ABSTRACT

MADARIAGA ROMAN, JAVIER IVAN. Stochastic Iterations in Hilbertian Nonlinear Analysis. (Under the direction of Dr. Patrick Combettes).

This dissertation focuses on the development and the analysis of stochastic iterative methods for solving nonlinear analysis problems in Hilbert spaces, including minimization, monotone inclusion, and fixed point problems. The methods under consideration are stochastic in the sense that, at every iteration, the update is constructed using random quantities, e.g., problem data defined over probability spaces, activation of random blocks of operators or coordinates, use of stochastic approximations of the operators, averaging of operators with random weights, and random relaxation parameters. Abstract frameworks are constructed to analyze in an unified fashion these methods and establish both almost sure and mean-square convergence results for the sequences of iterates they produce. They allow us to design novel stochastic algorithms for composite inclusion problems, fixed point iterations, gradient descent, and extrapolated parallel algorithms for feasibility problems. They also allow us to improve existing results on the convergence of stochastic methods such as the stochastic Krasnosel'skiĭ–Mann algorithm and the stochastic gradient algorithm. Finally, an abstract stochastic version of the Haugazeau method is proposed to compute the best approximation from a closed convex set by successive projections onto randomly generated stochastic outer approximations of that set. Throughout the dissertation, numerical experiments are provided to illustrate the theoretical findings.

This dissertation was supported by the U.S. National Science Foundation grants CCF-2211123 and DMS-2513409 (PI: P. L. Combettes).

Stochastic Iterations in Hilbertian Nonlinear Analysis

by
Javier Ivan Madariaga Roman

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Mathematics

Raleigh, North Carolina

2026

APPROVED BY:

Erik Bates	Tien Khai Nguyen
Andrew Papanicolaou	Sara Shashaani

Patrick Combettes
Chair of Advisory Committee

BIOGRAPHY

Javier received his degree in Mathematical Engineering and his Master's in Engineering Sciences, with a minor in Applied Mathematics, from the Universidad de Chile in December 2022. He began his Ph.D. studies in the Department of Mathematics at North Carolina State University under the supervision of Professor Patrick L. Combettes in January 2023, and was admitted to doctoral candidacy in December 2024.

DEDICATION

Para Catalina: *Sine qua non.*

Resumen. Iteraciones Estocásticas en el Análisis no lineal Hilbertiano (Bajo la dirección de Patrick L. Combettes.) Esta disertación se enfoca en el desarrollo y el análisis de métodos iterativos estocásticos para resolver problemas de análisis no lineal en espacios de Hilbert, incluyendo problemas de minimización, inclusión monótona, y punto fijo. Los métodos bajo consideración son estocásticos en el sentido de que, en cada iteración, la actualización es construida usando cantidades aleatorias, por ejemplo, datos del problema definidos sobre espacios de probabilidad, activación de bloques aleatorios de operadores o coordenadas, uso de aproximaciones estocásticas de los operadores, promediado de operadores con pesos aleatorios, y parámetros de relajación aleatorios. Esquemas abstractos son construidos para analizar de una manera unificada estos métodos y establecer resultados de convergencia tanto casi segura como en media cuadrática para las secuencias de iterados que ellos producen. Ellos nos permiten diseñar algoritmos estocásticos novedosos para problemas de inclusión compuesta, iteraciones de punto fijo, descenso de gradiente, y algoritmos paralelos extrapolados para problemas de factibilidad. Ellos también nos permiten mejorar resultados existentes sobre la convergencia de métodos estocásticos tales como el algoritmo estocástico de Krasnosel'skiĭ–Mann y el algoritmo de gradiente estocástico. Finalmente, una versión estocástica abstracta del método de Haugazeau es propuesta para computar la mejor aproximación desde un conjunto convexo cerrado por proyecciones sucesivas sobre aproximaciones exteriores estocásticas generadas aleatoriamente de ese conjunto. A lo largo de la disertación, experimentos numéricos son proporcionados para ilustrar los hallazgos teóricos.

ACKNOWLEDGMENTS

I am deeply grateful to Professor Patrick L. Combettes for his exceptional guidance during my doctoral studies. Beyond his academic expertise, I truly appreciate his patience and constant encouragement. I hope this final work reflects the high standard of his mentorship. Working alongside him has been a genuine privilege.

I thank Prof. Bates, Prof. Nguyen, Prof. Papanicolaou, and Prof. Shashaani for agreeing to serve on my advisory committee.

To my colleagues in Raleigh, Diego, Julien, Cristóbal, and every member of the nonlinear analysis group, thanks for the time we shared and the conversations we had along the way.

To my colleagues around the world, thank you for the time we've spent together online and for brightening my late nights working from home. A special thank you to my fellow Ph.D. candidates: P. Araya, M. Azocar, B. Barrientos, V. Saavedra, V. Salinas, L. Torrealba, F. Vásquez, and N. Valenzuela.

Raleigh became a home to us because of the many people who welcomed us as their own. I am deeply grateful to the Andales, Moise, Pahuamba, Ramos, Safley, and Shamblee families, as well as the entire RCC community. Special thanks to Pop for his kindness and for making us feel at home in his house.

I thank my lovely family for the support they have shown us throughout this journey, even from afar. I hope we can spend more time together soon.

I do not have enough words to thank my beloved Catalina, who sacrificed so much to be here with me. I look forward to spending the rest of my life making it up to you.

To the One who began a good work in me, and who is faithful to bring it to completion.

This dissertation was supported by the U.S. National Science Foundation grants CCF-2211123 and DMS-2513409 (PI: P. L. Combettes).

TABLE OF CONTENTS

LIST OF FIGURES

NOTATION AND DEFINITIONS

The following notation is used throughout this dissertation.

General notation

- $(\Omega, \mathcal{F}, \mathsf{P})$: Complete probability space.
- $(\Xi, \mathcal{G})$, $(\mathsf{K}, \mathcal{K})$: Measurable spaces.
- $\mathcal{X}$, $\mathcal{X}_n$: Sub σ-algebras of $\mathcal{F}$.
- α, β, x, y, T (sans-serif letters): Deterministic variables.
- α, β, x, y, T (italicized serif letters): Random variables.
- H, H_i, G, G_k: Separable real Hilbert spaces.
- $\langle \cdot \mid \cdot \rangle$: Scalar product of a real Hilbert space.
- $\|\cdot\|$: Norm.
- $\bigoplus_{i\in I} \mathsf{H}_i$: Hilbert direct sum of a finite family $(\mathsf{H}_i)_{i\in I}$ of real Hilbert spaces, that is,

$$\bigoplus_{i\in I} \mathsf{H}_i = \left\{ \mathsf{x} = (\mathsf{x}_i)_{i\in I} \;\middle|\; (\forall i \in I) \;\; \mathsf{x}_i \in \mathsf{H}_i \right\}$$

  equipped with the scalar product $(\mathsf{x}, \mathsf{y}) \mapsto \sum_{i\in I} \langle \mathsf{x}_i \mid \mathsf{y}_i \rangle_{\mathsf{H}_i}$.
- Id: Identity operator.
- 2^{H}: Power set of H.
- L^*: Adjoint of a bounded linear operator $\mathsf{L} \colon \mathsf{H} \to \mathsf{G}$.
- $\to$: Strong convergence.
- $\rightharpoonup$: Weak convergence.
- $\|\mathsf{L}\| = \sup\big\{ \|\mathsf{L}\mathsf{x}\| \mid \mathsf{x} \in \mathsf{H},\ \|\mathsf{x}\| \leqslant 1 \big\}$: Norm of a bounded linear operator $\mathsf{L} \colon \mathsf{H} \to \mathsf{G}$.

Notation and definitions relative to a function $\mathsf{f} \colon \mathsf{H} \to [-\infty, +\infty]$

- $\operatorname{dom} \mathsf{f} = \big\{ \mathsf{x} \in \mathsf{H} \mid \mathsf{f}(\mathsf{x}) < +\infty \big\}$: Domain of f.
- $\operatorname{epi} \mathsf{f} = \big\{ (\mathsf{x}, \xi) \in \mathsf{H} \times \mathbb{R} \mid \mathsf{f}(\mathsf{x}) \leqslant \xi \big\}$: Epigraph of f.
- f is proper if $-\infty \notin \mathsf{f}(\mathsf{H})$ and $\operatorname{dom} \mathsf{f} \neq \varnothing$.
- Suppose that f is proper. Then $\operatorname{Argmin} \mathsf{f} = \big\{ \mathsf{x} \in \mathsf{H} \mid \mathsf{f}(\mathsf{x}) = \inf \mathsf{f}(\mathsf{H}) \big\}$ is the set of minimizers of f over H.
- f is convex if $\operatorname{epi} \mathsf{f}$ is a convex subset of $\mathsf{H} \oplus \mathbb{R}$.
- f is lower semicontinuous if $\operatorname{epi} \mathsf{f}$ is a closed subset of $\mathsf{H} \oplus \mathbb{R}$.
- $\Gamma_0(\mathsf{H})$: Set of proper lower semicontinuous convex functions from H to $]-\infty, +\infty]$.

- Conjugate of f: $f^*\colon \mathcal{H} \to [-\infty, +\infty]\colon x^* \mapsto \sup_{x\in\mathcal{H}}(\langle x \mid x^*\rangle - f(x))$:
- Suppose that f is proper. Then

$$\partial f\colon \mathcal{H} \to 2^{\mathcal{H}}\colon x \mapsto \big\{x^* \in \mathcal{H} \mid (\forall y \in \mathcal{H})\ \ \langle y - x \mid x^*\rangle + f(x) \leqslant f(y)\big\} \tag{1}$$

  is the subdifferential of f.
- Suppose that $f \in \Gamma_0(\mathcal{H})$. Then, for every $x \in \mathcal{H}$, $\operatorname{prox}_f x$ denotes the unique minimizer of the function $f + (1/2)\|\cdot - x\|^2$. The proximity operator of f is $\operatorname{prox}_f\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto \operatorname{prox}_f x$.

Notation and definitions relative to a subset C of $\mathcal{H}$

- ι_C: Indicator function of C, that is,

$$\iota_C\colon \mathcal{H} \to [0, +\infty]\colon x \mapsto \begin{cases} 0, & \text{if } x \in C;\\ +\infty, & \text{if } x \notin C. \end{cases} \tag{2}$$

- $\operatorname{int} C$: Interior of C.
- $\overline{C}$: Closure of C.
- $\operatorname{sri} C$: Strong relative interior of C, that is,

$$\operatorname{sri} C = \Bigg\{x \in C \;\Bigg|\; \bigcup_{\lambda\in]0,+\infty[} \lambda(C - x) \text{ is a closed vector subspace of } \mathcal{H}\Bigg\}. \tag{3}$$

- $\operatorname{ri} C$: Relative interior of C, that is,

$$\operatorname{ri} C = \Bigg\{x \in C \;\Bigg|\; \bigcup_{\lambda\in]0,+\infty[} \lambda(C - x) \text{ is a vector subspace of } \mathcal{H}\Bigg\}. \tag{4}$$

- $d_C\colon \mathcal{H} \to [0, +\infty]\colon x \mapsto \inf\|C - x\|$: Distance function to C.
- Suppose that C is nonempty, closed, and convex. Then $\operatorname{proj}_C = \operatorname{prox}_{\iota_C}$.

Notation and definitions relative to an operator $T\colon \mathcal{H} \to \mathcal{H}$

- $\operatorname{Fix} T = \big\{x \in \mathcal{H} \mid Tx = x\big\}$: Set of fixed points of T.
- T is cocoercive with constant $\beta \in]0, +\infty[$ if

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \langle x - y \mid Tx - Ty\rangle \geqslant \beta\|Tx - Ty\|^2. \tag{5}$$

- T is Lipschitzian with constant $\beta \in [0, +\infty[$ if

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \|Tx - Ty\| \leqslant \beta\|x - y\|. \tag{6}$$

- T is nonexpansive if it is Lipschitzian with constant 1.
- T is firmly quasinonexpansive if

$$(\forall \mathsf{x} \in \mathsf{H})(\forall \mathsf{y} \in \operatorname{Fix} \mathsf{T}) \quad \|\mathsf{T}\mathsf{x} - \mathsf{y}\|^2 + \|\mathsf{T}\mathsf{x} - \mathsf{x}\|^2 \leqslant \|\mathsf{x} - \mathsf{y}\|^2. \tag{7}$$

Notation and definitions relative to a set-valued operator $\mathsf{M} \colon \mathsf{H} \to 2^{\mathsf{H}}$

- $\operatorname{dom} \mathsf{M} = \big\{\mathsf{x} \in \mathsf{H} \mid \mathsf{M}\mathsf{x} \neq \varnothing\big\}$: Domain of M.
- $\operatorname{ran} \mathsf{M} = \bigcup_{\mathsf{x} \in \mathsf{H}} \mathsf{M}\mathsf{x}$: Range of M.
- $\operatorname{zer} \mathsf{M} = \big\{\mathsf{x} \in \mathsf{H} \mid 0 \in \mathsf{M}\mathsf{x}\big\}$: Set of zeros of M.
- $\operatorname{gra} \mathsf{M} = \big\{(\mathsf{x}, \mathsf{x}^*) \in \mathsf{H} \times \mathsf{H} \mid \mathsf{x}^* \in \mathsf{M}\mathsf{x}\big\}$: Graph of M.
- M^{-1}: Inverse of M, that is,

$$\mathsf{M}^{-1} \colon \mathsf{H} \to 2^{\mathsf{H}} \colon \mathsf{x}^* \mapsto \big\{\mathsf{x} \in \mathsf{H} \mid \mathsf{x}^* \in \mathsf{M}\mathsf{x}\big\}. \tag{8}$$

- $\mathsf{J}_{\mathsf{M}} = (\mathrm{Id} + \mathsf{M})^{-1}$: Resolvent of M.
- M is monotone if

$$\big(\forall (\mathsf{x}, \mathsf{x}^*) \in \operatorname{gra} \mathsf{M}\big)\big(\forall (\mathsf{y}, \mathsf{y}^*) \in \operatorname{gra} \mathsf{M}\big) \quad \langle \mathsf{x} - \mathsf{y} \mid \mathsf{x}^* - \mathsf{y}^* \rangle \geqslant 0. \tag{9}$$

- M is maximally monotone if M is monotone and, for every monotone operator $\mathsf{M}' \colon \mathsf{H} \to 2^{\mathsf{H}}$, $\operatorname{gra} \mathsf{M} \subset \operatorname{gra} \mathsf{M}' \Rightarrow \mathsf{M} = \mathsf{M}'$.

Notation and definitions relative to a random variable $\xi \colon \Omega \to \mathbb{R}$

- $\mathsf{E}\xi = \displaystyle\int_{\Omega} \xi(\omega)\mathsf{P}(d\omega)$: Expectation of ξ.
- $\mathsf{E}(\xi \mid \mathcal{G})$: Conditional expectation of ξ given $\mathcal{G}$.
- $\sigma(\{\xi_i\}_{i \in I})$: The σ-algebra generated by the family of random variables $\{\xi_i\}_{i \in I}$.
- $(\forall \mathsf{p} \in [1, +\infty[)$ $L^{\mathsf{p}}(\Omega, \mathcal{F}, \mathsf{P}; \mathbb{R})$: Space of equivalence classes of P-a.s. equal random variables ξ such that $\mathsf{E}|\xi|^{\mathsf{p}} < +\infty$.
- $L^{\infty}(\Omega, \mathcal{F}, \mathsf{P}; \mathbb{R})$: Space of equivalence classes of P-a.s. equal random variables ξ such that $\inf\big\{\eta \in [0, +\infty[\mid |\xi| \leqslant \eta \ \mathsf{P}\text{-a.s.}\big\} < +\infty$.

AUTHORSHIP STATEMENT

Chapter 1

Javier I. Madariaga: sole author of Chapter 1.

Chapter 2

First Published in: P. L. Combettes and J. I. Madariaga, Almost-surely convergent randomly activated monotone operator splitting methods, *SIAM Journal on Imaging Sciences*, vol. 18, pp. 2177–2205, 2025, published by the Society for Industrial and Applied Mathematics (SIAM).

Chapter 3

P. L. Combettes and J. I. Madariaga, A geometric framework for stochastic iterations, *Mathematics of Computation*, to appear. First published in Mathematics of Computation (to appear), published by American Mathematical Society.

Chapter 4

P. L. Combettes and J. I. Madariaga, Asymptotic analysis of an abstract stochastic scheme for solving monotone inclusions, submitted.

Chapter 5

J. I. Madariaga, Convergence of the iterates of the stochastic proximal gradient method, submitted.

Chapter 6

J. I. Madariaga, An abstract stochastic Haugazeau method for best approximation, submitted.

Chapter 7

Javier I. Madariaga: sole author of Chapter 7.

Use of generative artificial intelligence: No generative artificial intelligence was used to write this dissertation.

Chapter 1

INTRODUCTION

1.1 Overview

The objective of this dissertation is to propose general algorithmic frameworks and convergence principles for dealing with stochasticity in a broad class of algorithms arising in optimization and nonlinear analysis. Throughout, H is a separable real Hilbert space with scalar product $\langle\cdot\mid\cdot\rangle$ and induced norm $\|\cdot\|$, and the underlying probability space $(\Omega, \mathcal{F}, \mathsf{P})$ is complete. We denote by $\mathsf{Z} \subset \mathsf{H}$ the set of solutions to the problem of interest and assume that it is nonempty, closed, and convex. This setting covers a broad range of models in nonlinear analysis, including the following ones in which the solution set Z is assumed to be nonempty.

- Let $\mathsf{f}\colon \mathsf{H} \to \left]-\infty, +\infty\right]$ be a lower semicontinuous convex function. The problem is to minimize f, where $\mathsf{Z} = \operatorname{Argmin} \mathsf{f}$.
- Let $\mathsf{T}\colon \mathsf{H} \to \mathsf{H}$ be a nonexpansive operator, i.e., 1-Lipschitzian. The problem is to find a fixed point of T, i.e., a point in $\mathsf{Z} = \operatorname{Fix} \mathsf{T} = \big\{\mathsf{z} \in \mathsf{H} \mid \mathsf{T}\mathsf{z} = \mathsf{z}\big\}$.
- Let $\mathsf{M}\colon \mathsf{H} \to 2^{\mathsf{H}}$ be a maximally monotone operator, i.e.,

$$\big(\forall(\mathsf{x},\mathsf{x}^*) \in \mathsf{H}\times\mathsf{H}\big) \quad \big[\, (\mathsf{x},\mathsf{x}^*) \in \operatorname{gra}\mathsf{M} \;\Leftrightarrow\; \big(\forall(\mathsf{y},\mathsf{y}^*) \in \operatorname{gra}\mathsf{M}\big)\;\; \langle \mathsf{x}-\mathsf{y} \mid \mathsf{x}^*-\mathsf{y}^*\rangle \geqslant 0 \,\big]. \tag{1.1}$$

  The problem is to find a zero of M, i.e., to find a point in $\mathsf{Z} = \operatorname{zer}\mathsf{M} = \big\{\mathsf{z} \in \mathsf{H} \mid 0 \in \mathsf{M}\mathsf{z}\big\}$.
- Let $(\mathsf{Z}_k)_{1\leqslant k\leqslant p}$ be a finite collection of closed convex subsets of H. The problem is to find a common point of the collection, where the solution set is $\mathsf{Z} = \bigcap_{1\leqslant k\leqslant p} \mathsf{Z}_k$.
- Let $\mathsf{B}\colon \mathsf{H} \to \mathsf{H}$ be a maximal monotone operator and let $\varphi\colon \mathsf{H} \to \left]-\infty, +\infty\right]$ be a lower semicontinuous convex function. The variational inequality problem consists in finding a point in

$$\mathsf{Z} = \Big\{\mathsf{z} \in \mathsf{H} \;\Big|\; (\forall \mathsf{x} \in \mathsf{H})\;\; \big\langle \mathsf{z}-\mathsf{x} \;\big|\; \mathsf{B}\mathsf{z}\big\rangle + \varphi(\mathsf{z}) \leqslant \varphi(\mathsf{x})\Big\}. \tag{1.2}$$

These examples are prominent in equilibria across diverse fields such as mechanics [46,58], dynamical systems [1], domain decomposition methods [3,6,7], machine learning [5,33,50,71], partial differential equations [16, 23, 49, 67], signal processing [35, 36, 52], image processing [22, 24, 47, 63] data science [32, 45], neural networks [12, 31, 72, 73], and optimization [4, 8, 9, 13, 28, 53]. A central notion for finding a point in Z is Fejér monotonicity: A sequence $(x_n)_{n\in\mathbb{N}}$ in H is Fejér monotone with respect to Z if

$$(\forall z \in Z)(\forall n \in \mathbb{N}) \quad \|x_{n+1} - z\| \leqslant \|x_n - z\|. \tag{1.3}$$

Provided that every weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ belongs to Z, the sequence converges weakly to a point in Z [11, Theorem 5.5]. This fact has been used to show the weak convergence of a wide range of deterministic algorithms in nonlinear analysis [26, 27].

When studying algorithms for solving nonlinear problems, stochasticity can arise at different levels.

- The problem itself is defined using probability spaces [14, 15, 19, 48, 54, 61].
- The activation of randomly selected blocks of operators [38, 51, 60, 62, 68].
- The randomized activation and updating of coordinate blocks [17, 29, 30, 59].
- The use of stochastic approximations for the underlying operators [20, 40, 43, 55, 64, 65].

State-of-the-art stochastic methods address only specific aspects of stochasticity, and there is no unified analysis for them. Moreover, their results for asymptotic behavior remain unstandardized. Indeed, a significant volume of research focuses exclusively on weaker forms of convergence, such as ergodic convergence or convergence of the values of the objective function in the context of minimization. The few studies concerning iterates convergence focus on almost sure convergence and L^p convergence, mostly in finite dimensional settings.

Our goal is to analyze stochasticity in iterative algorithms in full generality. Furthermore, we aim at exploring conditions that guarantee almost sure and L^p convergence of sequences constructed by stochastic methods. To achieve this, we leverage concepts analogous to those used in deterministic methods by applying (1.3). Unlike their deterministic counterparts, however, stochastic iterations are not generally Fejér monotone in an almost sure sense. In practice, random updates introduce perturbations that modify the trajectory of the algorithm. For example, the stochastic gradient descent method is not a "descent" method in the strict sense. Given this, the generated sequences often move away from the target set of solutions Z during individual steps. Thus, the condition of (1.3) is too restrictive for the analysis of these methods. To address this issue, the concept of stochastic quasi-Fejér monotonicity was introduced to analyze stochastic iterations, first in the late 1960s for Euclidean spaces in the works of Ermol'ev [41, 42, 44], and later extended to Hilbert spaces in [29, 30]. A sequence $(x_n)_{n\in\mathbb{N}}$ of H-valued random variables is stochastic quasi-Fejér monotone with respect to Z if

$$(\forall z \in Z)(\forall n \in \mathbb{N}) \quad \mathsf{E}\big(\|x_{n+1} - z\|^2 \,\big|\, \sigma(x_0, \ldots, x_n)\big) \leqslant \|x_n - z\|^2 + \eta_n(z) \ \ \mathsf{P}\text{-a.s.}, \tag{1.4}$$

where, for every $z \in Z$, $(\eta_n(z))_{n\in\mathbb{N}}$ is a sequence of $[0, +\infty[$-valued random variables satisfying $\sum_{n\in\mathbb{N}} \eta_n(z) < +\infty$ P-a.s. Conceptually, this means that for any solution $z \in Z$ and any iteration $n \in \mathbb{N}$, the update x_{n+1} is conditionally expected to be closer to every solution z than x_n, up to some stochastic error $\eta_n(z)$. If, for almost every $\omega \in \Omega$, the weak cluster points of $(x_n(\omega))_{n\in\mathbb{N}}$ belong to Z, then $(x_n)_{n\in\mathbb{N}}$ converges weakly P-a.s. to a Z-valued random variable [29, Proposition 2.3(iv)].

