Math. Program., Ser. B (2019) 174:433–451
https://doi.org/10.1007/s10107-018-1296-y

FULL LENGTH PAPER

Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping II: mean-square and linear convergence

Patrick L. Combettes[1] · **Jean-Christophe Pesquet**[2]

Received: 23 April 2017 / Accepted: 11 May 2018 / Published online: 26 May 2018

Abstract Combettes and Pesquet (SIAM J Optim 25:1221–1248, 2015) investigated the almost sure weak convergence of block-coordinate fixed point algorithms and discussed their applications to nonlinear analysis and optimization. This algorithmic framework features random sweeping rules to select arbitrarily the blocks of variables that are activated over the course of the iterations and it allows for stochastic errors in the evaluation of the operators. The present paper establishes results on the mean-square and linear convergence of the iterates. Applications to monotone operator splitting and proximal optimization algorithms are presented.

Keywords Block-coordinate algorithm · Fixed-point algorithm · Mean-square convergence · Monotone operator splitting · Linear convergence · Stochastic algorithm

Mathematics Subject Classification 47J25 · 46M10 · 65K10 · 90C15

The work of P. L. Combettes was partially supported by the National Science Foundation under Grant CCF-1715671.

✉ Patrick L. Combettes
plc@math.ncsu.edu

Jean-Christophe Pesquet
jean-christophe@pesquet.eu

1 Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205, USA

2 CentraleSupélec, Université Paris-Saclay, Center for Visual Computing, 92295 Châtenay-Malabry, France

1 Introduction

In [11], we investigated the asymptotic behavior of abstract stochastic quasi-Fejér fixed point iterations in a Hilbert space H and applied these results to establish almost sure convergence properties for randomly activated block-coordinate, stochastically perturbed extensions of algorithms employed in fixed point theory, monotone operator splitting, and optimization. The basic property of the operators used in the underlying model was that of quasinonexpansiveness. Recall that an operator $\mathsf{T}\colon\mathsf{H}\to\mathsf{H}$ with fixed point set $\operatorname{Fix}\mathsf{T}$ is quasinonexpansive if

$$(\forall \mathsf{z}\in\operatorname{Fix}\mathsf{T})(\forall \mathsf{x}\in\mathsf{H})\quad \|\mathsf{T}\mathsf{x}-\mathsf{z}\|\leqslant\|\mathsf{x}-\mathsf{z}\|, \tag{1.1}$$

and strictly quasinonexpansive if the above inequality is strict whenever $\mathsf{x}\notin\operatorname{Fix}\mathsf{T}$ [6]. The fixed point problem under investigation in [11] was the following.

Problem 1.1 Let $(\mathsf{H}_i)_{1\leqslant i\leqslant m}$ be separable real Hilbert spaces and let $\mathsf{H}=\mathsf{H}_1\oplus\cdots\oplus\mathsf{H}_m$ be their direct Hilbert sum. For every $n\in\mathbb{N}$, let $\mathsf{T}_n\colon\mathsf{H}\to\mathsf{H}\colon\mathsf{x}\mapsto(\mathsf{T}_{i,n}\,\mathsf{x})_{1\leqslant i\leqslant m}$ be a quasinonexpansive operator where, for every $i\in\{1,\ldots,m\}$, $\mathsf{T}_{i,n}\colon\mathsf{H}\to\mathsf{H}_i$ is measurable. Suppose that $\mathsf{F}=\bigcap_{n\in\mathbb{N}}\operatorname{Fix}\mathsf{T}_n\neq\varnothing$. The problem is to find a point in F.

In [11], Problem 1.1 was solved via the following block-coordinate algorithm. The main advantages of a block-coordinate strategy is to reduce the computational load and the memory requirements per iteration. In addition, our approach adopts random sweeping rules to select arbitrarily the blocks of variables that are activated at each iteration, and it allows for stochastic errors in the implementation of the operators.

Algorithm 1.2 Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,1]$ and set $\mathsf{D}=\{0,1\}^m\smallsetminus\{\mathbf{0}\}$. Let $\boldsymbol{x}_0$ and $(\boldsymbol{a}_n)_{n\in\mathbb{N}}$ be H-valued random variables, and let $(\boldsymbol{\varepsilon}_n)_{n\in\mathbb{N}}$ be identically distributed D-valued random variables. Iterate

$$\begin{array}{l}\text{for } n=0,1,\ldots\\ \left\lfloor\begin{array}{l}\text{for } i=1,\ldots,m\\ \left\lfloor\; x_{i,n+1}=x_{i,n}+\varepsilon_{i,n}\lambda_n\big(\mathsf{T}_{i,n}\,(x_{1,n},\ldots,x_{m,n})+a_{i,n}-x_{i,n}\big).\right.\end{array}\right.\end{array} \tag{1.2}$$

At iteration n of Algorithm 1.2, $\lambda_n\in\,]0,1]$ is a relaxation parameter, $a_{i,n}$ an H_i-valued random variable modeling some stochastic error in the application of the operator $\mathsf{T}_{i,n}$, and $\varepsilon_{i,n}$ an $\{0,1\}$-valued random variable that signals the activation of the ith block $\mathsf{T}_{i,n}$ of the operator T_n. Almost sure weak and strong convergence properties of this scheme were established in [11]. In the present paper, we complement these results by proving mean-square and linear convergence properties for the orbits of (1.2) under the additional assumption that each operator T_n in Problem 1.1 satisfies the property

$$(\exists\,\tau_n\in[0,1[)(\forall \mathsf{z}\in\operatorname{Fix}\mathsf{T}_n)(\forall \mathsf{x}\in\mathsf{H})\quad \|\mathsf{T}_n\mathsf{x}-\mathsf{z}\|\leqslant\sqrt{\tau_n}\|\mathsf{x}-\mathsf{z}\|, \tag{1.3}$$

which implies that T_n is strictly quasinonexpansive and that $\operatorname{Fix}\mathsf{T}_n$ is a singleton. Our results appear to be the first of this kind regarding the block-coordinate algorithm

(1.2), even in the case of a single-block, when it reduces to the stochastically perturbed iteration

$$\begin{array}{l}\text{for } n = 0, 1, \ldots \\ \lfloor x_{n+1} = x_n + \lambda_n \big(\mathsf{T}_n x_n + a_n - x_n \big),\end{array} \tag{1.4}$$

special cases of which are studied in [2,12,24].

The problem we address is more precisely described as follows.

Problem 1.3 Let $(\mathsf{H}_i)_{1 \leqslant i \leqslant m}$ be separable real Hilbert spaces, set $\mathbf{H} = \mathsf{H}_1 \oplus \cdots \oplus \mathsf{H}_m$, and let $\{\tau_{i,n}\}_{1 \leqslant i \leqslant m} \subset [0, 1[$. For every $n \in \mathbb{N}$, let $\mathbf{T}_n : \mathbf{H} \to \mathbf{H} : \mathbf{x} \mapsto (\mathsf{T}_{i,n}\, \mathbf{x})_{1 \leqslant i \leqslant m}$ be measurable and quasinonexpansive with common fixed point $\overline{\mathbf{x}} = (\overline{\mathsf{x}}_i)_{1 \leqslant i \leqslant m}$, and such that

$$(\forall n \in \mathbb{N})(\forall \mathbf{x} \in \mathbf{H}) \quad \|\mathbf{T}_n \mathbf{x} - \overline{\mathbf{x}}\|^2 \leqslant \sum_{i=1}^{m} \tau_{i,n} \|\mathsf{x}_i - \overline{\mathsf{x}}_i\|^2. \tag{1.5}$$

The problem is to find $\overline{\mathbf{x}}$.

The proposed mean-square convergence results are the most comprehensive available to date for stochastic block-iterative fixed point methods at the level of generality and flexibility of Algorithm 1.2. Special cases concerning finite-dimensional minimization problems involving a smooth function with restrictions in the implementation of (1.2) are discussed in [18,20,21].

The remainder of the paper consists of three sections. In Sect. 2, we provide our notation and preliminary results. Section 3 is dedicated to the mean-square convergence analysis of Algorithm 1.2 and it discusses its linear convergence properties. Applications are presented in Sect. 4.

2 Notation, background, and preliminary results

Notation H is a separable real Hilbert space with scalar product $\langle \cdot \mid \cdot \rangle$, associated norm $\|\cdot\|$, Borel σ-algebra $\mathcal{B}$, and identity operator Id. The underlying probability space is $(\Omega, \mathcal{F}, \mathsf{P})$. A H-valued random variable is a measurable map $x : (\Omega, \mathcal{F}) \to (\mathsf{H}, \mathcal{B})$ [14,15]. The σ-algebra generated by a family Φ of random variables is denoted by $\sigma(\Phi)$. Let $\mathscr{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ be a sequence of sub-sigma algebras of $\mathcal{F}$ such that $(\forall n \in \mathbb{N})\ \mathcal{F}_n \subset \mathcal{F}_{n+1}$. We denote by $\ell_+(\mathscr{F})$ 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 $\mathcal{F}_n$-measurable. We set

$$(\forall p \in]0, +\infty[) \quad \ell_+^p(\mathscr{F}) = \left\{ (\xi_n)_{n\in\mathbb{N}} \in \ell_+(\mathscr{F}) \;\middle|\; \sum_{n\in\mathbb{N}} \xi_n^p < +\infty \quad \mathsf{P}\text{-a.s.} \right\}. \tag{2.1}$$

Lemma 2.1 *Let $\mathscr{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ be a sequence of sub-sigma algebras of $\mathcal{F}$ such that $(\forall n \in \mathbb{N})\ \mathcal{F}_n \subset \mathcal{F}_{n+1}$. Let $(\alpha_n)_{n\in\mathbb{N}} \in \ell_+(\mathscr{F})$, let $(\vartheta_n)_{n\in\mathbb{N}} \in \ell_+(\mathscr{F})$, let*

