 36, 54]. More generally, proximal stochastic gradient methods have been applied to various problems; see for instance [1, 26, 45, 48, 55].

The objective of the present paper is to provide an analysis of the stochastic forward-backward method in the context of Algorithm 1.3. Almost sure convergence of the iterates $(x_n)_{n\in\mathbb{N}}$ to a solution to Problem 1.1 will be established under general conditions on the sequences $(u_n)_{n\in\mathbb{N}}$, $(a_n)_{n\in\mathbb{N}}$, $(\gamma_n)_{n\in\mathbb{N}}$, and $(\lambda_n)_{n\in\mathbb{N}}$. In particular, a feature of our analysis is that it allows for relaxation parameters and it does not require that the proximal parameter sequence $(\gamma_n)_{n\in\mathbb{N}}$ be vanishing. Our proofs are based on properties of stochastic quasi-Fejér iterations [18], for which we provide a novel convergence result.

The organization of the paper is as follows. The notation is introduced in Section 2. Section 3 provides an asymptotic principle which will be used in Section 4 to present the main result on the weak and strong convergence of the iterates of Algorithm 1.3. Finally, Section 5 deals with applications and features a novel stochastic primal-dual method.

2. Notation

Id denotes the identity operator on H and $\rightharpoonup$ and $\to$ denote, respectively, weak and strong convergence. The sets of weak and strong sequential cluster points of a sequence $(\mathsf{x}_n)_{n\in\mathbb{N}}$ in H are denoted by $\mathfrak{W}(\mathsf{x}_n)_{n\in\mathbb{N}}$ and $\mathfrak{S}(\mathsf{x}_n)_{n\in\mathbb{N}}$, respectively.

Let $\mathsf{A}\colon \mathsf{H}\to 2^{\mathsf{H}}$ be a set-valued operator. The domain of A is $\operatorname{dom}\mathsf{A} = \big\{\mathsf{x}\in\mathsf{H} \;\big|\; \mathsf{A}\mathsf{x}\neq\varnothing\big\}$ and the graph of A is $\operatorname{gra}\mathsf{A} = \big\{(\mathsf{x},\mathsf{u})\in\mathsf{H}\times\mathsf{H} \;\big|\; \mathsf{u}\in\mathsf{A}\mathsf{x}\big\}$. The inverse A^{-1} of A is defined via the equivalences $(\forall(\mathsf{x},\mathsf{u})\in\mathsf{H}^2)$ $\mathsf{x}\in\mathsf{A}^{-1}\mathsf{u}$ $\Leftrightarrow$ $\mathsf{u}\in\mathsf{A}\mathsf{x}$. The resolvent of A is $\mathsf{J}_{\mathsf{A}} = (\mathrm{Id}+\mathsf{A})^{-1}$. If A is maximally monotone, then J_{A} is single-valued and firmly nonexpansive, with $\operatorname{dom}\mathsf{J}_{\mathsf{A}} = \mathsf{H}$. An operator $\mathsf{A}\colon\mathsf{H}\to 2^{\mathsf{H}}$ is demiregular at $\mathsf{x}\in\operatorname{dom}\mathsf{A}$ if, for every sequence $(\mathsf{x}_n,\mathsf{u}_n)_{n\in\mathbb{N}}$ in $\operatorname{gra}\mathsf{A}$ and every $\mathsf{u}\in\mathsf{A}\mathsf{x}$ such that $\mathsf{x}_n \rightharpoonup \mathsf{x}$ and $\mathsf{u}_n\to\mathsf{u}$, we have $\mathsf{x}_n\to\mathsf{x}$ [2]. Let G be a real Hilbert space. We denote by $\mathcal{B}(\mathsf{H},\mathsf{G})$ the space of bounded linear operators from H to G, and we set $\mathcal{B}(\mathsf{H}) = \mathcal{B}(\mathsf{H},\mathsf{H})$. The adjoint of $\mathsf{L}\in\mathcal{B}(\mathsf{H},\mathsf{G})$ is denoted by L^*. For more details on convex analysis and monotone operator theory, see [4].

Let $(\Omega,\mathcal{F},\mathsf{P})$ denote the underlying probability space. The smallest σ-algebra generated by a family Φ of random variables is denoted by $\sigma(\Phi)$. Given a sequence $(x_n)_{n\in\mathbb{N}}$ of H-valued random variables, we denote by $\mathscr{X} = (\mathcal{X}_n)_{n\in\mathbb{N}}$ a sequence of sigma-algebras such that

$$(\forall n\in\mathbb{N})\quad \mathcal{X}_n\subset\mathcal{F}\quad\text{and}\quad \sigma(x_0,\ldots,x_n)\subset\mathcal{X}_n\subset\mathcal{X}_{n+1}. \tag{2.1}$$

Furthermore, we denote by $\ell_+(\mathscr{X})$ the set of sequences of $[0,+\infty[$-valued random variables $(\xi_n)_{n\in\mathbb{N}}$ such that, for every $n\in\mathbb{N}$, ξ_n is $\mathcal{X}_n$-measurable, and we define

$$(\forall p \in \left]0,+\infty\right[) \quad \ell_+^p(\mathscr{X}) = \left\{ (\xi_n)_{n\in\mathbb{N}} \in \ell_+(\mathscr{X}) \;\middle|\; \sum_{n\in\mathbb{N}} \xi_n^p < +\infty \;\; \mathsf{P}\text{-a.s.} \right\}, \tag{2.2}$$

and

$$\ell_+^\infty(\mathscr{X}) = \left\{ (\xi_n)_{n\in\mathbb{N}} \in \ell_+(\mathscr{X}) \;\middle|\; \sup_{n\in\mathbb{N}} \xi_n < +\infty \;\; \mathsf{P}\text{-a.s.} \right\}. \tag{2.3}$$

Equalities and inequalities involving random variables will always be understood to hold P-almost surely, although this will not always be expressly mentioned. Let $\mathcal{E}$ be a sub sigma-algebra of $\mathcal{F}$, let $x \in L^1(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$, and let $y \in L^1(\Omega,\mathcal{E},\mathsf{P};\mathsf{H})$. Then y is the conditional expectation of x with respect to $\mathcal{E}$ if $(\forall E \in \mathcal{E})$ $\int_E x d\mathsf{P} = \int_E y d\mathsf{P}$; in this case we write $y = \mathsf{E}(x\,|\,\mathcal{E})$. We have

$$\big(\forall x \in L^1(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})\big) \quad \|\mathsf{E}(x\,|\,\mathcal{E})\| \leqslant \mathsf{E}(\|x\|\,|\,\mathcal{E}). \tag{2.4}$$

In addition, $L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ is a Hilbert space and

$$\big(\forall x \in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})\big) \quad \begin{cases} \|\mathsf{E}(x\,|\,\mathcal{E})\|^2 \leqslant \mathsf{E}(\|x\|^2\,|\,\mathcal{E}) \\ (\forall \mathsf{u} \in \mathsf{H}) \quad \mathsf{E}(\langle x \mid \mathsf{u}\rangle\,|\,\mathcal{E}) = \langle \mathsf{E}(x\,|\,\mathcal{E}) \mid \mathsf{u}\rangle. \end{cases} \tag{2.5}$$

Geometrically, if $x \in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$, $\mathsf{E}(x \mid \mathcal{E})$ is the projection of x onto $L^2(\Omega,\mathcal{E},\mathsf{P};\mathsf{H})$. For background on probability in Hilbert spaces, see [32, 37].

3. An asymptotic principle

In this section, we establish an asymptotic principle which will lay the foundation for the convergence analysis of our stochastic forward-backward algorithm. First, we need the following result.

Proposition 3.1. *Let* F *be a nonempty closed subset of* H, *let* $\phi\colon \left[0,+\infty\right[\to \left[0,+\infty\right[$ *be a strictly increasing function such that* $\lim_{t\to+\infty}\phi(t) = +\infty$, *let* $(x_n)_{n\in\mathbb{N}}$ *be a sequence of* H*-valued random variables, and let* $(\mathcal{X}_n)_{n\in\mathbb{N}}$ *be a sequence of sub-sigma-algebras of* $\mathcal{F}$ *such that*

$$(\forall n \in \mathbb{N}) \quad \sigma(x_0,\ldots,x_n) \subset \mathcal{X}_n \subset \mathcal{X}_{n+1}. \tag{3.1}$$

Suppose that, for every $\mathsf{z} \in \mathsf{F}$, *there exist* $(\vartheta_n(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+(\mathscr{X})$, $(\chi_n(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^1(\mathscr{X})$, *and* $(\eta_n(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^1(\mathscr{X})$ *such that*

$$(\forall n \in \mathbb{N}) \quad \mathsf{E}(\phi(\|x_{n+1}-\mathsf{z}\|)\,|\,\mathcal{X}_n) + \vartheta_n(\mathsf{z}) \leqslant (1+\chi_n(\mathsf{z}))\phi(\|x_n-\mathsf{z}\|) + \eta_n(\mathsf{z}) \;\; \mathsf{P}\text{-a.s.} \tag{3.2}$$

Then the following hold:

(i) $(\forall \mathsf{z} \in \mathsf{F})$ $\big[\, \sum_{n\in\mathbb{N}} \vartheta_n(\mathsf{z}) < +\infty \;\; \mathsf{P}\text{-a.s.}\big]$

(ii) $(x_n)_{n\in\mathbb{N}}$ *is bounded* P-a.s.

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

(iv) *Suppose that* $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \mathsf{F}$ P-a.s. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges weakly* P-a.s. *to an* F*-valued random variable.*

(v) *Suppose that* $\mathfrak{S}(x_n)_{n\in\mathbb{N}} \cap \mathsf{F} \neq \varnothing$ P-a.s. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges strongly* P-a.s. *to an* F*-valued random variable.*

(vi) *Suppose that* $\mathfrak{S}(x_n)_{n\in\mathbb{N}} \neq \varnothing$ P-a.s. *and that* $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \mathsf{F}$ P-a.s. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges strongly* P-a.s. *to an* F*-valued random variable.*

Proof. This is [18, Proposition 2.3] in the case when $(\forall n \in \mathbb{N})\ \mathcal{X}_n = \sigma(x_0, \ldots, x_n)$. However, the proof remains the same in the more general setting of (2.1). □

The following result describes the asymptotic behavior of an abstract stochastic recursion in Hilbert spaces.

Theorem 3.2. *Let* F *be a nonempty closed subset of* H *and let* $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence in* $]0,1]$. *In addition, let* $(x_n)_{n\in\mathbb{N}}$, $(t_n)_{n\in\mathbb{N}}$, $(c_n)_{n\in\mathbb{N}}$, *and* $(d_n)_{n\in\mathbb{N}}$ *be sequences in* $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. *Suppose that the following are satisfied:*

(a) $\mathscr{X} = (\mathcal{X}_n)_{n\in\mathbb{N}}$ *is a sequence of sub-sigma-algebras of* $\mathcal{F}$ *such that* $(\forall n \in \mathbb{N})\ \sigma(x_0, \ldots, x_n) \subset \mathcal{X}_n \subset \mathcal{X}_{n+1}$.

(b) $(\forall n \in \mathbb{N})\ x_{n+1} = x_n + \lambda_n(t_n + c_n - x_n)$.

(c) $\sum_{n\in\mathbb{N}} \lambda_n \sqrt{\mathsf{E}(\|c_n\|^2 \,|\, \mathcal{X}_n)} < +\infty$ *and* $\sum_{n\in\mathbb{N}} \sqrt{\lambda_n \mathsf{E}(\|d_n\|^2 \,|\, \mathcal{X}_n)} < +\infty$.

(d) *For every* $\mathsf{z} \in \mathsf{F}$, *there exist a sequence* $(s_n(\mathsf{z}))_{n\in\mathbb{N}}$ *of* H*-valued random variables,* $(\theta_{1,n}(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+(\mathscr{X})$, $(\theta_{2,n}(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+(\mathscr{X})$, $(\mu_{1,n}(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^\infty(\mathscr{X})$, $(\mu_{2,n}(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^\infty(\mathscr{X})$, $(\nu_{1,n}(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^\infty(\mathscr{X})$, *and* $(\nu_{2,n}(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^\infty(\mathscr{X})$ *such that* $(\lambda_n\mu_{1,n}(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^1(\mathscr{X})$, $(\lambda_n\mu_{2,n}(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^1(\mathscr{X})$, $(\lambda_n\nu_{1,n}(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^{1/2}(\mathscr{X})$, $(\lambda_n\nu_{2,n}(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^{1/2}(\mathscr{X})$,

(3.3)
$$(\forall n \in \mathbb{N})\quad \mathsf{E}(\|t_n - \mathsf{z}\|^2 \,|\, \mathcal{X}_n) + \theta_{1,n}(\mathsf{z}) \leqslant (1+\mu_{1,n}(\mathsf{z}))\mathsf{E}(\|s_n(\mathsf{z}) + d_n\|^2 \,|\, \mathcal{X}_n) + \nu_{1,n}(\mathsf{z}),$$

and

$$(\forall n \in \mathbb{N})\quad \mathsf{E}(\|s_n(\mathsf{z})\|^2 \,|\, \mathcal{X}_n) + \theta_{2,n}(\mathsf{z}) \leqslant (1 + \mu_{2,n}(\mathsf{z}))\|x_n - \mathsf{z}\|^2 + \nu_{2,n}(\mathsf{z}). \tag{3.4}$$

Then the following hold:

(i) $(\forall \mathsf{z} \in \mathsf{F})\ \big[\ \sum_{n\in\mathbb{N}} \lambda_n\theta_{1,n}(\mathsf{z}) < +\infty$ *and* $\sum_{n\in\mathbb{N}} \lambda_n\theta_{2,n}(\mathsf{z}) < +\infty$ P-a.s. $\big]$.

(ii) $\sum_{n\in\mathbb{N}} \lambda_n(1-\lambda_n)\mathsf{E}(\|t_n - x_n\|^2 \,|\, \mathcal{X}_n) < +\infty$ P-a.s.