The analysis of stochastic iterative methods has seen significant development in the last decades. Yet, many important open questions remain unanswered. We list below the ones that will be addressed in this dissertation:

(Q1) Many stochastic splitting algorithms have been proposed in the literature for solving large-scale composite inclusion problems involving monotone and linear operators [2, 21, 39, 62, 69]. To establish convergence, these methods rely on restrictive assumptions, such as finite-dimensional spaces, strong monotonicity or cocoercivity of the operators, or knowledge of a bound on the norm of the linear operators $(L_k)_{1\leqslant k\leqslant p}$, among many others. Further, just a few of them guarantee weak P-a.s. convergence of the iterates. Can we develop stochastic splitting methods for solving the composite inclusion problem by avoiding the extra assumptions and that guarantee weak P-a.s. convergence of the constructed sequence $(x_n)_{n\in\mathbb{N}}$?

(Q2) A notable fact about Fejér monotonicity is that every Fejér monotone sequence $(x_n)_{n\in\mathbb{N}}$ can be constructed by an iterative process in which, at every iteration $n \in \mathbb{N}$, x_{n+1} is a relaxed projection of x_n onto a half-space H_n of H that contains Z [10, Proposition 2.7]. This fact is important for the design and the analysis of iterative methods for finding a point in Z. (see, e.g., [27, Section 4.2]) and leaves the following question: Does a geometric framework exist that covers the construction of stochastic quasi-Fejér monotone sequences and explains their asymptotic behavior?

(Q3) In [27, Section 4.4], a monotone inclusion problem is studied that considers the sum of two operators $W + C$, where $W\colon H \to 2^H$ is maximally monotone and $C\colon H \to H$ is cocoercive. An abstract framework to solve it is provided, which covers classical algorithms such as the proximal point algorithm [56, 66], the forward-backward-forward algorithm [70], the unrelaxed forward-backward algorithm [57], among others. Additionally, in [18], this framework is explored in the context of saddle block-iterative projective splitting methods. Can this framework be extended to the context of stochastic iterations?

(Q4) The problem of minimizing the sum of two convex functions when one of them is smooth, has been extensively studied via the proximal-gradient method; see [33, 34, 37]. In modern applications, such as those in machine learning, a common approach is to express those functions as the sum (or integral in the continuous case) of a collection of functions. These individual components are simpler in the sense that their proximity operator and gradient (for the smooth function) are known and easy to evaluate. Can we provide suitable assumptions that guarantee the convergence of the proximal-gradient method when the functions are selected randomly?

(Q5) Haugazeau's scheme has been used to solve the best approximation problem from closed convex sets in Hilbert spaces. Under mild conditions, one can construct a sequence that converges strongly to the solution [10, 25]. However, this deterministic scheme does not allow for the incorporation of randomness as in stochastic block-coordinate splitting iterations. Can we construct a version of Haugazeau's scheme to solve the best approximation problem through stochastic iterations?

1.2 Contributions and organization

The main contributions of the thesis are outlined below.

- In Chapter 2, we answer **(Q1)** positively by introducing three different algorithmic frameworks for solving large-scale composite inclusion problems involving monotone and linear operators in Hilbert spaces. Each framework guarantees weak almost sure convergence to a solution without extra assumptions.
- We address **(Q2)** in Chapter 3, where we propose a geometric framework to construct stochastic quasi-Fejér monotone sequences. This framework recovers the deterministic case and provides a template to create new stochastic methods. We establish conditions to achieve almost-sure convergence, convergence in expectation, and convergence in mean-square. We apply this framework to derive novel results in the context of the stochastic gradient descent method in Hilbert spaces, stochastic extrapolated parallel algorithms for feasibility problems, among others.
- In Chapter 4, we focus on question **(Q3)** and specialize the geometric framework to develop a stochastic scheme for solving a broad range of systems of composite inclusion problems. The proposed scheme uses, at each iteration, stochastic approximations to points in the graph of the operators to form the update. We apply the scheme to derive the convergence of versions of the proximal point algorithm as well as randomized block-iterative projective splitting methods.
- Question **(Q4)** is addressed in Chapter 5. The convergence of a sequence generated by the stochastic proximal gradient method is established under suitable assumptions on the subgradients. Unlike state-of-the-art methods, this result requires neither bounded variance of the random variables nor uniform boundedness of the subgradients.
- Chapter 6 is devoted to **(Q5)**. We introduce an abstract stochastic Haugazeau method to compute the best approximation to a point from a closed convex set in a Hilbert space. We also provide conditions to derive strong convergence in both the mean square and the almost sure senses. We apply the results to design stochastic extrapolated parallel algorithms for computing the best approximation from an arbitrary intersection of sets.
- We conclude the dissertation in Chapter 7 and we provide future research directions.

1.3 Publications

This work has produced the following two conference articles and five journal articles:

1. P. L. Combettes and J. I. Madariaga, Randomly activated proximal methods for nonsmooth convex minimization, *Proceedings of the European Signal Processing Conference*, pp. 2642–2646. Lyon, France, August 26–30, 2024.
2. P. L. Combettes and J. I. Madariaga, Stochastic block-iterative parallel subgradient projections method with super relaxations, *Proceedings of the European Signal Processing Conference*, pp. 2757–2761. Palermo, Italy, September 7–12, 2025.
3. P. L. Combettes and J. I. Madariaga, Almost-surely convergent randomly activated monotone operator splitting methods, *SIAM Journal on Imaging Sciences*, vol. 18, pp. 2177–2205, 2025.
4. P. L. Combettes and J. I. Madariaga, A geometric framework for stochastic iterations, *Mathematics of Computation*, to appear.
5. P. L. Combettes and J. I. Madariaga, Asymptotic analysis of an abstract stochastic scheme for solving monotone inclusions, submitted.
6. J. I. Madariaga, An abstract stochastic Haugazeau method for best approximation, submitted.
7. J. I. Madariaga, Convergence of the iterates of the stochastic proximal gradient method, submitted.

References

[1] S. Adly, A. Hantoute, and B. K. Le, Maximal monotonicity and cyclic monotonicity arising in nonsmooth Lur'e dynamical systems, *J. Math. Anal. Appl.*, vol. 448, pp. 691–706, 2017.

[2] A. Alacaoglu, O. Fercoq, and V. Cevher, On the convergence of stochastic primal-dual hybrid gradient, *SIAM J. Optim.*, vol. 32, pp. 1288–1318, 2022.

[3] G. Alduncin, Multidomain optimal control of variational subpotential mixed evolution inclusions, *Appl. Math. Optim.*, vol. 88, art. 35, 2023.

[4] H. C. Andrews and B. R. Hunt, *Digital Image Restoration*. Prentice-Hall, Englewood Cliffs, NJ, 1977.

[5] A. Argyriou, R. Foygel, and N. Srebro, Sparse prediction with the k-support norm, *Proc. Adv. Neural Inform. Process. Syst. Conf.*, vol. 25, pp. 1457–1465, 2012.

[6] H. Attouch, L. M. Briceño–Arias, and P. L. Combettes, A strongly convergent primal-dual method for nonoverlapping domain decomposition, *Numer. Math.*, vol. 133, pp. 433–470, 2016.

[7] H. Attouch, A. Cabot, P. Frankel, and J. Peypouquet, Alternating proximal algorithms for linearly constrained variational inequalities: application to domain decomposition for PDE's, *Nonlinear Anal.*, vol. 74, pp. 7455–7473, 2011.

[8] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, Optimization with sparsity-inducing penalties, *Found. Trends Machine Learn.*, vol. 4, pp. 1–106, 2012.

[9] S. Banert, A. Ringh, J. Adler, J. Karlsson, and O. Öktem, Data-driven nonsmooth optimization, *SIAM J. Optim.*, vol. 30, pp. 102–131, 2020.

[10] H. H. Bauschke and P. L. Combettes, A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces, *Math. Oper. Res.*, vol. 26, pp. 248–264, 2001.

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

[12] Y. Belkouchi, J.-C. Pesquet, A. Repetti, and G. Talbot, Learning truly monotone operators with applications to nonlinear inverse problems, *SIAM J. Imaging Sci.*, vol. 18, pp. 735–764, 2025.

[13] M. Benning and M. Burger, Modern regularization methods for inverse problems, *Acta Numer.*, vol. 27, pp. 1–111, 2018.

[14] P. Bianchi, Ergodic convergence of a stochastic proximal point algorithm, *SIAM J. Optim.*, vol. 26, pp. 2235–2260, 2016.

[15] J. R. Blum, Multidimensional stochastic approximation methods, *Ann. Math. Stat.*, vol. 25, pp. 737–744, 1954.

[16] H. Brézis and F. Browder, Partial differential equations in the 20th century, *Adv. Math.*, vol. 135, pp. 76–144, 1998.

[17] L. M. Briceño–Arias, G. Chierchia, E. Chouzenoux, and J.-C. Pesquet, A random block-coordinate Douglas–Rachford splitting method with low computational complexity for binary logistic regression, *Comput. Optim. Appl.*, vol. 72, pp. 707–726, 2019.

[18] M. N. Bùi and P. L. Combettes, Multivariate monotone inclusions in saddle form, *Math. Oper. Res.*, vol. 47, pp. 1082–1109, 2022.

[19] D. Butnariu and S. D. Flåm, Strong convergence of expected-projection methods in Hilbert spaces, *Numer. Funct. Anal. Optim.*, vol. 16, pp. 601–636, 1995.

[20] V. Cevher, B. C. Vũ, and A. Yurtsever, Stochastic forward Douglas–Rachford splitting method for monotone inclusions, in: P. Giselsson and A. Rantzer (eds.), *Large-Scale and Distributed Optimization*, pp. 149–179. Springer, Berlin, 2018.

[21] A. Chambolle, C. Delplancke, M. J. Ehrhardt, C.-B. Schönlieb, and J. Tang, Stochastic primal-dual hybrid gradient algorithm with adaptive step sizes, *J. Math. Imaging Vision*, vol. 66, pp. 294–313, 2024.

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

[23] C. Clason and T. Valkonen, Primal-dual extragradient methods for nonlinear nonsmooth PDE-constrained optimization, *SIAM J. Optim.*, vol. 27, pp. 1314–1339, 2017.

[24] P. L. Combettes, The convex feasibility problem in image recovery, in: *Advances in Imaging and Electron Physics* (P. Hawkes, Ed.), vol. 95, pp. 155–270. Academic Press, New York, 1996.

[25] P. L. Combettes, Strong convergence of block-iterative outer approximation methods for convex optimization, *SIAM J. Control Optim.*, vol. 38, pp. 538–565, 2000.

[26] P. L. Combettes, Fejér-monotonicity in convex optimization, in: *Encyclopedia of Optimization*, (C. A. Floudas and P. M. Pardalos, Eds.), vol. 2, Springer-Verlag, New York, 2001, pp. 106–114. (Also available in 2nd ed., pp. 1016–1024, 2009.)

[27] P. L. Combettes, The geometry of monotone operator splitting methods, *Acta Numer.*, vol. 33, pp. 487–632, 2024.

[28] P. L. Combettes and J.-C. Pesquet, Proximal splitting methods in signal processing, in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pp. 185–212. Springer, New York, 2011.

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

[30] P. L. Combettes and J.-C. Pesquet, Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping II: Mean-square and linear convergence, *Math. Program.*, vol. B174, pp. 433–451, 2019.

[31] P. L. Combettes and J.-C. Pesquet, Deep neural network structures solving variational inequalities, *Set-Valued Var. Anal.*, vol. 28, pp. 491–518, 2020.

[32] P. L. Combettes and J.-C. Pesquet, Fixed point strategies in data science, *IEEE Trans. Signal Process.*, vol. 69, pp. 3878–3905, 2021.

[33] P. L. Combettes, S. Salzo, and S. Villa, Consistent learning by composite proximal thresholding, *Math. Program.*, vol. B167, pp. 99–127, 2018.

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

[35] P. L. Combettes and Z. C. Woodstock, Reconstruction of functions from prescribed proximal points, *J. Approx. Theory*, vol. 268, art. 105606, 2021.

[36] P. L. Combettes and Z. C. Woodstock, A variational inequality model for the construction of signals from inconsistent nonlinear equations, *SIAM J. Imaging Sci.*, vol. 15, pp. 84–109, 2022.

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

[38] D. Davis, Variance reduction for root-finding problems, *Math. Program.*, vol. A197, pp. 375–410, 2023.

[39] A. Dieuleveut, G. Fort, E. Moulines, and H.-T. Wai, Stochastic approximation beyond gradient for signal processing and machine learning, *IEEE Trans. Signal Process.*, vol. 71, pp. 3117–3148, 2023.

[40] M. Duflo, *Méthodes Récursives Aléatoires.* Masson, Paris, 1990. English translation: *Random Iterative Models.* Springer, New York, 1997.

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

[42] Yu. M. Ermol'ev, On convergence of random quasi-Fejér sequences, *Cybernetics*, vol. 7, pp. 655–656, 1971.

[43] Yu. M. Ermol'ev and Z. V. Nekrylova, Some methods of stochastic optimization, *Kibernetika (Kiev)*, vol. 1966, pp. 96–98, 1966.

[44] Yu. M. Ermol'ev and A. D. Tuniev, Random Fejér and quasi-Fejér sequences, *Theory of Optimal Solutions – Akad. Nauk Ukrainskoĭ SSR Kiev*, vol. 2, pp. 76–83, 1968.

[45] F. Florez-Ospina, D. A. Jimenez-Sierra, H. D. Benitez-Restrepo, and G. R. Arce, Exploiting variational inequalities for generalized change detection on graphs, *IEEE Trans. Geosci. Remote Sensing*, vol. 61, pp. 1–16, 2023.

[46] M. Fortin and R. Glowinski (eds.), *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems.* North-Holland, Amsterdam, 1983.

[47] R. Glowinski, S. J. Osher, and W. Yin (eds.), *Splitting Methods in Communication, Imaging, Science, and Engineering*. Springer, New York, 2016.

[48] C. Geiersbach and G. C. Pflug, Projected stochastic gradients for convex constrained problems in Hilbert spaces, *SIAM J. Optim.*, vol. 29, pp. 2079–2099, 2019

[49] N. Ghoussoub, *Self-Dual Partial Differential Systems and Their Variational Principles*. Springer, New York, 2009.

[50] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach, Proximal methods for hierarchical sparse coding, *J. Machine Learn. Res.*, vol. 12, pp. 2297–2334, 2011.

[51] P. R. Johnstone, J. Eckstein, T. Flynn, and S. Yoo, Stochastic projective splitting, *Comput. Optim. Appl.*, vol. 87, pp. 397–437, 2024.

[52] A. B. Juditsky and A. S. Nemirovski, Signal recovery by stochastic optimization, *Autom. Remote Control*, vol. 80, pp. 1878–1893, 2019.

[53] A. B. Juditsky and A. S. Nemirovski, *Statistical Inference via Convex Optimization*. Princeton University Press, Princeton, NJ, 2020.

[54] J. Kiefer and J. Wolfowitz, Stochastic estimation of the maximum of a regression function, *Ann. Math. Stat.*, vol. 23, pp. 462–466, 1952.

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

[56] J. Lieutaud, Approximations d'opérateurs monotones par des méthodes de splitting, in: A. Ghizzetti (ed.), *Theory and Applications of Monotone Operators*, Proceedings of a NATO Advanced Study Institute held in Venice, Italy, June 17–30, 1968. Edizioni Oderisi, Gubbio, 1969, pp. 259–264.

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

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

[59] Yu. Nesterov, Efficiency of coordinate descent methods on huge-scale optimization problems, *SIAM J. Optim.*, vol. 22, pp. 314–362, 2012.

[60] V. D. Nguyen and B. C. Vũ, Convergence analysis of the stochastic reflected forward-backward splitting algorithm, *Optim. Lett.*, vol. 16, pp. 2649–2679, 2022.

[61] N. Pagliana and L. Rosasco, Implicit regularization of accelerated methods in Hilbert spaces, in: H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, vol. 32, pp. 14481–14491, 2019.

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

[63] J.-C. Pesquet, A. Repetti, M. Terris, and Y. Wiaux, Learning maximally monotone operators for image recovery, *SIAM J. Imaging Sci.*, vol. 14, pp. 1206–1237, 2021.

[64] B. T. Polyak and A. B. Juditsky, Acceleration of stochastic approximation by averaging, *SIAM J. Control Optim.*, vol. 30, pp. 838–855, 1992.

[65] H. Robbins and S. Monro, A stochastic approximation method, *Ann. Math. Statist.*, vol. 22, pp. 400–407, 1951.

[66] R. T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. Control Optim.*, vol. 14, pp. 877–898, 1976.

[67] R. E. Showalter, *Monotone Operators in Banach Space and Nonlinear Partial Differential Equations.* Amer. Math. Soc., Providence, RI, 1997.

[68] T. Strohmer and R. Vershynin, A randomized Kaczmarz algorithm with exponential convergence, *J. Fourier Anal. Appl.*, vol. 15, pp. 262, 2009

[69] C. Traoré, V. Apidopoulos, S. Salzo, and S. Villa, Variance reduction techniques for stochastic proximal point algorithms, *J. Optim. Theory Appl.*, vol. 203, pp. 1910–1939, 2024.

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

[71] S. Vaiter, G. Peyré, and J. Fadili, Model consistency of partly smooth regularizers, *IEEE Trans. Inform. Theory*, vol. 64, pp. 1725–1737, 2018.

[72] E. Winston and J. Z. Kolter, Monotone operator equilibrium networks, *Proc. Adv. Neural Inform. Process. Syst.*, vol. 22, pp. 10718–10728, 2020.

[73] P. Yi and S. Ching, Synthesis of recurrent neural dynamics for monotone inclusion with application to Bayesian inference, *Neural Networks*, vol. 131, pp. 231–241, 2020.

Chapter 2

ALMOST-SURELY CONVERGENT RANDOMLY ACTIVATED MONOTONE OPERATOR SPLITTING METHODS

2.1 Introduction and context

We address question **(Q1)** of Chapter 1 by introducing three algorithmic frameworks for solving large-scale composite inclusion problems involving monotone and linear operators.

This chapter presents the following journal article:

> P. L. Combettes and J. I. Madariaga, Almost-surely convergent randomly activated monotone operator splitting methods, *SIAM Journal on Imaging Sciences*, vol. 18, pp. 2177–2205, 2025.

2.2 Article: Almost-surely convergent randomly activated monotone operator splitting methods

Abstract. We propose stochastic splitting algorithms for solving large-scale composite inclusion problems involving monotone and linear operators. They activate at each iteration blocks of randomly selected resolvents of monotone operators and, unlike existing methods, achieve almost sure convergence of the iterates to a solution without any regularity assumptions or knowledge of the norms of the linear operators. Applications to image recovery and machine learning are provided.

2.2.1 Introduction

The problem of extracting information from data is at the core of many tasks in signal processing, inverse problems, and machine learning. A prevalent methodology to seek meaningful solutions is to build a mathematical model that incorporates the prior knowledge about the object of interest $\overline{x}$ and the data, which consist of observations mathematically or physically related to $\overline{x}$ (see Figure 2.1). Since the first mathematical formalizations of Euler [23] and Mayer [31] in the

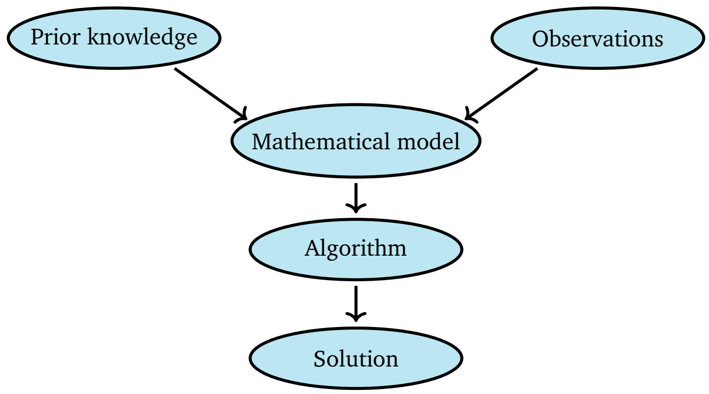


Figure 2.1 Data processing flowchart.

late 1740s, which contained the embryo of least-squares data fitting techniques, convex minimization formulations have been a tool of choice. The following problem encapsulates a broad range of minimization models found in data analysis problems [2, 4, 5, 7, 10, 12, 16, 25, 28, 39] (see Section 2.2.2.1 for notation).

Problem 2.1 H is a separable real Hilbert space and $f \in \Gamma_0(\mathsf{H})$. For every $k \in \{1, \ldots, p\}$, G_k is a separable real Hilbert space, $g_k \in \Gamma_0(\mathsf{G}_k)$, and $0 \neq L_k \colon \mathsf{H} \to \mathsf{G}_k$ is linear and bounded. It is assumed that $\operatorname{zer}(\partial f + \sum_{k=1}^{p} L_k^* \circ (\partial g_k) \circ L_k) \neq \varnothing$. The task is to

$$\underset{x \in \mathsf{H}}{\text{minimize}} \;\; f(x) + \sum_{k=1}^{p} g_k(L_k x). \tag{2.1}$$

In recent years, an increasing number of problem formulations have emerged, which cannot be naturally reduced to tractable minimization problems and which are best captured by more general notions of equilibria provided by inclusion problems [14, 15, 17, 18, 24, 27, 35, 41, 42]. A formulation covering such models, as well as Problem 2.1, is the following composite monotone inclusion formulation.

Problem 2.2 H is a separable real Hilbert space and $A \colon \mathsf{H} \to 2^{\mathsf{H}}$ is maximally monotone. For every $k \in \{1, \ldots, p\}$, G_k is a separable real Hilbert space, $B_k \colon \mathsf{G}_k \to 2^{\mathsf{G}_k}$ is maximally monotone, and $0 \neq L_k \colon H \to G_k$ is linear and bounded. It is assumed that $Z = \operatorname{zer}(A + \sum_{k=1}^{p} L_k^* \circ B_k \circ L_k) \neq \varnothing$. The task is to

$$\text{find } x \in H \text{ such that } 0 \in Ax + \sum_{k=1}^{p} L_k^*\big(B_k(L_k x)\big). \tag{2.2}$$

Splitting algorithms for solving Problem 2.2 operate on the principle that each nonlinear and linear operator is used separately over the course of the iterations. Since the nonlinear operators are general set-valued monotone operators, they must be activated through their resolvent. Various deterministic operator splitting methods are available to solve Problem 2.2, most of which require the activation of the resolvents of the $p + 1$ operators A and $(B_k)_{1 \leqslant k \leqslant p}$ at each iteration [11]. Our specific focus is on solving Problem 2.2 in instances when p is large, as is often the case in data analysis problems. In such scenarios, memory and computing power limitations make the execution of standard monotone operator splitting algorithms inefficient if not simply impossible. We aim at designing monotone splitting algorithms which are stochastic in the sense that they activate a randomly selected block of operators at each iteration and, in addition, allow for random errors in the implementation of these resolvent steps. Furthermore, the proposed algorithms satisfy the following requirements:

- **R1:** They guarantee the almost sure convergence of the sequence of iterates to a solution to Problem 2.2 (respectively Problem 2.1) without any additional assumptions on the nonlinear operators (respectively the functions), the linear operators, or the underlying Hilbert spaces.
- **R2:** At each iteration, more than one randomly selected resolvent of the operators $(A, B_1, \ldots, B_p)$ can be activated.
- **R3:** Knowledge of bounds on the norms of the linear operators is not required.
- **R4:** The operators are available only through a stochastic approximation.