$(\eta_n)_{n\in\mathbb{N}} \in \ell_+(\mathscr{F})$, and suppose that there exists a sequence $(\chi_n)_{n\in\mathbb{N}}$ in $[0,+\infty[$ such that $\overline{\lim}\,\chi_n < 1$ and

$$(\forall n \in \mathbb{N})\quad \mathsf{E}\left(\alpha_{n+1} \,\middle|\, \mathcal{F}_n\right) + \vartheta_n \leqslant \chi_n\alpha_n + \eta_n \quad \mathsf{P}\text{-a.s.} \tag{2.2}$$

Then the following hold:

(i) *Set $(\forall n\in\mathbb{N})\ \overline{\vartheta}_n = \sum_{k=0}^{n}\left(\prod_{\ell=k+1}^{n}\chi_\ell\right)\mathsf{E}\left(\vartheta_k \,\middle|\, \mathcal{F}_0\right)$ and $\overline{\eta}_n = \sum_{k=0}^{n}\left(\prod_{\ell=k+1}^{n}\chi_\ell\right)\mathsf{E}\left(\eta_k \,\middle|\, \mathcal{F}_0\right)$ (with the convention $\prod_{n+1}^{n}\cdot = 1$). Then*

$$(\forall n \in \mathbb{N})\quad \mathsf{E}\left(\alpha_{n+1} \,\middle|\, \mathcal{F}_0\right) + \overline{\vartheta}_n \leqslant \left(\prod_{k=0}^{n}\chi_k\right)\alpha_0 + \overline{\eta}_n \quad \mathsf{P}\text{-a.s.} \tag{2.3}$$

(ii) *Suppose that $\mathsf{E}\alpha_0 < +\infty$ and $\sum_{n\in\mathbb{N}}\mathsf{E}\eta_n < +\infty$. Then $\sum_{n\in\mathbb{N}}\mathsf{E}\alpha_n < +\infty$ and $\sum_{n\in\mathbb{N}}\mathsf{E}\vartheta_n < +\infty$.*

Proof (i) Let $n \in \mathbb{N}\backslash\{0\}$. We deduce from (2.2) that

$$\begin{aligned}\mathsf{E}\left(\mathsf{E}\left(\alpha_{n+1} \,\middle|\, \mathcal{F}_n\right) \,\middle|\, \mathcal{F}_{n-1}\right) + \mathsf{E}\left(\vartheta_n \,\middle|\, \mathcal{F}_{n-1}\right) &\leqslant \mathsf{E}\left(\chi_n\alpha_n \,\middle|\, \mathcal{F}_{n-1}\right) + \mathsf{E}\left(\eta_n \,\middle|\, \mathcal{F}_{n-1}\right)\\ &= \chi_n\mathsf{E}\left(\alpha_n \,\middle|\, \mathcal{F}_{n-1}\right) + \mathsf{E}\left(\eta_n \,\middle|\, \mathcal{F}_{n-1}\right) \quad \mathsf{P}\text{-a.s.}\end{aligned} \tag{2.4}$$

However, since $\mathcal{F}_{n-1} \subset \mathcal{F}_n$, we have $\mathsf{E}\left(\mathsf{E}\left(\alpha_{n+1} \,\middle|\, \mathcal{F}_n\right) \,\middle|\, \mathcal{F}_{n-1}\right) = \mathsf{E}\left(\alpha_{n+1} \,\middle|\, \mathcal{F}_{n-1}\right)$. Therefore (2.4) yields

$$\mathsf{E}\left(\alpha_{n+1} \,\middle|\, \mathcal{F}_{n-1}\right) \leqslant \chi_n\mathsf{E}\left(\alpha_n \,\middle|\, \mathcal{F}_{n-1}\right) + \mathsf{E}\left(\eta_n \,\middle|\, \mathcal{F}_{n-1}\right) - \mathsf{E}\left(\vartheta_n \,\middle|\, \mathcal{F}_{n-1}\right) \quad \mathsf{P}\text{-a.s.} \tag{2.5}$$

By proceeding by induction and observing that α_0 is $\mathcal{F}_0$-measurable, we obtain (2.3).

(ii) We derive from (2.3) that

$$\begin{aligned}(\forall n \in \mathbb{N})\quad \mathsf{E}\alpha_{n+1} + \mathsf{E}\overline{\vartheta}_n &\leqslant \left(\prod_{k=0}^{n}\chi_k\right)\mathsf{E}\alpha_0 + \mathsf{E}\overline{\eta}_n\\ &= \left(\prod_{k=0}^{n}\chi_k\right)\mathsf{E}\alpha_0 + \sum_{k=0}^{n}\left(\prod_{\ell=k+1}^{n}\chi_\ell\right)\mathsf{E}\eta_k.\end{aligned} \tag{2.6}$$

On the other hand, there exist $q \in \mathbb{N}$ and $\rho \in\,]0,1[$ such that, for every integer $n > q$, $\chi_n < \rho$ and, therefore,

$$\begin{aligned}\mathsf{E}\alpha_{n+1}+\mathsf{E}\overline{\vartheta}_n &\leqslant \left(\prod_{k=0}^{q}\chi_k\right)\rho^{n-q}\mathsf{E}\alpha_0+\sum_{k=0}^{q}\left(\prod_{\ell=k+1}^{q}\chi_\ell\right)\rho^{n-q}\mathsf{E}\eta_k+\sum_{k=q+1}^{n}\rho^{n-k}\mathsf{E}\eta_k\\ &\leqslant \left(\prod_{k=0}^{q}\chi_k\right)\rho^{n-q}\mathsf{E}\alpha_0+\max\left\{\left(\frac{\prod_{\ell=k+1}^{q}\chi_\ell}{\rho^{q-k}}\right)_{0\leqslant k\leqslant q},1\right\}\sum_{k=0}^{n}\rho^{n-k}\mathsf{E}\eta_k.\end{aligned}\tag{2.7}$$

Since $\sum_{n\in\mathbb{N}}\rho^n<+\infty$ and $\sum_{n\in\mathbb{N}}\mathsf{E}\eta_n<+\infty$, it follows from standard properties of the discrete convolution that $(\sum_{k=0}^{n}\rho^{n-k}\mathsf{E}\eta_k)_{n\in\mathbb{N}}$ is summable. We then deduce from (2.7) that $\sum_{n\in\mathbb{N}}\mathsf{E}\alpha_n<+\infty$ and $\sum_{n\in\mathbb{N}}\mathsf{E}\overline{\vartheta}_n<+\infty$. Thus, the inequalities

$$(\forall n\in\mathbb{N})\quad \mathsf{E}\vartheta_n\leqslant\sum_{k=0}^{n}\left(\prod_{\ell=k+1}^{n}\chi_\ell\right)\mathsf{E}\vartheta_k=\mathsf{E}\overline{\vartheta}_n\tag{2.8}$$

yield $\sum_{n\in\mathbb{N}}\mathsf{E}\vartheta_n<+\infty$. □

Lemma 2.2 *Let* $\phi:[0,+\infty[\,\to[0,+\infty[$ *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* $(\mathfrak{F}_n)_{n\in\mathbb{N}}$ *be a sequence of sub-sigma-algebras of* $\mathfrak{F}$ *such that*

$$(\forall n\in\mathbb{N})\quad \sigma(x_0,\ldots,x_n)\subset\mathfrak{F}_n\subset\mathfrak{F}_{n+1}.\tag{2.9}$$

Suppose that there exist $\mathsf{z}\in\mathsf{H}$, $(\vartheta_n)_{n\in\mathbb{N}}\in\ell_+(\mathscr{F})$, $(\eta_n)_{n\in\mathbb{N}}\in\ell_+(\mathscr{F})$, *and a sequence* $(\chi_n)_{n\in\mathbb{N}}$ *in* $[0,+\infty[$ *such that* $\overline{\lim}\,\chi_n<1$ *and*

$$(\forall n\in\mathbb{N})\quad \mathsf{E}\left(\phi(\|x_{n+1}-\mathsf{z}\|)\,\middle|\,\mathfrak{F}_n\right)+\vartheta_n\leqslant\chi_n\phi(\|x_n-\mathsf{z}\|)+\eta_n\quad \mathsf{P}\text{-a.s.}\tag{2.10}$$

Set $(\forall n\in\mathbb{N})$ $\overline{\vartheta}_n=\sum_{k=0}^{n}\left(\prod_{\ell=k+1}^{n}\chi_\ell\right)\mathsf{E}\left(\vartheta_k\,\middle|\,\mathfrak{F}_0\right)$ *and* $\overline{\eta}_n=\sum_{k=0}^{n}\left(\prod_{\ell=k+1}^{n}\chi_\ell\right)\mathsf{E}\left(\eta_k\,\middle|\,\mathfrak{F}_0\right)$. *Then the following hold:*

(i) $(\forall n\in\mathbb{N})$ $\mathsf{E}\left(\phi(\|x_{n+1}-\mathsf{z}\|)\,\middle|\,\mathfrak{F}_0\right)+\overline{\vartheta}_n\leqslant\left(\prod_{k=0}^{n}\chi_k\right)\phi(\|x_0-\mathsf{z}\|)+\overline{\eta}_n$ P-a.s.

(ii) *Let* $p\in\,]0,+\infty[$ *and set* $\phi=|\cdot|^p$. *Suppose that* $x_0\in L^p(\Omega,\mathfrak{F},\mathsf{P};\mathsf{H})$ *and that* $\sum_{n\in\mathbb{N}}\mathsf{E}\eta_n<+\infty$. *Then the following hold:*
   (a) $\mathsf{E}\|x_n-\mathsf{z}\|^p\to0$ *and* $\sum_{n\in\mathbb{N}}\mathsf{E}\vartheta_n<+\infty$.
   (b) *Suppose that* $(\eta_n)_{n\in\mathbb{N}}\in\ell_+^1(\mathscr{F})$. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges strongly* P*-a.s. to* z.

Proof We apply Lemma 2.1(i) with $(\forall n\in\mathbb{N})$ $\alpha_n=\phi(\|x_n-\mathsf{z}\|)$.

(i) See Lemma 2.1(i).