(iii) *Suppose that* $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \mathsf{F}$ P-a.s. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges weakly* P-a.s. *to an* F*-valued random variable.*

(iv) *Suppose that* $\mathfrak{S}(x_n)_{n\in\mathbb{N}} \cap \mathsf{F} \neq \varnothing$ P-a.s. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges strongly* P-a.s. *to an* F*-valued random variable.*

(v) *Suppose that* $\mathfrak{S}(x_n)_{n\in\mathbb{N}} \neq \varnothing$ P-a.s. *and that* $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \mathsf{F}$ P-a.s. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges strongly* P-a.s. *to an* F*-valued random variable.*

Proof. Let $\mathsf{z} \in \mathsf{F}$. By (2.5) and (3.3),

$$
\begin{aligned}
(\forall n \in \mathbb{N}) \quad & \mathsf{E}(\|t_n - \mathsf{z}\| \,|\, \mathfrak{X}_n) \\
& \leqslant \sqrt{\mathsf{E}(\|t_n - \mathsf{z}\|^2 \,|\, \mathfrak{X}_n)} \\
& \leqslant \sqrt{1+\mu_{1,n}(\mathsf{z})}\sqrt{\mathsf{E}(\|s_n(\mathsf{z}) + d_n\|^2 \,|\, \mathfrak{X}_n)} + \sqrt{\nu_{1,n}(\mathsf{z})} \\
& \leqslant \Big(1 + \frac{\mu_{1,n}(\mathsf{z})}{2}\Big)\sqrt{\mathsf{E}(\|s_n(\mathsf{z}) + d_n\|^2 \,|\, \mathfrak{X}_n)} + \sqrt{\nu_{1,n}(\mathsf{z})}.
\end{aligned} \tag{3.5}
$$

On the other hand, according to the triangle inequality and (3.4),

$$
\begin{aligned}
(\forall n \in \mathbb{N}) \quad & \sqrt{\mathsf{E}(\|s_n(\mathsf{z}) + d_n\|^2 \,|\, \mathfrak{X}_n)} \\
& \leqslant \sqrt{\mathsf{E}(\|s_n(\mathsf{z})\|^2 \,|\, \mathfrak{X}_n)} + \sqrt{\mathsf{E}(\|d_n\|^2 \,|\, \mathfrak{X}_n)} \\
& \leqslant \sqrt{1+\mu_{2,n}(\mathsf{z})}\|x_n - \mathsf{z}\| + \sqrt{\nu_{2,n}(\mathsf{z})} + \sqrt{\mathsf{E}(\|d_n\|^2 \,|\, \mathfrak{X}_n)} \\
& \leqslant \Big(1 + \frac{\mu_{2,n}(\mathsf{z})}{2}\Big)\|x_n - \mathsf{z}\| + \sqrt{\nu_{2,n}(\mathsf{z})} + \sqrt{\mathsf{E}(\|d_n\|^2 \,|\, \mathfrak{X}_n)}.
\end{aligned} \tag{3.6}
$$

Furthermore, (b) yields

$$
(\forall n \in \mathbb{N}) \quad \|x_{n+1} - \mathsf{z}\| \leqslant (1-\lambda_n)\|x_n - \mathsf{z}\| + \lambda_n\|t_n - \mathsf{z}\| + \lambda_n\|c_n\|. \tag{3.7}
$$

Consequently, (3.5) and (3.6) lead to

$$
\begin{aligned}
(\forall n \in \mathbb{N}) \quad & \mathsf{E}(\|x_{n+1} - \mathsf{z}\| \,|\, \mathfrak{X}_n) \\
& \leqslant (1-\lambda_n)\|x_n - \mathsf{z}\| + \lambda_n\mathsf{E}(\|t_n - \mathsf{z}\| \,|\, \mathfrak{X}_n) + \lambda_n\mathsf{E}(\|c_n\| \,|\, \mathfrak{X}_n) \\
& \leqslant (1+\rho_n(\mathsf{z}))\|x_n - \mathsf{z}\| + \zeta_n(\mathsf{z}),
\end{aligned} \tag{3.8}
$$

where

$$
\rho_n(\mathsf{z}) = \frac{\lambda_n}{2}\Big(\mu_{1,n}(\mathsf{z}) + \mu_{2,n}(\mathsf{z}) + \frac{\mu_{1,n}(\mathsf{z})\mu_{2,n}(\mathsf{z})}{2}\Big) \tag{3.9}
$$

and

$$
\begin{aligned}
\zeta_n(\mathsf{z}) = \lambda_n\sqrt{\nu_{1,n}(\mathsf{z})} + \lambda_n\Big(1 + \frac{\mu_{1,n}(\mathsf{z})}{2}\Big)\Big(&\sqrt{\nu_{2,n}(\mathsf{z})} \\
&+ \sqrt{\mathsf{E}(\|d_n\|^2 \,|\, \mathfrak{X}_n)}\Big) + \lambda_n\mathsf{E}(\|c_n\| \,|\, \mathfrak{X}_n).
\end{aligned} \tag{3.10}
$$

Now set

$$
\overline{\mu}_1(\mathsf{z}) = \sup_{n\in\mathbb{N}} \mu_{1,n}(\mathsf{z}). \tag{3.11}
$$

In view of (3.3) and (3.4), we have

$$
\begin{aligned}
2\sum_{n\in\mathbb{N}}\rho_n(\mathsf{z}) &= \sum_{n\in\mathbb{N}}\lambda_n\mu_{1,n}(\mathsf{z})+\sum_{n\in\mathbb{N}}\lambda_n\mu_{2,n}(\mathsf{z})+\frac{1}{2}\sum_{n\in\mathbb{N}}\lambda_n\mu_{1,n}(\mathsf{z})\mu_{2,n}(\mathsf{z})\\
&\leqslant \sum_{n\in\mathbb{N}}\lambda_n\mu_{1,n}(\mathsf{z})+\Big(1+\frac{\overline{\mu}_1(\mathsf{z})}{2}\Big)\sum_{n\in\mathbb{N}}\lambda_n\mu_{2,n}(\mathsf{z})\\
&<+\infty. \qquad (3.12)
\end{aligned}
$$

In addition, since (2.5) yields

$$
(\forall n\in\mathbb{N})\quad \mathsf{E}(\|c_n\|\,|\,\mathcal{X}_n)\leqslant\sqrt{\mathsf{E}(\|c_n\|^2\,|\,\mathcal{X}_n)}, \qquad (3.13)
$$

we derive from (c) and (d) that

$$
\begin{aligned}
\sum_{n\in\mathbb{N}}\zeta_n(\mathsf{z}) &\leqslant \sum_{n\in\mathbb{N}}\sqrt{\lambda_n\nu_{1,n}(\mathsf{z})}+\Big(1+\frac{\overline{\mu}_1(\mathsf{z})}{2}\Big)\Big(\sum_{n\in\mathbb{N}}\sqrt{\lambda_n\nu_{2,n}(\mathsf{z})}\\
&\quad+\sum_{n\in\mathbb{N}}\sqrt{\lambda_n\mathsf{E}(\|d_n\|^2\,|\,\mathcal{X}_n)}\Big)+\sum_{n\in\mathbb{N}}\lambda_n\sqrt{\mathsf{E}(\|c_n\|^2\,|\,\mathcal{X}_n)}\\
&<+\infty. \qquad (3.14)
\end{aligned}
$$

Using Proposition 3.1(ii), (3.8), (3.12), and (3.14), we obtain that

$$
\big(\|x_n-\mathsf{z}\|\big)_{n\in\mathbb{N}}\ \text{ is almost surely bounded.} \qquad (3.15)
$$

In turn, by (3.4),

$$
\big(\mathsf{E}(\|s_n(\mathsf{z})\|^2\,|\,\mathcal{X}_n)\big)_{n\in\mathbb{N}}\ \text{ is almost surely bounded.} \qquad (3.16)
$$

In addition, (3.3) implies that

$$
\begin{aligned}
(\forall n\in\mathbb{N})\quad \mathsf{E}(\|t_n-\mathsf{z}\|^2\,|\,\mathcal{X}_n)&\leqslant 2(1+\overline{\mu}_1(\mathsf{z}))\big(\mathsf{E}(\|s_n(\mathsf{z})\|^2\,|\,\mathcal{X}_n)\\
&\quad+\mathsf{E}(\|d_n\|^2\,|\,\mathcal{X}_n)\big)+\nu_{1,n}(\mathsf{z}), \qquad (3.17)
\end{aligned}
$$

from which we deduce that

$$
\big(\lambda_n\mathsf{E}(\|t_n-\mathsf{z}\|^2\,|\,\mathcal{X}_n)\big)_{n\in\mathbb{N}}\ \text{ is almost surely bounded.} \qquad (3.18)
$$

Next, we observe that (3.3) and (3.4) yield

$$
\begin{aligned}
(\forall n\in\mathbb{N})\quad &\mathsf{E}(\|t_n-\mathsf{z}\|^2\,|\,\mathcal{X}_n)+\theta_{1,n}(\mathsf{z})+(1+\mu_{1,n}(\mathsf{z}))\theta_{2,n}(\mathsf{z})\\
&\leqslant(1+\mu_{1,n}(\mathsf{z}))(1+\mu_{2,n}(\mathsf{z}))\|x_n-\mathsf{z}\|^2+\nu_{1,n}(\mathsf{z})\\
&\quad+(1+\mu_{1,n}(\mathsf{z}))\big(\nu_{2,n}(\mathsf{z})+2\mathsf{E}(\langle s_n(\mathsf{z})\mid d_n\rangle\,|\,\mathcal{X}_n)+\mathsf{E}(\|d_n\|^2\,|\,\mathcal{X}_n)\big). \qquad (3.19)
\end{aligned}
$$

Now set

$$
\begin{cases}
\theta_n(\mathsf{z})=\theta_{1,n}(\mathsf{z})+(1+\mu_{1,n}(\mathsf{z}))\theta_{2,n}(\mathsf{z})\\
\mu_n(\mathsf{z})=\mu_{1,n}(\mathsf{z})+(1+\overline{\mu}_1(\mathsf{z}))\mu_{2,n}(\mathsf{z})\\
\nu_n(\mathsf{z})=\nu_{1,n}(\mathsf{z})+(1+\overline{\mu}_1(\mathsf{z}))\big(\nu_{2,n}(\mathsf{z})\\
\qquad\qquad +2\sqrt{\mathsf{E}(\|s_n(\mathsf{z})\|^2\,|\,\mathcal{X}_n)}\sqrt{\mathsf{E}(\|d_n\|^2\,|\,\mathcal{X}_n)}+\mathsf{E}(\|d_n\|^2\,|\,\mathcal{X}_n)\big)\\
\xi_n(\mathsf{z})=2\lambda_n\|t_n-\mathsf{z}\|\,\|c_n\|+2(1-\lambda_n)\|x_n-\mathsf{z}\|\,\|c_n\|+\lambda_n\|c_n\|^2.
\end{cases}
$$

By the Cauchy-Schwarz inequality and (3.19),

$$(\forall n \in \mathbb{N}) \quad \mathsf{E}(\|t_n - \mathsf{z}\|^2 \,|\, \mathfrak{X}_n) + \theta_n(\mathsf{z}) \leqslant (1 + \mu_n(\mathsf{z}))\|x_n - \mathsf{z}\|^2 + \nu_n(\mathsf{z}). \tag{3.20}$$

On the other hand, by the conditional Cauchy-Schwarz inequality,

$$\begin{aligned}(\forall n \in \mathbb{N}) \quad & \lambda_n \mathsf{E}(\xi_n(\mathsf{z}) \,|\, \mathfrak{X}_n)\\ & \leqslant 2(1-\lambda_n)\lambda_n \|x_n - \mathsf{z}\|\, \mathsf{E}(\|c_n\| \,|\, \mathfrak{X}_n) + \lambda_n^2 \mathsf{E}(\|c_n\|^2 \,|\, \mathfrak{X}_n)\\ & \quad + 2\lambda_n \sqrt{\lambda_n \mathsf{E}(\|t_n - \mathsf{z}\|^2 \,|\, \mathfrak{X}_n)} \sqrt{\lambda_n \mathsf{E}(\|c_n\|^2 \,|\, \mathfrak{X}_n)}\\ & \leqslant 2\|x_n - \mathsf{z}\|\, \lambda_n \sqrt{\mathsf{E}(\|c_n\|^2 \,|\, \mathfrak{X}_n)} + \lambda_n^2 \mathsf{E}(\|c_n\|^2 \,|\, \mathfrak{X}_n)\\ & \quad + 2\sqrt{\lambda_n \mathsf{E}(\|t_n - \mathsf{z}\|^2 \,|\, \mathfrak{X}_n)}\lambda_n \sqrt{\mathsf{E}(\|c_n\|^2 \,|\, \mathfrak{X}_n)}.\end{aligned} \tag{3.21}$$

Thus, it follows from (3.15), (c), and (3.18) that

$$\sum_{n\in\mathbb{N}} \lambda_n \mathsf{E}(\xi_n(\mathsf{z}) \,|\, \mathfrak{X}_n) < +\infty. \tag{3.22}$$

Let us define

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \vartheta_n(\mathsf{z}) = \lambda_n \theta_n(\mathsf{z}) + \lambda_n(1-\lambda_n)\mathsf{E}(\|t_n - x_n\|^2 \,|\, \mathfrak{X}_n)\\ \chi_n(\mathsf{z}) = \lambda_n \mu_n(\mathsf{z})\\ \eta_n(\mathsf{z}) = \lambda_n \mathsf{E}(\xi_n(\mathsf{z}) \,|\, \mathfrak{X}_n) + \lambda_n \nu_n(\mathsf{z}). \end{cases} \tag{3.23}$$

It follows from (c), (d), (3.16), and the inclusion $\ell_+^{1/2}(\mathscr{X}) \subset \ell_+^1(\mathscr{X})$ that $(\theta_n(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+(\mathscr{X})$, $(\lambda_n\mu_n(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^1(\mathscr{X})$, and $(\lambda_n\nu_n(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^1(\mathscr{X})$. Therefore,

$$\big(\vartheta_n(\mathsf{z})\big)_{n\in\mathbb{N}} \in \ell_+(\mathscr{X}) \tag{3.24}$$

and

$$\big(\chi_n(\mathsf{z})\big)_{n\in\mathbb{N}} \in \ell_+^1(\mathscr{X}). \tag{3.25}$$

Furthermore, we deduce from (3.22) that

$$\big(\eta_n(\mathsf{z})\big)_{n\in\mathbb{N}} \in \ell_+^1(\mathscr{X}). \tag{3.26}$$

Next, we derive from (b), [4, Corollary 2.14], and (3.20) that

$$
\begin{aligned}
(\forall n \in \mathbb{N}) \quad & \mathsf{E}(\|x_{n+1} - \mathsf{z}\|^2 \,|\, \mathfrak{X}_n) \\
&= \mathsf{E}(\|(1-\lambda_n)(x_n - \mathsf{z}) + \lambda_n(t_n - \mathsf{z} + c_n)\|^2 \,|\, \mathfrak{X}_n) \\
&= (1-\lambda_n)\mathsf{E}(\|x_n - \mathsf{z}\|^2 \,|\, \mathfrak{X}_n) + \lambda_n \mathsf{E}(\|t_n - \mathsf{z} + c_n\|^2 \,|\, \mathfrak{X}_n) \\
&\quad - \lambda_n(1-\lambda_n)\mathsf{E}(\|t_n - x_n + c_n\|^2 \,|\, \mathfrak{X}_n) \\
&= (1-\lambda_n)\|x_n - \mathsf{z}\|^2 + \lambda_n \mathsf{E}(\|t_n - \mathsf{z}\|^2 \,|\, \mathfrak{X}_n) \\
&\quad + 2\lambda_n \mathsf{E}(\langle t_n - \mathsf{z} \mid c_n\rangle \,|\, \mathfrak{X}_n) - \lambda_n(1-\lambda_n)\mathsf{E}(\|t_n - x_n\|^2 \,|\, \mathfrak{X}_n) \\
&\quad - 2\lambda_n(1-\lambda_n)\mathsf{E}(\langle t_n - x_n \mid c_n\rangle \,|\, \mathfrak{X}_n) + \lambda_n^2 \mathsf{E}(\|c_n\|^2 \,|\, \mathfrak{X}_n) \\
&= (1-\lambda_n)\|x_n - \mathsf{z}\|^2 + \lambda_n \mathsf{E}(\|t_n - \mathsf{z}\|^2 \,|\, \mathfrak{X}_n) \\
&\quad - \lambda_n(1-\lambda_n)\mathsf{E}(\|t_n - x_n\|^2 \,|\, \mathfrak{X}_n) + 2\lambda_n^2 \mathsf{E}(\langle t_n - \mathsf{z} \mid c_n\rangle \,|\, \mathfrak{X}_n) \\
&\quad + 2\lambda_n(1-\lambda_n)\mathsf{E}(\langle x_n - z \mid c_n\rangle \,|\, \mathfrak{X}_n) + \lambda_n^2 \mathsf{E}(\|c_n\|^2 \,|\, \mathfrak{X}_n) \\
&\leqslant (1-\lambda_n)\|x_n - \mathsf{z}\|^2 + \lambda_n \mathsf{E}(\|t_n - \mathsf{z}\|^2 \,|\, \mathfrak{X}_n) \\
&\quad - \lambda_n(1-\lambda_n)\mathsf{E}(\|t_n - x_n\|^2 \,|\, \mathfrak{X}_n) + \lambda_n \mathsf{E}(\xi_n(\mathsf{z}) \,|\, \mathfrak{X}_n) \\
&\leqslant \big(1 + \chi_n(\mathsf{z})\big)\|x_n - \mathsf{z}\|^2 - \vartheta_n(\mathsf{z}) + \eta_n(\mathsf{z}). \qquad (3.27)
\end{aligned}
$$

We therefore recover (3.2) with $\phi\colon t \mapsto t^2$. Hence, appealing to (3.24), (3.25), (3.26), and Proposition 3.1(i), we obtain $(\vartheta_n(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^1(\mathscr{X})$, which establishes (i) and (ii). Finally, (iii)–(v) follow from Proposition 3.1(iv)–(vi). ☐

Remark 3.3.