Requirement **R1** imposes actual iterate convergence to a solution and not a weaker form of convergence such as ergodic convergence, vanishing step sizes, or, in the context of Problem 2.1, convergence of the values of the objective function. It also asks that Problems 2.2 and 2.1 be addressed in their generality, without restricting their scope by introducing additional assumptions. Requirement **R2** makes it possible to activate more than one operator, hence opening the way to matching efficiently the computational load of an iteration to the possibly parallel architecture at hand. Requirement **R3** broadens the scope of the methods by not assuming any knowledge of the norms of the linear operators present in the model. For instance, in domain decomposition methods, it is quite difficult to obtain tight upper bounds on the norms of the trace operators [3]. Finally, in the spirit of the classical stochastic iteration models of [8,22,37], **R4** addresses the robustness of the algorithm to stochastic errors affecting the implementation of the operators.

As will be seen in the literature review of Section 2.2.2.2, there does not seem to exist methods that satisfy simultaneously **R1–R4**. Our main contribution is presented in Section 2.2.3, where we propose three algorithmic frameworks that comply with **R1–R4**. Section 2.2.4 is devoted to the minimization setting of Problem 2.1. The last section of the paper is Section 2.2.5, where the proposed algorithms are applied to signal restoration, support vector machine, classification, and image reconstruction problems.

2.2.2 Notation and existing algorithms

2.2.2.1 Notation

Throughout, H is a separable real Hilbert space with power set 2^{H}, identity operator Id, scalar product $\langle \cdot \mid \cdot \rangle$, and associated norm $\|\cdot\|$.

Let $\mathsf{A}\colon \mathsf{H} \to 2^{\mathsf{H}}$. The graph of A is $\operatorname{gra}\mathsf{A} = \{(\mathsf{x},\mathsf{x}^*) \in \mathsf{H}\times\mathsf{H} \mid \mathsf{x}^* \in \mathsf{A}\mathsf{x}\}$ and the set of zeros of A is $\operatorname{zer}\mathsf{A} = \{\mathsf{x} \in \mathsf{H} \mid \mathsf{0} \in \mathsf{A}\mathsf{x}\}$. The inverse of A is the operator $\mathsf{A}^{-1}\colon \mathsf{H} \to 2^{\mathsf{H}}$ with graph $\operatorname{gra}\mathsf{A}^{-1} = \{(\mathsf{x}^*,\mathsf{x}) \in \mathsf{H}\times\mathsf{H} \mid \mathsf{x}^* \in \mathsf{A}\mathsf{x}\}$ and the resolvent of A is $\mathsf{J}_{\mathsf{A}} = (\mathsf{Id}+\mathsf{A})^{-1}$. Further, A is maximally monotone if

$$(\forall(\mathsf{x},\mathsf{x}^*) \in \mathsf{H}\times\mathsf{H}) \quad \big[\, (\mathsf{x},\mathsf{x}^*) \in \operatorname{gra}\mathsf{A} \;\Leftrightarrow\; (\forall(\mathsf{y},\mathsf{y}^*) \in \operatorname{gra}\mathsf{A})\;\; \langle \mathsf{x}-\mathsf{y} \mid \mathsf{x}^*-\mathsf{y}^*\rangle \geqslant 0 \,\big]. \tag{2.3}$$

An operator $\mathsf{F}\colon \mathsf{H} \to \mathsf{H}$ is firmly nonexpansive if

$$(\forall \mathsf{x} \in \mathsf{H})(\forall \mathsf{y} \in \mathsf{H}) \quad \langle \mathsf{x}-\mathsf{y} \mid \mathsf{F}\mathsf{x}-\mathsf{F}\mathsf{y}\rangle \geqslant \|\mathsf{F}\mathsf{x}-\mathsf{F}\mathsf{y}\|^2. \tag{2.4}$$

Lemma 2.3 *Let* $\mathsf{F}\colon \mathsf{H} \to \mathsf{H}$ *be firmly nonexpansive and let* $\gamma \in \left]0,+\infty\right[$. *Then there exists a maximally monotone operator* $\mathsf{A}\colon \mathsf{H} \to 2^{\mathsf{H}}$ *such that the following hold:*

(i) $\mathsf{F} = \mathsf{J}_{\mathsf{A}}$.

(ii) $\mathsf{J}_{\gamma\mathsf{F}} = \mathsf{Id} - \gamma\mathsf{J}_{(1+\gamma)^{-1}\mathsf{A}} \circ (1+\gamma)^{-1}\mathsf{Id}$.

Proof. (i): See [6, Corollary 23.9].

(ii): This follows from (i) and [6, Proposition 23.29]. ☐

$\Gamma_0(\mathsf{H})$ denotes the class of lower semicontinuous convex functions $\mathsf{f}\colon \mathsf{H} \to \left]-\infty,+\infty\right]$ such that $\operatorname{dom}\mathsf{f} = \{\mathsf{x} \in \mathsf{H} \mid \mathsf{f}(\mathsf{x}) < +\infty\} \neq \varnothing$. Let $\mathsf{f} \in \Gamma_0(\mathsf{H})$. The subdifferential of f is the maximally monotone operator

$$\partial\mathsf{f}\colon \mathsf{H} \to 2^{\mathsf{H}}\colon \mathsf{x} \mapsto \{\mathsf{x}^* \in \mathsf{H} \mid (\forall \mathsf{z} \in \mathsf{H})\; \langle \mathsf{z}-\mathsf{x} \mid \mathsf{x}^*\rangle + \mathsf{f}(\mathsf{x}) \leqslant \mathsf{f}(\mathsf{z})\} \tag{2.5}$$

and the proximity operator of f is

$$\operatorname{prox}_{\mathsf{f}} = \mathsf{J}_{\partial\mathsf{f}}\colon \mathsf{H} \to \mathsf{H}\colon \mathsf{x} \mapsto \underset{\mathsf{z}\in\mathsf{H}}{\operatorname{argmin}} \left(\mathsf{f}(\mathsf{z}) + \frac{1}{2}\|\mathsf{x}-\mathsf{z}\|^2\right). \tag{2.6}$$

Let C be a nonempty closed convex subset of H. Then ι_{C} denotes the indicator function of C, d_{C}

the distance function to C,

$$\mathsf{N}_\mathsf{C} = \partial\iota_\mathsf{C} \colon \mathsf{x} \mapsto \begin{cases} \big\{\mathsf{x}^* \in \mathsf{H} \mid (\forall \mathsf{y} \in \mathsf{C})\ \ \langle \mathsf{y} - \mathsf{x} \mid \mathsf{x}^* \rangle \leqslant 0\big\}, & \text{if } \mathsf{x} \in \mathsf{C};\\ \varnothing, & \text{otherwise} \end{cases} \tag{2.7}$$

the normal cone operator of C, and $\mathrm{proj}_\mathsf{C} = \mathrm{prox}_{\iota_\mathsf{C}} = \mathsf{J}_{\mathsf{N}_\mathsf{C}}$ the projection operator onto C. In particular, if V is a closed vector subspace of H, then

$$\mathsf{N}_\mathsf{V} \colon \mathsf{H} \to 2^\mathsf{H} \colon \mathsf{x} \mapsto \begin{cases} \mathsf{V}^\perp, & \text{if } \mathsf{x} \in \mathsf{V};\\ \varnothing, & \text{otherwise.} \end{cases} \tag{2.8}$$

The underlying probability space $(\Omega, \mathcal{F}, \mathsf{P})$ is assumed to be complete and $\mathcal{B}_\mathsf{H}$ denotes the Borel σ-algebra of H. An H-valued random variable is a measurable mapping $x \colon (\Omega, \mathcal{F}) \to (\mathsf{H}, \mathcal{B}_\mathsf{H})$. The σ-algebra generated by a family Φ of random variables is denoted by $\sigma(\Phi)$. Given $x \colon \Omega \to \mathsf{H}$ and $\mathsf{S} \subset \mathsf{H}$, we set $[x \in \mathsf{S}] = \big\{\omega \in \Omega \mid x(\omega) \in \mathsf{S}\big\}$. The reader is referred to [6] for background on monotone operators and convex analysis, and to [29] for background on probability in Hilbert spaces.

We use sans-serif letters to denote deterministic variables and italicized serif letters to denote random variables. Finally, in connection with Problem 2.2, we define the Hilbert direct sum

$$\mathsf{G} = \mathsf{G}_1 \oplus \cdots \oplus \mathsf{G}_\mathsf{p}, \tag{2.9}$$

as well as the subspace

$$\mathsf{W} = \left\{\mathsf{x} \in \mathsf{H} \oplus \mathsf{G} \;\middle|\; (\forall \mathsf{k} \in \{1, \ldots, \mathsf{p}\})\ \mathsf{x}_{\mathsf{k}+1} = \mathsf{L}_\mathsf{k} \mathsf{x}_1\right\}, \tag{2.10}$$

and note that

$$\mathsf{W}^\perp = \left\{\mathsf{x}^* \in \mathsf{H} \oplus \mathsf{G} \;\middle|\; \mathsf{x}_1^* = -\sum_{\mathsf{k}=1}^{\mathsf{p}} \mathsf{L}_\mathsf{k}^* \mathsf{x}_{\mathsf{k}+1}^*\right\}. \tag{2.11}$$

2.2.2.2 Existing algorithms

It seems that no algorithm satisfying requirements **R1–R4** has been explicitly proposed to solve Problem 2.2 — or even Problem 2.1 — in the literature. There is a vast body of papers on random activation algorithms in the special case of Problem 2.1 that consists in minimizing a sum of smooth functions $\sum_{\mathsf{k}=1}^{\mathsf{p}} \mathsf{g}_\mathsf{k}$ in $\mathsf{H} = \mathbb{R}^\mathsf{N}$ via so-called stochastic gradient descent methods. Their principle is to activate a randomly selected gradient in $(\nabla \mathsf{g}_\mathsf{k})_{1 \leqslant \mathsf{k} \leqslant \mathsf{p}}$ at each iteration; see [21] and its bibliography and [19, 40] for related work with random proximal activations for this type of problem. These methods focus on a very specific instance of Problem 2.1 and they do not satisfy **R1–R2**. The only random activation algorithm tailored to Problem 2.1 which guarantees almost sure convergence of the iterates without additional assumptions such as strong convexity is the following (see also [1] for a nonadaptive version).

Proposition 2.4 ([9, Theorem 2.1 and Algorithm 3.1]) *Consider the setting of Problem 2.1 and suppose that $\mathsf{H} = \mathbb{R}^N$ and that, for every $k \in \{1,\ldots,p\}$, $\mathsf{G}_k = \mathbb{R}^{M_k}$, all considered as standard Euclidean spaces. Let $(\pi_k)_{1\leqslant k\leqslant p}$ be real numbers in $]0,1]$ such that $\sum_{k=1}^p \pi_k = 1$, and let $(k_n)_{n\in\mathbb{N}}$ be identically distributed $\{1,\ldots,p\}$-valued random variables such that, for every $k \in \{1,\ldots,p\}$, $\mathsf{P}[k_0 = k] = \pi_k$. For every $k \in \{1,\ldots,p\}$ and every $n \in \mathbb{N}$, set $\varepsilon_{k,n} = 1_{[k_n=k]}$. Let $\tau_0 \in\,]0,+\infty[$ and $\sigma_0 \in\,]0,+\infty[$ be such that*

$$\tau_0\sigma_0 \max_{1\leqslant k\leqslant p} \frac{\|L_k\|^2}{\pi_k} < 1. \tag{2.12}$$

Further, let $\chi_0 \in [0,1[$, $\eta \in\,]0,1[$, and $\delta \in\,]1,+\infty[$, set $\rho_0 = 0$ and $\nu_0 = 0$, let $x_{1,0}$ be a H-valued random variable, and let $\boldsymbol{y}_0$ be a G-valued random variable. Set $z_0 = \boldsymbol{y}_0$ and $L\colon \mathsf{H} \to \mathsf{G}\colon x \mapsto (L_k x)_{1\leqslant k\leqslant p}$, and iterate

$$\begin{array}{l}
\text{for } n = 0,1,\ldots\\
\left\lfloor
\begin{array}{l}
(\tau_{n+1},\sigma_{n+1},\chi_{n+1}) = \begin{cases}
\Big(\dfrac{\tau_n}{1-\chi_n}, \sigma_n(1-\chi_n), \chi_n\eta\Big), & \text{if } \rho_n > \|L\|\nu_n\delta;\\
\Big(\tau_n(1-\chi_n), \dfrac{\sigma_n}{1-\chi_n}, \chi_n\eta\Big), & \text{if } \rho_n < \|L\|\nu_n\delta;\\
(\tau_n,\sigma_n,\chi_n), & \text{if } \dfrac{\|L\|\nu_n}{\delta} \leqslant \rho_n \leqslant \|L\|\nu_n\delta
\end{cases}\\
x_{1,n+1} = x_{1,n} + \operatorname{prox}_{\tau_{n+1}f}\big(x_{1,n} - \tau_{n+1}\sum_{k=1}^p L_k^* z_{k,n}\big)\\
\text{for } k = 1,\ldots,p\\
\left\lfloor
\begin{array}{l}
y_{k,n+1} = y_{k,n+1} + \varepsilon_{k,n}\Big(\operatorname{prox}_{\sigma_{n+1}g_k^*}\big(y_{k,n} + \sigma_{n+1}L_k x_{1,n+1}\big) - y_{k,n}\Big)\\[2mm]
z_{k,n+1} = y_{k,n} + \varepsilon_{k,n}\Big(y_{k,n+1} + \dfrac{1}{\pi_k}\big(y_{k,n+1} - y_{k,n}\big) - y_{k,n}\Big)\\[2mm]
\rho_{n+1} = \Big\|\dfrac{1}{\tau_{n+1}}(x_n - x_{n+1}) - \dfrac{1}{\pi_{k_n}}L_{k_n}^*\big(y_{k_n,n} - y_{k_n,n+1}\big)\Big\|_1\\[2mm]
\nu_{n+1} = \dfrac{1}{\pi_{k_n}}\Big\|L_{k_n}(x_n - x_{n+1}) - \dfrac{1}{\sigma_{n+1}}\big(y_{k_n,n} - y_{k_n,n+1}\big)\Big\|_1,
\end{array}
\right.
\end{array}
\right.
\end{array} \tag{2.13}$$

where $\|\cdot\|_1$ denotes the ℓ^1-norm. Then $(x_{1,n})_{n\in\mathbb{N}}$ converges P-a.s. *to an* $\operatorname{Argmin}(f + \sum_{k=1}^p g_k \circ L_k)$-*valued random variable.*

Algorithm (2.13) is of interest because it guarantees **R1** in a finite-dimensional setting. However, it does not satisfy **R2** since, at each iteration, f must be activated together with one of the functions $(g_k)_{1\leqslant k\leqslant p}$. It does not satisfy **R3** either since it requires the knowledge of the norms of linear operators in (2.12). We also note that it does not tolerate errors in the evaluation of the proximity operators, which means that **R4** is not satisfied.

Let us now turn to the general Problem 2.2. The only algorithm that satisfies **R1** is that of [34], which corresponds to an implementation of the random block-coordinate forward-backward algorithm of [13, Section 5.2] suggested in [13, Remark 5.10(iv)].

Proposition 2.5 ([34, Proposition 4.6]) *Consider the setting of Problem 2.2. Let* $\mathsf{W}\colon \mathsf{H} \to \mathsf{H}$ *and, for every* $\mathsf{k} \in \{1, \ldots, \mathsf{p}\}$*, let* $\mathsf{U}_\mathsf{k}\colon \mathsf{G}_\mathsf{k} \to \mathsf{G}_\mathsf{k}$ *be bounded linear strongly positive self-adjoint operators such that*

$$\sum_{\mathsf{k}=1}^{\mathsf{p}} \|\mathsf{U}_\mathsf{k}^{1/2}\mathsf{L}_\mathsf{k}\mathsf{W}^{1/2}\|^2 < \frac{1}{2}. \tag{2.14}$$

Let $(\lambda_\mathsf{n})_{\mathsf{n}\in\mathbb{N}}$ *be a sequence in* $]0,1]$ *such that* $\inf_{\mathsf{n}\in\mathbb{N}} \lambda_\mathsf{n} > 0$*, let* $x_{1,0}$ *and* $(a_{1,\mathsf{n}})_{\mathsf{n}\in\mathbb{N}}$ *be* H*-valued random variables, let* $\boldsymbol{v}_0$ *and* $(\boldsymbol{b}_\mathsf{n})_{\mathsf{n}\in\mathbb{N}}$ *be* G*-valued random variables, and let* $(\boldsymbol{\varepsilon}_\mathsf{n})_{\mathsf{n}\in\mathbb{N}}$ *be identically distributed* $\{0,1\}^\mathsf{p} \smallsetminus \{\mathbf{0}\}$*-valued random variables. Iterate*

$$\begin{array}{l}
\textit{for}\ \mathsf{n} = 0, 1, \ldots\\
\left\lfloor
\begin{array}{l}
y_{1,\mathsf{n}} = \mathsf{J}_{\mathsf{W}\mathsf{A}}\Big(x_{1,\mathsf{n}} - \mathsf{W}\big(\sum_{\mathsf{k}=1}^{\mathsf{p}} \mathsf{L}_\mathsf{k}^* v_{\mathsf{k},\mathsf{n}}\big)\Big) + a_{1,\mathsf{n}}\\
x_{1,\mathsf{n}+1} = x_{1,\mathsf{n}} + \lambda_\mathsf{n}(y_{1,\mathsf{n}} - x_{1,\mathsf{n}})\\
\textit{for}\ \mathsf{k} = 1, \ldots, \mathsf{p}\\
\left\lfloor
\begin{array}{l}
u_{\mathsf{k},\mathsf{n}} = \varepsilon_{\mathsf{k},\mathsf{n}}\Big(\mathsf{J}_{\mathsf{U}_\mathsf{k}\mathsf{B}_\mathsf{k}^{-1}}\Big(v_{\mathsf{k},\mathsf{n}} + \mathsf{U}_\mathsf{k}\big(\mathsf{L}_\mathsf{k}(2y_{1,\mathsf{n}} - x_{1,\mathsf{n}})\big)\Big) + b_{\mathsf{k},\mathsf{n}}\Big)\\
v_{\mathsf{k},\mathsf{n}+1} = v_{\mathsf{k},\mathsf{n}} + \varepsilon_{\mathsf{k},\mathsf{n}}\lambda_\mathsf{n}\big(u_{\mathsf{k},\mathsf{n}} - v_{\mathsf{k},\mathsf{n}}\big),
\end{array}
\right.
\end{array}
\right.
\end{array} \tag{2.15}$$

and set $(\forall \mathsf{n} \in \mathbb{N})$ $\mathcal{E}_\mathsf{n} = \sigma(\boldsymbol{\varepsilon}_\mathsf{n})$ *and* $\mathcal{X}_\mathsf{n} = \sigma(x_{1,\mathsf{l}}, \boldsymbol{v}_\mathsf{l})_{0\leqslant\mathsf{l}\leqslant\mathsf{n}}$*. In addition, assume that the following hold:*

(i) $\sum_{\mathsf{n}\in\mathbb{N}} \sqrt{\mathsf{E}(\|a_{1,\mathsf{n}}\|^2 \,|\, \mathcal{X}_\mathsf{n})} < +\infty$ *and* $\sum_{\mathsf{n}\in\mathbb{N}} \sqrt{\mathsf{E}(\|\boldsymbol{b}_\mathsf{n}\|^2 \,|\, \mathcal{X}_\mathsf{n})} < +\infty$.

(ii) *For every* $\mathsf{n} \in \mathbb{N}$, $\mathcal{E}_\mathsf{n}$ *and* $\mathcal{X}_\mathsf{n}$ *are independent.*

(iii) *For every* $\mathsf{l} \in \{1, \ldots, \mathsf{p}\}$, $\mathsf{P}[\varepsilon_{\mathsf{l},0} = 1] > 0$.

Then $(x_{1,\mathsf{n}})_{\mathsf{n}\in\mathbb{N}}$ *converges weakly* P-a.s. *to a* Z*-valued random variable.*

Let us note that algorithm (2.15) satisfies **R1** but not **R2** since it must activate A at each iteration, nor does it satisfy **R3** since it requires the knowledge of the norms of linear operators to implement (2.14). Another framework related to Problem 2.2 is that of [20], which allows for random activations in Problem 2.2 in a finite-dimensional setting when no linear operator is present and under the assumption that the operators $(\mathsf{B}_\mathsf{k})_{1\leqslant\mathsf{k}\leqslant\mathsf{p}}$ are cocoercive. It therefore does not satisfy several requirements of **R1** and activates only one operator per iteration, which violates **R2**. On the other hand, the recent work [38] solves Problem 2.2 in a finite-dimensional setting when no linear operator is present and under strong monotonicity of the nonlinear operators. Hence, it does not satisfy **R1** and, since it does not allow for multiple activations at each iteration, it does not satisfy **R2** either.

2.2.3 Proposed algorithms

2.2.3.1 Multivariate framework

Our strategy consists in embedding Problem 2.2 into multivariate problems that have the following general form studied in [13] and involve m agents $(\mathsf{x}_1, \ldots, \mathsf{x}_\mathsf{m})$.