(iia) Since $L^p(\Omega,\mathfrak{F},\mathsf{P};\mathsf{H})$ is a vector space [25, Théorème 5.8.8 and Proposition 5.8.9] that contains x_0 and z, it also contains $x_0-\mathsf{z}$. Hence $\mathsf{E}\alpha_0=\mathsf{E}\|x_0-$

$\mathsf{z}\|^p < +\infty$, and it follows from Lemma 2.1(ii) that $\sum_{n\in\mathbb{N}} \mathsf{E}\|x_n - \mathsf{z}\|^p < +\infty$ and $\sum_{n\in\mathbb{N}} \mathsf{E}\vartheta_n < +\infty$. Consequently,

$$\mathsf{E}\|x_n - \mathsf{z}\|^p \to 0. \tag{2.11}$$

(iib) In view of (2.10), since $(\eta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathscr{F})$, it follows from [12, Proposition 3.1(iii)] that $(\|x_n - \mathsf{z}\|)_{n\in\mathbb{N}}$ converges P-a.s. However, we derive from (2.11) that there exists a strictly increasing sequence $(k_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$ such that $\|x_{k_n} - \mathsf{z}\| \to 0$ P-a.s. [25, Corollaire 5.8.11]. Altogether $\|x_n - \mathsf{z}\| \to 0$ P-a.s. □

Theorem 2.3 *Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,1]$ such that $\inf_{n\in\mathbb{N}}\lambda_n > 0$, and let $(t_n)_{n\in\mathbb{N}}$, $(x_n)_{n\in\mathbb{N}}$, and $(e_n)_{n\in\mathbb{N}}$ be sequences of H-valued random variables. Further, let $(\mathfrak{F}_n)_{n\in\mathbb{N}}$ be a sequence of sub-sigma-algebras of $\mathfrak{F}$ such that*

$$(\forall n \in \mathbb{N}) \quad \sigma(x_0,\ldots,x_n) \subset \mathfrak{F}_n \subset \mathfrak{F}_{n+1}. \tag{2.12}$$

Suppose that the following are satisfied:

[a] $(\forall n \in \mathbb{N})$ $x_{n+1} = x_n + \lambda_n(t_n + e_n - x_n)$.

[b] *There exists a sequence $(\xi_n)_{n\in\mathbb{N}}$ in $[0,+\infty[$ such that*

$$\sum_{n\in\mathbb{N}} \sqrt{\xi_n} < +\infty \tag{2.13}$$

and $(\forall n \in \mathbb{N})$ $\mathsf{E}\left(\|e_n\|^2 \,\middle|\, \mathfrak{F}_n\right) \leqslant \xi_n$.

[c] *There exist $\mathsf{z} \in \mathsf{H}$, $(\theta_n)_{n\in\mathbb{N}} \in \ell_+(\mathscr{F})$, $(\nu_n)_{n\in\mathbb{N}} \in \ell_+(\mathscr{F})$, and a sequence $(\mu_n)_{n\in\mathbb{N}}$ in $[0,+\infty[$ such that $\overline{\lim}\,\mu_n < 1$ and*

$$(\forall n \in \mathbb{N}) \quad \mathsf{E}\left(\|t_n - \mathsf{z}\|^2 \,\middle|\, \mathfrak{F}_n\right) + \theta_n \leqslant \mu_n\|x_n - \mathsf{z}\|^2 + \nu_n \quad \mathsf{P}\text{-a.s.} \tag{2.14}$$

Set

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \chi_n = 1 - \lambda_n + \lambda_n\mu_n + \sqrt{\xi_n}\lambda_n\left(1 - \lambda_n + \lambda_n\sqrt{\mu_n}\right) \\ \overline{\vartheta}_n = \displaystyle\sum_{k=0}^{n}\left[\prod_{\ell=k+1}^{n}\chi_\ell\right]\lambda_k\left(\mathsf{E}\left(\theta_k \,\middle|\, \mathfrak{F}_0\right) + (1-\lambda_k)\mathsf{E}\left(\|t_k - x_k\|^2 \,\middle|\, \mathfrak{F}_0\right)\right) \\ \overline{\eta}_n = \displaystyle\sum_{k=0}^{n}\left[\prod_{\ell=k+1}^{n}\chi_\ell\right]\lambda_k\left(\mathsf{E}\left(\nu_k \,\middle|\, \mathfrak{F}_0\right)\right. \\ \left. +\left(1 - \lambda_k + \lambda_k\left(2\mathsf{E}\left(\sqrt{\nu_k} \,\middle|\, \mathfrak{F}_0\right) + \sqrt{\mu_k}\right)\right)\sqrt{\xi_k} + \lambda_k\xi_k\right). \end{cases} \tag{2.15}$$

Then the following hold:

(i) $(\forall n \in \mathbb{N})$ $\mathsf{E}\left(\|x_{n+1} - \mathsf{z}\|^2 \,\middle|\, \mathfrak{F}_0\right) + \overline{\vartheta}_n \leqslant \left(\displaystyle\prod_{k=0}^{n}\chi_k\right)\|x_0 - \mathsf{z}\|^2 + \overline{\eta}_n$ P-a.s.

(ii) *Suppose that* $x_0 \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ *and that*

$$\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}\nu_n} < +\infty. \tag{2.16}$$

Then the following hold:

(a) $\mathsf{E}\|x_n - \mathsf{z}\|^2 \to 0$.
(b) $\sum_{n\in\mathbb{N}} \mathsf{E}\theta_n < +\infty$.
(c) $\sum_{n\in\mathbb{N}} (1-\lambda_n)\mathsf{E}\|t_n - x_n\|^2 < +\infty$.
(d) *Suppose that* $(\nu_n)_{n\in\mathbb{N}} \in \ell_+^{1/2}(\mathscr{F})$. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges strongly* P-a.s. *to* z.

Proof (i) Set $\lambda = \inf_{n\in\mathbb{N}} \lambda_n$. Then

$$(\forall n \in \mathbb{N}) \quad \chi_n \leqslant 1 - (1-\mu_n)\lambda + \sqrt{\xi_n}\big(1 + \sqrt{\mu_n}\big). \tag{2.17}$$

Since $\overline{\lim}\, \mu_n < 1$ and $\lim \xi_n = 0$, we have $\overline{\lim}\, \chi_n < 1$. In addition, we derive from [a], [6, Corollary 2.15], and (2.14) that

$$\begin{aligned}
&(\forall n \in \mathbb{N}) \quad \|x_{n+1} - \mathsf{z}\|^2 \\
&\quad = \|(1-\lambda_n)(x_n - \mathsf{z}) + \lambda_n(t_n - \mathsf{z})\|^2 \\
&\qquad + 2\lambda_n\big\langle (1-\lambda_n)(x_n-\mathsf{z}) + \lambda_n\big(t_n - \mathsf{z}\big) \mid e_n\big\rangle + \lambda_n^2\|e_n\|^2 \\
&\quad = (1-\lambda_n)\|x_n - \mathsf{z}\|^2 + \lambda_n\|t_n - \mathsf{z}\|^2 - \lambda_n(1-\lambda_n)\|t_n - x_n\|^2 \\
&\qquad + 2\lambda_n\big\langle (1-\lambda_n)(x_n-\mathsf{z}) + \lambda_n\big(t_n - \mathsf{z}\big) \mid e_n\big\rangle + \lambda_n^2\|e_n\|^2 \quad \mathsf{P}\text{-a.s.}
\end{aligned} \tag{2.18}$$

Hence, [c] implies that

$$\begin{aligned}
&(\forall n \in \mathbb{N}) \quad \mathsf{E}\Big(\|x_{n+1} - \mathsf{z}\|^2 \,\Big|\, \mathfrak{F}_n\Big) \\
&\leqslant (1-\lambda_n)\|x_n - \mathsf{z}\|^2 + \lambda_n\mathsf{E}\Big(\|t_n - \mathsf{z}\|^2 \,\Big|\, \mathfrak{F}_n\Big) - \lambda_n(1-\lambda_n)\mathsf{E}\Big(\|t_n - x_n\|^2 \,\Big|\, \mathfrak{F}_n\Big) \\
&\quad + 2\lambda_n\left((1-\lambda_n)\|x_n - \mathsf{z}\| + \lambda_n\sqrt{\mathsf{E}\Big(\|t_n - \mathsf{z}\|^2 \,\Big|\, \mathfrak{F}_n\Big)}\right)\sqrt{\mathsf{E}\Big(\|e_n\|^2 \,\Big|\, \mathfrak{F}_n\Big)} + \lambda_n^2\mathsf{E}\Big(\|e_n\|^2 \,\Big|\, \mathfrak{F}_n\Big) \\
&\leqslant (1-\lambda_n)\|x_n - \mathsf{z}\|^2 + \lambda_n\big(\mu_n\|x_n - \mathsf{z}\|^2 + \nu_n - \theta_n\big) \\
&\quad - \lambda_n(1-\lambda_n)\mathsf{E}\Big(\|t_n - x_n\|^2 \,\Big|\, \mathfrak{F}_n\Big) \\
&\quad + 2\lambda_n\big((1-\lambda_n)\|x_n - \mathsf{z}\| + \lambda_n\sqrt{\mu_n\|x_n - \mathsf{z}\|^2 + \nu_n}\big)\sqrt{\mathsf{E}\Big(\|e_n\|^2 \,\Big|\, \mathfrak{F}_n\Big)} \\
&\quad + \lambda_n^2\mathsf{E}\Big(\|e_n\|^2 \,\Big|\, \mathfrak{F}_n\Big) \quad \mathsf{P}\text{-a.s.}
\end{aligned} \tag{2.19}$$

Now set

$$(\forall n\in\mathbb{N})\quad \begin{cases} \vartheta_n = \lambda_n\theta_n + \lambda_n(1-\lambda_n)\mathsf{E}\left(\|t_n - x_n\|^2 \,\middle|\, \mathfrak{F}_n\right) \\ \kappa_n = \lambda_n\nu_n + 2\lambda_n^2\sqrt{\nu_n}\sqrt{\mathsf{E}\left(\|e_n\|^2 \,\middle|\, \mathfrak{F}_n\right)} + \lambda_n^2\mathsf{E}\left(\|e_n\|^2 \,\middle|\, \mathfrak{F}_n\right) \\ \eta_n = \lambda_n\nu_n + \lambda_n\big(1-\lambda_n+\lambda_n(2\sqrt{\nu_n}+\sqrt{\mu_n})\big)\sqrt{\xi_n} + \lambda_n^2\xi_n. \end{cases} \tag{2.20}$$