(i) Theorem 3.2 extends [18, Theorem 2.5], which corresponds to the special case when, for every $n \in \mathbb{N}$ and every $\mathsf{z} \in \mathsf{F}$, $\mu_{1,n}(\mathsf{z}) = \nu_{1,n}(\mathsf{z}) = \theta_{2,n}(\mathsf{z}) = 0$ and $d_n = 0$. Note that the L^2 assumptions in Theorem 3.2 are just made to unify the presentation with the forthcoming results of Section 4. However, since we take only conditional expectations of $[0,+\infty[$-valued random variables, they are not necessary.

(ii) Suppose that $(\forall n \in \mathbb{N})$ $c_n = d_n = 0$. Then (3.20) and (3.23) imply that

$$
(\forall n \in \mathbb{N}) \quad \eta_n(\mathsf{z}) = \lambda_n\big(\nu_{1,n}(\mathsf{z}) + (1 + \overline{\mu}_1(\mathsf{z}))\nu_{2,n}(\mathsf{z})\big), \qquad (3.28)
$$

and it follows directly from (3.27) and Proposition 3.1 that the conditions on $(\nu_{1,n}(\mathsf{z}))_{n\in\mathbb{N}}$ and $(\nu_{1,n}(\mathsf{z}))_{n\in\mathbb{N}}$ can be weakened to $(\lambda_n\nu_{1,n}(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^1(\mathscr{X})$ and $(\lambda_n\nu_{2,n}(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+^1(\mathscr{X})$.

4. A stochastic forward-backward algorithm

We now state the main result of the paper.

Theorem 4.1. *Consider the setting of Problem 1.1, let* $(\tau_n)_{n\in\mathbb{N}}$ *be a sequence in* $[0,+\infty[$*, let* $\mathscr{X} = (\mathfrak{X}_n)_{n\in\mathbb{N}}$ *be a sequence of sub-sigma-algebras of*

$\mathcal{F}$, and let $(x_n)_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 1.3. Assume that the following are satisfied:

(a) $(\forall n\in\mathbb{N})\ \sigma(x_0,\ldots,x_n)\subset\mathcal{X}_n\subset\mathcal{X}_{n+1}$.
(b) $\sum_{n\in\mathbb{N}}\lambda_n\sqrt{\mathsf{E}(\|a_n\|^2\,|\,\mathcal{X}_n)}<+\infty$.
(c) $\sum_{n\in\mathbb{N}}\sqrt{\lambda_n}\|\mathsf{E}(u_n\,|\,\mathcal{X}_n)-\mathsf{B}x_n\|<+\infty$.
(d) *For every* $\mathsf{z}\in\mathsf{F}$, *there exists* $\big(\zeta_n(\mathsf{z})\big)_{n\in\mathbb{N}}\in\ell_+^\infty(\mathscr{X})$ *such that* $\big(\lambda_n\zeta_n(\mathsf{z})\big)_{n\in\mathbb{N}}\in\ell_+^{1/2}(\mathscr{X})$ *and*

$$(\forall n\in\mathbb{N})\quad \mathsf{E}(\|u_n-\mathsf{E}(u_n\,|\,\mathcal{X}_n)\|^2\,|\,\mathcal{X}_n)\leqslant\tau_n\|\mathsf{B}x_n-\mathsf{Bz}\|^2+\zeta_n(\mathsf{z}). \tag{4.1}$$

(e) $\inf_{n\in\mathbb{N}}\gamma_n>0$, $\sup_{n\in\mathbb{N}}\tau_n<+\infty$, *and* $\sup_{n\in\mathbb{N}}(1+\tau_n)\gamma_n<2\vartheta$.
(f) *Either* $\inf_{n\in\mathbb{N}}\lambda_n>0$ *or* $\big[\,\gamma_n\equiv\gamma$, $\sum_{n\in\mathbb{N}}\tau_n<+\infty$, *and* $\sum_{n\in\mathbb{N}}\lambda_n=+\infty\,\big]$.

Then the following hold for some F*-valued random variable* x*:*

(i) *Let* $\mathsf{z}\in\mathsf{F}$. *Then* $\sum_{n\in\mathbb{N}}\lambda_n\|\mathsf{B}x_n-\mathsf{Bz}\|^2<+\infty$ P-a.s.
(ii) *Let* $\mathsf{z}\in\mathsf{F}$. *Then* $\sum_{n\in\mathbb{N}}\lambda_n\|x_n-\gamma_n\mathsf{B}x_n-\mathsf{J}_{\gamma_n\mathsf{A}}(x_n-\gamma_n\mathsf{B}x_n)+\gamma_n\mathsf{Bz}\|^2<+\infty$ P-a.s.
(iii) $(x_n)_{n\in\mathbb{N}}$ *converges weakly* P-a.s. *to* x.
(iv) *Suppose that one of the following is satisfied:*
 (g) A *is demiregular at every* $\mathsf{z}\in\mathsf{F}$.
 (h) B *is demiregular at every* $\mathsf{z}\in\mathsf{F}$.
 Then $(x_n)_{n\in\mathbb{N}}$ *converges strongly* P-a.s. *to* x.

Proof. Set

$$(\forall n\in\mathbb{N})\quad \mathsf{R}_n=\mathrm{Id}-\gamma_n\mathsf{B},\ r_n=x_n-\gamma_nu_n,\ \text{and}\ t_n=\mathsf{J}_{\gamma_n\mathsf{A}}r_n. \tag{4.2}$$

Then it follows from (1.5) that assumption (b) in Theorem 3.2 is satisfied with

$$(\forall n\in\mathbb{N})\quad c_n=a_n. \tag{4.3}$$

In addition, for every $n\in\mathbb{N}$, $\mathsf{F}=\operatorname{Fix}(\mathsf{J}_{\gamma_n\mathsf{A}}\mathsf{R}_n)$ [4, Proposition 25.1(iv)] and we deduce from the firm nonexpansiveness of the operators $(\mathsf{J}_{\gamma_n\mathsf{A}})_{n\in\mathbb{N}}$ [4, Corollary 23.8] that

$$(\forall\mathsf{z}\in\mathsf{F})(\forall n\in\mathbb{N})\quad \|t_n-\mathsf{z}\|^2+\|r_n-\mathsf{J}_{\gamma_n\mathsf{A}}r_n-\mathsf{R}_n\mathsf{z}+\mathsf{z}\|^2\leqslant\|r_n-\mathsf{R}_n\mathsf{z}\|^2. \tag{4.4}$$

Now set

$$(\forall n\in\mathbb{N})\quad \widetilde{u}_n=u_n-\mathsf{E}(u_n\,|\,\mathcal{X}_n)+\mathsf{B}x_n. \tag{4.5}$$

Then we derive from (4.4) that (3.3) holds with

$$(\forall\mathsf{z}\in\mathsf{F})(\forall n\in\mathbb{N})\quad \begin{cases}\theta_{1,n}(\mathsf{z})=\mathsf{E}(\|r_n-\mathsf{J}_{\gamma_n\mathsf{A}}r_n-\mathsf{R}_n\mathsf{z}+\mathsf{z}\|^2\,|\,\mathcal{X}_n)\\ \mu_{1,n}(\mathsf{z})=\nu_{1,n}(\mathsf{z})=0\\ s_n(\mathsf{z})=x_n-\gamma_n\widetilde{u}_n-\mathsf{R}_n\mathsf{z}\\ d_n=-\gamma_n(\mathsf{E}(u_n\,|\,\mathcal{X}_n)-\mathsf{B}x_n).\end{cases} \tag{4.6}$$

Thus, (4.3), (4.6), (b), (c), and (e), imply that assumption (c) in Theorem 3.2 is satisfied since

$$\begin{aligned}\sum_{n\in\mathbb{N}}\sqrt{\lambda_n\mathsf{E}(\|d_n\|^2\,|\,\mathcal{X}_n)} &\leqslant 2(\tau_n+1)^{-1}\vartheta\sum_{n\in\mathbb{N}}\sqrt{\lambda_n\|\mathsf{E}(u_n\,|\,\mathcal{X}_n)-\mathsf{B}x_n\|^2}\\ &\leqslant 2\vartheta\sum_{n\in\mathbb{N}}\sqrt{\lambda_n}\|\mathsf{E}(u_n\,|\,\mathcal{X}_n)-\mathsf{B}x_n\|\\ &< +\infty. \end{aligned}\tag{4.7}$$

Moreover, for every $\mathsf{z}\in\mathsf{F}$ and $n\in\mathbb{N}$, we derive from (4.5), (1.1), and (4.1) that

$$\begin{aligned}&\mathsf{E}(\|s_n(\mathsf{z})\|^2\,|\,\mathcal{X}_n)\\ &=\mathsf{E}(\|x_n-\mathsf{z}-\gamma_n(\widetilde{u}_n-\mathsf{Bz})\|^2\,|\,\mathcal{X}_n)\\ &=\|x_n-\mathsf{z}\|^2-2\gamma_n\langle x_n-\mathsf{z}\mid \mathsf{E}(\widetilde{u}_n\,|\,\mathcal{X}_n)-\mathsf{Bz}\rangle+\gamma_n^2\mathsf{E}(\|\widetilde{u}_n-\mathsf{Bz}\|^2\,|\,\mathcal{X}_n)\\ &=\|x_n-\mathsf{z}\|^2-2\gamma_n\langle x_n-\mathsf{z}\mid \mathsf{B}x_n-\mathsf{Bz}\rangle\\ &\quad+\gamma_n^2\big(\mathsf{E}(\|u_n-\mathsf{E}(u_n\,|\,\mathcal{X}_n)\|^2\,|\,\mathcal{X}_n)\\ &\quad+2\langle u_n-\mathsf{E}(u_n\,|\,\mathcal{X}_n)\mid \mathsf{B}x_n-\mathsf{Bz}\rangle+\|\mathsf{B}x_n-\mathsf{Bz}\|^2\big)\\ &=\|x_n-\mathsf{z}\|^2-2\gamma_n\langle x_n-\mathsf{z}\mid \mathsf{B}x_n-\mathsf{Bz}\rangle\\ &\quad+\gamma_n^2\big(\mathsf{E}(\|u_n-\mathsf{E}(u_n\,|\,\mathcal{X}_n)\|^2\,|\,\mathcal{X}_n)+\|\mathsf{B}x_n-\mathsf{Bz}\|^2\big)\\ &\leqslant\|x_n-\mathsf{z}\|^2-\gamma_n(2\vartheta-\gamma_n)\|\mathsf{B}x_n-\mathsf{Bz}\|^2\\ &\quad+\gamma_n^2\mathsf{E}(\|u_n-\mathsf{E}(u_n\,|\,\mathcal{X}_n)\|^2\,|\,\mathcal{X}_n)\\ &\leqslant\|x_n-\mathsf{z}\|^2-\gamma_n\big(2\vartheta-(1+\tau_n)\gamma_n\big)\|\mathsf{B}x_n-\mathsf{Bz}\|^2+\gamma_n^2\zeta_n(\mathsf{z}). \end{aligned}\tag{4.8}$$

Thus, (3.4) is obtained by setting

$$(\forall n\in\mathbb{N})\quad \begin{cases}\theta_{2,n}(\mathsf{z})=\gamma_n\big(2\vartheta-(1+\tau_n)\gamma_n\big)\|\mathsf{B}x_n-\mathsf{Bz}\|^2\\ \mu_{2,n}(\mathsf{z})=0\\ \nu_{2,n}(\mathsf{z})=\gamma_n^2\zeta_n(\mathsf{z}).\end{cases}\tag{4.9}$$

Altogether, it follows from (d) and (e) that assumption (d) in Theorem 3.2 is also satisfied. By applying Theorem 3.2(i), we deduce from (e), (4.6), and (4.9) that

$$(\forall \mathsf{z}\in\mathsf{F})\quad \sum_{n\in\mathbb{N}}\lambda_n\|\mathsf{B}x_n-\mathsf{Bz}\|^2<+\infty\tag{4.10}$$

and

$$(\forall \mathsf{z}\in\mathsf{F})\quad \sum_{n\in\mathbb{N}}\lambda_n\mathsf{E}(\|r_n-\mathsf{J}_{\gamma_n\mathsf{A}}r_n-\mathsf{R}_n\mathsf{z}+\mathsf{z}\|^2\,|\,\mathcal{X}_n)<+\infty.\tag{4.11}$$

(i): See (4.10).

(ii): It follows from (4.2), (4.5), (2.5), and the nonexpansiveness of the operators $(\mathsf{J}_{\gamma_n\mathsf{A}})_{n\in\mathbb{N}}$ that

$$
\begin{aligned}
(\forall n\in\mathbb{N})\ &\|x_n-\gamma_n\mathsf{B}x_n-\mathsf{J}_{\gamma_n\mathsf{A}}(x_n-\gamma_n\mathsf{B}x_n)+\gamma_n\mathsf{B}\mathsf{z}\|^2\\
&=\|\mathsf{E}(x_n-\gamma_n\widetilde{u}_n\,|\,\mathcal{X}_n)-\mathsf{J}_{\gamma_n\mathsf{A}}(x_n-\gamma_n\mathsf{B}x_n)+\gamma_n\mathsf{B}\mathsf{z}\|^2\\
&\leqslant 3\big(\|\mathsf{E}(r_n-\mathsf{J}_{\gamma_n\mathsf{A}}r_n+\gamma_n\mathsf{B}\mathsf{z}\,|\,\mathcal{X}_n)\|^2+\gamma_n^2\|\mathsf{E}(u_n\,|\,\mathcal{X}_n)-\mathsf{B}x_n\|^2\\
&\quad+\|\mathsf{E}(\mathsf{J}_{\gamma_n\mathsf{A}}r_n\,|\,\mathcal{X}_n)-\mathsf{J}_{\gamma_n\mathsf{A}}(x_n-\gamma_n\mathsf{B}x_n)\|^2\big)\\
&\leqslant 3\big(\mathsf{E}(\|r_n-\mathsf{J}_{\gamma_n\mathsf{A}}r_n+\gamma_n\mathsf{B}\mathsf{z}\|^2\,|\,\mathcal{X}_n)+\gamma_n^2\mathsf{E}(\|u_n-\mathsf{B}x_n\|^2\,|\,\mathcal{X}_n)\\
&\quad+\mathsf{E}(\|\mathsf{J}_{\gamma_n\mathsf{A}}r_n-\mathsf{J}_{\gamma_n\mathsf{A}}(x_n-\gamma_n\mathsf{B}x_n)\|^2\,|\,\mathcal{X}_n)\big)\\
&\leqslant 3\big(\mathsf{E}(\|r_n-\mathsf{J}_{\gamma_n\mathsf{A}}r_n+\gamma_n\mathsf{B}\mathsf{z}\|^2\,|\,\mathcal{X}_n)+\gamma_n^2\mathsf{E}(\|u_n-\mathsf{B}x_n\|^2\,|\,\mathcal{X}_n)\\
&\quad+\mathsf{E}(\|r_n-(x_n-\gamma_n\mathsf{B}x_n)\|^2\,|\,\mathcal{X}_n)\big)\\
&=3\big(\mathsf{E}(\|r_n-\mathsf{J}_{\gamma_n\mathsf{A}}r_n-\mathsf{R}_n\mathsf{z}+\mathsf{z}\|^2\,|\,\mathcal{X}_n)\\
&\quad+2\gamma_n^2\mathsf{E}(\|u_n-\mathsf{B}x_n\|^2\,|\,\mathcal{X}_n)\big)\\
&\leqslant 3\big(\mathsf{E}(\|r_n-\mathsf{J}_{\gamma_n\mathsf{A}}r_n-\mathsf{R}_n\mathsf{z}+\mathsf{z}\|^2\,|\,\mathcal{X}_n)\\
&\quad+8\vartheta^2\mathsf{E}(\|u_n-\mathsf{B}x_n\|^2\,|\,\mathcal{X}_n)\big).
\end{aligned}
\tag{4.12}
$$

However, by (4.1),

$$
\begin{aligned}
(\forall n\in\mathbb{N})\quad &\mathsf{E}(\|u_n-\mathsf{B}x_n\|^2\,|\,\mathcal{X}_n)\\
&\leqslant 2\mathsf{E}(\|u_n-\mathsf{E}(u_n\,|\,\mathcal{X}_n)\|^2+\|\mathsf{E}(u_n\,|\,\mathcal{X}_n)-\mathsf{B}x_n\|^2\,|\,\mathcal{X}_n)\\
&\leqslant 2\big(\tau_n\|\mathsf{B}x_n-\mathsf{B}\mathsf{z}\|^2+\zeta_n+\|\mathsf{E}(u_n\,|\,\mathcal{X}_n)-\mathsf{B}x_n\|^2\big).
\end{aligned}
\tag{4.13}
$$

Since $\sup_{n\in\mathbb{N}}\tau_n<+\infty$ by (e), we therefore derive from (i), (c), and (d) that

$$
\sum_{n\in\mathbb{N}}\lambda_n\mathsf{E}(\|u_n-\mathsf{B}x_n\|^2\,|\,\mathcal{X}_n)<+\infty. \tag{4.14}
$$

Altogether, the claim follows from (4.11), (4.12), and (4.14).