Problem 2.6 Let $(\mathsf{X}_i)_{1\leqslant i\leqslant m}$ and $(\mathsf{Y}_j)_{1\leqslant j\leqslant r}$ be families of separable real Hilbert spaces with Hilbert direct sums $\mathsf{X} = \mathsf{X}_1\oplus\cdots\oplus\mathsf{X}_m$ and $\mathsf{Y} = \mathsf{Y}_1\oplus\cdots\oplus\mathsf{Y}_r$. For every $i \in \{1,\ldots,m\}$ and every $j \in \{1,\ldots,r\}$, let $\mathsf{C}_i\colon \mathsf{X}_i \to 2^{\mathsf{X}_i}$ and $\mathsf{D}_j\colon \mathsf{Y}_j \to 2^{\mathsf{Y}_j}$ be maximally monotone, and let $\mathsf{M}_{ji}\colon \mathsf{X}_i \to \mathsf{Y}_j$ be linear and bounded. Set

$$\begin{cases}\mathsf{M}\colon \mathsf{X}\to\mathsf{Y}\colon \mathsf{x}\mapsto \big(\sum_{i=1}^m \mathsf{M}_{1i}\mathsf{x}_i,\ldots,\sum_{i=1}^m \mathsf{M}_{ri}\mathsf{x}_i\big)\\ \mathsf{C}\colon \mathsf{X}\to 2^{\mathsf{X}}\colon \mathsf{x}\mapsto \mathsf{C}_1\mathsf{x}_1\times\cdots\times\mathsf{C}_m\mathsf{x}_m\\ \mathsf{D}\colon \mathsf{Y}\to 2^{\mathsf{Y}}\colon \mathsf{y}\mapsto \mathsf{D}_1\mathsf{y}_1\times\cdots\times\mathsf{D}_r\mathsf{y}_r.\end{cases}\tag{2.16}$$

The task is to

$$\text{find}\;\; \mathsf{x}\in\mathsf{X}\;\;\text{such that}\;\; \mathbf{0}\in\mathsf{C}\mathsf{x}+\mathsf{M}^*\big(\mathsf{D}(\mathsf{M}\mathsf{x})\big).\tag{2.17}$$

The set of solutions to (2.17) is denoted by Z and assumed to be nonempty. Further, the projection operator onto the subspace

$$\mathsf{V} = \big\{(\mathsf{x},\mathsf{y})\in\mathsf{X}\oplus\mathsf{Y} \mid \mathsf{y}=\mathsf{M}\mathsf{x}\big\}\tag{2.18}$$

is decomposed as

$$\mathrm{proj}_{\mathsf{V}}\colon (\mathsf{x},\mathsf{y})\mapsto \big(\mathsf{Q}_l(\mathsf{x},\mathsf{y})\big)_{1\leqslant l\leqslant m+r},\quad \text{where}\quad \begin{cases}(\forall i\in\{1,\ldots,m\})\;\; \mathsf{Q}_i\colon \mathsf{X}\oplus\mathsf{Y}\to\mathsf{X}_i\\ (\forall j\in\{1,\ldots,r\})\;\; \mathsf{Q}_{m+j}\colon \mathsf{X}\oplus\mathsf{Y}\to\mathsf{Y}_j.\end{cases}\tag{2.19}$$

Our approach is ultimately based on the Douglas–Rachford algorithm implemented in $\mathsf{X}\oplus\mathsf{Y}$. Define

$$\mathsf{A}\colon \mathsf{X}\oplus\mathsf{Y}\to 2^{\mathsf{X}\oplus\mathsf{Y}}\colon (\mathsf{x},\mathsf{y})\mapsto \mathsf{C}\mathsf{x}\times\mathsf{D}\mathsf{y}\quad\text{and}\quad \mathsf{B}=N_{\mathsf{V}}.\tag{2.20}$$

Then it follows from [13, Eq. (5.23)] that $(\mathsf{x},\mathsf{y})\in\mathrm{zer}(\mathsf{A}+\mathsf{B})$ if and only if $\mathsf{x}\in\mathrm{zer}(\mathsf{C}+\mathsf{M}^*\circ\mathsf{D}\circ\mathsf{M})$ and $\mathsf{y}=\mathsf{M}\mathsf{x}$. We can construct a point in $\mathrm{zer}(\mathsf{A}+\mathsf{B})$ iteratively by the Douglas–Rachford algorithm [6, Section 26.3], which requires the resolvents of A and B. By [6, Proposition 23.18], J_{A} can be decomposed in terms of $(J_{\mathsf{C}_1},\ldots,J_{\mathsf{C}_m},J_{\mathsf{D}_1},\ldots,J_{\mathsf{D}_r})$. On the other hand, $J_{\mathsf{B}}=\mathrm{proj}_{\mathsf{V}}$ and it follows from (2.18) and [6, Example 29.19(ii)] that

$$(\forall\mathsf{x}\in\mathsf{X})(\forall\mathsf{y}\in\mathsf{Y})\quad \mathrm{proj}_{\mathsf{V}}(\mathsf{x},\mathsf{y})=(\mathsf{p},\mathsf{M}\mathsf{p}),\quad\text{where}\quad \mathsf{p}=(\mathrm{Id}+\mathsf{M}^*\circ\mathsf{M})^{-1}\big(\mathsf{x}+\mathsf{M}^*\mathsf{y}\big).\tag{2.21}$$

This operator is decomposed in terms of the operators $(\mathsf{Q}_l)_{1\leqslant l\leqslant m+r}$ in (2.19). The following result provides a randomly block-activated implementation of this product space version of the Douglas–Rachford algorithm.

Theorem 2.7 ([13, Corollary 5.3]) *Consider the setting of Problem 2.6. Set* $\mathsf{O}=\{0,1\}^{m+r}\smallsetminus\{\mathbf{0}\}$, *let* $\gamma\in\,]0,+\infty[$, *let* $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence in* $]0,2[$ *such that* $\inf_{n\in\mathbb{N}}\lambda_n>0$ *and* $\sup_{n\in\mathbb{N}}\lambda_n<2$, *let* $\boldsymbol{x}_0$, $\boldsymbol{z}_0$, $(\boldsymbol{a}_n)_{n\in\mathbb{N}}$, *and* $(\boldsymbol{c}_n)_{n\in\mathbb{N}}$ *be* X*-valued random variables, let* $\boldsymbol{y}_0$, $\boldsymbol{w}_0$, $(\boldsymbol{b}_n)_{n\in\mathbb{N}}$, *and* $(\boldsymbol{d}_n)_{n\in\mathbb{N}}$ *be* Y*-valued* *random variables, and let* $(\boldsymbol{\varepsilon}_n)_{n\in\mathbb{N}}$ *be identically distributed* O*-valued random variables. Iterate*

$$\begin{array}{l}\text{for } n = 0, 1, \ldots\\ \left\lfloor\begin{array}{l}\text{for } i = 1, \ldots, m\\ \left\lfloor\begin{array}{l} x_{i,n+1} = x_{i,n} + \varepsilon_{i,n}\big(Q_i(z_n, \boldsymbol{w}_n) + a_{i,n} - x_{i,n}\big)\\ z_{i,n+1} = z_{i,n} + \varepsilon_{i,n}\lambda_n\big(J_{\gamma C_i}(2x_{i,n+1} - z_{i,n}) + c_{i,n} - x_{i,n+1}\big)\end{array}\right.\\ \text{for } j = 1, \ldots, r\\ \left\lfloor\begin{array}{l} y_{j,n+1} = y_{j,n} + \varepsilon_{m+j,n}\big(Q_{m+j}(z_n, \boldsymbol{w}_n) + b_{j,n} - y_{j,n}\big)\\ w_{j,n+1} = w_{j,n} + \varepsilon_{m+j,n}\lambda_n\big(J_{\gamma D_j}(2y_{j,n+1} - w_{j,n}) + d_{j,n} - y_{j,n+1}\big),\end{array}\right.\end{array}\right.\end{array} \tag{2.22}$$

and set $(\forall n \in \mathbb{N})$ $\mathcal{E}_n = \sigma(\boldsymbol{\varepsilon}_n)$ *and* $\mathcal{S}_n = \sigma(z_l, \boldsymbol{w}_l)_{0\leqslant l\leqslant n}$*. In addition, assume that the following are satisfied:*

(i) $\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|\boldsymbol{a}_n\|^2 \,|\, \mathcal{S}_n)} < +\infty$, $\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|\boldsymbol{b}_n\|^2 \,|\, \mathcal{S}_n)} < +\infty$, $\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|\boldsymbol{c}_n\|^2 \,|\, \mathcal{S}_n)} < +\infty$, $\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|\boldsymbol{d}_n\|^2 \,|\, \mathcal{S}_n)} < +\infty$, $\boldsymbol{a}_n \rightharpoonup \boldsymbol{0}$ P-a.s.*, and* $\boldsymbol{b}_n \rightharpoonup \boldsymbol{0}$ P-a.s.

(ii) *For every* $n \in \mathbb{N}$*,* $\mathcal{E}_n$ *and* $\mathcal{S}_n$ *are independent.*

(iii) *For every* $l \in \{1, \ldots, m + r\}$*,* $\mathsf{P}[\varepsilon_{l,0} = 1] > 0$.

Then $(\boldsymbol{x}_n)_{n\in\mathbb{N}}$ *converges weakly* P-a.s. *to a* Z*-valued random variable.*

Remark 2.8 The measurability of the weak limit in [13, Corollary 5.3] relies on [13, Proposition 2.3], which involves Pettis' theorem [36, Corollary 1.13]. The applicability of the latter follows from the separability of H and the fact that $(\Omega, \mathcal{F}, \mathsf{P})$ is a complete probability space; see [26, Sections 1.1a–b] for details.

Remark 2.9 At iteration n, the random variables $(\varepsilon_{i,n})_{1\leqslant i\leqslant m}$ and $(\varepsilon_{m+j,n})_{1\leqslant j\leqslant r}$ act as switches which control which components are updated, while the random variables $a_{i,n}$, $b_{j,n}$, $c_{i,n}$, and $d_{j,n}$ model approximations in the implementation of the operators Q_i, Q_j, $J_{\gamma C_i}$, and $J_{\gamma D_j}$, respectively.

We now present three frameworks for solving Problem 2.2 which are based on specializations of Theorem 2.7.

2.2.3.2 Framework 1

The first approach stems from the observation that Problem 2.6 reduces to Problem 2.2 when $m = 1$, $r = p$, $X_1 = \mathsf{H}$, $C_1 = A$, and $(\forall k \in \{1, \ldots, p\})$ $Y_k = G_k$, $M_{k,1} = L_k$, and $D_k = B_k$. Surprisingly, this basic observation does not seem to have been exploited in attempts to design random block activation algorithms for solving Problem 2.1 or Problem 2.2 (or special cases thereof) using the stochastic quasi-Fejér framework of [13]; see for instance [9, 30, 32, 33].

We derive from Theorem 2.7 the following convergence result.

Proposition 2.10 *Consider the setting of Problem 2.2. Set* $\mathsf{O} = \{0, 1\}^{1+p} \smallsetminus \{\boldsymbol{0}\}$*, let* $\gamma \in \left]0, +\infty\right[$*, let* $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence in* $\left]0, 2\right[$ *such that* $\inf_{n\in\mathbb{N}} \lambda_n > 0$ *and* $\sup_{n\in\mathbb{N}} \lambda_n < 2$*, let* $x_{1,0}$*,* $z_{1,0}$*,* $(c_{1,n})_{n\in\mathbb{N}}$*,* *and* $(e_n)_{n\in\mathbb{N}}$ *be* H*-valued random variables, let* $\boldsymbol{y}_0$*,* $\boldsymbol{w}_0$*, and* $(\boldsymbol{d}_n)_{n\in\mathbb{N}}$ *be* G*-valued random variables, and let* $(\boldsymbol{\varepsilon}_n)_{n\in\mathbb{N}}$ *be identically distributed* O*-valued random variables. Set* $\mathsf{Q} = (\mathrm{Id} + \sum_{k=1}^{p} \mathsf{L}_k^* \circ \mathsf{L}_k)^{-1}$ *and iterate*

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
s_n = \mathsf{Q}\big(z_{1,n} + \sum_{k=1}^{p} \mathsf{L}_k^* w_{k,n}\big) + e_n \\
x_{1,n+1} = x_{1,n} + \varepsilon_{1,n}(s_n - x_{1,n}) \\
z_{1,n+1} = z_{1,n} + \varepsilon_{1,n}\lambda_n\big(\mathsf{J}_{\gamma\mathsf{A}}(2x_{1,n+1} - z_{1,n}) + c_{1,n} - x_{1,n+1}\big) \\
\text{for } k = 1, \ldots, p \\
\left\lfloor
\begin{array}{l}
y_{k,n+1} = y_{k,n} + \varepsilon_{1+k,n}(\mathsf{L}_k s_n - y_{k,n}) \\
w_{k,n+1} = w_{k,n} + \varepsilon_{1+k,n}\lambda_n\big(\mathsf{J}_{\gamma\mathsf{B}_k}(2y_{k,n+1} - w_{k,n}) + d_{k,n} - y_{k,n+1}\big).
\end{array}
\right.
\end{array}
\right.
\end{array}
\tag{2.23}
$$

In addition, assume that the following are satisfied:

(i) $\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|c_{1,n}\|^2 \,|\, \sigma(z_{1,l}, \boldsymbol{v}_l)_{0\leqslant l\leqslant n})} < +\infty$*,* $\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|\boldsymbol{d}_n\|^2 \,|\, \sigma(z_{1,l}, \boldsymbol{v}_l)_{0\leqslant l\leqslant n})} < +\infty$*,* $\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|e_n\|^2 \,|\, \sigma(z_{1,l}, \boldsymbol{v}_l)_{0\leqslant l\leqslant n})} < +\infty$*, and* $e_n \rightharpoonup 0$*.*

(ii) *For every* $n \in \mathbb{N}$*,* $\sigma(\boldsymbol{\varepsilon}_n)$ *and* $\sigma(z_{1,l}, \boldsymbol{w}_l)_{0\leqslant l\leqslant n}$ *are independent.*

(iii) *For every* $l \in \{1, \ldots, p+1\}$*,* $\mathsf{P}[\varepsilon_{l,0} = 1] > 0$*.*

Then $(x_{1,n})_{n\in\mathbb{N}}$ *converges weakly* P*-a.s. to a* Z*-valued random variable.*

Proof. In Problem 2.6, set $m = 1$, $r = p$, $\mathsf{X}_1 = \mathsf{H}$, $\mathsf{C}_1 = \mathsf{A}$, and, for every $k \in \{1, \ldots, p\}$, $\mathsf{Y}_k = \mathsf{G}_k$, $\mathsf{M}_{k,1} = \mathsf{L}_k$, and $\mathsf{D}_k = \mathsf{B}_k$. Further, for every $n \in \mathbb{N}$, set $a_{1,n} = e_n$ and, for every $k \in \{1, \ldots, p\}$, set $b_{k,n} = \mathsf{L}_k e_n$. Then it follows from (i) that $a_{1,n} \rightharpoonup 0$ P-a.s., $\boldsymbol{b}_n \rightharpoonup \boldsymbol{0}$ P-a.s., and

$$
\begin{aligned}
\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|\boldsymbol{b}_n\|^2 \,|\, \sigma(z_l, \boldsymbol{v}_l)_{0\leqslant l\leqslant n})} &\leqslant \sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}\left(\left(\sum_{k=1}^{p} \|\mathsf{L}_k\|^2\right)\|e_n\|^2 \,\middle|\, \sigma(z_l, \boldsymbol{v}_l)_{0\leqslant l\leqslant n}\right)} \\
&= \sqrt{\sum_{k=1}^{p} \|\mathsf{L}_k\|^2} \sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|e_n\|^2 \,|\, \sigma(z_l, \boldsymbol{v}_l)_{0\leqslant l\leqslant n})} \\
&< +\infty.
\end{aligned}
\tag{2.24}
$$

The assertion therefore results from Theorem 2.7. ☐

2.2.3.3 Framework 2

In Framework 1, Problem 2.6 collapses to Problem 2.2 by reducing the number of agents to $m = 1$. Here, we use $m = p + 1$ agents in Problem 2.6 and capture Problem 2.2 by forcing these agents $(\mathsf{x}_1, \ldots, \mathsf{x}_{p+1})$ to lie in the subspace W of (2.10).

Proposition 2.11 *Consider the setting of Problem 2.2. Set* $\mathsf{O} = \{0,1\}^{p+2} \smallsetminus \{\boldsymbol{0}\}$*, let* $\gamma \in \left]0, +\infty\right[$*, let* $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence in* $\left]0, 2\right[$ *such that* $\inf_{n\in\mathbb{N}} \lambda_n > 0$ *and* $\sup_{n\in\mathbb{N}} \lambda_n < 2$*, let* $\boldsymbol{x}_0$*,* $\boldsymbol{z}_0$*,* $\boldsymbol{u}_0$*,* $\boldsymbol{v}_0$*, and* $(\boldsymbol{c}_n)_{n\in\mathbb{N}}$ *be* $\mathsf{H} \oplus \mathsf{G}$*-valued random variables, let* $(e_n)_{n\in\mathbb{N}}$ *be* H*-valued random variables, and let*

$(\boldsymbol{\varepsilon}_n)_{n\in\mathbb{N}}$ *be identically distributed* O*-valued random variables. Set* $\mathsf{Q} = (\mathrm{Id} + \sum_{k=1}^{p} \mathsf{L}_k^* \circ \mathsf{L}_k)^{-1}$. *Iterate*

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\text{for } i = 1, \ldots, p+1 \\
\left\lfloor\; x_{i,n+1} = x_{i,n} + \varepsilon_{i,n}\left(\dfrac{z_{i,n} + v_{i,n}}{2} - x_{i,n}\right)\right. \\
z_{1,n+1} = z_{1,n} + \varepsilon_{1,n}\lambda_n\big(J_{\gamma A}(2x_{1,n+1} - z_{1,n}) + c_{1,n} - x_{1,n+1}\big) \\
\text{for } k = 1, \ldots, p \\
\left\lfloor\; z_{k+1,n+1} = z_{k+1,n} + \varepsilon_{k+1,n}\lambda_n\big(J_{\gamma B_k}(2x_{k+1,n+1} - z_{k+1,n}) + c_{k+1,n} - x_{k+1,n+1}\big)\right. \\
\text{for } k = 1, \ldots, p+1 \\
\left\lfloor\; u_{k,n+1} = u_{k,n} + \varepsilon_{p+2,n}\left(\dfrac{z_{k,n} + v_{k,n}}{2} - u_{k,n}\right)\right. \\
s_n = \varepsilon_{p+2,n}\Big(\mathsf{Q}\big(2u_{1,n+1} - v_{1,n} + \sum_{k=1}^{p} \mathsf{L}_k^*(2u_{k+1,n+1} - v_{k+1,n})\big) + e_n\Big) \\
v_{1,n+1} = v_{1,n} + \varepsilon_{p+2,n}\lambda_n(s_n - u_{1,n+1}) \\
\text{for } k = 1, \ldots, p \\
\left\lfloor\; v_{k+1,n+1} = v_{k+1,n} + \varepsilon_{p+2,n}\lambda_n(\mathsf{L}_k s_n - u_{k+1,n+1}).\right.
\end{array}
\right.
\end{array}
\tag{2.25}
$$

In addition, assume that the following are satisfied:

(i) $\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|\boldsymbol{c}_n\|^2 \mid \sigma(\boldsymbol{z}_l, \boldsymbol{v}_l)_{0\leqslant l\leqslant n})} < +\infty$ *and* $\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|\boldsymbol{e}_n\|^2 \mid \sigma(\boldsymbol{z}_l, \boldsymbol{v}_l)_{0\leqslant l\leqslant n})} < +\infty$.

(ii) *For every* $n \in \mathbb{N}$, $\sigma(\boldsymbol{\varepsilon}_n)$ *and* $\sigma(\boldsymbol{z}_l, \boldsymbol{v}_l)_{0\leqslant l\leqslant n}$ *are independent.*

(iii) *For every* $l \in \{1, \ldots, p+2\}$, $\mathsf{P}[\varepsilon_{l,0} = 1] > 0$.

Then $(x_{1,n})_{n\in\mathbb{N}}$ *converges weakly* P*-a.s. to a* Z*-valued random variable.*

Proof. In Problem 2.6, set $m = p+1$, $r = 1$, $\mathsf{X}_1 = \mathsf{H}$, $(\mathsf{X}_i)_{2\leqslant i\leqslant m} = (\mathsf{G}_{i-1})_{2\leqslant i\leqslant m}$, and $\mathsf{Y}_1 = \mathsf{H}\oplus\mathsf{G}$. Thus, $\mathsf{Y} = \mathsf{Y}_1 = \mathsf{H}\oplus\mathsf{G} = \mathsf{X}$. Moreover, for every $i \in \{1, \ldots, p+1\}$, set

$$
\mathsf{M}_{1i} \colon \mathsf{X}_i \to \mathsf{H}\oplus\mathsf{G} \colon \mathsf{x}_i \mapsto \big(0, \ldots, 0, \underbrace{\mathsf{x}_i}_{\text{ith position}}, 0, \ldots, 0\big),
\tag{2.26}
$$

which yields

$$
\mathsf{M}_{1i}^* \colon \mathsf{H}\oplus\mathsf{G} \to \mathsf{X}_i \colon (\mathsf{x}_1^*, \ldots, \mathsf{x}_{p+1}^*) \mapsto \mathsf{x}_i^*.
\tag{2.27}
$$

Further, denote by $\mathsf{x} = (\mathsf{x}_1, \ldots, \mathsf{x}_{p+1})$ a generic element in $\mathsf{H}\oplus\mathsf{G}$ and define $\mathsf{D}_1 = \mathsf{N}_{\mathsf{W}}$, where W is the subspace of (2.10). In this configuration, (2.17) reduces to

$$
\text{find } \mathsf{x} \in \mathsf{H}\oplus\mathsf{G} \text{ such that } \mathbf{0} \in \bigtimes_{i=1}^{p+1} \mathsf{C}_i\mathsf{x}_i + \mathsf{N}_{\mathsf{W}}\mathsf{x}.
\tag{2.28}
$$

We observe that

$$(\forall i \in \{1,\ldots,p+1\})(\forall l \in \{1,\ldots,p+1\}) \quad M_{1i}^* \circ M_{1l} = \begin{cases} \mathrm{Id}, & \text{if } i = l; \\ 0, & \text{if } i \neq l. \end{cases} \tag{2.29}$$

As a result, $(\mathrm{Id} + M^* \circ M)^{-1} = (1/2)\mathrm{Id}$ and we derive from (2.19), (2.21), and (2.27) that

$$Q_{p+2} \colon (\boldsymbol{z}, \boldsymbol{v}) \mapsto \frac{\boldsymbol{z} + \boldsymbol{v}}{2} \quad \text{and} \quad (\forall l \in \{1,\ldots,p+1\}) \quad Q_l \colon (\boldsymbol{z}, \boldsymbol{v}) \mapsto \frac{z_l + v_l}{2}. \tag{2.30}$$

Altogether, (2.22) with variables $y_{1,n} = \boldsymbol{u}_n \in H \oplus G$ and $w_{1,n} = \boldsymbol{v}_n \in H \oplus G$ becomes

$$\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor \begin{array}{l}
\text{for } i = 1, \ldots, p+1 \\
\left\lfloor \begin{array}{l}
x_{i,n+1} = x_{i,n} + \varepsilon_{i,n}\left(\dfrac{z_{i,n} + v_{i,n}}{2} - x_{i,n}\right) \\
z_{i,n+1} = z_{i,n} + \varepsilon_{i,n}\lambda_n\big(J_{\gamma C_i}(2x_{i,n+1} - z_{i,n}) + c_{i,n} - x_{i,n+1}\big)
\end{array} \right. \\
\boldsymbol{u}_{n+1} = \boldsymbol{u}_n + \varepsilon_{p+2,n}\left(\dfrac{\boldsymbol{z}_n + \boldsymbol{v}_n}{2} - \boldsymbol{u}_n\right) \\
\boldsymbol{v}_{n+1} = \boldsymbol{v}_n + \varepsilon_{p+2,n}\lambda_n\big(\mathrm{proj}_{\boldsymbol{W}}(2\boldsymbol{u}_{n+1} - \boldsymbol{v}_n) + \boldsymbol{d}_{1,n} - \boldsymbol{u}_{n+1}\big),
\end{array} \right.
\end{array} \tag{2.31}$$

where $\boldsymbol{d}_{1,n}$ is the error incurred when projecting onto $\boldsymbol{W}$ at iteration n. We derive from (2.10) and [6, Example 29.19(ii)] that

$$\begin{aligned}
&\mathrm{proj}_{\boldsymbol{W}} \colon (x, y_1, \ldots, y_p) \mapsto (s, L_1 s, \ldots, L_p s), \\
&\qquad\qquad \text{where} \quad s = \left(\mathrm{Id} + \sum_{k=1}^{p} L_k^* \circ L_k\right)^{-1}\left(x + \sum_{k=1}^{p} L_k^* y_k\right).
\end{aligned} \tag{2.32}$$

Set $(\forall n \in \mathbb{N})$ $\boldsymbol{d}_{1,n} = (e_n, L_1 e_n, \ldots, L_p e_n)$. Then we infer from (i) that

$$\begin{aligned}
\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|\boldsymbol{d}_{1,n}\|^2 \,|\, \sigma(\boldsymbol{z}_l, \boldsymbol{v}_l)_{0\leqslant l\leqslant n})} &\leqslant \sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}\left(\left(1 + \sum_{k=1}^{p} \|L_k\|^2\right)\|e_n\|^2 \,\middle|\, \sigma(\boldsymbol{z}_l, \boldsymbol{v}_l)_{0\leqslant l\leqslant n}\right)} \\
&= \sqrt{1 + \sum_{k=1}^{p} \|L_k\|^2} \sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|e_n\|^2 \,|\, \sigma(\boldsymbol{z}_l, \boldsymbol{v}_l)_{0\leqslant l\leqslant n})} \\
&< +\infty.
\end{aligned} \tag{2.33}$$