It follows from [b] that

$$\begin{aligned} &(\forall n\in\mathbb{N})\quad \mathsf{E}\left(\|x_{n+1}-\mathsf{z}\|^2 \,\middle|\, \mathfrak{F}_n\right) \\ &\quad\leqslant (1-\lambda_n+\lambda_n\mu_n)\|x_n-\mathsf{z}\|^2 + 2\lambda_n(1-\lambda_n+\lambda_n\sqrt{\mu_n})\|x_n-\mathsf{z}\|\sqrt{\mathsf{E}\left(\|e_n\|^2 \,\middle|\, \mathfrak{F}_n\right)} \\ &\qquad - \vartheta_n + \kappa_n \\ &\quad\leqslant (1-\lambda_n+\lambda_n\mu_n)\|x_n-\mathsf{z}\|^2 + \lambda_n(1-\lambda_n+\lambda_n\sqrt{\mu_n})(\|x_n-\mathsf{z}\|^2+1) \\ &\qquad \times\sqrt{\mathsf{E}\left(\|e_n\|^2 \,\middle|\, \mathfrak{F}_n\right)} - \vartheta_n + \kappa_n \\ &\quad\leqslant \chi_n\|x_n-\mathsf{z}\|^2 - \vartheta_n + \eta_n \quad \mathsf{P}\text{-a.s.} \end{aligned} \tag{2.21}$$

The result then follows by applying Lemma 2.2(i) with $\phi = |\cdot|^2$.

(iia) According to (2.20), for every $n\in\mathbb{N}$,

$$\begin{aligned} \mathsf{E}\eta_n &= \lambda_n\mathsf{E}\nu_n + \lambda_n\big(1-\lambda_n+\lambda_n(2\mathsf{E}\sqrt{\nu_n}+\sqrt{\mu_n})\big)\sqrt{\xi_n} + \lambda_n^2\xi_n \\ &= \lambda_n\mathsf{E}\nu_n + (1-\lambda_n+\sqrt{\mu_n})\lambda_n\sqrt{\xi_n} + 2\lambda_n^2\sqrt{\xi_n}\,\mathsf{E}\sqrt{\nu_n} + \big(\lambda_n\sqrt{\xi_n}\big)^2 \\ &\leqslant \mathsf{E}\nu_n + (1+\sqrt{\mu_n})\sqrt{\xi_n} + 2\Big(\sup_{k\in\mathbb{N}}\sqrt{\xi_k}\Big)\sqrt{\mathsf{E}\nu_n} + \big(\sqrt{\xi_n}\big)^2, \end{aligned} \tag{2.22}$$

where we have used the fact that $\lambda_n\in\,]0,1]$ and Jensen's inequality. We deduce from (2.22), (2.13), and (2.16) that $\sum_{n\in\mathbb{N}}\mathsf{E}\eta_n < +\infty$. Hence it follows from (2.21) and Lemma 2.2(iia) that $\mathsf{E}\|x_n-\mathsf{z}\|^2\to 0$ and that $\sum_{n\in\mathbb{N}}\mathsf{E}\vartheta_n < +\infty$. In view of (2.20), we obtain (iib) and (iic).

(iid) In view of (2.20), if $(\nu_n)_{n\in\mathbb{N}}\in\ell_+^{1/2}(\mathscr{F})$, then $(\eta_n)_{n\in\mathbb{N}}\in\ell_+^1(\mathscr{F})$ and the strong convergence claim follows from Lemma 2.2(iib). □

Remark 2.4 (i) Under the assumptions of Theorem 2.3, if $\nu_n\equiv 0$ and $\xi_n\equiv 0$, then $\overline{\eta}_n\equiv 0$ and it follows from (i) that $(\mathsf{E}\left(\|x_n-\mathsf{z}\|^2 \,\middle|\, \mathfrak{F}_0\right))_{n\in\mathbb{N}}$ converges linearly to 0.

(ii) The weak and strong almost sure convergences of a sequence $(x_n)_{n\in\mathbb{N}}$ governed by [a] and (2.14) were established in [11, Theorem 2.5] under different assumptions on $(\mu_n)_{n\in\mathbb{N}}$, $(\nu_n)_{n\in\mathbb{N}}$, and $(e_n)_{n\in\mathbb{N}}$.

3 Mean-square and linear convergence of Algorithm 1.2

We complement the almost sure weak and strong convergence results of [11] on the convergence of the orbits of Algorithm 1.2 by establishing mean-square and linear convergence properties.

3.1 Main results

The next theorem constitutes our main result in terms of mean-square convergence. For added flexibility, this convergence will be evaluated in a norm $|||\cdot|||$ on H parameterized by weights $(\omega_i)_{1\leqslant i\leqslant m} \in]0,+\infty[^m$ and defined by

$$(\forall \mathbf{x} \in \mathsf{H}) \quad |||\mathbf{x}|||^2 = \sum_{i=1}^{m} \omega_i \|\mathsf{x}_i\|^2. \tag{3.1}$$

Theorem 3.1 *Consider the setting of Problem 1.3 and Algorithm 1.2, and let $(\mathcal{F}_n)_{n\in\mathbb{N}}$ be a sequence of sub-sigma-algebras of $\mathcal{F}$ such that*

$$(\forall n \in \mathbb{N}) \quad \sigma(\boldsymbol{x}_0,\ldots,\boldsymbol{x}_n) \subset \mathcal{F}_n \subset \mathcal{F}_{n+1}. \tag{3.2}$$

Assume that the following are satisfied:

[a] $\inf_{n\in\mathbb{N}} \lambda_n > 0$.
[b] *There exists a sequence $(\alpha_n)_{n\in\mathbb{N}}$ in $[0,+\infty[$ such that $\sum_{n\in\mathbb{N}} \sqrt{\alpha_n} < +\infty$ and, for every $n \in \mathbb{N}$, $\mathsf{E}\left(\|\boldsymbol{a}_n\|^2 \,\middle|\, \mathcal{F}_n\right) \leqslant \alpha_n$.*
[c] *For every $n \in \mathbb{N}$, $\mathcal{E}_n = \sigma(\boldsymbol{\varepsilon}_n)$ and $\mathcal{F}_n$ are independent.*
[d] *For every $i \in \{1,\ldots,m\}$, $\mathsf{p}_i = \mathsf{P}[\varepsilon_{i,0} = 1] > 0$.*

Then the following hold:

(i) *Let $(\omega_i)_{1\leqslant i\leqslant m} \in]0,+\infty[^m$ be such that*

$$\begin{cases} (\forall i \in \{1,\ldots,m\}) \quad \overline{\lim}\, \tau_{i,n} < \omega_i \mathsf{p}_i \\ \max\limits_{1\leqslant i\leqslant m} \omega_i \mathsf{p}_i = 1, \end{cases} \tag{3.3}$$

set

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \xi_n = \alpha_n \max\limits_{1\leqslant i\leqslant m} \omega_i \\ \mu_n = 1 - \min\limits_{1\leqslant i\leqslant m} \left(\mathsf{p}_i - \dfrac{\tau_{i,n}}{\omega_i}\right), \end{cases} \tag{3.4}$$

and define

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \chi_n = 1 - \lambda_n(1-\mu_n) + \sqrt{\xi_n}\lambda_n(1-\lambda_n+\lambda_n\sqrt{\mu_n}) \\ \overline{\eta}_n = \sum_{k=0}^{n}\left[\prod_{\ell=k+1}^{n}\chi_\ell\right]\lambda_k\big(1-\lambda_k+\lambda_k\sqrt{\mu_k}+\lambda_k\sqrt{\xi_k}\big)\sqrt{\xi_k}. \end{cases} \tag{3.5}$$

Then

$$(\forall n \in \mathbb{N}) \quad \sum_{i=1}^{m}\omega_i \mathsf{E}\left(\|x_{i,n+1}-\overline{\mathsf{x}}_i\|^2 \,\middle|\, \mathcal{F}_0\right) \leqslant \left(\prod_{k=0}^{n}\chi_k\right)\left(\sum_{i=1}^{m}\omega_i\|x_{i,0}-\overline{\mathsf{x}}_{i,0}\|^2\right) + \overline{\eta}_n \quad \mathsf{P}\text{-a.s.} \tag{3.6}$$

(ii) *Suppose that* $\boldsymbol{x}_0 \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathbf{H})$ *and* $(\forall i \in \{1,\ldots,m\})$ $\overline{\lim}\,\tau_{i,n} < 1$. *Then* $\mathsf{E}\|\boldsymbol{x}_n - \overline{\mathbf{x}}\|^2 \to 0$ *and* $\boldsymbol{x}_n \to \overline{\mathbf{x}}$ P-a.s.