(iii)–(iv): Let $\mathsf{z}\in\mathsf{F}$. We consider the two cases separately.

- Suppose that $\inf_{n\in\mathbb{N}}\lambda_n>0$. We derive from (i), (ii), and (e) that there exists $\widetilde{\Omega}\in\mathcal{F}$ such that $\mathsf{P}(\widetilde{\Omega})=1$,

$$
(\forall\omega\in\widetilde{\Omega})\quad x_n(\omega)-\mathsf{J}_{\gamma_n\mathsf{A}}\big(x_n(\omega)-\gamma_n\mathsf{B}x_n(\omega)\big)\to 0, \tag{4.15}
$$

and

$$
(\forall\omega\in\widetilde{\Omega})\quad \mathsf{B}x_n(\omega)\to\mathsf{B}\mathsf{z}. \tag{4.16}
$$

Now set

$$
(\forall n\in\mathbb{N})\quad y_n=\mathsf{J}_{\gamma_n\mathsf{A}}(x_n-\gamma_n\mathsf{B}x_n)\quad\text{and}\quad v_n=\gamma_n^{-1}(x_n-y_n)-\mathsf{B}x_n. \tag{4.17}
$$

It follows from (e), (4.15), and (4.16) that

$$
(\forall\omega\in\widetilde{\Omega})\quad y_n(\omega)-x_n(\omega)\to 0\quad\text{and}\quad v_n(\omega)\to-\mathsf{B}\mathsf{z}. \tag{4.18}
$$

Let $\omega\in\widetilde{\Omega}$. Assume that there exist $\mathsf{x}\in\mathsf{H}$ and a strictly increasing sequence $(k_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$ such that $x_{k_n}(\omega)\rightharpoonup\mathsf{x}$. Since $\mathsf{B}x_{k_n}(\omega)\to\mathsf{B}\mathsf{z}$

by (4.16) and since B is maximally monotone [4, Example 20.28], [4, Proposition 20.33(ii)] yields $\mathsf{Bx} = \mathsf{Bz}$. In addition, (4.18) implies that $y_{k_n}(\omega) \rightharpoonup \mathsf{x}$ and $v_{k_n}(\omega) \to -\mathsf{Bz} = -\mathsf{Bx}$. Since (4.17) entails that $(y_{k_n}(\omega), v_{k_n}(\omega))_{n\in\mathbb{N}}$ lies in the graph of A, [4, Proposition 20.33(ii)] asserts that $-\mathsf{Bx} \in \mathsf{Ax}$, i.e., $\mathsf{x} \in \mathsf{F}$. It therefore follows from Theorem 3.2(iii) that

$$x_n(\omega) \rightharpoonup x(\omega) \tag{4.19}$$

for every ω in some $\widehat{\Omega} \in \mathcal{F}$ such that $\widehat{\Omega} \subset \widetilde{\Omega}$ and $\mathsf{P}(\widehat{\Omega}) = 1$. We now turn to the strong convergence claims. To this end, take $\omega \in \widehat{\Omega}$. First, suppose that (g) holds. Then A is demiregular at $x(\omega)$. In view of (4.18) and (4.19), $y_n(\omega) \rightharpoonup x(\omega)$. Furthermore, $v_n(\omega) \to -\mathsf{B}x(\omega)$ and $(y_n(\omega), v_n(\omega))_{n\in\mathbb{N}}$ lies in the graph of A. Altogether $y_n(\omega) \to x(\omega)$ and therefore $x_n(\omega) \to x(\omega)$. Next, suppose that (h) holds. Then, since (4.16) yields $\mathsf{B}x_n(\omega) \to \mathsf{B}x(\omega)$, (4.19) implies that $x_n(\omega) \to x(\omega)$.

- Suppose that $\sum_{n\in\mathbb{N}} \tau_n < +\infty$, $\sum_{n\in\mathbb{N}} \lambda_n = +\infty$, and $(\forall n \in \mathbb{N})$ $\gamma_n = \gamma$. Let $\mathsf{T} = \mathsf{J}_{\gamma\mathsf{A}} \circ (\mathrm{Id} - \gamma\mathsf{B})$. We deduce from (i) that

$$(\forall \mathsf{z} \in \mathsf{F}) \quad \underline{\lim}\, \|\mathsf{B}x_n - \mathsf{Bz}\| = 0 \tag{4.20}$$

and from (ii) that

$$(\forall \mathsf{z} \in \mathsf{F}) \quad \underline{\lim}\, \|x_n - \mathsf{T}x_n - \gamma(\mathsf{B}x_n - \mathsf{Bz})\| = 0. \tag{4.21}$$

In view of (e), we obtain

$$\underline{\lim}\, \|\mathsf{T}x_n - x_n\| = 0. \tag{4.22}$$

In addition, since (e) and [4, Proposition 4.33] imply that T is nonexpansive, we derive from (1.5) that

$$\begin{aligned}
(\forall n \in \mathbb{N}) \quad & \|\mathsf{T}x_{n+1} - x_{n+1}\| \\
&= \|\mathsf{T}x_{n+1} - (1-\lambda_n)x_n - \lambda_n(\mathsf{J}_{\gamma\mathsf{A}}(x_n - \gamma u_n) + a_n)\| \\
&= \|\mathsf{T}x_{n+1} - \mathsf{T}x_n - (1-\lambda_n)(x_n - \mathsf{T}x_n) \\
&\quad - \lambda_n(\mathsf{J}_{\gamma\mathsf{A}}(x_n - \gamma u_n) - \mathsf{J}_{\gamma\mathsf{A}}(x_n - \gamma\mathsf{B}x_n)) - \lambda_n a_n\| \\
&\leqslant \|\mathsf{T}x_{n+1} - \mathsf{T}x_n\| + (1-\lambda_n)\|\mathsf{T}x_n - x_n\| \\
&\quad + \lambda_n\|\mathsf{J}_{\gamma\mathsf{A}}(x_n - \gamma u_n) - \mathsf{J}_{\gamma\mathsf{A}}(x_n - \gamma\mathsf{B}x_n)\| + \lambda_n\|a_n\| \\
&\leqslant \|x_{n+1} - x_n\| + (1-\lambda_n)\|\mathsf{T}x_n - x_n\| + \lambda_n\gamma\|u_n - \mathsf{B}x_n\| + \lambda_n\|a_n\| \\
&= \lambda_n\|\mathsf{J}_{\gamma\mathsf{A}}(x_n - \gamma u_n) + a_n - x_n\| + (1-\lambda_n)\|\mathsf{T}x_n - x_n\| \\
&\quad + \lambda_n\gamma\|u_n - \mathsf{B}x_n\| + \lambda_n\|a_n\| \\
&\leqslant \|\mathsf{T}x_n - x_n\| + \lambda_n\|\mathsf{J}_{\gamma\mathsf{A}}(x_n - \gamma u_n) - \mathsf{J}_{\gamma\mathsf{A}}(x_n - \gamma\mathsf{B}x_n)\| \\
&\quad + \lambda_n\gamma\|u_n - \mathsf{B}x_n\| + 2\lambda_n\|a_n\| \\
&\leqslant \|\mathsf{T}x_n - x_n\| + 2\lambda_n\big(\gamma\|u_n - \mathsf{B}x_n\| + \|a_n\|\big).
\end{aligned} \tag{4.23}$$

Now set

$$(\forall n\in\mathbb{N})\quad \xi_n=\gamma\sqrt{\lambda_n\mathsf{E}(\|u_n-\mathsf{B}x_n\|^2\,|\,\mathcal{X}_n)}+\lambda_n\sqrt{\mathsf{E}(\|a_n\|^2\,|\,\mathcal{X}_n)}. \tag{4.24}$$

Using (4.1), we get

$$\begin{aligned}\xi_n &\leqslant \gamma\sqrt{\lambda_n\mathsf{E}(\|u_n-\mathsf{E}(u_n\,|\,\mathcal{X}_n)\|^2\,|\,\mathcal{X}_n)}+\gamma\sqrt{\lambda_n\|\mathsf{E}(u_n\,|\,\mathcal{X}_n)-\mathsf{B}x_n\|^2}\\ &\quad+\lambda_n\sqrt{\mathsf{E}(\|a_n\|^2\,|\,\mathcal{X}_n)}\\ &\leqslant \gamma\sqrt{\lambda_n\tau_n}\|\mathsf{B}x_n-\mathsf{Bz}\|+\gamma\sqrt{\lambda_n\zeta_n(\mathsf{z})}+\gamma\sqrt{\lambda_n}\|\mathsf{E}(u_n\,|\,\mathcal{X}_n)-\mathsf{B}x_n\|\\ &\quad+\lambda_n\sqrt{\mathsf{E}(\|a_n\|^2\,|\,\mathcal{X}_n)}.\end{aligned} \tag{4.25}$$

Thus, (4.23) and (2.4) yield

$$\begin{aligned}(\forall n\in\mathbb{N})\quad &\mathsf{E}(\|\mathsf{T}x_{n+1}-x_{n+1}\|\,|\,\mathcal{X}_n)\\ &\leqslant \|\mathsf{T}x_n-x_n\|+2\lambda_n\big(\gamma\mathsf{E}(\|u_n-\mathsf{B}x_n\|\,|\,\mathcal{X}_n)+\mathsf{E}(\|a_n\|\,|\,\mathcal{X}_n)\big)\\ &\leqslant \|\mathsf{T}x_n-x_n\|+2\xi_n.\end{aligned} \tag{4.26}$$

In addition, according to the Cauchy-Schwarz inequality and (i),

$$\sum_{n\in\mathbb{N}}\sqrt{\lambda_n\tau_n}\|\mathsf{B}x_n-\mathsf{Bz}\|\leqslant\sqrt{\sum_{n\in\mathbb{N}}\tau_n}\sqrt{\sum_{n\in\mathbb{N}}\lambda_n\|\mathsf{B}x_n-\mathsf{Bz}\|^2}<+\infty. \tag{4.27}$$

Thus, it follows from assumptions (b)-(d) that $(\xi_n)_{n\in\mathbb{N}}\in\ell_+^1(\mathscr{X})$, and we deduce from Proposition 3.1(iii) and (4.26) that $(\|\mathsf{T}x_n-x_n\|)_{n\in\mathbb{N}}$ converges almost surely. We then derive from (4.22) that there exists $\widetilde{\Omega}\in\mathcal{F}$ such that $\mathsf{P}(\widetilde{\Omega})=1$ and (4.15) holds. Let $\omega\in\widetilde{\Omega}$. Suppose that there exist $\mathsf{x}\in\mathsf{H}$ and a strictly increasing sequence $(k_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$ such that $x_{k_n}(\omega)\rightharpoonup\mathsf{x}$. Since $x_{k_n}(\omega)\rightharpoonup\mathsf{x}$ and $\mathsf{T}x_{k_n}(\omega)-x_{k_n}(\omega)\to 0$, the demiclosedness principle [4, Corollary 4.18] asserts that $\mathsf{x}\in\mathsf{F}$. Hence, the weak convergence claim follows from Theorem 3.2(iii). To establish the strong convergence claims, set $\mathsf{w}=\mathsf{z}-\gamma\mathsf{Bz}$, and set $(\forall n\in\mathbb{N})$ $w_n=x_n-\gamma\mathsf{B}x_n$. Then $\mathsf{T}x_n=\mathsf{J}_{\gamma\mathsf{A}}w_n$ and $\mathsf{z}=\mathsf{Tz}=\mathsf{J}_{\gamma\mathsf{A}}\mathsf{w}$. Hence, appealing to the firm nonexpansiveness of $\mathsf{J}_{\gamma\mathsf{A}}$, we obtain

$$\begin{aligned}(\forall n\in\mathbb{N})\quad &\langle\mathsf{T}x_n-\mathsf{z}\mid x_n-\mathsf{T}x_n-\gamma(\mathsf{B}x_n-\mathsf{Bz})\rangle\\ &=\langle\mathsf{T}x_n-\mathsf{z}\mid w_n-\mathsf{T}x_n+\mathsf{z}-\mathsf{w}\rangle\\ &=\langle\mathsf{J}_{\gamma\mathsf{A}}w_n-\mathsf{J}_{\gamma\mathsf{A}}\mathsf{w}\mid(\mathrm{Id}-\mathsf{J}_{\gamma\mathsf{A}})w_n-(\mathrm{Id}-\mathsf{J}_{\gamma\mathsf{A}})\mathsf{w}\rangle\\ &\geqslant 0\end{aligned} \tag{4.28}$$

and therefore

$$(\forall n\in\mathbb{N})\quad\langle\mathsf{T}x_n-\mathsf{z}\mid x_n-\mathsf{T}x_n\rangle\geqslant\gamma\langle\mathsf{T}x_n-\mathsf{z}\mid\mathsf{B}x_n-\mathsf{Bz}\rangle. \tag{4.29}$$

Consequently, since T is nonexpansive and B satisfies (1.1),

$$\begin{aligned}(\forall n\in\mathbb{N})\quad &\|x_n-\mathsf{z}\|\,\|\mathsf{T}x_n-x_n\|\\ &\geqslant \|\mathsf{T}x_n-\mathsf{z}\|\,\|\mathsf{T}x_n-x_n\|\\ &\geqslant \langle \mathsf{T}x_n-\mathsf{z}\mid x_n-\mathsf{T}x_n\rangle\\ &\geqslant \gamma\langle \mathsf{T}x_n-\mathsf{z}\mid \mathsf{B}x_n-\mathsf{B}\mathsf{z}\rangle\\ &=\gamma\big(\langle \mathsf{T}x_n-x_n\mid \mathsf{B}x_n-\mathsf{B}\mathsf{z}\rangle+\langle x_n-\mathsf{z}\mid \mathsf{B}x_n-\mathsf{B}\mathsf{z}\rangle\big)\\ &\geqslant -\gamma\|\mathsf{T}x_n-x_n\|\,\|\mathsf{B}x_n-\mathsf{B}\mathsf{z}\|+\gamma\vartheta\|\mathsf{B}x_n-\mathsf{B}\mathsf{z}\|^2\\ &\geqslant -\frac{\gamma}{\vartheta}\|\mathsf{T}x_n-x_n\|\,\|x_n-\mathsf{z}\|+\gamma\vartheta\|\mathsf{B}x_n-\mathsf{B}\mathsf{z}\|^2 \end{aligned}\tag{4.30}$$

and hence

$$(\forall n\in\mathbb{N})\quad \|\mathsf{B}x_n-\mathsf{B}\mathsf{z}\|^2\leqslant\frac{1}{\gamma\vartheta}\Big(1+\frac{\gamma}{\vartheta}\Big)\|x_n-\mathsf{z}\|\,\|\mathsf{T}x_n-x_n\|.\tag{4.31}$$

Since, P-a.s., $(x_n)_{n\in\mathbb{N}}$ is bounded and $\mathsf{T}x_n-x_n\to 0$, we infer that $\mathsf{B}x_n\to\mathsf{B}\mathsf{z}$ P-a.s. Thus there exists $\widehat{\Omega}\in\mathcal{F}$ such that $\widehat{\Omega}\subset\widetilde{\Omega}$, $\mathsf{P}(\widehat{\Omega})=1$, and

$$(\forall\omega\in\widehat{\Omega})\quad x_n(\omega)\rightharpoonup x(\omega)\quad\text{and}\quad \mathsf{B}x_n(\omega)\to\mathsf{B}x(\omega).\tag{4.32}$$

Thus, (h) $\Rightarrow x_n(\omega)\to x(\omega)$. Finally, if (g) holds, the strong convergence of $(x_n(\omega))_{n\in\mathbb{N}}$ follows from the same arguments as in the previous case.