Thus, it follows from Theorem 2.7 that, with Z denoting the set of solutions to (2.28),

$$\begin{aligned}
&\text{the sequence } (x_{1,n}, x_{2,n}, \ldots, x_{p+1,n})_{n\in\mathbb{N}} \text{ in (2.31) converges weakly P-a.s.} \\
&\qquad \text{to a } Z\text{-valued random variable } \overline{x} = (\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_{p+1}) \text{ if } Z \neq \varnothing.
\end{aligned} \tag{2.34}$$

Next, we specialize (2.28) to

$$C_1 = A \quad \text{and} \quad (\forall i \in \{2, \dots, p+1\}) \quad C_i = B_{i-1}. \tag{2.35}$$

In this context, (2.31) reduces to (2.25). Recalling that Z denotes the set of solutions to Problem 2.2, in view of (2.34), it remains to show that

$$\boldsymbol{Z} = \big\{(x_1, L_1x_1, \dots, L_px_1) \mid x_1 \in Z\big\}. \tag{2.36}$$

Let $\boldsymbol{x} \in \boldsymbol{H} \oplus \boldsymbol{G}$. We have

$$\begin{aligned}
\boldsymbol{x} \in \boldsymbol{Z} &\Leftrightarrow \boldsymbol{x} \text{ solves (2.28)} \\
&\Leftrightarrow \begin{cases} \boldsymbol{x} \in \boldsymbol{W} \\ (\exists \boldsymbol{x}^* \in \boldsymbol{W}^\perp) \quad \boldsymbol{0} \in \displaystyle\bigtimes_{i=1}^{p+1} C_ix_i + \boldsymbol{x}^* \end{cases} \\
&\Leftrightarrow \begin{cases} (\exists x_1 \in H) \quad \boldsymbol{x} = (x_1, L_1x_1, \dots, L_px_1) \\ (\exists \boldsymbol{x}^* \in \boldsymbol{W}^\perp) \quad \boldsymbol{0} \in Ax_1 \times B_1(L_1x_1) \times \cdots \times B_p(L_px_1) + \boldsymbol{x}^* \end{cases} \\
&\Leftrightarrow \begin{cases} (\exists x_1 \in H) \quad \boldsymbol{x} = (x_1, L_1x_1, \dots, L_px_1) \\ (\exists (y_1^*, \dots, y_p^*) \in \boldsymbol{G}) \quad (0, 0, \dots, 0) \in \\ \qquad\qquad Ax_1 \times B_1(L_1x_1) \times \cdots \times B_p(L_px_1) + \left(\displaystyle\sum_{k=1}^p L_k^*y_k^*, -y_1^*, \dots, -y_p^*\right) \end{cases} \\
&\Leftrightarrow \begin{cases} (\exists x_1 \in H) \quad \boldsymbol{x} = (x_1, L_1x_1, \dots, L_px_1) \\ (\exists (y_1^*, \dots, y_p^*) \in \boldsymbol{G}) \quad \begin{cases} 0 \in Ax_1 + \sum_{k=1}^p L_k^*y_k^* \\ (\forall k \in \{1, \dots, p\}) \quad y_k^* \in B_k(L_kx_1). \end{cases} \end{cases} \\
&\Leftrightarrow \begin{cases} (\exists x_1 \in H) \quad \boldsymbol{x} = (x_1, L_1x_1, \dots, L_px_1) \\ 0 \in Ax_1 + \sum_{k=1}^p L_k^*(B_k(L_kx_1)) \end{cases} \\
&\Leftrightarrow (\exists x_1 \in Z) \quad \boldsymbol{x} = (x_1, L_1x_1, \dots, L_px_1),
\end{aligned} \tag{2.37}$$

which completes the proof. ☐

2.2.3.4 Framework 3

Our last algorithm connects Problem 2.2 to Problem 2.6 by means of a coupling operator $\boldsymbol{E}$ mapping to an auxiliary space $\boldsymbol{K}$ and such that $\ker \boldsymbol{E}$ coincides with the space $\boldsymbol{W}$ of (2.10).

Proposition 2.12 *Consider the setting of Problem 2.2, let $(K_j)_{1 \leqslant j \leqslant r}$ be separable real Hilbert* *spaces, set*

$$\mathsf{K} = \bigoplus_{j=1}^{r} \mathsf{K}_j, \tag{2.38}$$

and let

$$\mathsf{E}\colon \mathsf{H}\oplus\mathsf{G} \to \mathsf{K}\colon x \mapsto \left(\sum_{i=1}^{p+1} \mathsf{E}_{ji} x_i\right)_{1\leqslant j\leqslant r} \tag{2.39}$$

be linear, bounded, and such that $\ker \mathsf{E} = \mathsf{W}$. *Define* V *as in* (2.18), *where* X *is replaced with* $\mathsf{H}\oplus\mathsf{G}$, Y *with* K, *and* M *with* E, *and decompose its projection operator as* $\mathrm{proj}_{\mathsf{V}}\colon x \mapsto (\mathsf{R}_j x)_{1\leqslant j\leqslant p+1+r}$, *where* $\mathsf{R}_1\colon \mathsf{H}\oplus\mathsf{G}\oplus\mathsf{K} \to \mathsf{H}$, $(\forall i \in \{1,\ldots,p\})$ $\mathsf{R}_{1+i}\colon \mathsf{H}\oplus\mathsf{G}\oplus\mathsf{K} \to \mathsf{G}_i$, *and* $(\forall k \in \{1,\ldots,r\})$ $\mathsf{R}_{p+1+k}\colon \mathsf{H}\oplus\mathsf{G}\oplus\mathsf{K} \to \mathsf{K}_k$. *Set* $\mathsf{O} = \{0,1\}^{p+1+r} \smallsetminus \{\mathbf{0}\}$, *let* $\gamma \in \left]0,+\infty\right[$, *let* $(\lambda_n)_{n\in\mathbb{N}}$ *be a sequence in* $\left]0,2\right[$ *such that* $\inf_{n\in\mathbb{N}} \lambda_n > 0$ *and* $\sup_{n\in\mathbb{N}} \lambda_n < 2$, *let* $\boldsymbol{x}_0$, $\boldsymbol{z}_0$, $(\boldsymbol{a}_n)_{n\in\mathbb{N}}$, *and* $(\boldsymbol{c}_n)_{n\in\mathbb{N}}$ *be* $\mathsf{H}\oplus\mathsf{G}$*-valued random variables, let* $\boldsymbol{y}_0$, $\boldsymbol{w}_0$, *and* $(\boldsymbol{b}_n)_{n\in\mathbb{N}}$ *be* K*-valued random variables, and let* $(\boldsymbol{\varepsilon}_n)_{n\in\mathbb{N}}$ *be identically distributed* O*-valued random variables. Iterate*

$$
\begin{array}{l}
\textit{for}\ n = 0,1,\ldots\\
\left\lfloor
\begin{array}{l}
x_{1,n+1} = x_{1,n} + \varepsilon_{1,n}\big(\mathsf{R}_1(\boldsymbol{z}_n,\boldsymbol{w}_n) + a_{1,n} - x_{1,n}\big)\\
z_{1,n+1} = z_{1,n} + \varepsilon_{1,n}\lambda_n\big(\mathsf{J}_{\gamma\mathsf{A}}\big(2x_{1,n+1} - z_{1,n}\big) + c_{1,n} - x_{1,n+1}\big)\\
\textit{for}\ k = 1,\ldots,p\\
\left\lfloor
\begin{array}{l}
x_{k+1,n+1} = x_{k+1,n} + \varepsilon_{k+1,n}\big(\mathsf{R}_{k+1}(\boldsymbol{z}_n,\boldsymbol{w}_n) + a_{k+1,n} - x_{k+1,n}\big)\\
z_{k+1,n+1} = z_{k+1,n} + \varepsilon_{k+1,n}\lambda_n\big(\mathsf{J}_{\gamma\mathsf{B}_k}\big(2x_{k+1,n+1} - z_{k+1,n}\big) + c_{k+1,n} - x_{k+1,n+1}\big)
\end{array}
\right.\\
\textit{for}\ j = 1,\ldots,r\\
\left\lfloor
\begin{array}{l}
y_{j,n+1} = y_{j,n} + \varepsilon_{p+1+j,n}\big(\mathsf{R}_{p+1+j}(\boldsymbol{z}_n,\boldsymbol{w}_n) + b_{j,n} - y_{j,n}\big)\\
w_{j,n+1} = w_{j,n} - \varepsilon_{p+1+j,n}\lambda_n y_{j,n+1}.
\end{array}
\right.
\end{array}
\right.
\end{array}
\tag{2.40}
$$

In addition, assume that the following are satisfied:

(i) $\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|\boldsymbol{a}_n\|^2 \,|\, \sigma(\boldsymbol{z}_l,\boldsymbol{w}_l)_{0\leqslant l\leqslant n})} < +\infty$, $\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|\boldsymbol{b}_n\|^2 \,|\, \sigma(\boldsymbol{z}_l,\boldsymbol{w}_l)_{0\leqslant l\leqslant n})} < +\infty$, $\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}(\|\boldsymbol{c}_n\|^2 \,|\, \sigma(\boldsymbol{z}_l,\boldsymbol{w}_l)_{0\leqslant l\leqslant n})} < +\infty$, $\boldsymbol{a}_n \rightharpoonup \mathbf{0}$ P-a.s., *and* $\boldsymbol{b}_n \rightharpoonup \mathbf{0}$ P-a.s.

(ii) *For every* $n \in \mathbb{N}$, $\sigma(\boldsymbol{\varepsilon}_n)$ *and* $\sigma(\boldsymbol{z}_l,\boldsymbol{w}_l)_{0\leqslant l\leqslant n}$ *are independent.*

(iii) *For every* $l \in \{1,\ldots,p+1+r\}$, $\mathsf{P}[\varepsilon_{l,0} = 1] > 0$.

Then $(x_{1,n})_{n\in\mathbb{N}}$ *converges weakly* P-a.s. *to a* Z*-valued random variable.*

Proof. In Problem 2.6, set $m = p+1$, $\mathsf{X}_1 = \mathsf{H}$, $(\mathsf{X}_i)_{2\leqslant i\leqslant m} = (\mathsf{G}_{i-1})_{2\leqslant i\leqslant m}$, $\mathsf{Y} = \mathsf{K}$, for every $j \in \{1,\ldots,r\}$, $\mathsf{D}_j = \mathsf{N}_{\{0\}}$, and, for every $i \in \{1,\ldots,m\}$, $\mathsf{M}_{ji} = \mathsf{E}_{ji}$. Thus, the subspace V of (2.18) becomes

$$\mathsf{V} = \left\{(x,y) \in \mathsf{X}\oplus\mathsf{Y} \;\middle|\; (\forall j \in \{1,\ldots,r\})\ y_j = \sum_{i=1}^{p+1} \mathsf{E}_{ji} x_i\right\}, \tag{2.41}$$

Further, denote by $x = (x_1, \ldots, x_{p+1})$ a generic element in $H \oplus G$. In this configuration, (2.17) reduces to

$$\text{find } x \in H \oplus G \text{ such that } \mathbf{0} \in \bigtimes_{i=1}^{p+1} C_i x_i + E^*\big(N_{\{\mathbf{0}\}}(Ex)\big). \tag{2.42}$$

We note that Proposition 2.12 is the application of Theorem 2.7 to (2.42) when

$$C_1 = A \quad \text{and} \quad (\forall i \in \{2, \ldots, p+1\}) \quad C_i = B_{i-1}. \tag{2.43}$$

Let $\boldsymbol{Z}$ be the set of solutions to (2.42) in the context of (2.43). Recalling that Z denotes the set of solutions to Problem 2.2, it remains to show that

$$\boldsymbol{Z} = \big\{(x_1, L_1x_1, \ldots, L_px_1) \mid x_1 \in Z\big\}. \tag{2.44}$$

Let $x \in H \oplus G$. It follows at once from (2.39) that

$$\iota_W(x) = \iota_{\{\mathbf{0}\}}(Ex). \tag{2.45}$$

Hence, we deduce from [6, Corollary 16.53] that

$$N_W x = E^*\big(N_{\{\mathbf{0}\}}(Ex)\big). \tag{2.46}$$

Note that the set in (2.46) is nonempty if and only if $x \in W$. Consequently,

$$x \in \boldsymbol{Z} \Leftrightarrow x \text{ solves (2.42)} \Leftrightarrow x \text{ solves (2.28)} \tag{2.47}$$

and the claim follows from (2.37). ☐

Let us provide some examples of implementations of Proposition 2.12.

Example 2.13 In Proposition 2.12, set $r = p$, $K = G$, and, for every $k \in \{1, \ldots, p\}$ and every $i \in \{1, \ldots, p+1\}$,

$$E_{ki} = \begin{cases} L_k, & \text{if } i = 1; \\ -\mathrm{Id}, & \text{if } i = k+1; \\ 0, & \text{otherwise.} \end{cases} \tag{2.48}$$

Let $x \in H \oplus G$, let $y \in G$, and set $q = (2\mathrm{Id} + \sum_{k=1}^{p} L_k^* \circ L_k)^{-1}(2x_1 + \sum_{k=1}^{p} L_k^*(x_{k+1} + y_k))$. Then, for every $i \in \{1, \ldots, p+1\}$,

$$R_i(x, y) = \begin{cases} q, & \text{if } i = 1; \\ \dfrac{1}{2}\big(L_{i-1}q + x_i - y_{i-1}\big), & \text{if } 2 \leqslant i \leqslant p+1; \\ \dfrac{1}{2}\big(L_{i-p-1}q - x_{i-p} + y_{i-p-1}\big), & \text{if } p+2 \leqslant i \leqslant 2p+1. \end{cases} \tag{2.49}$$

Let $(e_n)_{n\in\mathbb{N}}$ be H-valued random variables such that $\sum_{n\in\mathbb{N}}\sqrt{\mathsf{E}(\|e_n\|^2\,|\,\sigma(z_l,w_l)_{0\leqslant l\leqslant n})} < +\infty$ and $e_n \rightharpoonup 0$ P-a.s., and set

$$Q = \left(2\mathrm{Id} + \sum_{k=1}^{p} L_k^* \circ L_k\right)^{-1}. \tag{2.50}$$

Then (2.40) becomes

$$\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
q_n = Q\Big(2z_{1,n} + \sum_{k=1}^{p} L_k^*(z_{k+1,n} + w_{k,n})\Big) + e_n \\
x_{1,n+1} = x_{1,n} + \varepsilon_{1,n}(q_n - x_{1,n}) \\
z_{1,n+1} = z_{1,n} + \varepsilon_{1,n}\lambda_n\big(J_{\gamma A}\,(2x_{1,n+1} - z_{1,n}) + c_{1,n} - x_{1,n+1}\big) \\
\text{for } k = 1, \ldots, p \\
\left\lfloor
\begin{array}{l}
x_{k+1,n+1} = x_{k+1,n} + \varepsilon_{k+1,n}\left(\dfrac{L_k q_n + z_{k+1,n} - w_{k,n}}{2} - x_{k+1,n}\right) \\
z_{k+1,n+1} = z_{k+1,n} + \varepsilon_{k+1,n}\lambda_n\big(J_{\gamma B_k}\,(2x_{k+1,n+1} - z_{k+1,n}) + c_{k+1,n} - x_{k+1,n+1}\big)
\end{array}
\right. \\
\text{for } k = 1, \ldots, p \\
\left\lfloor
\begin{array}{l}
y_{k,n+1} = y_{k,n} + \varepsilon_{p+1+k,n}\left(\dfrac{L_k q_n - z_{k+1,n} + w_{k,n}}{2} - y_{k,n}\right) \\
w_{k,n+1} = w_{k,n} - \varepsilon_{p+1+k,n}\lambda_n y_{k,n+1}
\end{array}
\right.
\end{array}
\right.
\end{array} \tag{2.51}$$

and Proposition 2.12 asserts that $(x_{1,n})_{n\in\mathbb{N}}$ converges weakly P-a.s. to a solution to Problem 2.2.

The next examples focus on the special case of Problem 2.2 in which, for every $k \in \{1, \ldots, p\}$, $G_k = H$ and $L_k = \mathrm{Id}$, that is,

$$\text{find } \ x \in H \ \text{ such that } \ 0 \in Ax + \sum_{k=1}^{p} B_k x. \tag{2.52}$$

Example 2.14 Consider the setting of Example 2.13 where, for every $k \in \{1, \ldots, p\}$, $G_k = H$ and $L_k = \mathrm{Id}$. Then, in view of (2.48), the operator E is defined by setting, for every $k \in \{1, \ldots, p\}$ and every $i \in \{1, \ldots, p+1\}$,

$$E_{ki} = \begin{cases} \mathrm{Id}, & \text{if } \ i = 1; \\ -\mathrm{Id}, & \text{if } \ i = k+1; \\ 0, & \text{otherwise.} \end{cases} \tag{2.53}$$

Further, the operator Q of (2.50) is just $(p+2)^{-1}\mathrm{Id}$. Thus, (2.40) becomes

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
q_n = \dfrac{1}{p+2}\Big(2z_{1,n} + \displaystyle\sum_{k=1}^{p}(z_{k+1,n} + w_{k,n})\Big) \\
x_{1,n+1} = x_{1,n} + \varepsilon_{1,n}(q_n - x_{1,n}) \\
z_{1,n+1} = z_{1,n} + \varepsilon_{1,n}\lambda_n\big(J_{\gamma A}\,(2x_{1,n+1} - z_{1,n}) + c_{1,n} - x_{1,n+1}\big) \\
\text{for } k = 1, \ldots, p \\
\left\lfloor
\begin{array}{l}
x_{k+1,n+1} = x_{k+1,n} + \varepsilon_{k+1,n}\Big(\dfrac{q_n + z_{k+1,n} - w_{k,n}}{2} - x_{k+1,n}\Big) \\
z_{k+1,n+1} = z_{k+1,n} + \varepsilon_{k+1,n}\lambda_n\big(J_{\gamma B_k}\,(2x_{k+1,n+1} - z_{k+1,n}) + c_{k+1,n} - x_{k+1,n+1}\big)
\end{array}
\right. \\
\text{for } k = 1, \ldots, p \\
\left\lfloor
\begin{array}{l}
y_{k,n+1} = y_{k,n} + \varepsilon_{p+1+k,n}\Big(\dfrac{q_n - z_{k+1,n} + w_{k,n}}{2} - y_{k,n}\Big) \\
w_{k,n+1} = w_{k,n} - \varepsilon_{p+1+k,n}\lambda_n y_{k,n+1}
\end{array}
\right.
\end{array}
\right.
\end{array}
\tag{2.54}
$$

and Proposition 2.12 asserts that $(x_{1,n})_{n\in\mathbb{N}}$ converges weakly P-a.s. to a solution to (2.52).

Example 2.15 In Proposition 2.12, set $r = p+1$, $K = H^{p+1}$, and, for every $k \in \{1, \ldots, p+1\}$ and every $i \in \{1, \ldots, p+1\}$,

$$
E_{ki} = \begin{cases} \dfrac{p}{p+1}\mathrm{Id}, & \text{if } k = i; \\ -\dfrac{1}{p+1}\mathrm{Id}, & \text{if } k \neq i. \end{cases}
\tag{2.55}
$$

Then $\ker E$ is the subspace of all the vectors $x \in H^{p+1}$ such that, for every $i \in \{1, \ldots, p+1\}$, $x_i = (p+1)^{-1}\sum_{j=1}^{p+1} x_j$. Hence, for every $i \in \{1, \ldots, 2p+2\}$, every $x \in H^{p+1}$, and every $y \in H^{p+1}$,

$$
R_i(x, y) = \begin{cases} \dfrac{1}{2}(x_i + y_i) + \dfrac{1}{2(p+1)}\displaystyle\sum_{j=1}^{p+1}(x_j - y_j), & \text{if } i \leqslant p+1; \\ \dfrac{1}{2}(x_i + y_i) - \dfrac{1}{2(p+1)}\displaystyle\sum_{j=1}^{p+1}(x_j + y_j), & \text{if } p+2 \leqslant i \leqslant 2p+2. \end{cases}
\tag{2.56}
$$

Then (2.40) becomes

$$
\begin{array}{l}
\text{for } n = 0, 1, \dots \\
\left\lfloor
\begin{array}{l}
x_{1,n+1} = x_{1,n} + \varepsilon_{1,n}\Big(\dfrac{z_{1,n} + w_{1,n}}{2} + \dfrac{1}{2(p+1)}\displaystyle\sum_{l=1}^{p+1}(z_{l,n} - w_{l,n}) - x_{1,n}\Big) \\
z_{1,n+1} = z_{1,n} + \varepsilon_{1,n}\lambda_n\big(J_{\gamma A}\,(2x_{1,n+1} - z_{1,n}) + c_{1,n} - x_{1,n+1}\big) \\
\text{for } k = 1, \dots, p \\
\left\lfloor
\begin{array}{l}
x_{k+1,n+1} = x_{k+1,n} + \varepsilon_{k+1,n}\Big(\dfrac{z_{k+1,n} + w_{k+1,n}}{2} + \dfrac{1}{2(p+1)}\displaystyle\sum_{l=1}^{p+1}(z_{l,n} - w_{l,n}) - x_{k+1,n}\Big) \\
z_{k+1,n+1} = z_{k+1,n} + \varepsilon_{k+1,n}\lambda_n\big(J_{\gamma B_k}\,(2x_{k+1,n+1} - z_{k+1,n}) + c_{k+1,n} - x_{k+1,n+1}\big)
\end{array}
\right. \\
\text{for } j = 1, \dots, p+1 \\
\left\lfloor
\begin{array}{l}
y_{j,n+1} = y_{j,n} + \varepsilon_{p+1+j,n}\Big(\dfrac{z_{j,n} + w_{j,n}}{2} - \dfrac{1}{2(p+1)}\displaystyle\sum_{l=1}^{p+1}(z_{l,n} + w_{l,n}) - y_{j,n}\Big) \\
w_{j,n+1} = w_{j,n} - \varepsilon_{p+1+j,n}\lambda_n y_{j,n+1}
\end{array}
\right.
\end{array}
\right.
\tag{2.57}
$$

and Proposition 2.12 asserts that $(x_{1,n})_{n\in\mathbb{N}}$ converges weakly P-a.s. to a solution to (2.52).

Remark 2.16 In Example 2.14, the operator E applied to $x \in H^{p+1}$ couples each agent in $(x_2, \dots, x_{p+1})$ with x_1. In Example 2.15 the operator E applied to $x \in H^{p+1}$ couples each agent in $(x_1, \dots, x_{p+1})$ with the average of all the agents. Various alternative coupling operators E can be considered to enforce the condition $x_1 = \cdots = x_{p+1}$.