Proof (i) We are going to apply Theorem 2.3 in the Hilbert space $(\mathbf{H}, |||\cdot|||)$ defined by (3.1). Set

$$(\forall n \in \mathbb{N}) \quad \boldsymbol{t}_n = \big(x_{i,n} + \varepsilon_{i,n}(\mathsf{T}_{i,n}\boldsymbol{x}_n - x_{i,n})\big)_{1\leqslant i\leqslant m} \quad \text{and} \quad \boldsymbol{e}_n = (\varepsilon_{i,n}a_{i,n})_{1\leqslant i\leqslant m}. \tag{3.7}$$

Then it follows from (1.2) that

$$(\forall n \in \mathbb{N}) \quad \boldsymbol{x}_{n+1} = \boldsymbol{x}_n + \lambda_n\big(\boldsymbol{t}_n + \boldsymbol{e}_n - \boldsymbol{x}_n\big), \tag{3.8}$$

while [b] implies that

$$(\forall n \in \mathbb{N}) \qquad \mathsf{E}\left(|||\boldsymbol{e}_n|||^2 \,\middle|\, \mathcal{F}_n\right) \leqslant \mathsf{E}\left(|||\boldsymbol{a}_n|||^2 \,\middle|\, \mathcal{F}_n\right) \leqslant \alpha_n \max_{1\leqslant i\leqslant m}\omega_i = \xi_n. \tag{3.9}$$

We note that it also follows from [b] that $\sum_{n\in\mathbb{N}}\sqrt{\xi_n} < +\infty$. Now define

$$(\forall n \in \mathbb{N})(\forall i \in \{1,\ldots,m\}) \quad \mathsf{q}_{i,n} : \mathbf{H}\times\mathsf{D} \to \mathbb{R} : (\mathbf{x}, \boldsymbol{\epsilon}) \mapsto \|\mathsf{x}_i - \overline{\mathsf{x}}_i + \epsilon_i(\mathsf{T}_{i,n}\mathbf{x} - \mathsf{x}_i)\|^2. \tag{3.10}$$

Then, for every $n \in \mathbb{N}$ and every $i \in \{1,\ldots,m\}$, the measurability of $\mathsf{T}_{i,n}$ implies that of the functions $(\mathsf{q}_{i,n}(\cdot,\boldsymbol{\epsilon}))_{\boldsymbol{\epsilon}\in\mathsf{D}}$. However, for every $n \in \mathbb{N}$, [c] asserts that the events $([\boldsymbol{\varepsilon}_n = \boldsymbol{\epsilon}])_{\boldsymbol{\epsilon}\in\mathsf{D}}$ constitute an almost sure partition of Ω and are independent from $\mathcal{F}_n$, while the random variables $(\mathsf{q}_{i,n}(\boldsymbol{x}_n, \boldsymbol{\epsilon}))_{1\leqslant i\leqslant m,\ \boldsymbol{\epsilon}\in\mathsf{D}}$ are $\mathcal{F}_n$-measurable. Therefore, we derive from [16, Section 28.2] that

$$
\begin{aligned}
(\forall n \in \mathbb{N})(\forall i \in \{1,\ldots,m\}) \quad & \mathsf{E}\left(\|x_{i,n} + \varepsilon_{i,n}(\mathsf{T}_{i,n}\,\boldsymbol{x}_n - x_{i,n}) - \mathsf{x}_i\|^2 \,\middle|\, \mathcal{F}_n\right) \\
&= \mathsf{E}\Bigg(\mathsf{q}_{i,n}(\boldsymbol{x}_n, \boldsymbol{\varepsilon}_n) \sum_{\boldsymbol{\epsilon}\in\mathsf{D}} 1_{[\boldsymbol{\varepsilon}_n=\boldsymbol{\epsilon}]} \,\Bigg|\, \mathcal{F}_n\Bigg) \\
&= \sum_{\boldsymbol{\epsilon}\in\mathsf{D}} \mathsf{E}\left(\mathsf{q}_{i,n}(\boldsymbol{x}_n, \boldsymbol{\epsilon}) 1_{[\boldsymbol{\varepsilon}_n=\boldsymbol{\epsilon}]} \,\middle|\, \mathcal{F}_n\right) \\
&= \sum_{\boldsymbol{\epsilon}\in\mathsf{D}} \mathsf{E}\left(1_{[\boldsymbol{\varepsilon}_n=\boldsymbol{\epsilon}]} \,\middle|\, \mathcal{F}_n\right) \mathsf{q}_{i,n}(\boldsymbol{x}_n, \boldsymbol{\epsilon}) \\
&= \sum_{\boldsymbol{\epsilon}\in\mathsf{D}} \mathsf{P}[\boldsymbol{\varepsilon}_n = \boldsymbol{\epsilon}]\mathsf{q}_{i,n}(\boldsymbol{x}_n, \boldsymbol{\epsilon}) \quad \mathsf{P}\text{-a.s.}
\end{aligned}
\tag{3.11}
$$

Combining this identity with (3.1), (3.7), [d], (3.3), and (1.5) yields

$$
\begin{aligned}
(\forall n \in \mathbb{N}) \quad & \mathsf{E}\left(|||\boldsymbol{t}_n - \overline{\mathbf{x}}|||^2 \,\middle|\, \mathcal{F}_n\right) \\
&= \sum_{i=1}^m \omega_i \mathsf{E}\Big(\big\|x_{i,n} + \varepsilon_{i,n}(\mathsf{T}_{i,n}\,\boldsymbol{x}_n - x_{i,n}) - \overline{\mathsf{x}}_i\big\|^2 \,\Big|\, \mathcal{F}_n\Big) \\
&= \sum_{i=1}^m \omega_i \sum_{\boldsymbol{\epsilon}\in\mathsf{D}} \mathsf{P}[\boldsymbol{\varepsilon}_n = \boldsymbol{\epsilon}]\mathsf{q}_{i,n}(\boldsymbol{x}_n, \boldsymbol{\epsilon}) \\
&= \sum_{i=1}^m \omega_i \left(\sum_{\boldsymbol{\epsilon}\in\mathsf{D},\epsilon_i=1} \mathsf{P}[\boldsymbol{\varepsilon}_n = \boldsymbol{\epsilon}]\,\|\mathsf{T}_{i,n}\,\boldsymbol{x}_n - \overline{\mathsf{x}}_i\|^2 + \sum_{\boldsymbol{\epsilon}\in\mathsf{D},\,\epsilon_i=0} \mathsf{P}[\boldsymbol{\varepsilon}_n = \boldsymbol{\epsilon}]\,\|x_{i,n} - \overline{\mathsf{x}}_i\|^2\right) \\
&= \sum_{i=1}^m \omega_i \mathsf{p}_i \|\mathsf{T}_{i,n}\,\boldsymbol{x}_n - \overline{\mathsf{x}}_i\|^2 + \sum_{i=1}^m \omega_i(1-\mathsf{p}_i)\|x_{i,n} - \overline{\mathsf{x}}_i\|^2 \\
&\leqslant \left(\max_{1\leqslant i\leqslant m} \omega_i \mathsf{p}_i\right) \sum_{i=1}^m \|\mathsf{T}_{i,n}\,\boldsymbol{x}_n - \overline{\mathsf{x}}_i\|^2 + \sum_{i=1}^m \omega_i(1-\mathsf{p}_i)\|x_{i,n} - \overline{\mathsf{x}}_i\|^2 \\
&= |||\boldsymbol{x}_n - \overline{\mathbf{x}}|||^2 + \|\mathbf{T}_n\boldsymbol{x}_n - \overline{\mathbf{x}}\|^2 - \sum_{i=1}^m \omega_i \mathsf{p}_i \|x_{i,n} - \overline{\mathsf{x}}_i\|^2 \\
&\leqslant |||\boldsymbol{x}_n - \overline{\mathbf{x}}|||^2 + \sum_{i=1}^m (\tau_{i,n} - \omega_i \mathsf{p}_i)\|x_{i,n} - \overline{\mathsf{x}}_i\|^2 \\
&= \sum_{i=1}^m \omega_i \left(1 + \frac{\tau_{i,n}}{\omega_i} - \mathsf{p}_i\right)\|x_{i,n} - \overline{\mathsf{x}}_i\|^2 \\
&\leqslant \left(1 - \min_{1\leqslant i\leqslant m}\left(\mathsf{p}_i - \frac{\tau_{i,n}}{\omega_i}\right)\right) |||\boldsymbol{x}_n - \overline{\mathbf{x}}|||^2 \quad \mathsf{P}\text{-a.s.}
\end{aligned}
\tag{3.12}
$$

Altogether, properties [a]–[c] of Theorem 2.3 are satisfied with

$$
(\forall n \in \mathbb{N}) \quad \theta_n = \nu_n = 0. \tag{3.13}
$$

On the other hand, it follows from (3.3) and (3.4) that $\overline{\lim}\, \mu_n < 1$. Hence, we derive from Theorem 2.3(i) that

$$(\forall n \in \mathbb{N}) \quad \mathsf{E}\left(|||\boldsymbol{x}_{n+1} - \overline{\mathbf{x}}|||^2 \,\middle|\, \mathcal{F}_0\right) \leqslant \left(\prod_{k=0}^{n} \chi_k\right) |||\boldsymbol{x}_0 - \overline{\mathbf{x}}|||^2 + \overline{\eta}_n \quad \mathsf{P}\text{-a.s.} \tag{3.14}$$

(ii) Consider (i) when $(\forall i \in \{1, \ldots, m\})\ \omega_i = 1/\mathsf{p}_i$. The convergence then follows from the inequalities

$$(\forall \mathbf{x} \in \mathbf{H}) \quad \min_{1\leqslant i\leqslant m} \mathsf{p}_i\, |||\mathbf{x}||| \leqslant \|\mathbf{x}\| \leqslant \max_{1\leqslant i\leqslant m} \mathsf{p}_i\, |||\mathbf{x}||| \tag{3.15}$$

and Theorem 2.3(ii). □

3.2 Linear convergence

As an offspring of the results in Sect. 3.1, we obtain the following perturbed linear convergence result.