□

Remark 4.2. The demiregularity property in Theorem 4.1(iv) is satisfied by a wide class of operators, e.g., uniformly monotone operators or subdifferentials of proper lower semicontinuous uniformly convex functions; further examples are provided in [2, Proposition 2.4].

Remark 4.3. To place our analysis in perspective, we comment on results of the literature that seem the most pertinently related to Theorem 4.1.

(i) In the deterministic case, Theorem 4.1(iii) can be found in [14, Corollary 6.5].

(ii) In [1, Corollary 8], Problem 1.2 is considered in the special case when $\mathsf{H}=\mathbb{R}^N$ and solved via (1.5). Almost sure convergence properties are established under the following assumptions: $(\gamma_n)_{n\in\mathbb{N}}$ is a decreasing sequence in $]0,\vartheta]$ such that $\sum_{n\in\mathbb{N}}\gamma_n=+\infty$, $\lambda_n\equiv 1$, $a_n\equiv 0$, and the sequence $(x_n)_{n\in\mathbb{N}}$ is bounded *a priori.*

(iii) In [46], Problem 1.1 is addressed using Algorithm 1.3. The authors make the additional assumptions that

$$(\forall n\in\mathbb{N})\quad \mathsf{E}(u_n\,|\,\mathcal{X}_n)=\mathsf{B}x_n\quad\text{and}\quad a_n=0.\tag{4.33}$$

Furthermore they employ vanishing proximal parameters $(\gamma_n)_{n\in\mathbb{N}}$. Almost sure convergence properties of the sequence $(x_n)_{n\in\mathbb{N}}$ are then established under the additional assumption that B is uniformly monotone.

(iv) The recently posted paper [47] employs tools from [18] to investigate the convergence of a variant of (1.5) in which no errors $(a_n)_{n\in\mathbb{N}}$ are allowed in the implementation of the resolvents, and an inertial term is added, namely,

$$(\forall n\in\mathbb{N})\quad x_{n+1}=x_n+\lambda_n\big(\mathsf{J}_{\gamma_n\mathsf{A}}(x_n+\rho_n(x_n-x_{n-1})-\gamma_n u_n)-x_n\big),\quad \text{where}\quad \rho_n\in[0,1[\,. \tag{4.34}$$

In the case when $\rho_n\equiv 0$, assertions (iii) and (iv)(h) of Theorem 4.1 are obtained under the additional hypothesis that $\inf\lambda_n>0$ and that the stochastic approximations which can be performed are constrained by (4.33).

Next, we provide a version of Theorem 3.2 in which a variant of (1.5) featuring approximations $(\mathsf{A}_n)_{n\in\mathbb{N}}$ of the operator A is used. In the deterministic forward-backward method, such approximations were first used in [39, Proposition 3.2] (see also [14, Proposition 6.7]).

Proposition 4.4. *Consider the setting of Problem 1.1. Let x_0, $(u_n)_{n\in\mathbb{N}}$, and $(a_n)_{n\in\mathbb{N}}$ be random variables in $L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,1]$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $]0,2\vartheta[$, and let $(\mathsf{A}_n)_{n\in\mathbb{N}}$ be a sequence of maximally monotone operators from H to 2^{H}. Set*

$$(\forall n\in\mathbb{N})\quad x_{n+1}=x_n+\lambda_n\big(\mathsf{J}_{\gamma_n\mathsf{A}_n}(x_n-\gamma_n u_n)+a_n-x_n\big). \tag{4.35}$$

Suppose that assumptions (a)–(f) *in Theorem 4.1 are satisfied, as well as the following:*

(k) *There exist sequences $(\alpha_n)_{n\in\mathbb{N}}$ and $(\beta_n)_{n\in\mathbb{N}}$ in $[0,+\infty[$ such that $\sum_{n\in\mathbb{N}}\sqrt{\lambda_n}\alpha_n<+\infty$, $\sum_{n\in\mathbb{N}}\lambda_n\beta_n<+\infty$, and*

$$(\forall n\in\mathbb{N})(\forall \mathsf{x}\in\mathsf{H})\quad \|\mathsf{J}_{\gamma_n\mathsf{A}_n}\mathsf{x}-\mathsf{J}_{\gamma_n\mathsf{A}}\mathsf{x}\|\leqslant\alpha_n\|\mathsf{x}\|+\beta_n. \tag{4.36}$$

Then the conclusions of Theorem 4.1 remain valid.

Proof. Let $\mathsf{z}\in\mathsf{F}$. We have

$$(\forall n\in\mathbb{N})\quad \|x_{n+1}-\mathsf{z}\|\leqslant(1-\lambda_n)\|x_n-\mathsf{z}\|+\lambda_n\|\mathsf{J}_{\gamma_n\mathsf{A}_n}(x_n-\gamma_n u_n)-\mathsf{z}\|+\lambda_n\|a_n\|. \tag{4.37}$$

In addition,

$$\begin{aligned}(\forall n\in\mathbb{N})\ &\|\mathsf{J}_{\gamma_n\mathsf{A}_n}(x_n-\gamma_n u_n)-\mathsf{z}\|\\
&\leqslant\|\mathsf{J}_{\gamma_n\mathsf{A}_n}(x_n-\gamma_n u_n)-\mathsf{J}_{\gamma_n\mathsf{A}_n}(\mathsf{z}-\gamma_n\mathsf{B}\mathsf{z})\|\\
&\quad+\|\mathsf{J}_{\gamma_n\mathsf{A}_n}(\mathsf{z}-\gamma_n\mathsf{B}\mathsf{z})-\mathsf{J}_{\gamma_n\mathsf{A}}(\mathsf{z}-\gamma_n\mathsf{B}\mathsf{z})\|\\
&\leqslant\|x_n-\gamma_n u_n-\mathsf{z}+\gamma_n\mathsf{B}\mathsf{z}\|+\|\mathsf{J}_{\gamma_n\mathsf{A}_n}(\mathsf{z}-\gamma_n\mathsf{B}\mathsf{z})-\mathsf{J}_{\gamma_n\mathsf{A}}(\mathsf{z}-\gamma_n\mathsf{B}\mathsf{z})\|\\
&\leqslant\|x_n-\mathsf{z}-\gamma_n(\mathsf{B}x_n-\mathsf{B}\mathsf{z})-\gamma_n(u_n-\mathsf{E}(u_n\,|\,\mathcal{X}_n))\|\\
&\quad+\gamma_n\|\mathsf{E}(u_n\,|\,\mathcal{X}_n)-\mathsf{B}x_n\|+\|\mathsf{J}_{\gamma_n\mathsf{A}_n}(\mathsf{z}-\gamma_n\mathsf{B}\mathsf{z})-\mathsf{J}_{\gamma_n\mathsf{A}}(\mathsf{z}-\gamma_n\mathsf{B}\mathsf{z})\|.\end{aligned} \tag{4.38}$$

On the other hand, using assumptions (d) and (e) in Theorem 4.1 as well as (1.1), we obtain as in (4.8)

$$\begin{aligned}(\forall n\in\mathbb{N})\quad & \mathsf{E}(\|x_n-\mathsf{z}-\gamma_n(\mathsf{B}x_n-\mathsf{B}\mathsf{z})-\gamma_n(u_n-\mathsf{E}(u_n\,|\,\mathcal{X}_n))\|^2\,|\,\mathcal{X}_n)\\ &\leqslant \|x_n-\mathsf{z}\|^2-\gamma_n\big(2\vartheta-(1+\tau_n)\gamma_n\big)\|\mathsf{B}x_n-\mathsf{B}\mathsf{z}\|^2+\gamma_n^2\zeta_n(\mathsf{z})\\ &\leqslant \|x_n-\mathsf{z}\|^2+\gamma_n^2\zeta_n(\mathsf{z}),\end{aligned}\tag{4.39}$$

which implies that

$$\begin{aligned}(\forall n\in\mathbb{N})\quad & \mathsf{E}(\|x_n-\mathsf{z}-\gamma_n(\mathsf{B}x_n-\mathsf{B}\mathsf{z})-\gamma_n(u_n-\mathsf{E}(u_n\,|\,\mathcal{X}_n))\|\,|\,\mathcal{X}_n)\\ &\leqslant \|x_n-\mathsf{z}\|+\gamma_n\sqrt{\zeta_n(\mathsf{z})}.\end{aligned}\tag{4.40}$$

Combining (4.37), (4.38), and (4.40) yields

$$\begin{aligned}(\forall n\in\mathbb{N})\quad & \mathsf{E}(\|x_{n+1}-\mathsf{z}\|\,|\,\mathcal{X}_n)\\ &\leqslant \|x_n-\mathsf{z}\|+\lambda_n\gamma_n\sqrt{\zeta_n(\mathsf{z})}+\lambda_n\gamma_n\|\mathsf{E}(u_n\,|\,\mathcal{X}_n)-\mathsf{B}x_n\|\\ &\quad+\lambda_n\|\mathsf{J}_{\gamma_n\mathsf{A}_n}(\mathsf{z}-\gamma_n\mathsf{B}\mathsf{z})-\mathsf{J}_{\gamma_n\mathsf{A}}(\mathsf{z}-\gamma_n\mathsf{B}\mathsf{z})\|\\ &\quad+\lambda_n\mathsf{E}(\|a_n\|\,|\,\mathcal{X}_n)\\ &\leqslant \|x_n-\mathsf{z}\|+\gamma_n\sqrt{\lambda_n\zeta_n(\mathsf{z})}+\gamma_n\sqrt{\lambda_n}\|\mathsf{E}(u_n\,|\,\mathcal{X}_n)-\mathsf{B}x_n\|\\ &\quad+\lambda_n\|\mathsf{J}_{\gamma_n\mathsf{A}_n}(\mathsf{z}-\gamma_n\mathsf{B}\mathsf{z})-\mathsf{J}_{\gamma_n\mathsf{A}}(\mathsf{z}-\gamma_n\mathsf{B}\mathsf{z})\|\\ &\quad+\lambda_n\sqrt{\mathsf{E}(\|a_n\|^2\,|\,\mathcal{X}_n)}.\end{aligned}\tag{4.41}$$

Since [4, Proposition 4.33] asserts that

$$\text{the operators } (\mathrm{Id}-\gamma_n\mathsf{B})_{n\in\mathbb{N}} \text{ are nonexpansive,}\tag{4.42}$$

it follows from (k) that

$$\begin{aligned}(\forall n\in\mathbb{N})\quad & \lambda_n\|\mathsf{J}_{\gamma_n\mathsf{A}_n}(\mathsf{z}-\gamma_n\mathsf{B}\mathsf{z})-\mathsf{J}_{\gamma_n\mathsf{A}}(\mathsf{z}-\gamma_n\mathsf{B}\mathsf{z})\|\\ &\leqslant \sqrt{\lambda_n}\alpha_n\|\mathsf{z}-\gamma_n\mathsf{B}\mathsf{z}\|+\lambda_n\beta_n\\ &\leqslant \sqrt{\lambda_n}\alpha_n\|\mathsf{z}\|+\lambda_n\beta_n.\end{aligned}\tag{4.43}$$

Thus,

$$\sum_{n\in\mathbb{N}}\lambda_n\|\mathsf{J}_{\gamma_n\mathsf{A}_n}(\mathsf{z}-\gamma_n\mathsf{B}\mathsf{z})-\mathsf{J}_{\gamma_n\mathsf{A}}(\mathsf{z}-\gamma_n\mathsf{B}\mathsf{z})\|<+\infty.\tag{4.44}$$

In view of assumptions (a)-(e) in Theorem 4.1 and (4.44), we deduce from (4.41) and Proposition 3.1(ii) that $(x_n)_{n\in\mathbb{N}}$ is almost surely bounded. In turn, (4.42) asserts that $(x_n-\gamma_n\mathsf{B}x_n)_{n\in\mathbb{N}}$ is likewise. Now set

$$(\forall n\in\mathbb{N})\quad \widetilde{a}_n=\mathsf{J}_{\gamma_n\mathsf{A}_n}(x_n-\gamma_n u_n)-\mathsf{J}_{\gamma_n\mathsf{A}}(x_n-\gamma_n u_n)+a_n.\tag{4.45}$$

Then (4.35) can be rewritten as

$$(\forall n\in\mathbb{N})\quad x_{n+1}=x_n+\lambda_n\big(\mathsf{J}_{\gamma_n\mathsf{A}}(x_n-\gamma_n u_n)+\widetilde{a}_n-x_n\big).\tag{4.46}$$

However,

$$
\begin{aligned}
(\forall n\in\mathbb{N})\quad &\sqrt{\mathsf{E}(\|\widetilde{a}_n\|^2\,|\,\mathcal{X}_n)}\\
&\leqslant \sqrt{\mathsf{E}(\|\mathsf{J}_{\gamma_n\mathsf{A}_n}(x_n-\gamma_nu_n)-\mathsf{J}_{\gamma_n\mathsf{A}}(x_n-\gamma_nu_n)\|^2\,|\,\mathcal{X}_n)}\\
&\quad+\sqrt{\mathsf{E}(\|a_n\|^2\,|\,\mathcal{X}_n)}.
\end{aligned}
\tag{4.47}
$$

On the other hand, according to (k), assumption (d) in Theorem 4.1, and (4.42),

$$
\begin{aligned}
(\forall n\in\mathbb{N})\quad &\lambda_n\sqrt{\mathsf{E}(\|\mathsf{J}_{\gamma_n\mathsf{A}_n}(x_n-\gamma_nu_n)-\mathsf{J}_{\gamma_n\mathsf{A}}(x_n-\gamma_nu_n)\|^2\,|\,\mathcal{X}_n)}\\
&\leqslant \lambda_n\sqrt{\mathsf{E}((\alpha_n\|x_n-\gamma_nu_n\|+\beta_n)^2\,|\,\mathcal{X}_n)}\\
&\leqslant \lambda_n\sqrt{\mathsf{E}((\alpha_n\|x_n-\gamma_n\mathsf{B}x_n\|+\gamma_n\|u_n-\mathsf{B}x_n\|+\beta_n)^2\,|\,\mathcal{X}_n)}\\
&\leqslant \lambda_n\alpha_n\big(\|x_n-\gamma_n\mathsf{B}x_n\|+\gamma_n\sqrt{\mathsf{E}(\|u_n-\mathsf{B}x_n\|^2\,|\,\mathcal{X}_n)}\big)+\lambda_n\beta_n\\
&\leqslant \lambda_n\alpha_n\big(\|x_n-\gamma_n\mathsf{B}x_n\|+\gamma_n\|\mathsf{E}(u_n\,|\,\mathcal{X}_n)-\mathsf{B}x_n\|\\
&\quad+\gamma_n\sqrt{\mathsf{E}(\|u_n-\mathsf{E}(u_n\,|\,\mathcal{X}_n)\|^2\,|\,\mathcal{X}_n)}\big)+\lambda_n\beta_n\\
&\leqslant \lambda_n\alpha_n\big(\|x_n-\gamma_n\mathsf{B}x_n\|+\gamma_n\|\mathsf{E}(u_n\,|\,\mathcal{X}_n)-\mathsf{B}x_n\|\\
&\quad+\gamma_n\sqrt{\tau_n}\|\mathsf{B}x_n-\mathsf{Bz}\|+\gamma_n\sqrt{\zeta_n(\mathsf{z})}\big)+\lambda_n\beta_n\\
&\leqslant \sqrt{\lambda_n}\alpha_n\big(\|x_n-\gamma_n\mathsf{B}x_n\|+\gamma_n\sqrt{\lambda_n}\|\mathsf{E}(u_n\,|\,\mathcal{X}_n)-\mathsf{B}x_n\|\\
&\quad+\gamma_n\sqrt{\tau_n}\|\mathsf{B}x_n-\mathsf{Bz}\|+\gamma_n\sqrt{\lambda_n\zeta_n(\mathsf{z})}\big)+\lambda_n\beta_n.
\end{aligned}
\tag{4.48}
$$