2.2.3.5 Computation of inverse operators

The existing algorithms presented in Section 2.2.2.2 require the computation of norms of arbitrary linear operators whereas the proposed algorithms of Sections 2.2.3.2–2.2.3.4 require the inversion of strongly positive Hermitian operators of the type $\mathrm{Id} + L^* \circ L$. Note that, because of the strongly positive Hermitian structure of $\mathrm{Id}+L^*\circ L$, the computation of the inverse is typically much cheaper than the computation of the norm of L in (2.12) or those of $(U_k^{1/2} L_k W^{1/2})_{1\leqslant k\leqslant p}$ in (2.14). In a finite-dimensional setting, in full generality, if $\mathrm{Id} + L^* \circ L$ has size N, then its inversion via the Cholesky decomposition method requires about $N^3/6$ multiplications. However, this complexity can be reduced in several standard scenarios. Here are two examples in $H = \mathbb{R}^N$ that will be used in Section 2.2.5.

Example 2.17

(i) If, for every $k \in \{1, \dots, p\}$, $L_k = \mathrm{Id}$, then

$$
\begin{cases}
\Big(\mathrm{Id} + \sum_{k=1}^{p} L_k^* \circ L_k\Big)^{-1} = \dfrac{1}{1+p}\mathrm{Id} \\
\Big(2\mathrm{Id} + \sum_{k=1}^{p} L_k^* \circ L_k\Big)^{-1} = \dfrac{1}{2+p}\mathrm{Id}.
\end{cases}
\tag{2.58}
$$

The cost of the inversion is $\mathcal{O}(1)$.

(ii) Suppose that, for every $\mathsf{k} \in \{1, \ldots, \mathsf{p}\}$, L_k is a block-Toeplitz. Then, following a standard argument [2], each L_k can be approximated by a block-circulant matrix with convolution kernel ℓ_k and

$$\begin{cases} \left(\mathrm{Id} + \displaystyle\sum_{k=1}^{p} L_k^* \circ L_k\right)^{-1} & : x \mapsto \mathfrak{F}^{-1}\left(\mathfrak{F}(x) \div \left(1 + \displaystyle\sum_{k=1}^{p} |\mathfrak{F}(\ell_k)|^2\right)\right) \\ \left(2\mathrm{Id} + \displaystyle\sum_{k=1}^{p} L_k^* \circ L_k\right)^{-1} & : x \mapsto \mathfrak{F}^{-1}\left(\mathfrak{F}(x) \div \left(2 + \displaystyle\sum_{k=1}^{p} |\mathfrak{F}(\ell_k)|^2\right)\right). \end{cases} \tag{2.59}$$

where $\mathfrak{F}$ denotes the discrete Fourier transform and $\div$ denotes pointwise division. The cost of the inversion using the fast Fourier transform is $\mathcal{O}(\mathsf{N}\log(\mathsf{N}))$ [2].

(iii) The worst case is if the operators $(\mathsf{L}_\mathsf{k})_{1\leqslant \mathsf{k}\leqslant \mathsf{p}}$ do not present a special structure. Even so, the composed operators $\mathrm{Id} + \sum_{k=1}^{p} L_k^* \circ L_k$ and $2\mathrm{Id} + \sum_{k=1}^{p} L_k^* \circ L_k$ are symmetric and positive-definite. Hence they admit a Cholesky decomposition. The cost of computing the Cholesky decomposition is $\mathcal{O}(\mathsf{N}^3)$ (one time) and the cost of solving the linear system using the Cholesky decomposition is $\mathcal{O}(\mathsf{N}^2)$. It will be shown in the numerical experiments that in the case when no special structure is present, Framework 2 is preferred since the application of the inverse operator does not occur at every iteration.

2.2.4 Minimization problems

We dedicate this section to the minimization setting of Problem 2.1. Let us first formalize the connection between Problems 2.1 and 2.2.

Proposition 2.18 *In Problem 2.2, set* $\mathsf{A} = \partial \mathsf{f}$ *and* $(\forall \mathsf{k} \in \{1, \ldots, \mathsf{p}\})$ $\mathsf{B}_\mathsf{k} = \partial \mathsf{g}_\mathsf{k}$. *Then every solution to (2.2) solves Problem 2.1.*

Proof. Set $\mathsf{L} \colon \mathsf{H} \to \mathsf{G} \colon x \mapsto (\mathsf{L}_1 x, \ldots, \mathsf{L}_\mathsf{p} x)$ and $\mathsf{g} \colon \mathsf{G} \to \left]-\infty, +\infty\right] \colon y \mapsto \sum_{k=1}^{p} g_k(y_k)$. Then $\mathsf{L}^* \colon \mathsf{G} \to \mathsf{H} \colon y \mapsto \sum_{k=1}^{p} L_k^* y_k$. Hence, it follows from [6, Proposition 16.9] that

$$x \in \operatorname{zer}\left(\partial f + \sum_{k=1}^{p} L_k^* \circ (\partial g_k) \circ L_k\right) = \operatorname{zer}\big(\partial f + L^* \circ (\partial g) \circ L\big). \tag{2.60}$$

However, [6, Proposition 27.5(i)] asserts that

$$\operatorname{zer}\left(\partial f + \sum_{k=1}^{p} L^* \circ (\partial g) \circ L\right) \subset \operatorname{Argmin}(f + g \circ L) = \operatorname{Argmin}\left(f + \sum_{k=1}^{p} g_k \circ L_k\right), \tag{2.61}$$

which confirms the claim. ☐

Problem 2.1 relies on the assumption that $\operatorname{zer}(\partial f + \sum_{k=1}^{p} L_k^* \circ (\partial g_k) \circ L_k) \neq \varnothing$. Let us provide sufficient conditions that guarantee it.

Proposition 2.19 *Let* H *be a separable real Hilbert space and* $\mathsf{f} \in \Gamma_0(\mathsf{H})$. *For every* $\mathsf{k} \in \{1, \ldots, \mathsf{p}\}$, *let* G_k *be a separable real Hilbert space, let* $\mathsf{g}_\mathsf{k} \in \Gamma_0(\mathsf{G}_\mathsf{k})$, *and let* $0 \neq \mathsf{L}_\mathsf{k} \colon \mathsf{H} \to \mathsf{G}_\mathsf{k}$ *be linear and bounded. Set*

$$\mathsf{S} = \big\{(\mathsf{L}_\mathsf{k}\mathsf{x} - \mathsf{y}_\mathsf{k})_{1\leqslant \mathsf{k}\leqslant \mathsf{p}} \mid \mathsf{x} \in \operatorname{dom}\mathsf{f} \text{ and } (\forall \mathsf{k} \in \{1, \ldots, \mathsf{p}\})\ \ \mathsf{y}_\mathsf{k} \in \operatorname{dom}\mathsf{g}_\mathsf{k}\big\}. \tag{2.62}$$

Then $\operatorname{zer}(\partial\mathsf{f} + \sum_{\mathsf{k}=1}^{\mathsf{p}} \mathsf{L}_\mathsf{k}^* \circ (\partial\mathsf{g}_\mathsf{k}) \circ \mathsf{L}_\mathsf{k}) \neq \varnothing$ *if the following hold:*

(i) $\mathsf{f}(\mathsf{x}) + \sum_{\mathsf{k}=1}^{\mathsf{p}} \mathsf{g}_\mathsf{k}(\mathsf{L}_\mathsf{k}\mathsf{x}) \to +\infty$ *as* $\|\mathsf{x}\| \to +\infty$.

(ii) *Any of the following is satisfied:*

  (a) *The cone generated by* S *is a closed vector subspace of* G.

  (b) *For every* $\mathsf{k} \in \{1, \ldots, \mathsf{p}\}$, g_k *is real-valued.*

  (c) H *and* $(\mathsf{G}_\mathsf{k})_{1\leqslant \mathsf{k}\leqslant \mathsf{p}}$ *are finite-dimensional, and there exists* $\mathsf{x} \in \operatorname{ri}\operatorname{dom}\mathsf{f}$ *such that*

$$(\forall \mathsf{k} \in \{1, \ldots, \mathsf{p}\}) \quad \mathsf{L}_\mathsf{k}\mathsf{x} \in \operatorname{ri}\operatorname{dom}\mathsf{g}_\mathsf{k}, \tag{2.63}$$

  where ri *stands for the relative interior.*

Proof. Set $\mathsf{L} \colon \mathsf{H} \to \mathsf{G} \colon \mathsf{x} \mapsto (\mathsf{L}_1\mathsf{x}, \ldots, \mathsf{L}_\mathsf{p}\mathsf{x})$ and $\mathsf{g} \colon \mathsf{G} \to \left]-\infty, +\infty\right] \colon \mathsf{y} \mapsto \sum_{\mathsf{k}=1}^{\mathsf{p}} \mathsf{g}_\mathsf{k}(\mathsf{y}_\mathsf{k})$. Then L is linear and bounded, $\mathsf{g} \in \Gamma_0(\mathsf{G})$, $\mathsf{S} = \big\{\mathsf{L}\mathsf{x} - \mathsf{y} \mid \mathsf{x} \in \operatorname{dom}\mathsf{f} \text{ and } \mathsf{y} \in \operatorname{dom}\mathsf{g}\big\}$, and $\mathsf{f} + \mathsf{g} \circ \mathsf{L} = \mathsf{f} + \sum_{\mathsf{k}=1}^{\mathsf{p}} \mathsf{g}_\mathsf{k} \circ \mathsf{L}_\mathsf{k}$. On the other hand, it follows from (ii) that $\mathbf{0} \in \mathsf{S}$, which implies that $\operatorname{dom}(\mathsf{f} + \mathsf{g} \circ \mathsf{L}) \neq \varnothing$. Thus, because $\mathsf{f} + \mathsf{g} \circ \mathsf{L}$ is also lower semicontinuous and convex, we have $\mathsf{f} + \mathsf{g} \circ \mathsf{L} \in \Gamma_0(\mathsf{H})$. Hence, since (i) states that $\mathsf{f}(\mathsf{x}) + \mathsf{g}(\mathsf{L}\mathsf{x}) \to +\infty$ as $\|\mathsf{x}\| \to +\infty$, it follows from [6, Proposition 11.15(i)] that

$$\operatorname{Argmin}(\mathsf{f} + \mathsf{g} \circ \mathsf{L}) \neq \varnothing. \tag{2.64}$$

However, (ii) and [6, Proposition 27.5(iii)] guarantee that

$$\operatorname{Argmin}(\mathsf{f} + \mathsf{g} \circ \mathsf{L}) = \operatorname{zer}\big(\partial\mathsf{f} + \mathsf{L}^* \circ (\partial\mathsf{g}) \circ \mathsf{L}\big), \tag{2.65}$$

which completes the proof. ☐

In view of Proposition 2.18 and (2.6), we obtain the following solution methods for Problem 2.1.

Corollary 2.20 *Consider the setting of Problem* 2.1 *and set* $\mathsf{F} = \operatorname{Argmin}(\mathsf{f} + \sum_{\mathsf{k}=1}^{\mathsf{p}} \mathsf{g}_\mathsf{k} \circ \mathsf{L}_\mathsf{k})$. *In* (2.23), (2.25), *and* (2.40), *replace the resolvent operators* $(\mathsf{J}_{\gamma\mathsf{A}}, \mathsf{J}_{\gamma\mathsf{B}_1}, \ldots, \mathsf{J}_{\gamma\mathsf{B}_\mathsf{p}})$ *by the proximity operators* $(\operatorname{prox}_{\gamma\mathsf{f}}, \operatorname{prox}_{\gamma\mathsf{g}_1}, \ldots, \operatorname{prox}_{\gamma\mathsf{g}_\mathsf{p}})$. *Then Propositions* 2.10, 2.11, *and* 2.12 *provide sequences* $(x_{1,n})_{n\in\mathbb{N}}$ *which converge weakly* P-a.s. *to an* F*-valued random variable.*

2.2.5 Numerical experiments

2.2.5.1 Preamble

We present four experiments to illustrate the numerical behavior of the three algorithmic frameworks presented in Section 2.2.3. These algorithms are initialized by setting $\boldsymbol{x}_0$, z_0, $\boldsymbol{y}_0$, and $\boldsymbol{w}_0$ to $\mathbf{0}$, and they use the proximal parameter $\gamma = 1.0$ and the relaxation strategy $(\forall n \in \mathbb{N})\ \lambda_n = 1.9$. The random variable ε_0 activates operator indices in $\{1, \ldots, p+1\}$ (Framework 1), $\{1, \ldots, p+2\}$ (Framework 2), and $\{1, \ldots, 2p+1\}$ (Framework 3 using Example 2.13), with a uniform distribution.

We also provide comparisons with the existing methods of Section 2.2.2.2 when applicable, because they do provide almost sure iterate convergence to a solution, although they do not satisfy the requirements **R2**–**R3**:

- Algorithm (2.13) is initialized with $x_{1,0} = 0$ and $\boldsymbol{y}_0 = \mathbf{0}$. Further, for every $k \in \{1, \ldots, p\}$, $\pi_k = 1/p$ and, to enforce (2.12), we set $\tau_0 = 0.9/\sqrt{p}$ and $\sigma_0 = 1/\big(\sqrt{p} \max_{1 \leqslant k \leqslant p} \|L_k\|^2\big)$. In addition we set $\chi_0 = 0.5$, $\eta = 0.5$, and $\delta = 1.5$. We recall that algorithm (2.13) can activate only one operator at each iteration and does not satisfy **R2**–**R4**.
- Algorithm (2.15) is initialized with $x_{1,0} = 0$ and $\boldsymbol{v}_0 = \mathbf{0}$. Further, $W = 0.9\tau\mathrm{Id}$ and, for every $k \in \{1, \ldots, p\}$, $U_k = (\tau/\|L_k\|^2)\mathrm{Id}$, where $\lambda_n \equiv 1$ and, to enforce (2.14), $\tau = 1/\sqrt{2p}$. We recall that algorithm (2.15) does not satisfy **R2**–**R3**.

These parameters were found to enhance the performance of these two algorithms. The first three experiments (Sections 2.2.5.2–2.2.5.4) correspond to minimization problems fitting the format of Problem 2.1. The last experiment (Section 2.2.5.5) is a nonminimization problem that fits the format of Problem 2.2, and algorithm (2.13) is therefore not applicable.

2.2.5.2 Signal restoration

The goal is to recover the original signal $\overline{x} \in H = \mathbb{R}^N$ ($N = 1000$) shown in Figure 2.2(a) from $M = 10$ noisy observations $(r_l)_{1 \leqslant l \leqslant M}$ given by

$$(\forall l \in \{1, \ldots, M\}) \quad r_l = L_l \overline{x} + w_l \tag{2.66}$$

where, for every $l \in \{1, \ldots, M\}$, $L_l \colon \mathbb{R}^N \to \mathbb{R}^N$ is a known linear operator, $\eta_l \in \left]0, +\infty\right[$, and $w_l \in [-\eta_l, \eta_l]^N$ is the realization of a bounded random noise vector. The parameters $(\eta_l)_{1 \leqslant l \leqslant M} \in \left]0, +\infty\right[^M$ are not known exactly and are underestimated by $(\xi_l)_{1 \leqslant l \leqslant M} \in \left]0, +\infty\right[^M$. For every $l \in \{1, \ldots, M\}$, L_l is a Gaussian convolution filter with zero mean and standard deviation taken uniformly in $[20, 40]$, $\eta_l = 0.1$, w_l is taken uniformly in $[-\eta_l, \eta_l]^N$, and $\xi_l = 0.07$. For every $l \in \{1, \ldots, M\}$ and every $j \in \{1, \ldots, N\}$, set $Z_{l,j} = [\langle r_l \mid e_j \rangle - \xi_l, \langle r_l \mid e_j \rangle + \xi_l]$. Since the intersection of these sets is empty, we cannot recover the signal by solving the associated convex feasibility problem. Instead, our objective is to solve an instantiation of Problem 2.1 with $p = MN$, to wit,

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \;\; \alpha\|x\| + \sum_{l=1}^{M}\sum_{j=1}^{N} d_{Z_{l,j}}\big(\langle L_l x \mid e_j\rangle\big), \tag{2.67}$$

where $\alpha = 0.05$. Since, for every $x \in \mathbb{R}^N$, $\alpha\|x\| \leqslant \alpha\|x\| + \sum_{l=1}^{M}\sum_{j=1}^{N} d_{Z_{l,j}}(\langle L_l x \mid e_j\rangle)$, condition (i) in Proposition 2.19 holds. In addition, for every $l \in \{1,\ldots,M\}$ and every $j \in \{1,\ldots,N\}$, $d_{Z_{l,j}}$ is real-valued. Hence, condition (ii)(b) in Proposition 2.19 holds as well, which confirms that (2.67) is an instance of Problem 2.1. We can thus invoke Corollary 2.20. The three frameworks of Sections 2.2.3.2–2.2.3.4 are used to solve (2.67), where the operator E in Proposition 2.12 is that of Example 2.13. Two experiments are conducted: the random variable ε_0 produces (a) 1 activation with 1 core; and (b) 8 activations with 8 cores. Given $\gamma \in \left]0,+\infty\right[$, the operators $(\mathrm{prox}_{\gamma d_{Z_{l,j}}})_{1\leqslant l\leqslant M, 1\leqslant j\leqslant N}$ are computed via [6, Example 24.28] and $\mathrm{prox}_{\gamma\|\cdot\|}$ via [6, Example 24.20]. Furthermore, the convolutions and the inversions of linear operators are implemented using the fast Fourier transform [2]; see Example 2.17(ii). As mentioned in Section 2.2.5.1, we also compare with:

- Algorithm (2.13), which can activate only one operator at each iteration.
- Algorithm (2.15), where the random variable ε_0 activates (a) 1; and (b) 8 indices in $\{1,\ldots,p\}$ with a uniform distribution at each iteration.

The solution produced by Framework 2 is shown in Figure 2.3. We display in Figure 2.4 the normalized error versus execution time.

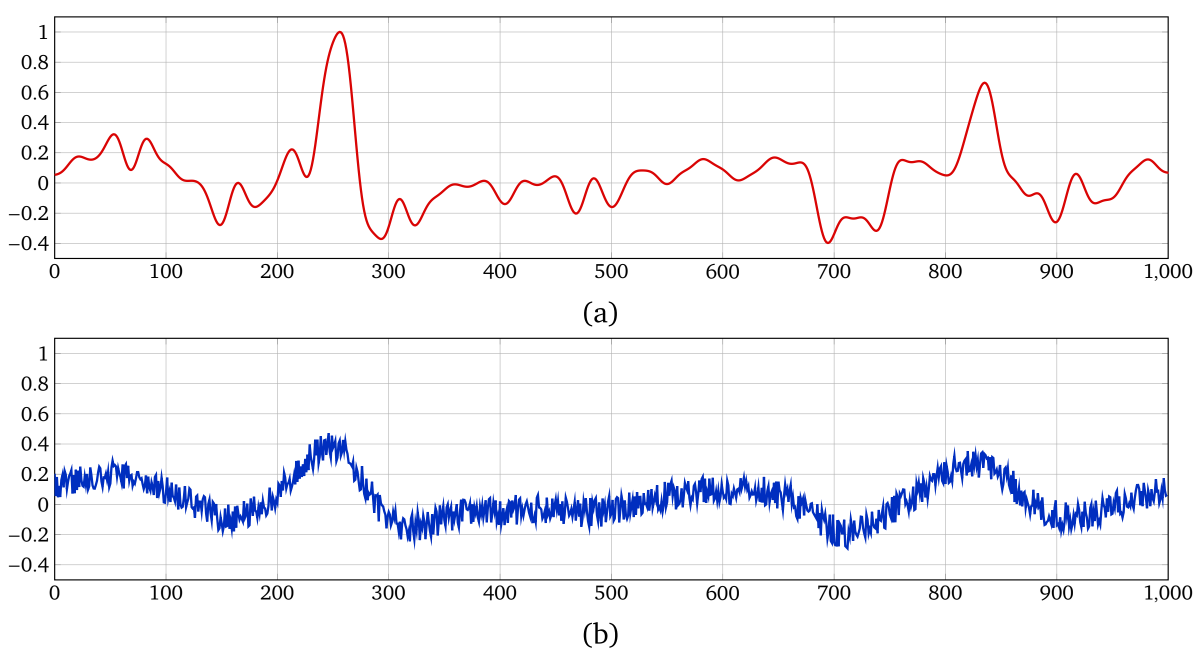


Figure 2.2 Experiment of Section 2.2.5.2. (a): Original signal $\overline{x}$. (b): Noisy observation r_1.

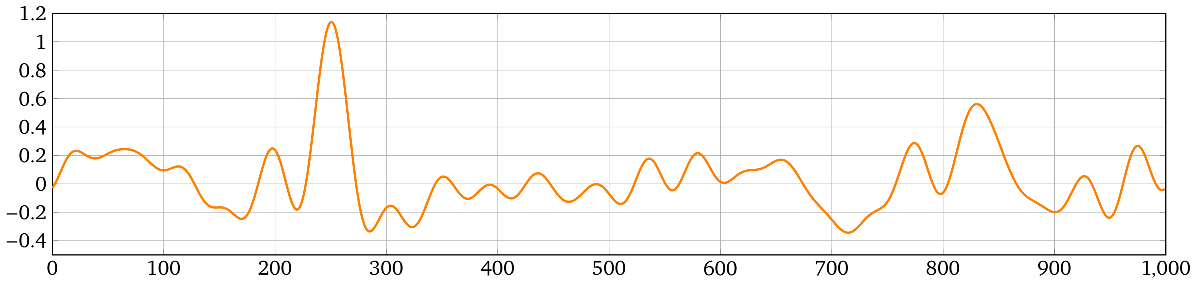


Figure 2.3 Experiment of Section 2.2.5.2. Solution produced by Framework 2.