Corollary 3.2 *Consider the setting of Problem 1.3 and Algorithm 1.2, suppose that [a]–[d] in Theorem 3.1 are satisfied, and define $(\chi_n)_{n\in\mathbb{N}}$ and $(\overline{\eta}_n)_{n\in\mathbb{N}}$ as in (3.5), where*

$$\max_{1\leqslant i\leqslant m} \overline{\lim}\, \tau_{i,n} < 1 \quad \textit{and} \quad (\forall n \in \mathbb{N}) \quad \begin{cases} \xi_n = \dfrac{\alpha_n}{\min\limits_{1\leqslant i\leqslant m} \mathsf{p}_i} \\ \mu_n = 1 - \min\limits_{1\leqslant i\leqslant m} \mathsf{p}_i\big(1 - \tau_{i,n}\big). \end{cases} \tag{3.16}$$

Then

$$(\forall n \in \mathbb{N}) \quad \mathsf{E}\left(\|\boldsymbol{x}_{n+1} - \overline{\mathbf{x}}\|^2 \,\middle|\, \mathcal{F}_0\right) \leqslant \frac{\max\limits_{1\leqslant i\leqslant m} \mathsf{p}_i}{\min\limits_{1\leqslant i\leqslant m} \mathsf{p}_i} \left(\prod_{k=0}^{n} \chi_k\right) \|\boldsymbol{x}_0 - \overline{\mathbf{x}}\|^2 + \overline{\eta}_n \quad \mathsf{P}\text{-a.s.} \tag{3.17}$$

Proof In view of (3.15), the claim follows from Theorem 3.1(i) applied with $(\forall i \in \{1, \ldots, m\})\ \omega_i = 1/\mathsf{p}_i$. □

Let us now make some observations to assess the consequences of Corollary 3.2 in terms of bounds on convergence rates, and the potential impact of the activation probabilities of the blocks $(\mathsf{p}_i)_{1\leqslant i\leqslant m}$ on them. Let us consider the case when $\alpha_n \equiv 0$, i.e., when there are no errors. Set

$$(\forall n \in \mathbb{N}) \quad \chi_n = 1 - \lambda_n \min_{1\leqslant i\leqslant m} \mathsf{p}_i(1 - \tau_{i,n}). \tag{3.18}$$

Then we derive from (3.5) and (3.16) that

$$(\forall n\in\mathbb{N})\quad \mathsf{E}\Big(\|\boldsymbol{x}_{n+1}-\overline{\mathbf{x}}\|^2\,\Big|\,\mathcal{F}_0\Big)\leqslant\frac{\max\limits_{1\leqslant i\leqslant m}\mathsf{p}_i}{\min\limits_{1\leqslant i\leqslant m}\mathsf{p}_i}\left(\prod_{k=0}^{n}\chi_k\right)\|\boldsymbol{x}_0-\overline{\mathbf{x}}\|^2\quad \mathsf{P}\text{-a.s.} \tag{3.19}$$

Since (3.16) yields $\sup_{n\in\mathbb{N}}\chi_n<1$, a linear convergence rate is thus obtained.

For simplicity, let us further assume that the blocks are processed uniformly in the sense that $(\forall i\in\{1,\ldots,m\})$ $\mathsf{p}_i=\mathsf{p}$. Set

$$\chi=1-\inf_{n\in\mathbb{N}}\Big(\lambda_n\Big(1-\max_{1\leqslant i\leqslant m}\tau_{i,n}\Big)\Big)\in[0,1[. \tag{3.20}$$

Then

$$(\forall n\in\mathbb{N})\quad \chi_n=1-\lambda_n\mathsf{p}\Big(1-\max_{1\leqslant i\leqslant m}\tau_{i,n}\Big)\leqslant 1-(1-\chi)\mathsf{p}. \tag{3.21}$$

When $\mathsf{p}=1$, the upper bound in (3.21) on the convergence rate is minimal and equal to χ. This is consistent with the intuition that frequently activating the coordinates should favor the convergence speed as a function of the iteration number. On the other hand, activating the blocks less frequently induces a reduction of the computational load per iteration. In large scale problems, this reduction may actually be imposed by limited computing or memory resources. In Algorithm 1.2, the cost of computing $\mathsf{T}_{i,n}(x_{1,n},\ldots,x_{m,n})$ is on the average p times smaller than in the standard non block-coordinate approach. Hence, if we assume that this cost is independent of i and the iteration number n, N iterations of the block-coordinate algorithm have the same computational cost as $\mathsf{p}N$ iterations of a non block-coordinate approach. In view of (3.21), let us introduce the quantity

$$\varrho(\mathsf{p})=-\frac{\ln\big(1-(1-\chi)\mathsf{p}\big)}{\mathsf{p}} \tag{3.22}$$

to evaluate the convergence rate normalized by the probability p accounting for computational cost. Under the above assumptions, (3.21) yields

$$(\forall n\in\mathbb{N})\quad \prod_{k=0}^{n}\chi_k\leqslant\exp\big(-\varrho(\mathsf{p})\mathsf{p}(n+1)\big). \tag{3.23}$$

Elementary calculations show that, if $\chi\neq 0$,

$$-\frac{1-\chi}{\ln\chi}\leqslant\frac{\varrho(\mathsf{p})}{\varrho(1)}\leqslant 1. \tag{3.24}$$

For example, if $\chi>0.2$, then $\varrho(\mathsf{p})/\varrho(1)\in[0.49,1]$. This shows that, for values of χ not too small, the decrease in the normalized convergence rate remains limited with respect to a deterministic approach in which all the blocks are activated. This fact is illustrated by Fig. 1, where the graph of ϱ is plotted for several values of χ.

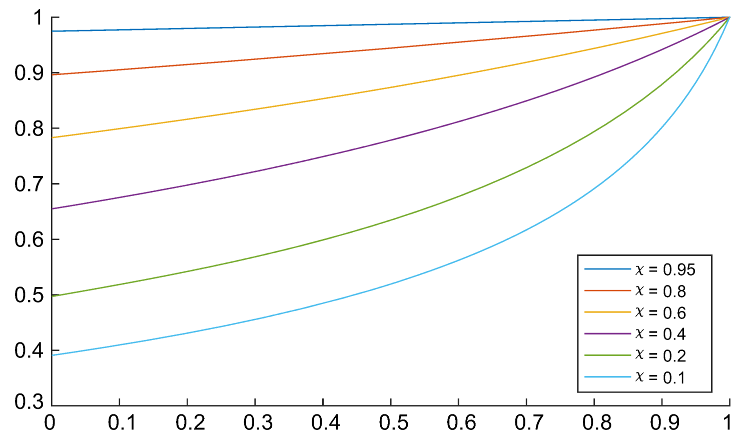

Fig. 1 Variations of $\varrho(\mathsf{p})/\varrho(1)$ as a function of p for various values of χ

Remark 3.3 Let us consider the special case in which, for every $i \in \{1, \ldots, m\}$, $\tau_{i,n} \equiv \tau_i$. Then (3.18) becomes

$$(\forall n \in \mathbb{N}) \quad \chi_n = 1 - \lambda_n \min_{1 \leqslant i \leqslant m} \mathsf{p}_i (1 - \tau_i). \tag{3.25}$$

Now, let us further assume that, at each iteration n, only one of the operators $(\mathsf{T}_{i,n})_{1 \leqslant i \leqslant m}$ is activated randomly. In this case, $\sum_{i=1}^m \mathsf{p}_i = 1$ and choosing

$$(\forall i \in \{1, \ldots, m\}) \quad \mathsf{p}_i = \frac{(1 - \tau_i)^{-1}}{\sum_{j=1}^m (1 - \tau_j)^{-1}} \tag{3.26}$$

leads to a minimum value of χ_n.

4 Applications

In variational analysis, commonly encountered operators include resolvent of monotone operators, projection operators, proximity operators of convex functions, gradient operators, and various compositions and combinations thereof [6,23]. Specific instances of such operators used in iterative processes which satisfy property (1.5) can be found in [5,6,8–10,13,19,22,23,26]. In this section we highlight a couple of examples in the area of splitting methods for systems of monotone inclusions. The notation is that used in Problem 1.3. In addition, let $\mathsf{A} : \mathsf{H} \to 2^{\mathsf{H}}$ be a set-valued operator. We denote by $\operatorname{zer} \mathsf{A} = \{\mathsf{x} \in \mathsf{H} \mid 0 \in \mathsf{A}\mathsf{x}\}$ the set of zeros of A and by $\mathsf{J}_\mathsf{A} = (\mathsf{Id} + \mathsf{A})^{-1}$ the resolvent of A. Recall that, if A is maximally monotone, then J_A is defined everywhere on H and nonexpansive [6]. In the particular case when A is the Moreau subdifferential $\partial \mathsf{f}$ of a proper lower semicontinuous convex function $\mathsf{f} : \mathsf{H} \to]-\infty, +\infty]$, J_A is the proximity operator $\operatorname{prox}_\mathsf{f}$ of f [6,17].

Example 4.1 For every $i \in \{1, \ldots, m\}$, let $\mathsf{A}_i : \mathsf{H} \to 2^{\mathsf{H}}$ be a maximally monotone operator, and consider the coupled inclusion problem

$$\text{find } \mathbf{x} = (\mathsf{x}_i)_{1 \leqslant i \leqslant m} \in \mathbf{H} \text{ such that } \begin{cases} 0 \in \mathsf{A}_1\mathsf{x}_1 + \mathsf{x}_1 - \mathsf{x}_2 \\ 0 \in \mathsf{A}_2\mathsf{x}_2 + \mathsf{x}_2 - \mathsf{x}_3 \\ \quad\vdots \\ 0 \in \mathsf{A}_{m-1}\mathsf{x}_{m-1} + \mathsf{x}_{m-1} - \mathsf{x}_m \\ 0 \in \mathsf{A}_m\mathsf{x}_m + \mathsf{x}_m - \mathsf{x}_1. \end{cases} \tag{4.1}$$

For instance, in the case when each A_i is the normal cone operator to a nonempty closed convex set, (4.1) models limit cycles in the method of periodic projections [4]. Another noteworthy instance is when $m = 2$, $\mathsf{A}_1 = \partial\mathsf{f}_1$, and $\mathsf{A}_2 = \partial\mathsf{f}_2$, where f_1 and f_2 are proper lower semicontinuous functions from H to $]-\infty, +\infty]$. Then (4.1) reduces to the joint minimization problem

$$\underset{(\mathsf{x}_1,\mathsf{x}_2)\in\mathsf{H}^2}{\text{minimize}} \;\; \mathsf{f}_1(\mathsf{x}_1) + \mathsf{f}_2(\mathsf{x}_2) + \frac{1}{2}\|\mathsf{x}_1 - \mathsf{x}_2\|^2, \tag{4.2}$$

studied in [1]. Now set

$$\mathbf{A} : \mathbf{x} \mapsto (\mathsf{A}_1\mathsf{x}_1, \ldots, \mathsf{A}_m\mathsf{x}_m) \quad \text{and} \quad \mathbf{B} : \mathbf{x} \mapsto (\mathsf{x}_1 - \mathsf{x}_2, \mathsf{x}_2 - \mathsf{x}_3, \ldots, \mathsf{x}_m - \mathsf{x}_1). \tag{4.3}$$