However, assumptions (c) and (d) in Theorem 4.1 guarantee that the sequences $(\sqrt{\lambda_n}\|\mathsf{E}(u_n\,|\,\mathcal{X}_n)-\mathsf{B}x_n\|)_{n\in\mathbb{N}}$ and $(\sqrt{\lambda_n\zeta_n(\mathsf{z})})_{n\in\mathbb{N}}$ are P-a.s. bounded. Since $(\mathsf{B}x_n)_{n\in\mathbb{N}}$ and $(x_n-\gamma_n\mathsf{B}x_n)_{n\in\mathbb{N}}$ are likewise, it follows from (k) and (4.42) that

$$
\sum_{n\in\mathbb{N}}\lambda_n\sqrt{\mathsf{E}(\|\mathsf{J}_{\gamma_n\mathsf{A}_n}(x_n-\gamma_nu_n)-\mathsf{J}_{\gamma_n\mathsf{A}}(x_n-\gamma_nu_n)\|^2\,|\,\mathcal{X}_n)}<+\infty,
\tag{4.49}
$$

and consequently that

$$
\sum_{n\in\mathbb{N}}\lambda_n\sqrt{\mathsf{E}(\|\widetilde{a}_n\|^2\,|\,\mathcal{X}_n)}<+\infty.
\tag{4.50}
$$

Applying Theorem 4.1 to algorithm (4.46) then yields the claims. □

5. Applications

As discussed in the Introduction, the forward-backward algorithm is quite versatile and it can be applied in various forms. Many standard applications of Theorem 4.1 can of course be recovered for specific choices of A and B, in particular Problem 1.2. Using the product space framework of [2], it can also be applied to solve systems of coupled monotone inclusions. On the other hand, using the approach proposed in [16, 20], it can be used to solve strongly monotone composite inclusions (in particular, strongly convex composite minimization problems), say,

$$\text{(5.1)} \qquad \text{find } \mathsf{x} \in \mathsf{H} \text{ such that } \mathsf{z} \in \mathsf{A}\mathsf{x} + \sum_{k=1}^{q} \mathsf{L}_k^*\big((\mathsf{B}_k \,\square\, \mathsf{D}_k)(\mathsf{L}_k\mathsf{x} - \mathsf{r}_k)\big) + \rho\mathsf{x},$$

since their dual problems assume the general form of Problem 1.1 and the primal solution can trivially be recovered from any dual solution. In (5.1), $\mathsf{z} \in \mathsf{H}$, $\rho \in \,]0,+\infty[$ and, for every $k \in \{1,\ldots,q\}$, r_k lies in a real Hilbert space G_k, $\mathsf{B}_k \colon \mathsf{G}_k \to 2^{\mathsf{G}_k}$ is maximally monotone, $\mathsf{D}_k \colon \mathsf{G}_k \to 2^{\mathsf{G}_k}$ is maximally monotone and strongly monotone, $\mathsf{B}_k \,\square\, \mathsf{D}_k = (\mathsf{B}_k^{-1} + \mathsf{D}_k^{-1})^{-1}$, and $\mathsf{L}_k \in \mathcal{B}(\mathsf{H},\mathsf{G}_k)$. In such instances the forward-backward algorithm actually yields a primal-dual method which produces a sequence converging to the primal solution (see [20, Section 5] for details). Now suppose that, in addition, $\mathsf{C} \colon \mathsf{H} \to \mathsf{H}$ is cocoercive. As in [17], consider the primal problem

$$\text{(5.2)} \qquad \text{find } \mathsf{x} \in \mathsf{H} \text{ such that } \mathsf{z} \in \mathsf{A}\mathsf{x} + \sum_{k=1}^{q} \mathsf{L}_k^*\big((\mathsf{B}_k \,\square\, \mathsf{D}_k)(\mathsf{L}_k\mathsf{x} - \mathsf{r}_k)\big) + \mathsf{C}\mathsf{x},$$

together with the dual problem

$$\text{(5.3)} \quad \text{find } \mathsf{v}_1 \in \mathsf{G}_1, \ldots, \mathsf{v}_q \in \mathsf{G}_q \quad \text{such that}$$
$$(\forall k \in \{1,\ldots,q\}) \quad -\mathsf{r}_k \in -\mathsf{L}_k^*(\mathsf{A}+\mathsf{C})^{-1}\left(\mathsf{z} - \sum_{l=1}^{q} \mathsf{L}_l^*\mathsf{v}_l\right) + \mathsf{B}_k^{-1}\mathsf{v}_k + \mathsf{D}_k^{-1}\mathsf{v}_k.$$

Using renorming techniques in the primal-dual space going back to [34] in the context of finite-dimensional minimization problems, the primal-dual problem (5.2)–(5.3) can be reduced to an instance of Problem 1.1 [20, 53] (see also [23]) and therefore solved via Theorem 4.1. Next, we explicitly illustrate an application of this approach in the special case when (5.2)–(5.3) is a minimization problem.

5.1. **A stochastic primal-dual minimization method.** We denote by $\Gamma_0(\mathsf{H})$ the class of proper lower semicontinuous convex functions. The Moreau subdifferential of $\mathsf{f} \in \Gamma_0(\mathsf{H})$ is the maximally monotone operator

$$\text{(5.4)} \qquad \partial\mathsf{f} \colon \mathsf{H} \to 2^{\mathsf{H}} \colon \mathsf{x} \mapsto \big\{\mathsf{u} \in \mathsf{H} \;\big|\; (\forall \mathsf{y} \in \mathsf{H}) \;\; \langle \mathsf{y}-\mathsf{x} \mid \mathsf{u}\rangle + \mathsf{f}(\mathsf{x}) \leqslant \mathsf{f}(\mathsf{y})\big\}.$$

The inf-convolution of $\mathsf{f} \colon \mathsf{H} \to \,]-\infty,+\infty]$ and $\mathsf{h} \colon \mathsf{H} \to \,]-\infty,+\infty]$ is defined as $\mathsf{f} \,\square\, \mathsf{h} \colon \mathsf{H} \to [-\infty,+\infty] \colon \mathsf{x} \mapsto \inf_{\mathsf{y}\in\mathsf{H}}\big(\mathsf{f}(\mathsf{y}) + \mathsf{h}(\mathsf{x}-\mathsf{y})\big)$. The conjugate of a function $\mathsf{f} \in \Gamma_0(\mathsf{H})$ is the function $\mathsf{f}^* \in \Gamma_0(\mathsf{H})$ defined by $(\forall \mathsf{u} \in \mathsf{H})$ $\mathsf{f}^*(\mathsf{u}) = \sup_{\mathsf{x}\in\mathsf{H}}(\langle \mathsf{x} \mid \mathsf{u}\rangle - \mathsf{f}(\mathsf{x}))$. Let U be a strongly positive self-adjoint operator in $\mathcal{B}(\mathsf{H})$. The proximity operator of $\mathsf{f} \in \Gamma_0(\mathsf{H})$ relative to the metric induced by U is

$$\text{(5.5)} \qquad \operatorname{prox}_{\mathsf{f}}^{\mathsf{U}} \colon \mathsf{H} \to \mathsf{H} \colon \mathsf{x} \to \underset{\mathsf{y}\in\mathsf{H}}{\operatorname{argmin}} \left(\mathsf{f}(\mathsf{y}) + \frac{1}{2}\|\mathsf{x}-\mathsf{y}\|_{\mathsf{U}}^2\right),$$

where

$$(\forall \mathsf{x} \in \mathsf{H}) \qquad \|\mathsf{x}\|_{\mathsf{U}} = \sqrt{\langle \mathsf{x} \mid \mathsf{U}\mathsf{x}\rangle}. \tag{5.6}$$

We have $\operatorname{prox}_{\mathsf{f}}^{\mathsf{U}} = \mathsf{J}_{\mathsf{U}^{-1}\partial\mathsf{f}}$.

We apply Theorem 4.1 to derive a stochastic version of a primal-dual optimization algorithm for solving a multivariate optimization problem which was first proposed in [17, Section 4].

Problem 5.1. Let $\mathsf{f} \in \Gamma_0(\mathsf{H})$, let $\mathsf{h}\colon \mathsf{H} \to \mathbb{R}$ be convex and differentiable with a Lipschitz-continuous gradient, and let q be a strictly positive integer. For every $k \in \{1,\ldots,q\}$, let G_k be a separable Hilbert space, let $\mathsf{g}_k \in \Gamma_0(\mathsf{G}_k)$, let $\mathsf{j}_k \in \Gamma_0(\mathsf{G}_k)$ be strongly convex, and let $\mathsf{L}_k \in \mathcal{B}(\mathsf{H},\mathsf{G}_k)$. Let $\mathbf{G} = \mathsf{G}_1 \oplus \cdots \oplus \mathsf{G}_q$ be the direct Hilbert sum of $\mathsf{G}_1,\ldots,\mathsf{G}_q$, and suppose that there exists $\overline{\mathsf{x}} \in \mathsf{H}$ such that

$$0 \in \partial\mathsf{f}(\overline{\mathsf{x}}) + \sum_{k=1}^{q} \mathsf{L}_k^*(\partial\mathsf{g}_k \,\square\, \partial\mathsf{j}_k)(\mathsf{L}_k\overline{\mathsf{x}}) + \nabla\mathsf{h}(\overline{\mathsf{x}}). \tag{5.7}$$

Let F be the set of solutions to the problem

$$\underset{\mathsf{x}\in\mathsf{H}}{\text{minimize}}\;\; \mathsf{f}(\mathsf{x}) + \sum_{k=1}^{q}(\mathsf{g}_k \,\square\, \mathsf{j}_k)(\mathsf{L}_k\mathsf{x}) + \mathsf{h}(\mathsf{x}) \tag{5.8}$$

and let $\mathbf{F}^*$ be the set of solutions to the dual problem

$$\underset{\mathbf{v}\in\mathbf{G}}{\text{minimize}}\;\; (\mathsf{f}^* \,\square\, \mathsf{h}^*)\bigg(-\sum_{k=1}^{q}\mathsf{L}_k^*\mathsf{v}_k\bigg) + \sum_{k=1}^{q}\big(\mathsf{g}_k^*(\mathsf{v}_k) + \mathsf{j}_k^*(\mathsf{v}_k)\big), \tag{5.9}$$

where we denote by $\mathbf{v} = (\mathsf{v}_1,\ldots,\mathsf{v}_q)$ a generic point in $\mathbf{G}$. The problem is to find a point in $\mathsf{F} \times \mathbf{F}^*$.

We address the case when only stochastic approximations of the gradients of h and $(\mathsf{j}_k^*)_{1\leqslant k\leqslant q}$ and approximations of the function f are available to solve Problem 5.1.

Algorithm 5.2. *Consider the setting of Problem 5.1 and let* $\mathsf{W} \in \mathcal{B}(\mathsf{H})$ *be strongly positive and self-adjoint. Let* $(\mathsf{f}_n)_{n\in\mathbb{N}}$ *be a sequence in* $\Gamma_0(\mathsf{H})$*, let* $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence in* $]0,1]$ *such that* $\sum_{n\in\mathbb{N}}\lambda_n = +\infty$*, and, for every* $k \in \{1,\ldots,q\}$*, let* $\mathsf{U}_k \in \mathcal{B}(\mathsf{G}_k)$ *be strongly positive and self-adjoint. Let* x_0*,* $(u_n)_{n\in\mathbb{N}}$*, and* $(b_n)_{n\in\mathbb{N}}$ *be random variables in* $L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$*, and let* $\boldsymbol{v}_0$*,* $(\boldsymbol{s}_n)_{n\in\mathbb{N}}$*, and* $(\boldsymbol{c}_n)_{n\in\mathbb{N}}$ *be random variables in* $L^2(\Omega,\mathcal{F},\mathsf{P};\mathbf{G})$*. Iterate*

$$
\begin{array}{l}
\text{for}\; n = 0,1,\ldots\\
\left\lfloor
\begin{array}{l}
y_n = \operatorname{prox}_{\mathsf{f}_n}^{\mathsf{W}^{-1}}\bigg(x_n - \mathsf{W}\bigg(\displaystyle\sum_{k=1}^{q}\mathsf{L}_k^* v_{k,n} + u_n\bigg)\bigg) + b_n\\
x_{n+1} = x_n + \lambda_n(y_n - x_n)\\
\text{for}\; k = 1,\ldots,q\\
\left\lfloor
\begin{array}{l}
w_{k,n} = \operatorname{prox}_{\mathsf{g}_k^*}^{\mathsf{U}_k^{-1}}\big(v_{k,n} + \mathsf{U}_k(\mathsf{L}_k(2y_n - x_n) - s_{k,n})\big) + c_{k,n}\\
v_{k,n+1} = v_{k,n} + \lambda_n(w_{k,n} - v_{k,n}).
\end{array}
\right.
\end{array}
\right.
\end{array}
\tag{5.10}
$$

Proposition 5.3. *Consider the setting of Problem 5.1, let $\mathscr{X} = (\mathfrak{X}_n)_{n\in\mathbb{N}}$ be a sequence of sub-sigma-algebras of $\mathcal{F}$, and let $(x_n)_{n\in\mathbb{N}}$ and $(\boldsymbol{v}_n)_{n\in\mathbb{N}}$ be sequences generated by Algorithm 5.2. Let $\mu \in \left]0,+\infty\right[$ be a Lipschitz constant of the gradient of $\mathsf{h}\circ\mathsf{W}^{1/2}$ and, for every $k \in \{1,\ldots,q\}$, let $\nu_k \in \left]0,+\infty\right[$ be a Lipschitz constant of the gradient of $\mathsf{j}_k^*\circ\mathsf{U}_k^{1/2}$. Assume that the following are satisfied:*

(a) $(\forall n\in\mathbb{N})\ \sigma(x_{n'},\boldsymbol{v}_{n'})_{0\leqslant n'\leqslant n} \subset \mathfrak{X}_n \subset \mathfrak{X}_{n+1}$.
(b) $\sum_{n\in\mathbb{N}} \lambda_n\sqrt{\mathsf{E}(\|b_n\|^2\,|\,\mathfrak{X}_n)} < +\infty$ *and* $\sum_{n\in\mathbb{N}} \lambda_n\sqrt{\mathsf{E}(\|\boldsymbol{c}_n\|^2\,|\,\mathfrak{X}_n)} < +\infty$.
(c) $\sum_{n\in\mathbb{N}} \sqrt{\lambda_n}\|\mathsf{E}(u_n\,|\,\mathfrak{X}_n) - \nabla\mathsf{h}(x_n)\| < +\infty$.
(d) *For every* $k \in \{1,\ldots,q\}$, $\sum_{n\in\mathbb{N}} \sqrt{\lambda_n}\|\mathsf{E}(s_{k,n}\mid\mathfrak{X}_n) - \nabla\mathsf{j}_k^*(v_{k,n})\| < +\infty$.
(e) *There exists a summable sequence $(\tau_n)_{n\in\mathbb{N}}$ in $\left[0,+\infty\right[$ such that, for every $(\mathsf{x},\mathsf{v}) \in \mathsf{F}\times\mathbf{F}^*$, there exists $\big(\zeta_n(\mathsf{x},\mathsf{v})\big)_{n\in\mathbb{N}} \in \ell_+^\infty(\mathscr{X})$ such that $\big(\lambda_n\zeta_n(\mathsf{x},\mathsf{v})\big)_{n\in\mathbb{N}} \in \ell_+^{1/2}(\mathscr{X})$ and*

$$
\begin{aligned}
&(\forall n\in\mathbb{N})\quad \mathsf{E}(\|u_n - \mathsf{E}(u_n\,|\,\mathfrak{X}_n)\|^2\,|\,\mathfrak{X}_n) + \mathsf{E}(\|\boldsymbol{s}_n - \mathsf{E}(\boldsymbol{s}_n\,|\,\mathfrak{X}_n)\|^2\,|\,\mathfrak{X}_n)\\
&\leqslant \tau_n\left(\|\nabla\mathsf{h}(x_n) - \nabla\mathsf{h}(\mathsf{x})\|^2 + \sum_{k=1}^q \|\nabla\mathsf{j}_k^*(v_{k,n}) - \nabla\mathsf{j}_k^*(\mathsf{v}_k)\|^2\right) + \zeta_n(\mathsf{x},\mathsf{v}).
\end{aligned}
\tag{5.11}
$$

(f) *There exist sequences $(\alpha_n)_{n\in\mathbb{N}}$ and $(\beta_n)_{n\in\mathbb{N}}$ in $\left[0,+\infty\right[$ such that $\sum_{n\in\mathbb{N}}\sqrt{\lambda_n}\alpha_n < +\infty$, $\sum_{n\in\mathbb{N}}\lambda_n\beta_n < +\infty$, and*

$$
(\forall n\in\mathbb{N})(\forall \mathsf{x}\in\mathsf{H})\quad \|\mathrm{prox}_{\mathsf{f}_n}^{\mathsf{W}^{-1}}\mathsf{x} - \mathrm{prox}_{\mathsf{f}}^{\mathsf{W}^{-1}}\mathsf{x}\| \leqslant \alpha_n\|\mathsf{x}\| + \beta_n. \tag{5.12}
$$

(g) $\max\{\mu,\nu_1,\ldots,\nu_q\} < 2\left(1 - \sqrt{\sum_{k=1}^q \|\mathsf{U}_k^{1/2}\mathsf{L}_k\mathsf{W}^{1/2}\|^2}\right)$.