2.2.5.3 Overlapping group lasso regression

We address the overlapping group lasso regression problem of [44]. Here $H = \mathbb{R}^N$ and q groups of indices $(I_k)_{1 \leqslant k \leqslant q}$ in $\{1, \ldots, N\}$ are present, with $\bigcup_{k=1}^{q} I_k = \{1, \ldots, N\}$. In addition, for every $k \in \{1, \ldots, q\}$,

$$S_k \colon \mathbb{R}^N \to \mathbb{R}^{\operatorname{card} I_k} \colon x = (\xi_j)_{1 \leqslant j \leqslant N} \mapsto (\xi_j)_{j \in I_k}. \tag{2.68}$$

The goal is to

$$\underset{x \in \mathbb{R}^N}{\text{minimize}} \;\; \frac{\alpha}{2}\|Ax - b\|^2 + \frac{1}{q}\sum_{k=1}^{q}\|S_k x\|, \tag{2.69}$$

where $A \in \mathbb{R}^{M \times N}$, $b = (\beta_l)_{1 \leqslant l \leqslant M} \in \mathbb{R}^M$, and $\alpha \in \,]0, +\infty[$. In the experiment, $M = 1200$, $N = 3610$, $q = 40$, and, as in [44], $\alpha = 5/q^2$. The entries of A are independent and identically distributed (i.i.d.) samples from a $\mathcal{N}(1, 10)$ distribution. The entries of the reference vector $\bar{x} \in \mathbb{R}^N$ are i.i.d. samples from a uniform distribution on $[0, 10]$, and $b = A\bar{x} + w$, where $w \in \mathbb{R}^M$ has entries that are i.i.d. samples from a $\mathcal{N}(0, 0.1)$ distribution. We split the term $\|Ax - b\|^2$ into a sum of 30

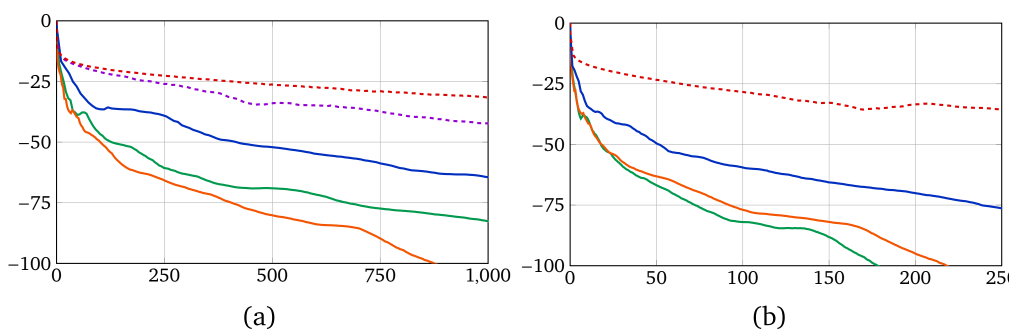


Figure 2.4 Experiment of Section 2.2.5.2. Normalized error $20\log_{10}(\|x_{1,n} - x_\infty\|/\|x_{1,0} - x_\infty\|)$ (dB) versus execution time (s). (a): Block size 1 with 1 core. (b): Block size 8 with 8 cores. Green: Framework 1. Orange: Framework 2. Blue: Framework 3 with Example 2.13. Dashed violet: Algorithm (2.13). Dashed red: Algorithm (2.15).

blocks of 40 entries each. Finally, the groups are defined by

$$(\forall k \in \{1,\ldots,q\}) \quad \mathrm{I}_k = \{90k-89,\ldots,90k+10\}. \tag{2.70}$$

Let $(a_l)_{1\leqslant l\leqslant M}$ be the rows of A. Then (2.69) is equivalent to

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \quad \sum_{k=1}^{p} g_k(L_k x), \tag{2.71}$$

where $p = 70$,

$$(\forall k \in \{1,\ldots,30\}) \quad \begin{cases} L_k\colon \mathbb{R}^N \to \mathbb{R}^{40}\colon x \mapsto \big(\langle x \mid a_l\rangle\big)_{40(k-1)+1\leqslant l\leqslant 40k} \\ g_k\colon \mathbb{R}^{40} \to \mathbb{R}\colon y \mapsto \dfrac{\alpha}{2}\|y-(\beta_l)_{40(k-1)+1\leqslant l\leqslant 40k}\|^2, \end{cases} \tag{2.72}$$

and

$$(\forall k \in \{31,\ldots,70\}) \quad \begin{cases} L_k\colon \mathbb{R}^N \to \mathbb{R}^{100}\colon x \mapsto S_{k-30}x \\ g_k\colon \mathbb{R}^{100} \to \mathbb{R}\colon y \mapsto \dfrac{1}{q}\|y\|. \end{cases} \tag{2.73}$$

Let $x = (\xi_l)_{1\leqslant l\leqslant N} \in \mathbb{R}^N$ and $j \in \{1,\ldots,N\}$. Since $\bigcup_{k=1}^{q} \mathrm{I}_k = \{1,\ldots,N\}$,

$$\frac{1}{q}\sum_{k=1}^{q}\|S_k x\| = \frac{1}{q}\sum_{k=1}^{q}\|(\xi_l)_{l\in \mathrm{I}_k}\| = \frac{1}{q}\sum_{k=1}^{q}\sqrt{\sum_{l\in \mathrm{I}_k}|\xi_l|^2} \geqslant \frac{1}{q}|\xi_j|. \tag{2.74}$$

In turn,

$$\sum_{k=1}^{70} g_k(L_k x) = \|Ax-b\|^2 + \frac{1}{q}\sum_{k=1}^{q}\|S_k x\| \geqslant \frac{1}{qN}\sum_{j=1}^{N}|\xi_j| \geqslant \frac{1}{qN}\sqrt{\sum_{j=1}^{N}|\xi_j|^2} = \frac{1}{qN}\|x\|, \tag{2.75}$$

which ensures that condition (i) in Proposition 2.19 holds. In addition, for every $k \in \{1,\ldots,70\}$, g_k is real-valued. Hence, condition (ii)(b) in Proposition 2.19 holds as well. Therefore Proposition 2.19 guarantees that (2.71) is an instance of Problem 2.1 and we invoke Corollary 2.20 to justify the convergence of the algorithms. We employ the three frameworks of Sections 2.2.3.2–2.2.3.4 to solve (2.69), where the operator E in Proposition 2.12 is that defined in Example 2.13. Two experiments are conducted: the random variable $\boldsymbol{\varepsilon}_0$ produces (a) 1 activation with 1 core; and (b) 8 activations with 8 cores. Given $\gamma \in \left]0,+\infty\right[$ and $z \in \mathbb{R}^{40}$, we compute $\mathrm{prox}_{\gamma\|\cdot\|}$ via [6, Example 24.20], $\mathrm{prox}_{\gamma\|\cdot-z\|^2}$ via [6, Proposition 24.8(i)], and the inverse operators by solving the linear systems with Example 2.17(iii). We also compare with:

- Algorithm (2.13).
- Algorithm (2.15), where the random variable $\boldsymbol{\varepsilon}_0$ activates (a) 1; and (b) 8 indices in $\{1,\ldots,p\}$ with a uniform distribution at each iteration.

We display in Figure 2.5 the normalized error versus execution time.

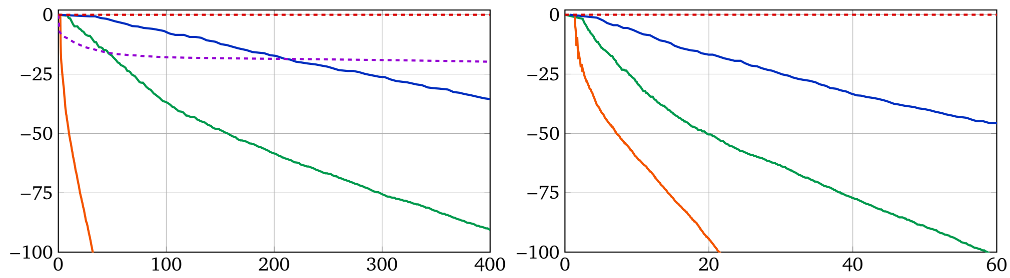


Figure 2.5 Experiment of Section 2.2.5.3. Normalized error $20\log_{10}(\|x_{1,n}-x_\infty\|/\|x_{1,0}-x_\infty\|)$ (dB) versus execution time (s). (a): Block size 1 with 1 core. (b): Block size 8 with 8 cores. Green: Framework 1. Orange: Framework 2. Blue: Framework 3 with Example 2.13. Dashed violet: Algorithm (2.13). Dashed red: Algorithm (2.15).

2.2.5.4 Classification using the hinge loss

We address a binary classification problem. The training data samples $(u_k,\xi_k)_{1\leqslant k\leqslant p}$ are in $\mathbb{R}^N\times\{-1,1\}$ and the goal is to learn a linear classifier $x\in H=\mathbb{R}^N$. For this purpose, we solve the instance of Problem 2.1 corresponding to the support vector machine model

$$\underset{x\in\mathbb{R}^N}{\text{minimize}}\;\;\frac{\alpha}{2}\|x\|^2+\frac{1}{p}\sum_{k=1}^{p}g_k(x), \tag{2.76}$$

where $\alpha\in\left]0,+\infty\right[$ and, for every $k\in\{1,\ldots,p\}$, $g_k\colon x\mapsto\max\{0,1-\xi_k\langle x\mid u_k\rangle\}$. In the experiment, $N=1500$, $\alpha=1$, $p=750$, and, for every $k\in\{1,\ldots,p\}$, the entries of u_k are i.i.d. samples from a $\mathcal{N}(100,10)$ distribution, and $(\xi_k)_{1\leqslant k\leqslant p}$ are i.i.d. samples from a uniform distribution on $\{-1,1\}$. Since, for every $x\in\mathbb{R}^N$, $(\alpha/2)\|x\|^2\leqslant(\alpha/2)\|x\|^2+\sum_{k=1}^{p}g_k(x)$, condition (i) in Proposition 2.19 holds. In addition, for every $k\in\{1,\ldots,p\}$, g_k is real-valued, so that condition (ii)(b) in Proposition 2.19 holds as well. This guarantees that (2.76) is an instance of Problem 2.1 and we can therefore invoke Corollary 2.20. We employ four methods to solve this problem: Framework 1, Framework 2, and Framework 3 using the operators E defined in Examples 2.14 and 2.15. In the case of Example 2.15 in Framework 3, the random variable ε_0 activates indices uniformly in $\{1,\ldots,2p+2\}$. Three experiments are conducted: the random variable ε_0 produces (a) 1 activation with 1 core; (b) 8 activations with 8 cores, and (c) 32 activations with 32 cores. Given $\gamma\in\left]0,+\infty\right[$, the operators $(\operatorname{prox}_{\gamma g_k})_{1\leqslant k\leqslant p}$ are computed via [6, Example 24.37]. The inverse operators are explicitly computed in Example 2.17(i).

We also compare with:

- Algorithm (2.13), which can activate only one operator at each iteration.

- Algorithm (2.15), where the random variable ε_0 activates (a) 1; (b) 8; and (c) 32 indices in $\{1,\ldots,\mathsf{p}\}$ with a uniform distribution at each iteration.

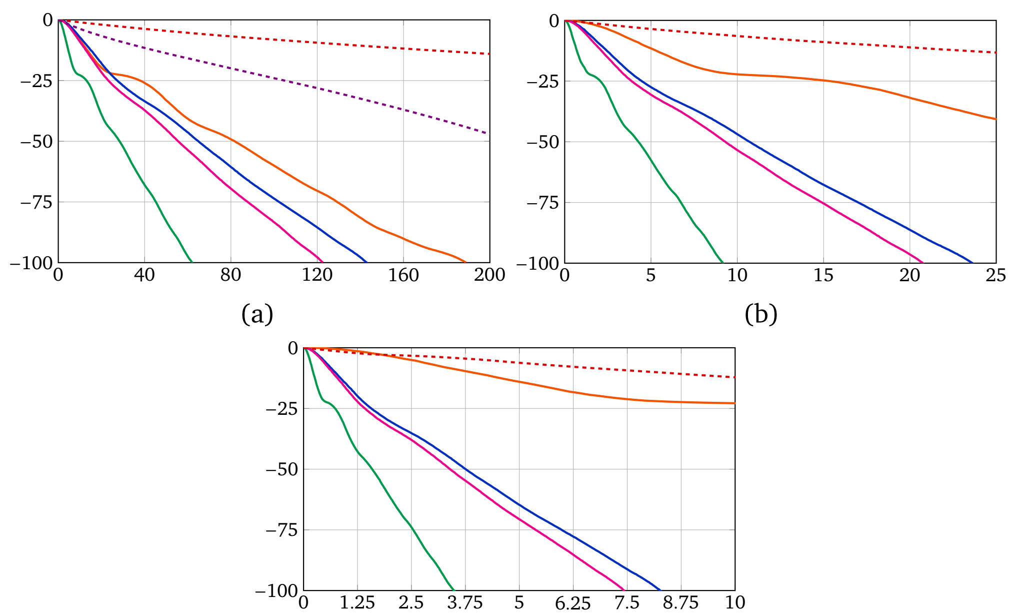
We display in Figure 2.6 the normalized error versus execution time for each instance. The execution time is evaluated based on the assumption that the computation corresponding to each selected index is assigned to a dedicated core and that all the cores are working in parallel.

2.2.5.5 Image reconstruction from phase

In contrast with the previous examples, we consider a data analysis framework, first proposed in [18], which requires the monotone inclusion format of Problem 2.2 and is not reducible to the minimization setting of Problem 2.1. The goal is to recover an image in a nonempty closed convex subset C of H from p nonlinear observations $(\mathsf{r}_k)_{1\leqslant k\leqslant \mathsf{p}}$ produced by Wiener models, namely,

$$\text{find } \mathsf{x}\in\mathsf{C} \text{ such that } (\forall k\in\{1,\ldots,\mathsf{p}\})\;\; \mathsf{r}_k=\mathsf{F}_k(\mathsf{L}_k\mathsf{x}), \tag{2.77}$$

where each operator $\mathsf{F}_k\colon\mathsf{G}_k\to\mathsf{G}_k$ is firmly nonexpansive and each operator $\mathsf{L}_k\colon\mathsf{H}\to\mathsf{G}_k$ is linear and bounded. In many instances, the operators $(\mathsf{F}_k)_{1\leqslant k\leqslant \mathsf{p}}$ or $(\mathsf{L}_k)_{1\leqslant k\leqslant \mathsf{p}}$ may be imperfectly known or the model may be corrupted by perturbations and, as a result, (2.77) may not have solutions. A classical approach would be to relax it into a minimization problem such as the



(a) (b) (c)

Figure 2.6 Experiment of Section 2.2.5.4. Normalized error $20\log_{10}(\|x_{1,n}-x_\infty\|/\|x_{1,0}-x_\infty\|)$ (dB) versus execution time (s). (a): Block size 1 with 1 core. (b): Block size 8 with 8 cores. (c): Block size 32 with 32 cores. Green: Framework 1. Orange: Framework 2. Blue: Framework 3 with Example 2.14. Magenta: Framework 3 with Example 2.15. Dashed violet: Algorithm (2.13). Dashed red: Algorithm (2.15).

least-squares model

$$\underset{x\in C}{\text{minimize}} \quad \sum_{k=1}^{p} \|F_k(L_k x) - r_k\|^2. \tag{2.78}$$

However, because of the nonlinearity of the operators $(F_k)_{1\leqslant k\leqslant p}$, the resulting optimization problem is nonconvex and usually intractable. The strategy of [18] consists in relaxing (2.77) into the variational inequality problem

$$\text{find} \;\; x \in C \;\; \text{such that} \;\; (\forall y \in C) \;\; \sum_{k=1}^{p} \alpha_k \langle L_k(y - x) \mid F_k(L_k x) - r_k \rangle \geqslant 0, \tag{2.79}$$

where the weights $(\alpha_k)_{1\leqslant k\leqslant p}$ are in $]0, +\infty[$. As shown there, (2.79) is an exact relaxation of (2.77) in the sense that, if (2.77) happens to have solutions, they are the same as those of (2.79). Let us introduce the operators

$$(\forall k \in \{1, \ldots, p\}) \quad B_k = \alpha_k(F_k - r_k), \tag{2.80}$$

which are maximally monotone by [6, Example 20.30]. Then, in terms of the normal cone operator of (2.7), (2.79) is equivalent to

$$\text{find} \;\; x \in H \;\; \text{such that} \;\; 0 \in N_C x + \sum_{k=1}^{p} L_k^*\big(B_k(L_k x)\big). \tag{2.81}$$

This inclusion problem is now in the format of Problem 2.2 with $A = N_C$, which allows us to apply the algorithms proposed in Sections 2.2.3.2–2.2.3.4 to solve it with guaranteed almost sure convergence of the iterates to a solution.

The specific image recovery problem under consideration is similar to that of [18, Section 5.1]. The goal is to recover the original image $\overline{x} \in H = \mathbb{R}^N$ ($N = 256^2$) of Figure 2.7(a) from the following prior knowledge and $p = 62$ observations:

(i) Bounds on pixel values: $\overline{x} \in C = [0, 255]^N$.

(ii) The degraded images $(r_k)_{1\leqslant k\leqslant 20}$ in $\mathbb{R}^N$ are obtained via a blurring process, addition of noise, and finally clipping. In terms of the model (2.77), for every $k \in \{1, \ldots, 20\}$, $G_k = \mathbb{R}^N$, $r_k = F_k(L_k\overline{x} + w_k)$, where L_k performs convolution with a Gaussian kernel with a standard deviation of 3, $w_k \in \mathbb{R}^N$ is a noise vector with i.i.d. entries uniformly distributed in $[-50, 50]$, and

$$F_k \colon \mathbb{R}^N \to \mathbb{R}^N \colon y \mapsto \operatorname{proj}_{C_1} y, \quad \text{where} \quad C_1 = [0, 60]^N \tag{2.82}$$

models a hard clipping process. This nonlinear measurement process models a low-quality image acquired by a device which saturates at photon counts beyond a certain threshold. As an example, the first degraded image r_1 is shown in Figure 2.7(b).

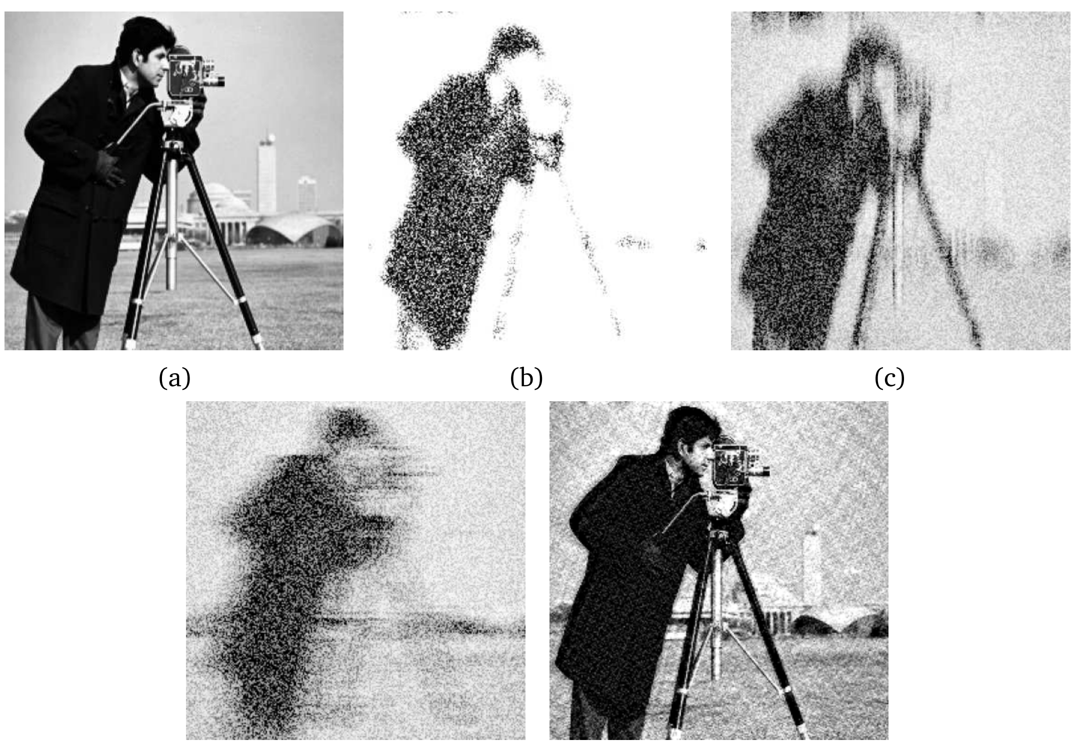
(iii) The degraded images $(r_k)_{21\leqslant k\leqslant 40}$ in $\mathbb{R}^N$ are obtained by a process similar to (ii). Here, for every $k \in \{21,\ldots,40\}$, the blurring operator L_k performs a convolution in the vertical direction with a uniform kernel of length 20, the entries of the noise vector $w_k \in \mathbb{R}^N$ are i.i.d. and uniformly distributed in $[-70,70]$, and pixel values beyond 90 are soft-clipped by

$$F_k\colon \mathbb{R}^N \to \mathbb{R}^N\colon (\eta_j)_{1\leqslant j\leqslant N} \mapsto \left(\frac{90\max\{0,\eta_j\}}{90+|\eta_j|}\right)_{1\leqslant j\leqslant N}. \tag{2.83}$$

As an example, the degraded image r_{21} is shown in Figure 2.7(c).

(iv) The degraded images $(r_k)_{41\leqslant k\leqslant 60}$ in $\mathbb{R}^N$ are obtained through an image formation process similar to that of (iii). For every $k \in \{41,\ldots,60\}$, the blurring operator L_k now performs a convolution in the horizontal direction with a uniform kernel of length 24, and the entries of the noise vector $w_k \in \mathbb{R}^N$ are i.i.d. and uniformly distributed in $[-90,90]$. For every $k \in \{41,\ldots,60\}$, pixel values beyond 90 are soft-clipped by the same operator F_k as in (2.83).

(v) The mean pixel value $\rho = 137$ of $\overline{x}$ is known. This information is imposed on a candidate solution $x \in \mathbb{R}^N$ via the equation $\langle x \mid \mathbf{1}\rangle = N\rho$, where $\mathbf{1} = (1,\ldots,1) \in \mathbb{R}^N$, which corresponds to the model $r_{61} = F_{61}(L_{61}x)$, with $G_{61} = \mathbb{R}$, $L_{61} = \langle\cdot \mid \mathbf{1}\rangle$, $r_{61} = N\rho$, and $F_{61} = \mathrm{Id}$.



(a) (b) (c)

(d) (e)

Figure 2.7 Experiment of Section 2.2.5.5: (a): Original image $\overline{x}$. (b): Degraded image r_1. (c): Degraded image r_{21}. (d): Degraded image r_{41}. (e): Recovered image.

(vi) The phase $\theta \in [-\pi, \pi]^N$ of the two-dimensional discrete Fourier transform of a noise-corrupted version of $\overline{x}$, i.e., $\theta = \angle\,\mathrm{DFT}(\overline{x} + w_{62})$, where $w_{62} \in \mathbb{R}^N$ is uniformly distributed in $[-3, 3]$. This information is enforced by forcing a candidate solution to lie in the closed convex set $C_{62} = \big\{x \in \mathbb{R}^N \mid \angle\,\mathrm{DFT}(x) = \theta\big\}$, i.e., by enforcing the constraint $x = \mathrm{proj}_{C_{62}} x$. This constraint corresponds to the model $r_{62} = F_{62}(L_{62}x)$, with $G_{62} = \mathbb{R}^N$, $L_{62} = \mathrm{Id}$, $r_{62} = 0$, and $F_{62} = \mathrm{Id} - \mathrm{proj}_{C_{62}}$, that is [43],

$$F_{62} \colon \mathbb{R}^N \to \mathbb{R}^N \colon x \mapsto x - \mathrm{IDFT}\Big(\big|\mathrm{DFT}\,x\big| \max\big\{\cos\big(\angle(\mathrm{DFT}\,x) - \theta\big), 0\big\} \exp(\imath\theta)\Big). \tag{2.84}$$

Due to the presence of the measurement errors $(w_k)_{1\leqslant k\leqslant 60}$ and w_{62}, problem (2.77) is inconsistent and we approximate it by (2.80)–(2.81), where $\alpha_1 = \cdots = \alpha_{62} = 1$. To implement the algorithms of Sections 2.2.3.2–2.2.3.4, we require the expressions of the resolvent of the operators N_C and $(B_k)_{1\leqslant k\leqslant p}$. The former is just

$$J_{N_C} = \mathrm{proj}_C \colon (\xi_j)_{1\leqslant j\leqslant N} \mapsto \big(\min\{\max\{0, \xi_j\}, 255\}\big)_{1\leqslant j\leqslant N}. \tag{2.85}$$

For the remaining cases, it follows from (2.80) that the operators $(B_k)_{1\leqslant k\leqslant p}$ are firmly nonexpansive. We therefore invoke Lemma 2.3 to compute their resolvents. Let $\gamma \in \left]0, +\infty\right[$ and note that [6, Proposition 23.17(ii)] entails that

$$(\forall k \in \{1, \ldots, p\}) \quad J_{\gamma B_k} = J_{\gamma F_k}(\cdot + \gamma r_k). \tag{2.86}$$

First, set $k \in \{1, \ldots, 20\}$. Then $F_k = \mathrm{proj}_{C_1} = J_{N_{C_1}}$. Hence, upon setting $r_k = (\rho_{k,j})_{1\leqslant j\leqslant N}$, we deduce from Lemma 2.3(ii) and (2.86) that

$$J_{\gamma B_k} \colon (\xi_j)_{1\leqslant j\leqslant N} \mapsto \left(\xi_j + \gamma\rho_{k,j} - \gamma \min\left\{\max\left\{0, \frac{\xi_j + \gamma\rho_{k,j}}{1+\gamma}\right\}, 60\right\}\right)_{1\leqslant j\leqslant N}. \tag{2.87}$$

On the other hand, for $k \in \{21, \ldots, 60\}$, $J_{\gamma F_k} \colon (\eta_j)_{1\leqslant j\leqslant N} \mapsto (\zeta_j)_{1\leqslant j\leqslant N}$, where

$$(\forall j \in \{1, \ldots, N\}) \quad \zeta_j = \begin{cases} \dfrac{\eta_j - 90(1+\gamma) + \sqrt{|\eta_j - 90(1+\gamma)|^2 + 360\eta_j}}{2}, & \text{if } \eta_j \geqslant 0;\\[2ex] \eta_j, & \text{otherwise.} \end{cases} \tag{2.88}$$

Thus, we derive from (2.86) the expressions for $J_{\gamma B_k}$. Next, we have $J_{\gamma B_{61}} = (1+\gamma)^{-1}(\cdot + \gamma N\rho)$ as a result of $J_{\gamma F_{61}} = (1+\gamma)^{-1}\mathrm{Id}$ and (2.86). Finally, we deduce from [6, Proposition 23.20] that

$$F_{62} = \mathrm{Id} - \mathrm{proj}_{C_{62}} = J_{N_{C_{62}}^{-1}} \quad \text{and} \quad J_{(1+\gamma)^{-1}N_{C_{62}}^{-1}} \circ (1+\gamma)^{-1}\mathrm{Id} = \frac{\mathrm{Id} - J_{N_{C_{62}}}}{1+\gamma} = \frac{\mathrm{Id} - \mathrm{proj}_{C_{62}}}{1+\gamma}. \tag{2.89}$$

Hence, it follows from Lemma 2.3(ii) that $J_{\gamma B_{62}} = J_{\gamma F_{62}} = (1+\gamma)^{-1}(\mathrm{Id} + \gamma\,\mathrm{proj}_{C_{62}})$, i.e.,

$$J_{\gamma B_{62}} \colon y \mapsto \frac{y}{1+\gamma} + \frac{\gamma}{1+\gamma}\,\mathrm{IDFT}\Big(\big|\mathrm{DFT}\,y\big|\max\big\{\cos\big(\angle(\mathrm{DFT}\,y) - \theta\big), 0\big\}\exp(\imath\theta)\Big). \tag{2.90}$$

Finally, we implement the inversions of linear operators using the fast Fourier transform and Example 2.17(ii).