Then it follows from [6, Proposition 20.23] that $\mathbf{A}$ is maximally monotone. On the other hand, $\mathbf{B}$ is linear, bounded, and monotone since

$$(\forall \mathbf{x} \in \mathbf{H}) \quad \langle \mathbf{Bx} \mid \mathbf{x} \rangle = \frac{\|\mathbf{Bx}\|^2}{2} \geqslant 0. \tag{4.4}$$

It is therefore maximally monotone [6, Example 20.34]. Altogether, $\mathbf{A}+\mathbf{B}$ is maximally monotone by [6, Corollary 25.5(i)]. In addition, suppose that each A_i is strongly monotone with constant $\delta_i \in]0, +\infty[$. Then $\mathbf{A}$ is strongly monotone with constant $\min_{1\leqslant i\leqslant m} \delta_i$, and so is $\mathbf{A} + \mathbf{B}$. We therefore deduce from [6, Corollary 23.37(ii)] that it possesses a unique zero $\overline{\mathbf{x}}$, which is the unique solution to (4.1). Let us also note that, for every $i \in \{1, \ldots, m\}$, the resolvent $\mathsf{J}_{\mathsf{A}_i}$ is Lipschitz continuous with constant $\eta_i = 1/(1 + \delta_i) \in]0, 1[$ [6, Proposition 23.13]. Next, define $\mathbf{T} : \mathbf{H} \to \mathbf{H} : \mathbf{x} \mapsto (\mathsf{T}_i\, \mathbf{x})_{1\leqslant i\leqslant m}$, where, for every $i \in \{1, \ldots, m\}$, $\mathsf{T}_i : \mathbf{H} \to \mathsf{H}_i : \mathbf{x} \mapsto \mathsf{J}_{\mathsf{A}_i}\mathsf{x}_{i+1}$, with the convention $\mathsf{x}_{m+1} = \mathsf{x}_1$. Then we derive from (4.1) that $\mathbf{T}\overline{\mathbf{x}} = \overline{\mathbf{x}}$. Moreover,

$$\begin{aligned} (\forall n \in \mathbb{N})(\forall \mathbf{x} \in \mathbf{H}) \quad \|\mathbf{Tx} - \overline{\mathbf{x}}\|^2 &= \sum_{i=1}^m \|\mathsf{J}_{\mathsf{A}_i}\mathsf{x}_{i+1} - \overline{\mathsf{x}}_i\|^2 \\ &= \sum_{i=1}^m \|\mathsf{J}_{\mathsf{A}_i}\mathsf{x}_{i+1} - \mathsf{J}_{\mathsf{A}_i}\overline{\mathsf{x}}_{i+1}\|^2 \\ &\leqslant \sum_{i=1}^m \eta_i^2\|\mathsf{x}_{i+1} - \overline{\mathsf{x}}_{i+1}\|^2, \end{aligned} \tag{4.5}$$

which shows that (1.5) is satisfied upon choosing $\mathbf{T}_n \equiv \mathbf{T}$ and, for every $i \in \{1, \dots, m\}$, $\tau_{i,n} \equiv \eta_i^2$. In this scenario, Algorithm 1.2 becomes

$$\begin{array}{l}\text{for } n = 0, 1, \dots \\ \left\lfloor \begin{array}{l} \text{for } i = 1, \dots, m-1 \\ \left\lfloor x_{i,n+1} = x_{i,n} + \varepsilon_{i,n}\lambda_n\big(\mathsf{J}_{\mathsf{A}_i} x_{i+1,n} + a_{i,n} - x_{i,n}\big) \right. \\ x_{m,n+1} = x_{m,n} + \varepsilon_{m,n}\lambda_n\big(\mathsf{J}_{\mathsf{A}_m} x_{1,n} + a_{m,n} - x_{m,n}\big), \end{array}\right.\end{array} \tag{4.6}$$

and Theorem 3.1 describes its asymptotic behavior. In the particular case of (4.2), for f_1 and f_2 strongly convex, (4.6) with $\lambda_n \equiv 1$ and no error, reduces to

$$\begin{array}{l}\text{for } n = 0, 1, \dots \\ \left\lfloor \begin{array}{l} x_{1,n+1} = x_{1,n} + \varepsilon_{1,n}\big(\mathrm{prox}_{\mathsf{f}_1} x_{2,n} - x_{1,n}\big) \\ x_{2,n+1} = x_{2,n} + \varepsilon_{2,n}\big(\mathrm{prox}_{\mathsf{f}_2} x_{1,n} - x_{2,n}\big). \end{array}\right.\end{array} \tag{4.7}$$

In the deterministic setting in which $\varepsilon_{1,n} \equiv 1$ and $\varepsilon_{2,n} \equiv 1$, the resulting sequence $(x_{2,n})_{n\in\mathbb{N}}$ is that produced by the alternating proximity operator method of [1], further studied in [7].

Example 4.2 We consider an m-agent model investigated in [3]. For every $i \in \{1, \dots, m\}$, let $\mathsf{A}_i : \mathsf{H}_i \to 2^{\mathsf{H}_i}$ be a maximally monotone operator modeling some abstract utility of agent i and let $\mathsf{B}_i : \mathbf{H} \to \mathsf{H}_i$ be a coupling operator. It is assumed that the operator $\mathbf{B} : \mathbf{H} \to \mathbf{H} : \mathbf{x} \mapsto (\mathsf{B}_i\,\mathbf{x})_{1\leqslant i\leqslant m}$ is β-cocoercive [6] for some $\beta \in\,]0, +\infty[$, that is,

$$(\forall \mathbf{x} \in \mathbf{H})(\forall \mathbf{y} \in \mathbf{H}) \quad \langle \mathbf{x} - \mathbf{y} \mid \mathbf{B}\mathbf{x} - \mathbf{B}\mathbf{y}\rangle \geqslant \beta\|\mathbf{B}\mathbf{x} - \mathbf{B}\mathbf{y}\|^2. \tag{4.8}$$

The equilibrium problem is to

$$\text{find } \mathbf{x} \in \mathbf{H} \ \text{ such that } \ (\forall i \in \{1, \dots, m\}) \ \ 0 \in \mathsf{A}_i \mathsf{x}_i + \mathsf{B}_i\big(\mathsf{x}_1, \dots, \mathsf{x}_m\big). \tag{4.9}$$

For every $i \in \{1, \dots, m\}$, let us further assume that A_i is δ_i-strongly monotone for some $\delta_i \in\,]0, +\infty[$ or, equivalently, that $\mathsf{M}_i = \mathsf{A}_i - \delta_i \mathrm{Id}$ is monotone. Since $\mathbf{B}$ is maximally monotone [6, Example 20.31], arguing as in Example 4.1, we arrive at the conclusion that $\mathbf{A}+\mathbf{B}$ has exactly one zero $\overline{\mathbf{x}}$, and that $\overline{\mathbf{x}}$ is the unique solution to (4.9). Let

$$\delta = \min_{1\leqslant i\leqslant m} \delta_i, \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad \theta_n \in [0, \delta] \quad \text{and} \quad \gamma_n \in\,]0, +\infty[. \tag{4.10}$$

Set

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \mathbf{C}_n : \mathbf{H} \to 2^{\mathbf{H}} : \mathbf{x} \mapsto \displaystyle\bigtimes_{i=1}^{m} \big(\mathsf{M}_i + (\delta_i - \theta_n)\mathrm{Id}\big)\mathsf{x}_i \\ \mathbf{D}_n = \mathbf{B} + \theta_n \mathbf{Id} \\ \mathbf{T}_n = \mathbf{J}_{\gamma_n \mathbf{C}_n} \circ (\mathbf{Id} - \gamma_n \mathbf{D}_n). \end{cases} \tag{4.11}$$

Now let $n \in \mathbb{N}$. We first observe that

$$\operatorname{zer}\big(\gamma_n \mathbf{C}_n + \gamma_n \mathbf{D}_n\big) = \operatorname{zer}\big(\mathbf{A} + \mathbf{B}\big) = \{\overline{\mathbf{x}}\} = \operatorname{Fix} \mathbf{T}_n, \tag{4.12}$$

and derive from [6, Proposition 23.17(i)] that

$$\mathbf{J}_{\gamma_n \mathbf{C}_n} \colon \mathbf{x} \mapsto \left(\mathsf{J}_{\frac{\gamma_n \mathsf{M}_i}{1+\gamma_n(\delta_i - \theta_n)}} \left(\frac{\mathsf{x}_i}{1 + \gamma_n(\delta_i - \theta_n)} \right) \right)_{1 \leqslant i \leqslant m}. \tag{4.13}$$

Hence (4.10) entails that $\mathbf{J}_{\gamma_n \mathbf{C}_n}$ is Lipschitz continuous with constant $1/(1+\gamma_n(\delta-\theta_n))$. On the other hand, since $\mathbf{B}$ is β-cocoercive, there exists a nonexpansive operator $\mathbf{R} \colon \mathbf{H} \to \mathbf{H}$ such that $\beta \mathbf{B} = (\mathbf{Id} + \mathbf{R})/2$ [6, Remark 4.34(iv)]. We have

$$\mathbf{Id} - \gamma_n \mathbf{D}_n = \left(1 - \gamma_n \theta_n - \frac{\gamma_n}{2\beta}\right) \mathbf{Id} - \frac{\gamma_n}{2\beta} \mathbf{R}. \tag{4.14}$$

In turn, a Lipschitz constant of $\mathbf{Id} - \gamma_n \mathbf{D}_n$ is $|1 - \gamma_n(\theta_n + 1/(2\beta))| + \gamma_n/(2\beta)$, and hence one for $\mathbf{T}_n$ is

$$\zeta_n = \frac{\big|1 - \gamma_n\big(\theta_n + 1/(2\beta)\big)\big| + \gamma_n/(2\beta)}{1 + \gamma_n(\delta - \theta_n)}. \tag{4.15}$$