Then, the following hold for some F-valued random variable x and some $\mathbf{F}^*$-valued random variable $\boldsymbol{v}$:

(i) *$(x_n)_{n\in\mathbb{N}}$ converges weakly* P-a.s. *to x and $(\boldsymbol{v}_n)_{n\in\mathbb{N}}$ converges weakly almost surely to $\boldsymbol{v}$.*
(ii) *Suppose that $\nabla\mathsf{h}$ is demiregular at every $\mathsf{x}\in\mathsf{F}$. Then $(x_n)_{n\in\mathbb{N}}$ converges strongly almost surely to x.*
(iii) *Suppose that there exists $k\in\{1,\ldots,q\}$ such that, for every $\mathsf{v}\in\mathbf{F}^*$, $\nabla\mathsf{j}_k^*$ is demiregular at v_k. Then $(v_{k,n})_{n\in\mathbb{N}}$ converges strongly almost surely to v_k.*

Proof. The proof relies on the ability to employ a constant proximal parameter in algorithm (4.35). Let us define $\mathbf{K} = \mathsf{H}\oplus\mathbf{G}$, $\mathbf{g}\colon \mathbf{G}\to\left]-\infty,+\infty\right] : \mathbf{v}\mapsto \sum_{k=1}^q \mathsf{g}_k(\mathsf{v}_k)$, $\mathbf{j}\colon \mathbf{G}\to\left]-\infty,+\infty\right] : \mathbf{v}\mapsto\sum_{k=1}^q \mathsf{j}_k(\mathsf{v}_k)$, $\mathbf{L}\colon \mathsf{H}\to\mathbf{G}\colon \mathsf{x}\mapsto \big(\mathsf{L}_k\mathsf{x}\big)_{1\leqslant k\leqslant q}$, and $\mathbf{U}\colon\mathbf{G}\to\mathbf{G}\colon\mathbf{v}\mapsto(\mathsf{U}_1\mathsf{v}_1,\ldots,\mathsf{U}_q\mathsf{v}_q)$. Let us now introduce the set-valued operator

$$
\mathbf{A}\colon \mathbf{K}\to 2^{\mathbf{K}}\colon (\mathsf{x},\mathbf{v})\mapsto \big(\partial\mathsf{f}(\mathsf{x}) + \mathbf{L}^*\mathbf{v}\big)\times\big(-\mathbf{L}\mathsf{x} + \partial\mathbf{g}^*(\mathbf{v})\big), \tag{5.13}
$$

the single-valued operator

$$\mathbf{B}\colon \mathbf{K} \to \mathbf{K}\colon (\mathsf{x}, \mathbf{v}) \mapsto \big(\nabla \mathsf{h}(\mathsf{x}), \nabla \mathbf{j}^*(\mathbf{v})\big), \tag{5.14}$$

and the bounded linear operator

$$\mathbf{V}\colon \mathbf{K} \to \mathbf{K}\colon (\mathsf{x}, \mathbf{v}) \mapsto \big(\mathsf{W}^{-1}\mathsf{x} - \mathbf{L}^*\mathbf{v}, -\mathbf{L}\mathsf{x} + \mathbf{U}^{-1}\mathbf{v}\big). \tag{5.15}$$

Further, set

$$\vartheta = \left(1 - \sqrt{\sum_{k=1}^{q} \|\mathsf{U}_k^{1/2}\mathsf{L}_k\mathsf{W}^{1/2}\|^2}\right) \min\{\mu^{-1}, \nu_1^{-1}, \ldots, \nu_q^{-1}\} \tag{5.16}$$

and

$$(\forall n \in \mathbb{N}) \quad \widetilde{\tau}_n = \|\mathbf{V}^{-1}\|\,\|\mathbf{V}\|\tau_n. \tag{5.17}$$

Since (e) imposes that $\sum_{n\in\mathbb{N}} \widetilde{\tau}_n < +\infty$, we assume without loss of generality that

$$\sup_{n\in\mathbb{N}} \widetilde{\tau}_n < 2\vartheta - 1. \tag{5.18}$$

In the renormed space $(\mathbf{K}, \|\cdot\|_{\mathbf{V}})$, $\mathbf{V}^{-1}\mathbf{A}$ is maximally monotone and $\mathbf{V}^{-1}\mathbf{B}$ is cocoercive [20, Lemma 3.7] with cocoercivity constant ϑ [43, Lemma 4.3]. In addition, finding a zero of the sum of these operators is equivalent to finding a point in $\mathsf{F} \times \mathbf{F}^*$, and algorithm (4.35) with $\gamma_n \equiv 1$ for solving this monotone inclusion problem specializes to (5.10) (see [20, 43] for details), which can thus be rewritten as

$$\begin{aligned}(\forall n \in \mathbb{N})\ \ (x_{n+1}, \boldsymbol{v}_{n+1}) &= (x_n, \boldsymbol{v}_n)\\ &+ \lambda_n\big(\mathsf{J}_{\mathbf{V}^{-1}\mathbf{A}_n}\big((x_n, \boldsymbol{v}_n) - \mathbf{V}^{-1}(u_n, \boldsymbol{s}_n)\big) + \boldsymbol{a}_n - (x_n, \boldsymbol{v}_n)\big),\end{aligned} \tag{5.19}$$

where

$$(\forall n \in \mathbb{N}) \quad \mathbf{a}_n = (b_n, \boldsymbol{c}_n) \tag{5.20}$$

and

$$(\forall n \in \mathbb{N}) \quad \mathbf{A}_n\colon \mathbf{K} \to 2^{\mathbf{K}}\colon (\mathsf{x}, \mathbf{v}) \mapsto \big(\partial \mathsf{f}_n(\mathsf{x}) + \mathbf{L}^*\mathbf{v}\big) \times \big(-\mathbf{L}\mathsf{x} + \partial \mathbf{g}^*(\mathbf{v})\big). \tag{5.21}$$

Then

$$\begin{aligned}(\forall n \in \mathbb{N})(\forall (\mathsf{x}, \mathbf{v}) \in \mathbf{K}) \quad \mathsf{J}_{\mathbf{V}^{-1}\mathbf{A}_n}(\mathsf{x}, \mathbf{v}) &= \Big(\mathsf{y}, \mathrm{prox}_{\mathbf{g}^*}^{\mathbf{U}^{-1}}\big(\mathbf{v} + \mathbf{U}\mathbf{L}(2\mathsf{y} - \mathsf{x})\big)\Big),\\ \text{where} \quad \mathsf{y} &= \mathrm{prox}_{\mathsf{f}_n}^{\mathsf{W}^{-1}}(\mathsf{x} - \mathsf{W}\mathbf{L}^*\mathbf{v}).\end{aligned} \tag{5.22}$$

Assumption (b) is equivalent to $\sum_{n\in\mathbb{N}} \lambda_n \sqrt{\mathsf{E}(\|\boldsymbol{a}_n\|_{\mathbf{V}}^2 \,|\, \mathfrak{X}_n)} < +\infty$, and assumptions (c) and (d) imply that

$$\sum_{n\in\mathbb{N}} \sqrt{\lambda_n}\|\mathsf{E}(\mathbf{V}^{-1}(u_n, \boldsymbol{s}_n) \,|\, \mathfrak{X}_n) - \mathbf{V}^{-1}\mathbf{B}(u_n, \boldsymbol{s}_n)\|_{\mathbf{V}} < +\infty. \tag{5.23}$$

For every $(\mathsf{x},\mathbf{v}) \in \mathsf{F} \times \mathbf{F}^*$, assumption (e) yields

$$
\begin{aligned}
(\forall n \in \mathbb{N})\quad & \mathsf{E}(\|\mathbf{V}^{-1}(u_n,\boldsymbol{s}_n) - \mathsf{E}(\mathbf{V}^{-1}(u_n,\boldsymbol{s}_n)\,|\,\mathfrak{X}_n)\|^2_{\mathbf{V}}\,|\,\mathfrak{X}_n) \\
& \leqslant \|\mathbf{V}^{-1}\|\big(\mathsf{E}(\|u_n - \mathsf{E}(u_n\,|\,\mathfrak{X}_n)\|^2\,|\,\mathfrak{X}_n) + \mathsf{E}(\|\boldsymbol{s}_n - \mathsf{E}(\boldsymbol{s}_n\,|\,\mathfrak{X}_n)\|^2\,|\,\mathfrak{X}_n)\big) \\
& \leqslant \|\mathbf{V}^{-1}\|\big(\tau_n\big(\|\nabla\mathsf{h}(x_n) - \nabla\mathsf{h}(\mathsf{x})\|^2 + \|\nabla\mathbf{j}^*(\boldsymbol{v}_n) - \nabla\mathbf{j}^*(\mathbf{v})\|^2\big) \\
& \quad + \zeta_n(\mathsf{x},\mathbf{v})\big) \\
& \leqslant \widetilde{\tau}_n\|\mathbf{V}^{-1}\mathbf{B}(x_n,\boldsymbol{v}_n) - \mathbf{V}^{-1}\mathbf{B}(\mathsf{x},\mathbf{v})\|^2_{\mathbf{V}} + \widetilde{\zeta}_n(\mathsf{x},\mathbf{v}),
\end{aligned}
\tag{5.24}
$$

where

$$
(\forall n \in \mathbb{N})\quad \widetilde{\zeta}_n(\mathsf{x},\mathbf{v}) = \|\mathbf{V}^{-1}\|\,\zeta_n(\mathsf{x},\mathbf{v}). \tag{5.25}
$$

According to assumption (e), $\big(\widetilde{\zeta}_n(\mathsf{x},\mathbf{v})\big)_{n\in\mathbb{N}} \in \ell^\infty_+(\mathscr{X})$, and $\big(\lambda_n\widetilde{\zeta}_n(\mathsf{x},\mathbf{v})\big)_{n\in\mathbb{N}} \in \ell^{1/2}_+(\mathscr{X})$. Now, let $n \in \mathbb{N}$, let $(\mathsf{x},\mathbf{v}) \in \mathbf{K}$, and set $\widetilde{\mathsf{y}} = \mathrm{prox}^{\mathsf{W}^{-1}}_{\mathsf{f}}(\mathsf{x} - \mathsf{W}\mathbf{L}^*\mathbf{v})$. By (5.22) and the nonexpansiveness of $\mathrm{prox}^{\mathbf{U}^{-1}}_{\mathbf{g}^*}$ in $(\mathbf{G}, \|\cdot\|_{\mathbf{U}^{-1}})$, we obtain

$$
\begin{aligned}
& \|\mathsf{J}_{\mathbf{V}^{-1}\mathbf{A}_n}(\mathsf{x},\mathbf{v}) - \mathsf{J}_{\mathbf{V}^{-1}\mathbf{A}}(\mathsf{x},\mathbf{v})\|^2_{\mathbf{V}} \\
& \leqslant \|\mathbf{V}\|\big(\|\mathsf{y} - \widetilde{\mathsf{y}}\|^2 \\
& \quad + \big\|\mathrm{prox}^{\mathbf{U}^{-1}}_{\mathbf{g}^*}(\mathbf{v} + \mathbf{UL}(2\mathsf{y} - \mathsf{x})) - \mathrm{prox}^{\mathbf{U}^{-1}}_{\mathbf{g}^*}(\mathbf{v} + \mathbf{UL}(2\widetilde{\mathsf{y}} - \mathsf{x}))\big\|^2\big) \\
& \leqslant \|\mathbf{V}\|\big(\|\mathsf{y} - \widetilde{\mathsf{y}}\|^2 + 4\|\mathbf{UL}(\mathsf{y} - \widetilde{\mathsf{y}})\|^2_{\mathbf{U}^{-1}}\big) \\
& \leqslant \|\mathbf{V}\|(1 + 4\|\mathbf{U}\|\|\mathbf{L}\|^2)\|\mathsf{y} - \widetilde{\mathsf{y}}\|^2.
\end{aligned}
\tag{5.26}
$$

It follows from (f) that

$$
\begin{aligned}
& \|\mathsf{J}_{\mathbf{V}^{-1}\mathbf{A}_n}(\mathsf{x},\mathbf{v}) - \mathsf{J}_{\mathbf{V}^{-1}\mathbf{A}}(\mathsf{x},\mathbf{v})\|_{\mathbf{V}} \\
& \leqslant \|\mathbf{V}\|^{1/2}\|(1 + 2\|\mathbf{U}\|^{1/2}\|\mathbf{L}\|) \\
& \quad \times \|\mathrm{prox}^{\mathsf{W}^{-1}}_{\mathsf{f}_n}(\mathsf{x} - \mathsf{W}\mathbf{L}^*\mathbf{v}) - \mathrm{prox}^{\mathsf{W}^{-1}}_{\mathsf{f}}(\mathsf{x} - \mathsf{W}\mathbf{L}^*\mathbf{v})\| \\
& \leqslant \|\mathbf{V}\|^{1/2}\|(1 + 2\|\mathbf{U}\|^{1/2}\|\mathbf{L}\|)(\alpha_n\|\mathsf{x} - \mathsf{W}\mathbf{L}^*\mathbf{v}\| + \beta_n) \\
& \leqslant \|\mathbf{V}\|^{1/2}\|(1 + 2\|\mathbf{U}\|^{1/2}\|\mathbf{L}\|)\big(\alpha_n(\|\mathsf{x}\| + \|\mathsf{W}\mathbf{L}^*\|\|\mathbf{v}\|) + \beta_n\big) \\
& \leqslant \widetilde{\alpha}_n\|(\mathsf{x},\mathbf{v})\|_{\mathbf{V}} + \widetilde{\beta}_n,
\end{aligned}
\tag{5.27}
$$

where

$$
\begin{cases}
\widetilde{\alpha}_n = \sqrt{2}\|\mathbf{V}\|^{1/2}\|(1 + 2\|\mathbf{U}\|^{1/2}\|\mathbf{L}\|)\max\{1, \|\mathsf{W}\mathbf{L}^*\|\}\|\mathbf{V}^{-1}\|^{1/2}\alpha_n \\
\widetilde{\beta}_n = \|\mathbf{V}\|^{1/2}\|(1 + 2\|\mathbf{U}\|^{1/2}\|\mathbf{L}\|)\beta_n.
\end{cases}
\tag{5.28}
$$

Thus, $\sum_{n\in\mathbb{N}}\sqrt{\lambda_n}\widetilde{\alpha}_n < +\infty$ and $\sum_{n\in\mathbb{N}}\lambda_n\widetilde{\beta}_n < +\infty$. Finally, since $\gamma_n \equiv 1$, (5.18) implies that $\sup_{n\in\mathbb{N}}(1 + \widetilde{\tau}_n)\gamma_n < 2\vartheta$. All the assumptions of Proposition 4.4 are therefore satisfied for algorithm (5.19). ☐

Remark 5.4.

(i) Algorithm 5.10 can be viewed as a stochastic version of the primal-dual algorithm investigated in [20, Example 6.4] when the metric is fixed in the latter. Particular cases of such fixed metric primal-algorithm can be found in [12, 15, 30, 34, 35].

(ii) The same type of primal-dual algorithm is investigated in [5, 43] in a different context since in those papers the stochastic nature of the algorithms stems from the random activation of blocks of variables.

5.2. **Example.** We illustrate an implementation of Algorithm 5.2 in a simple scenario with $\mathsf{H} = \mathbb{R}^N$ by constructing an example in which the gradient of h is available only through the observation of stochastic data and the approximation conditions are fulfilled.

For every $k \in \{1, \ldots, q\}$ and every $n \in \mathbb{N}$, set $s_{k,n} = \nabla \mathsf{j}_k^*(v_{k,n})$ and suppose that $(y_n)_{n\in\mathbb{N}}$ is almost surely bounded. This assumption is satisfied, in particular, if $\operatorname{dom} \mathsf{f}$ and $(b_n)_{n\in\mathbb{N}}$ are bounded. In addition, let

$$(5.29)\qquad (\forall n \in \mathbb{N}) \quad \boldsymbol{\mathcal{X}}_n = \sigma\big(x_0, \boldsymbol{v}_0, (K_{n'}, z_{n'})_{0\leqslant n' < m_n}, (b_{n'}, \boldsymbol{c}_{n'})_{1\leqslant n'<n}\big),$$

where $(m_n)_{n\in\mathbb{N}}$ is a strictly increasing sequence in $\mathbb{N}$ such that $m_n = O(n^{1+\delta})$ with $\delta \in \left]0, +\infty\right[$, $(K_n)_{n\in\mathbb{N}}$ is a sequence of independent and identically distributed (i.i.d.) random matrices of $\mathbb{R}^{M\times N}$, and $(z_n)_{n\in\mathbb{N}}$ is a sequence of i.i.d. random vectors of $\mathbb{R}^M$. For example, in signal recovery, $(K_n)_{n\in\mathbb{N}}$ may model a stochastic degradation operator [19], while $(z_n)_{n\in\mathbb{N}}$ are observations related to an unknown signal that we want to estimate. The variables $(K_n, z_n)_{n\in\mathbb{N}}$ are supposed to be i.i.d., independent of $(b_n, \boldsymbol{c}_n)_{n\in\mathbb{N}}$, and such that $\mathsf{E}\|K_0\|^4 < +\infty$ and $\mathsf{E}\|z_0\|^4 < +\infty$. Set

$$(5.30)\qquad (\forall \mathsf{x} \in \mathsf{H}) \quad \mathsf{h}(\mathsf{x}) = \frac{1}{2}\mathsf{E}\|K_0\mathsf{x} - z_0\|^2$$

and, for every $n \in \mathbb{N}$, let

$$(5.31)\qquad u_n = \frac{1}{m_{n+1}} \sum_{n'=0}^{m_{n+1}-1} K_{n'}^\top (K_{n'}x_n - z_{n'})$$

be an empirical estimate of $\nabla\mathsf{h}(x_n)$. We assume that $\lambda_n = O(n^{-\kappa})$, where $\kappa \in \left]1-\delta, 1\right] \cap [0,1]$. We have

$$(5.32)\qquad (\forall n \in \mathbb{N}) \quad \mathsf{E}(u_n \,|\, \boldsymbol{\mathcal{X}}_n) - \nabla\mathsf{h}(x_n) = \frac{1}{m_{n+1}}\big(Q_{0,m_n}x_n - r_{0,m_n}\big)$$

where, for every $(n_1, n_2) \in \mathbb{N}^2$ such that $n_1 < n_2$,

(5.33)
$$Q_{n_1,n_2} = \sum_{n'=n_1}^{n_2-1} \big(K_{n'}^\top K_{n'} - \mathsf{E}(K_0^\top K_0)\big) \text{ and } r_{n_1,n_2} = \sum_{n'=n_1}^{n_2-1} \big(K_{n'}^\top z_{n'} - \mathsf{E}(K_0^\top z_0)\big).$$

From the law of iterated logarithm [24, Section 25.8], we have almost surely

(5.34)
$$\varlimsup_{n\to+\infty} \frac{\|Q_{0,m_n}\|}{\sqrt{m_n\log(\log(m_n))}} < +\infty \quad \text{and} \quad \varlimsup_{n\to+\infty} \frac{\|r_{0,m_n}\|}{\sqrt{m_n\log(\log(m_n))}} < +\infty.$$

Since $(y_n)_{n\in\mathbb{N}}$ is assumed to be bounded, there exists a $[0,+\infty[$-valued random variable η such that, for every $n \in \mathbb{N}$, $\sup_{n\in\mathbb{N}} \|y_n\| \leqslant \eta$. Therefore,

$$(\forall n \in \mathbb{N}) \quad \|x_n\| \leqslant \|x_0\| + \eta. \tag{5.35}$$

Altogether, (5.32)–(5.35) yield

$$\lambda_n \|\mathsf{E}(u_n \,|\, \mathfrak{X}_n) - \nabla \mathsf{h}(x_n)\|^2 = O\Big(\frac{\lambda_n m_n \log(\log(m_n))}{m_{n+1}^2}\Big) = O\Big(\frac{\log(\log(n))}{n^{1+\delta+\kappa}}\Big). \tag{5.36}$$

Consequently, assumption (c) in Proposition 5.3 holds. In addition, for every $n \in \mathbb{N}$,

$$u_n - \mathsf{E}(u_n \,|\, \mathfrak{X}_n) = \frac{1}{m_{n+1}}\big(Q_{m_n,m_{n+1}} x_n - r_{m_n,m_{n+1}}\big) \tag{5.37}$$

which, by the triangle inequality, implies that

$$\begin{aligned}
&\mathsf{E}(\|u_n - \mathsf{E}(u_n \,|\, \mathfrak{X}_n)\|^2 \,|\, \mathfrak{X}_n)\\
&\qquad \leqslant \frac{1}{m_{n+1}^2}\mathsf{E}((\|Q_{m_n,m_{n+1}}\|\,\|x_n\| + \|r_{m_n,m_{n+1}}\|)^2 \,|\, \mathfrak{X}_n)\\
&\qquad \leqslant \frac{2}{m_{n+1}^2}\big(\mathsf{E}\|Q_{m_n,m_{n+1}}\|^2\,\|x_n\|^2 + \mathsf{E}\|r_{m_n,m_{n+1}}\|^2\big).
\end{aligned} \tag{5.38}$$

Upon invoking the i.i.d. assumptions, we obtain

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \mathsf{E}\|Q_{m_n,m_{n+1}}\|^2 = (m_{n+1} - m_n)\mathsf{E}\|K_0^\top K_0 - \mathsf{E}(K_0^\top K_0)\|^2\\ \mathsf{E}\|r_{m_n,m_{n+1}}\|^2 = (m_{n+1} - m_n)\mathsf{E}\|K_0^\top z_0 - \mathsf{E}(K_0^\top z_0)\|^2 \end{cases} \tag{5.39}$$

and it therefore follows from (5.35) that

$$\zeta_n = \mathsf{E}(\|u_n - \mathsf{E}(u_n \,|\, \mathfrak{X}_n)\|^2 \,|\, \mathfrak{X}_n) = O\Big(\frac{m_{n+1} - m_n}{m_{n+1}^2}\Big) = O\Big(\frac{1}{n^{2+\delta}}\Big) \tag{5.40}$$

and

$$\lambda_n \zeta_n = O\Big(\frac{1}{n^{2+\delta+\kappa}}\Big). \tag{5.41}$$

Thus, assumption (e) in Proposition 5.3 holds with $\tau_n \equiv 0$.

References

[1] Y. F. Atchadé, G. Fort, and E. Moulines, *On stochastic proximal gradient algorithms,* (2014). http://arxiv.org/abs/1402.2365

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

[3] F. Bach and E. Moulines, *Non-asymptotic analysis of stochastic approximation algorithms for machine learning,* in Proc. Adv. Neur. Inform. Process. Syst., Granada, Spain (2011), 451–459.

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

[5] P. Bianchi, W. Hachem, and F. Iutzeler, *A stochastic coordinate descent primal-dual algorithm and applications to large-scale composite optimization,* (2014). http://arxiv.org/abs/1407.0898

[6] R. I. Boţ and E. R. Csetnek, *On the convergence rate of a forward-backward type primal-dual splitting algorithm for convex optimization problems,* Optimization **64** (2015), 5–23.

[7] L. M. Briceño-Arias and P. L. Combettes, *Monotone operator methods for Nash equilibria in non-potential games*, in Computational and Analytical Mathematics, (D. Bailey *et. al.* eds.) (2013), 143–159. Springer, New York.

[8] L. M. Briceño-Arias, P. L. Combettes, J.-C. Pesquet, and N. Pustelnik, *Proximal algorithms for multicomponent image recovery problems*, J. Math. Imaging Vision **41** (2011), 3–22.

[9] C. L. Byrne, *Iterative Optimization in Inverse Problems.* CRC Press, Boca Raton, FL, 2014.

[10] L. Chaâri, J.-C. Pesquet, A. Benazza-Benyahia, and P. Ciuciu, *A wavelet-based regularized reconstruction algorithm for SENSE parallel MRI with applications to neuroimaging*, Med. Image Anal., **15** (2011), 185–201.

[11] A. Chambolle and C. Dossal, *On the convergence of the iterates of the "fast iterative shrinkage/thresholding algorithm,"* J. Optim. Theory Appl. **166** (2015), 968–982.

[12] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision **40** (2011), 120–145.

[13] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs, *A variational formulation for frame-based inverse problems*, Inverse Problems **23** (2007), 1495–1518.

[14] P. L. Combettes, *Solving monotone inclusions via compositions of nonexpansive averaged operators*, Optimization **53** (2004), 475–504.

[15] 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*, Proc. IEEE Int. Conf. Image Process. (2014), 4141–4145. Paris, France.

[16] P. L. Combettes, Đinh Dũng, and B. C. Vũ, *Dualization of signal recovery problems*, Set-Valued Anal. **18** (2010), 373–404.

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

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

[19] P. L. Combettes and H. J. Trussell, *Methods for digital restoration of signals degraded by a stochastic impulse response*, IEEE Trans. Acoustics, Speech, Signal Process. **37** (1989), 393–401.

[20] P. L. Combettes and B. C. Vũ, *Variable metric forward-backward splitting with applications to monotone inclusions in duality*, Optimization **63** (2014), 1289–1318.

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

[22] P. L. Combettes and I. Yamada, *Compositions and convex combinations of averaged nonexpansive operators*, J. Math. Anal. Appl. **425** (2015), 55–70.

[23] L. Condat, *A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms,* J. Optim. Theory Appl. **158** (2013), 460–479.

[24] J. Davidson, *Stochastic Limit Theory.* Oxford University Press, New York, 1994.

[25] E. De Vito, V. Umanità, and S. Villa, *A consistent algorithm to solve Lasso, elastic-net and Tikhonov regularization,* J. Complexity **27** (2011), 188–200.

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

[27] Yu. M. Ermol'ev, *On the method of generalized stochastic gradients and quasi-Fejér sequences*, Cybernetics **5** (1969), 208–220.

[28] Yu. M. Ermoliev and Z. V. Nekrylova, *Some methods of stochastic optimization*, Kibernetika (Kiev) **1966** (1966), 96–98.

[29] Yu. M. Ermoliev and Z. V. Nekrylova, *The method of stochastic gradients and its application*, in Seminar: Theory of Optimal Solutions, no. 1, Akad. Nauk Ukrain. SSR, Kiev, (1967), 24–47.

[30] E. Esser, X. Zhang, and T. Chan, *A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science*, SIAM J. Imaging Sci. **3** (2010), 1015–1046.

[31] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems.* Springer, New York, 2003.

[32] R. M. Fortet, *Vecteurs, Fonctions et Distributions Aléatoires dans les Espaces de Hilbert.* Hermès, Paris, 1995.

[33] O. V. Guseva, *The rate of convergence of the method of generalized stochastic gradients*, Kibernetika (Kiev) **1971** (1971), 143–145.

[34] B. He and X. Yuan, *Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective*, SIAM J. Imaging Sci. **5** (2012), 119–149.

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

[36] H. J. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, 2nd ed. Springer, New York, 2003.

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

[38] B. Lemaire, *Stability of the iteration method for nonexpansive mappings*, Serdica Math. J. **22** (1996), 331–340.

[39] B. Lemaire, *Which fixed point does the iteration method select?* Lecture Notes in Economics and Mathematical Systems, **452** (1997), 154–167. Springer, New York.

[40] B. Mercier, *Topics in Finite Element Solution of Elliptic Problems* (Lectures on Mathematics, no. 63). Tata Institute of Fundamental Research, Bombay, 1979.

[41] B. Mercier, *Inéquations Variationnelles de la Mécanique* (Publications Mathématiques d'Orsay, no. 80.01). Université de Paris-XI, Orsay, France, 1980.

[42] Z. V. Nekrylova, *Solution of certain variational problems and control problems by the stochastic gradient method*, Kibernetika (Kiev) **1974** (1974), 62–66; translated in Cybernetics **10** (1976), 622–626.

[43] J.-C. Pesquet and A. Repetti, *A class of randomized primal-dual algorithms for distributed optimization*, J. Nonlinear Convex Anal. **16** (2015), 2453–2490.

[44] H. Robbins and S. Monro, *A stochastic approximation method*, Ann. Math. Statist. **22** (1951), 400–407.

[45] L. Rosasco, S. Villa, and B. C. Vũ, *Convergence of stochastic proximal gradient algorithm*, (2014). `http://arxiv.org/abs/1403.5074`

[46] L. Rosasco, S. Villa, and B. C. Vũ, *A stochastic forward-backward splitting method for solving monotone inclusions in Hilbert spaces*, (2014). `http://arxiv.org/abs/1403.7999`

[47] L. Rosasco, S. Villa, and B. C. Vũ, *A stochastic inertial forward-backward splitting algorithm for multivariate monotone inclusions*, (2015). `http://arxiv.org/abs/1507.00848`

[48] S. Shalev-Shwartz and T. Zhang, *Stochastic dual coordinate ascent methods for regularized loss minimization*, J. Mach. Learn. Res. **14** (2013), 567–599.

[49] N. Z. Shor, *Minimization Methods for Non-Differentiable Functions.* Springer, New York, 1985.

[50] C. Talischi and G. H. Paulino, *A closer look at consistent operator splitting and its extensions for topology optimization*, Comput. Methods Appl. Mech. Engrg. **283** (2015), 573–598.

[51] P. Tseng, *Further applications of a splitting algorithm to decomposition in variational inequalities and convex programming*, Math. Programming **48** (1990), 249–263.

[52] P. Tseng, *Applications of a splitting algorithm to decomposition in convex programming and variational inequalities*, SIAM J. Control Optim. **29** (1991), 119–138.
[53] B. C. Vũ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators*, Adv. Comput. Math. **38** (2013), 667–681.
[54] B. Widrow and S. D. Stearns, *Adaptive Signal Processing.* Prentice-Hall, Englewood Cliffs, NJ, 1985.
[55] L. Xiao and T. Zhang, *A proximal stochastic gradient method with progressive variance reduction*, SIAM J. Optim. **24** (2014), 2057–2075.

(P. L. Combettes) Sorbonne Universités – UPMC Univ. Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, F-75005 Paris, France
E-mail address: plc@ljll.math.upmc.fr

(J.-C. Pesquet) Université Paris-Est, Laboratoire d'Informatique Gaspard Monge – CNRS UMR 8049, F-77454, Marne la Vallée Cedex 2, France
E-mail address: jean-christophe.pesquet@univ-paris-est.fr