Note that

$$\zeta_n = \begin{cases} \dfrac{1 - \gamma_n \theta_n}{1 + \gamma_n(\delta - \theta_n)} < 1, & \text{if } \gamma_n \leqslant \dfrac{2\beta}{1 + 2\beta\theta_n}; \\[2ex] \dfrac{\gamma_n(\theta_n + 1/\beta) - 1}{1 + \gamma_n(\delta - \theta_n)} < 1, & \text{if } \dfrac{2\beta}{1 + 2\beta\theta_n} < \gamma_n < \dfrac{2\beta}{1 + \beta(2\theta_n - \delta)}. \end{cases} \tag{4.16}$$

Consequently, imposing

$$\gamma_n < \frac{2\beta}{1 + \beta(2\theta_n - \delta)} \tag{4.17}$$

places us in the framework of Problem 1.3 with $(\forall i \in \{1, \ldots, m\})\ \tau_{i,n} = \zeta_n^2$. Algorithm 1.2 for solving (4.9), that is,

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} \text{for } i = 1, \ldots, m \\ \left\lfloor x_{i,n+1} = x_{i,n} + \varepsilon_{i,n} \lambda_n \left(\mathsf{J}_{\frac{\gamma_n \mathsf{M}_i}{1+\gamma_n(\delta_i - \theta_n)}} \left(\dfrac{(1 - \gamma_n \theta_n) x_{i,n} - \gamma_n \mathsf{B}_i \boldsymbol{x}_n}{1 + \gamma_n(\delta_i - \theta_n)} \right) + a_{i,n} - x_{i,n} \right), \right. \end{array} \right. \end{array} \tag{4.18}$$

is then an instance of the block-coordinate forward-backward algorithm of [11, Section 5.2]. Its convergence properties in the present setting are given in Theorem 3.1.

Remark 4.3 In view of (4.4), (4.1) constitutes a special case of (4.9) and it can also be solved via (4.18). In Example 4.1, we have exploited the special structure of $\mathbf{B}$ to obtain tighter coefficients $(\tau_{i,n})_{1\leqslant i\leqslant m, n\in\mathbb{N}}$ in (1.5).

Example 4.4 Let $\mathbf{g}\colon \mathbf{H}\to\mathbb{R}$ be a convex function which is differentiable with a β^{-1}-Lipschitzian gradient for some $\beta\in\,]0,+\infty[$ and, for every $i\in\{1,\ldots,m\}$, let $\mathsf{f}_i\colon \mathsf{H}_i\to\,]-\infty,+\infty]$ be a proper lower semicontinuous δ_i-strongly convex function for some $\delta_i\in\,]0,+\infty[$. We consider the optimization problem

$$\underset{\mathsf{x}_1\in\mathsf{H}_1,\ldots,\mathsf{x}_m\in\mathsf{H}_m}{\text{minimize}}\quad \sum_{i=1}^{m}\mathsf{f}_i(\mathsf{x}_i)+\mathbf{g}(\mathsf{x}_1,\ldots,\mathsf{x}_m). \tag{4.19}$$

Then it results from standard facts [6, Section 28.5] that this problem is the special case of Example 4.2 in which $\mathbf{B}=\nabla\mathbf{g}$ and, for every $i\in\{1,\ldots,m\}$, $\mathsf{A}_i=\partial\mathsf{f}_i$. Now set $(\forall i\in\{1,\ldots,m\})$ $\mathsf{h}_i=\mathsf{f}_i-\delta_i\|\cdot\|^2/2$. Then (4.18) assumes the form

$$\begin{array}{l}\text{for } n=0,1,\ldots\\ \left\lfloor\begin{array}{l}\text{for } i=1,\ldots,m\\ \left\lfloor\begin{array}{l} x_{i,n+1}=x_{i,n}+\varepsilon_{i,n}\lambda_n\\ \left(\operatorname{prox}_{\frac{\gamma_n\mathsf{h}_i}{1+\gamma_n(\delta_i-\theta_n)}}\left(\dfrac{(1-\gamma_n\theta_n)x_{i,n}-\gamma_n\nabla_i\,\mathbf{g}(\boldsymbol{x}_n)}{1+\gamma_n(\delta_i-\theta_n)}\right)+a_{i,n}-x_{i,n}\right),\end{array}\right.\end{array}\right.\end{array} \tag{4.20}$$

where $\nabla_i\,\mathbf{g}\colon\mathbf{H}\to\mathsf{H}_i$ is the ith component of $\nabla\mathbf{g}$.

Remark 4.5 In the case of a non block-coordinate implementation, i.e., $m=1$, a mean-square convergence result for the forward-backward algorithm can be found in [24] under different assumptions than ours and, in particular, the requirement that the proximal parameters $(\gamma_n)_{n\in\mathbb{N}}$ must go to 0.

Remark 4.6 In connection with the linear convergence of (4.20) deriving from Corollary 3.2, let us note that a similar result was obtained in [20] by imposing the restrictions

$$(\forall i\in\{1,\ldots,m\})\quad \mathsf{H}_i=\mathbb{R}^{N_i},\quad \mathsf{p}_i=\frac{1}{m},\quad\text{and}\quad(\forall n\in\mathbb{N})\quad\lambda_n=1\ \text{ and }\ a_{i,n}=0. \tag{4.21}$$

In this specific setting the proximal parameter in [20] was chosen differently for each block: it is not allowed to vary with the iteration n as in (4.20), but it can be chosen differently for each i. In the case when, for every $i\in\{1,\ldots,m\}$, $\mathsf{f}_i=0$, more freedom was given to the choice of $(\mathsf{p}_i)_{1\leqslant i\leqslant m}$ in [20], but by still activating only one block at each iteration. Further narrowing the problem to the minimization of a smooth strongly convex function on $\mathbb{R}^N$, a coordinate descent method is proposed in [21] which requires, for every $i\in\{1,\ldots,m\}$, $\mathsf{H}_i=\mathbb{R}$ and allows for multiple coordinates to be randomly updated at each iteration, as in (4.20).

References

1. Acker, F., Prestel, M.A.: Convergence d'un schéma de minimisation alternée. Ann. Fac. Sci. Toulouse V. Sér. Math. **2**, 1–9 (1980)
2. Atchadé, Y.F., Fort, G., Moulines, E.: On perturbed proximal gradient algorithms. J. Mach. Learn. Res. **18**, 1–33 (2017)
3. Attouch, H., Briceño-Arias, L.M., Combettes, P.L.: A parallel splitting method for coupled monotone inclusions. SIAM J. Control Optim. **48**, 3246–3270 (2010)
4. Baillon, J.-B., Combettes, P.L., Cominetti, R.: There is no variational characterization of the cycles in the method of periodic projections. J. Funct. Anal. **262**, 400–408 (2012)
5. Bauschke, H.H., Bello Cruz, J.Y., Nghia, T.T.A., Phan, H.M., Wang, X.: Optimal rates of linear convergence of relaxed alternating projections and generalized Douglas–Rachford methods for two subspaces. Numer. Algorithms **73**, 33–76 (2016)
6. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces, 2nd edn. Springer, New York (2017)
7. Bauschke, H.H., Combettes, P.L., Reich, S.: The asymptotic behavior of the composition of two resolvents. Nonlinear Anal. **60**, 283–301 (2005)
8. Bauschke, H.H., Moffat, S.M., Wang, X.: Firmly nonexpansive mappings and maximally monotone operators: correspondence and duality. Set Valued Var. Anal. **20**, 131–153 (2012)
9. Boţ, R.I., Csetnek, E.R.: Convergence rates for forward–backward dynamical systems associated with strongly monotone inclusions. J. Math. Anal. Appl. **457**, 1135–1152 (2018)
10. Boţ, R.I., Csetnek, E.R., Heinrich, A., Hendrich, C.: On the convergence rate improvement of a primal-dual splitting algorithm for solving monotone inclusion problems. Math. Program. **150**, 251–279 (2015)
11. Combettes, P.L., Pesquet, J.-C.: Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping. SIAM J. Optim. **25**, 1221–1248 (2015)
12. Combettes, P.L., Pesquet, J.-C.: Stochastic approximations and perturbations in forward–backward splitting for monotone operators. Pure Appl. Funct. Anal. **1**, 13–37 (2016)
13. Combettes, P.L., Vũ, B.C.: Variable metric forward–backward splitting with applications to monotone inclusions in duality. Optimization **63**, 1289–1318 (2014)
14. Fortet, R.M.: Vecteurs, Fonctions et Distributions Aléatoires dans les Espaces de Hilbert. Hermès, Paris (1995)
15. Ledoux, M., Talagrand, M.: Probability in Banach Spaces: Isoperimetry and Processes. Springer, New York (1991)
16. Loève, M.: Probability Theory II, 4th edn. Springer, New York (1978)
17. Moreau, J.J.: Fonctions convexes duales et points proximaux dans un espace hilbertien. C. R. Acad. Sci. Paris Sér. A Math. **255**, 2897–2899 (1962)
18. Nesterov, Yu.: Efficiency of coordinate descent methods on huge-scale optimization problems. SIAM J. Optim. **22**, 341–362 (2012)
19. Pesquet, J.-C., Pustelnik, N.: A parallel inertial proximal optimization method. Pac. J. Optim. **8**, 273–305 (2012)
20. Richtárik, P., Takáč, M.: Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function. Math. Program. **A144**, 1–38 (2014)
21. Richtárik, P., Takáč, M.: On optimal probabilities in stochastic coordinate descent methods. Optim. Lett. **10**, 1233–1243 (2016)
22. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control Optim. **14**, 877–898 (1976)
23. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis, 3rd edn. Springer, New York (2009)
24. Rosasco, L., Villa, S., Vũ, B.C.: Stochastic forward–backward splitting method for solving monotone inclusions in Hilbert spaces. J. Optim. Theory Appl. **169**, 388–406 (2016)
25. Schwartz, L.: Analyse III—Calcul Intégral. Hermann, Paris (1993)
26. Sibony, M.: Méthodes itératives pour les équations et inéquations aux dérivées partielles non linéaires de type monotone. Calcolo **7**, 65–183 (1970)