We employ the three frameworks of Sections 2.2.3.2–2.2.3.4 to solve (2.81), where Proposition 2.12 uses the operator E defined in Example 2.13. Two experiments are conducted: the random variable ε_0 produces (a) 1 activation with 1 core; and (b) 8 activations with 8 cores. We compare with algorithm (2.15), where the random variable ε_0 activates (a) 1; and (b) 8 indices in $\{1,\ldots,p\}$ with a uniform distribution. The solution produced by Framework 3 is shown in Figure 2.7(e). We display in Figure 2.8 the normalized error versus execution time on a single-processor machine.

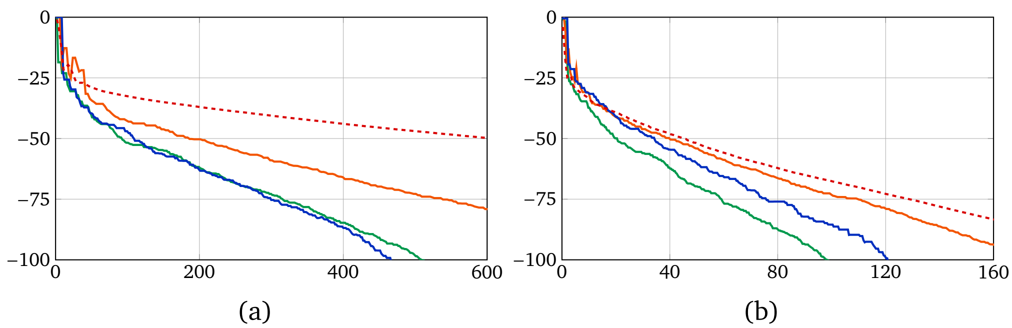


Figure 2.8 Experiment of Section 2.2.5.5: Normalized error $20\log_{10}(\|x_{1,n}-x_\infty\|/\|x_{1,0}-x_\infty\|)$ (dB) versus execution time (s). (a): Block size 1 with 1 core. (b): Block size 8 with 8 cores. Green: Framework 1. Orange: Framework 2. Blue: Framework 3 with Example 2.13. Dashed red: Algorithm (2.15).

2.2.5.6 Discussion

The three proposed frameworks differ in terms of storage requirements, use of resolvent operators, and use of linear operators.

- Framework 1: It stores $2p+3$ vectors. In addition, for each of the $p+1$ random activation indices, there is one resolvent evaluation.
- Framework 2: It stores $4p+5$ vectors. Out of the $p+2$ random activation indices, those in $\{1,\ldots,p+1\}$ involve the evaluation of a resolvent. In addition, the linear operators are used only if index $p+2$ is activated.
- Framework 3: It stores $2p+2r+2$ vectors. Moreover, out of the $p+r+1$ random activation indices, those in $\{1,\ldots,p+1\}$ involve the evaluation of a resolvent operator, while those in $\{p+2,\ldots,p+r+1\}$ do not require a resolvent evaluation.

Although Framework 1 is the most efficient in terms of storage, it may not always be the fastest, especially when resolvents are computationally expensive. For instance, in Section 2.2.5.5, where it is the case, Framework 3 is the fastest. Framework 2 has an advantage when the linear operators are costly, which is the case in Section 2.2.5.3. Finally, we observe that the existing algorithms (2.13) and (2.15) which, as discussed in Section 2.2.2.2, do not satisfy conditions **R2**–**R3**, are consistently slower than the methods proposed in Sections 2.2.3.2–2.2.3.4.

Acknowledgement

This work was supported by the National Science Foundation under grant CCF-2211123.

References

[1] A. Alacaoglu, O. Fercoq, and V. Cevher, On the convergence of stochastic primal-dual hybrid gradient, *SIAM J. Optim.*, vol. 32, pp. 1288–1318, 2022.

[2] H. C. Andrews and B. R. Hunt, *Digital Image Restoration*. Prentice-Hall, Englewood Cliffs, NJ, 1977.

[3] H. Attouch, L. M. Briceño–Arias, and P. L. Combettes, A strongly convergent primal-dual method for nonoverlapping domain decomposition, *Numer. Math.*, vol. 133, pp. 433–470, 2016.

[4] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, Optimization with sparsity-inducing penalties, *Found. Trends Machine Learn.*, vol. 4, pp. 1–106, 2012.

[5] S. Banert, A. Ringh, J. Adler, J. Karlsson, and O. Öktem, Data-driven nonsmooth optimization, *SIAM J. Optim.*, vol. 30, pp. 102–131, 2020.

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

[7] M. Benning and M. Burger, Modern regularization methods for inverse problems, *Acta Numer.*, vol. 27, pp 1–111, 2018.

[8] J. R. Blum, Multidimensional stochastic approximation methods, *Ann. Math. Stat.*, vol. 25, pp. 737–744, 1954.

[9] A. Chambolle, C. Delplancke, M. J. Ehrhardt, C.-B. Schönlieb, and J. Tang, Stochastic primal-dual hybrid gradient algorithm with adaptive step sizes, *J. Math. Imaging Vision*, vol. 66, pp. 294–313, 2024.

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

[11] P. L. Combettes, The geometry of monotone operator splitting methods, *Acta Numer.*, vol. 33, pp. 487–632, 2024.

[12] P. L. Combettes and J.-C. Pesquet, Proximal splitting methods in signal processing, in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pp. 185–212. Springer, New York, 2011.

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

[14] P. L. Combettes and J.-C. Pesquet, Deep neural network structures solving variational inequalities, *Set-Valued Var. Anal.*, vol. 28, pp. 491–518, 2020.

[15] P. L. Combettes and J.-C. Pesquet, Fixed point strategies in data science, *IEEE Trans. Signal Process.*, vol. 69, pp. 3878–3905, 2021.

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

[17] P. L. Combettes and Z. C. Woodstock, Reconstruction of functions from prescribed proximal points, *J. Approx. Theory*, vol. 268, art. 105606, 2021.

[18] P. L. Combettes and Z. C. Woodstock, A variational inequality model for the construction of signals from inconsistent nonlinear equations, *SIAM J. Imaging Sci.*, vol. 15, pp. 84–109, 2022.

[19] L. Condat and P. Richtárik, Randprox: Primal-dual optimization algorithms with randomized proximal updates, *Proc. Int. Conf. Learn. Represent.* Kigali, Rwanda, May 1–5, 2023.

[20] D. Davis, Variance reduction for root-finding problems, *Math. Program.*, vol. A197, pp. 375–410, 2023.

[21] A. Dieuleveut, G. Fort, E. Moulines, and H.-T. Wai, Stochastic approximation beyond gradient for signal processing and machine learning, *IEEE Trans. Signal Process.*, vol. 71, pp. 3117–3148, 2023.

[22] M. Duflo, *Méthodes Récursives Aléatoires.* Masson, Paris, 1990. English translation: *Random Iterative Models.* Springer, New York, 1997.

[23] L. Euler, *Recherches sur la Question des Inégalités du Mouvement de Saturne et de Jupiter.* Martin, Coignard et Guerin, Paris, 1749.

[24] F. Florez-Ospina, D. A. Jimenez-Sierra, H. D. Benitez-Restrepo, and G. R. Arce, Exploiting variational inequalities for generalized change detection on graphs, *IEEE Trans. Geosci. Remote Sensing*, vol. 61, pp. 1–16, 2023.

[25] R. Glowinski, S. J. Osher, and W. Yin (eds.), *Splitting Methods in Communication, Imaging, Science, and Engineering.* Springer, New York, 2016.

[26] T. Hytönen, J. van Neerven, M. Veraar, and L. Weis, *Analysis in Banach Spaces. Volume I: Martingales and Littlewood–Paley Theory.* Springer, New York, 2016.

[27] A. B. Juditsky and A. S. Nemirovski, Signal recovery by stochastic optimization, *Autom. Remote Control*, vol. 80, pp. 1878–1893, 2019.

[28] A. B. Juditsky and A. S. Nemirovski, *Statistical Inference via Convex Optimization.* Princeton University Press, Princeton, NJ, 2020.

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

[30] Y. Malitsky and M. K. Tam, A forward-backward splitting method for monotone inclusions without cocoercivity, *SIAM J. Optim.*, vol. 30, pp. 1451–1472, 2020.

[31] J. T. Mayer, Abhandlung über die Umwälzung des Monds um seine Axe und die scheinbare Bewegung der Mondflecken, *Kosmographische Nachrichten und Sammlungen*, vol. 1, pp. 52–183, 1748.

[32] D. W. Mimouni, P. Malisani, J. Zhu, and W. de Oliveira, Computing Wasserstein barycenters via operator splitting: The method of averaged marginals, *SIAM J. Data Sci.*, vol. 6, pp. 1000–1026, 2024.

[33] Z. Peng, Y. Xu, M. Yan, and W. Yin, ARock: An algorithmic framework for asynchronous parallel coordinate updates, *SIAM J. Sci. Comput.*, vol. 38, pp. A2851–A2879, 2016.

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

[35] J.-C. Pesquet, A. Repetti, M. Terris, and Y. Wiaux, Learning maximally monotone operators for image recovery, *SIAM J. Imaging Sci.*, vol. 14, pp. 1206–1237, 2021.

[36] B. J. Pettis, On integration in vector spaces, *Trans. Amer. Math. Soc.*, vol. 44, pp. 277–304, 1938.

[37] H. Robbins and S. Monro, A stochastic approximation method, *Ann. Math. Statist.*, vol. 22, pp. 400–407, 1951.

[38] A. Sadiev, L. Condat, and P. Richtárik, Stochastic proximal point methods for monotone inclusions under expected similarity, arxiv, 2024. `https://arxiv.org/pdf/2405.14255`

[39] S. Theodoridis, *Machine Learning: A Bayesian and Optimization Perspective*, 2nd ed. Amsterdam: Elsevier, 2020.

[40] C. Traoré, V. Apidopoulos, S. Salzo, and S. Villa, Variance reduction techniques for stochastic proximal point algorithms, *J. Optim. Theory Appl.*, vol. 203, pp. 1910–1939, 2024.

[41] E. Winston and J. Z. Kolter, Monotone operator equilibrium networks, *Proc. Adv. Neural Inform. Process. Syst.*, vol. 22, pp. 10718–10728, 2020.

[42] P. Yi and S. Ching, Synthesis of recurrent neural dynamics for monotone inclusion with application to Bayesian inference, *Neural Networks*, vol. 131, pp. 231–241, 2020.

[43] D. C. Youla, Mathematical theory of image restoration by the method of convex projections, in: H. Stark (ed.) *Image Recovery: Theory and Application*, pp. 29–77. Academic Press, San Diego, CA, 1987.

[44] Y.-L. Yu, Better approximation and faster algorithm using the proximal average, *Proc. Conf. Adv. Neural Inform. Process. Syst.*, pp. 458–466, 2013.

Chapter 3

A GEOMETRIC FRAMEWORK FOR STOCHASTIC ITERATIONS

3.1 Introduction and context

This chapter is dedicated to answering question **(Q2)** of Chapter 1. We propose a geometric framework to construct stochastic quasi-Fejér monotone sequences in the sense of (1.4), and we derive asymptotic results concerning the convergence of the generated sequences of random variables.

This chapter presents the following journal article:

P. L. Combettes and J. I. Madariaga, A geometric framework for stochastic iterations, *Mathematics of Computation*, to appear.

3.2 Article: A geometric framework for stochastic iterations

Abstract. This paper concerns models and convergence principles for dealing with stochasticity in a wide range of algorithms arising in nonlinear analysis and optimization in Hilbert spaces. It proposes a flexible geometric framework within which existing solution methods can be recast and improved, and new ones can be designed. Almost sure weak, strong, and linear convergence results are established in particular for stochastic fixed point iterations, the stochastic gradient descent method, and stochastic extrapolated parallel algorithms for feasibility problems. In these areas, the proposed algorithms exceed the features and convergence guarantees of the state of the art. Numerical applications to signal and image recovery are provided.

3.2.1 Introduction

The objective of this paper is to propose a general algorithmic framework and convergence principles for dealing with stochasticity in a broad class of algorithms arising in optimization and numerical nonlinear analysis. Throughout, H is a separable real Hilbert space and the underlying probability space $(\Omega, \mathcal{F}, \mathsf{P})$ is complete. We denote by $\mathsf{Z} \subset \mathsf{H}$ the set of solutions to the problem of interest and assume that it is nonempty and closed.

In the recent paper [18], we showed that a simple geometry underlies most deterministic monotone operator splitting algorithms and that, by exploiting this geometry, the convergence analysis of existing methods could be simplified and improved. It was also argued that this geometric framework provides a flexible template to create new algorithms. The basic idea is to construct the update at iteration n of a deterministic algorithm for finding a point in the solution set Z as a relaxed projection $\mathsf{x}_{n+1} = \mathsf{x}_n + \lambda_n(\mathrm{proj}_{\mathsf{H}_n} \mathsf{x}_n - \mathsf{x}_n)$ onto a half-space $\mathsf{H}_n = \{\mathsf{z} \in \mathsf{H} \mid \langle \mathsf{z} \mid \mathsf{t}_n^* \rangle_{\mathsf{H}} \leqslant \eta_n\}$ containing Z as follows (see Fig. 3.1(a)).

Algorithm 3.1 Let $\mathsf{x}_0 \in \mathsf{H}$ and iterate

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\text{take } \mathsf{t}_n^* \in \mathsf{H} \text{ and } \eta_n \in \mathbb{R} \text{ such that } (\forall \mathsf{z} \in \mathsf{Z}) \ \ \langle \mathsf{z} \mid \mathsf{t}_n^* \rangle_{\mathsf{H}} \leqslant \eta_n \\
\alpha_n = \begin{cases} \dfrac{\langle \mathsf{x}_n \mid \mathsf{t}_n^* \rangle_{\mathsf{H}} - \eta_n}{\|\mathsf{t}_n^*\|_{\mathsf{H}}^2} & \text{if } \ \langle \mathsf{x}_n \mid \mathsf{t}_n^* \rangle_{\mathsf{H}} > \eta_n; \\ 0, & \text{otherwise} \end{cases} \\
\mathsf{d}_n = \alpha_n \mathsf{t}_n^* \\
\text{take } \lambda_n \in \left]0, 2\right[\\
\mathsf{x}_{n+1} = \mathsf{x}_n - \lambda_n \mathsf{d}_n.
\end{array}
\right.
\end{array}
\tag{3.1}
$$

Our approach consists in extending the above geometric construction to a general stochastic environment by making the following changes at iteration n:

- The deterministic quantities t_n^* and η_n are replaced by random ones.
- A stochastic tolerance is added in the construction of the outer approximation.
- The relaxation parameter λ_n is now random and no longer restricted to the interval $\left]0, 2\right[$.

This leads to the following algorithmic scheme (see Section 3.2.2.1 for notation).

Algorithm 3.2 Let $x_0 \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ and iterate

$$
\begin{array}{l}
\text{for } \mathsf{n} = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\mathcal{X}_\mathsf{n} = \sigma(x_0, \ldots, x_\mathsf{n}) \\
\text{take } t_\mathsf{n}^* \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H}) \text{ and } \eta_\mathsf{n} \in L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathbb{R}) \text{ such that} \\
\quad \begin{cases}
\dfrac{1_{[t_\mathsf{n}^* \neq 0]} 1_{[\langle x_\mathsf{n} \mid t_\mathsf{n}^* \rangle_\mathsf{H} > \eta_\mathsf{n}]} \eta_\mathsf{n}}{\|t_\mathsf{n}^*\|_\mathsf{H} + 1_{[t_\mathsf{n}^* = 0]}} \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathbb{R}); \\
\alpha_\mathsf{n} = \dfrac{1_{[t_\mathsf{n}^* \neq 0]} 1_{[\langle x_\mathsf{n} \mid t_\mathsf{n}^* \rangle_\mathsf{H} > \eta_\mathsf{n}]} \big(\langle x_\mathsf{n} \mid t_\mathsf{n}^* \rangle_\mathsf{H} - \eta_\mathsf{n}\big)}{\|t_\mathsf{n}^*\|_\mathsf{H}^2 + 1_{[t_\mathsf{n}^* = 0]}}; \\
(\forall \mathsf{z} \in \mathsf{Z}) \;\; \langle \mathsf{z} \mid \mathsf{E}(\alpha_\mathsf{n} t_\mathsf{n}^* \mid \mathcal{X}_\mathsf{n}) \rangle_\mathsf{H} \leqslant \mathsf{E}(\alpha_\mathsf{n} \eta_\mathsf{n} \mid \mathcal{X}_\mathsf{n}) + \varepsilon_\mathsf{n}(\cdot, \mathsf{z}) \text{ P-a.s.} \\
\qquad\qquad\qquad\qquad \text{where } \varepsilon_\mathsf{n}(\cdot, \mathsf{z}) \in [0, +\infty[\text{ P-a.s.}
\end{cases} \\
d_\mathsf{n} = \alpha_\mathsf{n} t_\mathsf{n}^* \\
\text{take } \lambda_\mathsf{n} \in L^\infty(\Omega, \mathcal{F}, \mathsf{P};]0, +\infty[) \\
x_{\mathsf{n}+1} = x_\mathsf{n} - \lambda_\mathsf{n} d_\mathsf{n}.
\end{array}
\right.
\end{array}
\tag{3.2}
$$

Implicitly, Algorithm 3.2 constructs a random outer approximation S_n to Z, namely

$$
\mathsf{Z} \subset S_\mathsf{n} = \big\{\mathsf{z} \in \mathsf{H} \mid \langle \mathsf{z} \mid \mathsf{E}(\alpha_\mathsf{n} t_\mathsf{n}^* \mid \mathcal{X}_\mathsf{n}) \rangle_\mathsf{H} \leqslant \mathsf{E}(\alpha_\mathsf{n} \eta_\mathsf{n} \mid \mathcal{X}_\mathsf{n}) + \varepsilon_\mathsf{n}(\cdot, \mathsf{z}) \big\} \text{ P-a.s.} \tag{3.3}
$$

and the update $x_{\mathsf{n}+1}$ is obtained by performing a relaxed projection of the current iterate x_n onto the simpler set

$$
H_\mathsf{n} = \big\{\mathsf{z} \in \mathsf{H} \mid \langle \mathsf{z} \mid t_\mathsf{n}^* \rangle_\mathsf{H} \leqslant \eta_\mathsf{n} \big\}, \tag{3.4}
$$

which is a random affine half-space. It should be noted that, while $\mathsf{Z} \subset S_\mathsf{n}$, the more restrictive inclusion $\mathsf{Z} \subset H_\mathsf{n}$ does not hold in general (see Fig. 3.1(b)). In terms of modeling, choosing t_n^* and η_n such that $\mathsf{Z} \subset H_\mathsf{n}$ would restrict the scope of the processes we intend to model, whereas the more general inclusion $\mathsf{Z} \subset S_\mathsf{n}$ offers more flexibility. Let us observe that, if $\varepsilon_\mathsf{n} = 0$, S_n is also a random half-space. However, as the following example shows, projecting onto it is not judicious.

Example 3.3 For every $\mathsf{k} \in \{1, \ldots, \mathsf{p}\}$, let C_k be a closed convex subset of H. Suppose that $\mathsf{Z} = \bigcap_{\mathsf{k}=1}^{\mathsf{p}} \mathsf{C}_\mathsf{k} \neq \varnothing$ and implement Algorithm 3.2 with $\lambda_\mathsf{n} = 1$, $\varepsilon_\mathsf{n} = 0$, $t_\mathsf{n}^* = x_\mathsf{n} - \mathrm{proj}_{\mathsf{C}_k} x_\mathsf{n}$, and $\eta_\mathsf{n} = \langle \mathrm{proj}_{\mathsf{C}_k} x_\mathsf{n} \mid t_\mathsf{n}^* \rangle_\mathsf{H}$, where the random variable k is uniformly distributed in $\{1, \ldots, \mathsf{p}\}$. Then $\mathsf{E}(t_\mathsf{n}^* \mid \mathcal{X}_\mathsf{n}) = x_\mathsf{n} - \mathsf{p}^{-1} \sum_{\mathsf{k}=1}^{\mathsf{p}} \mathrm{proj}_{\mathsf{C}_\mathsf{k}} x_\mathsf{n}$ and therefore

$$
\begin{aligned}
\mathsf{Z} &\subset \left\{ \mathsf{z} \in \mathsf{H} \;\middle|\; \sum_{\mathsf{k}=1}^{\mathsf{p}} \big\langle \mathsf{z} - \mathrm{proj}_{\mathsf{C}_\mathsf{k}} x_\mathsf{n} \,\big|\, x_\mathsf{n} - \mathrm{proj}_{\mathsf{C}_\mathsf{k}} x_\mathsf{n} \big\rangle_\mathsf{H} \leqslant 0 \right\} \\
&= \big\{ \mathsf{z} \in \mathsf{H} \,\big|\, \big\langle \mathsf{z} \,\big|\, \mathsf{E}(t_\mathsf{n}^* \mid \mathcal{X}_\mathsf{n}) \big\rangle_\mathsf{H} \leqslant \mathsf{E}\big(\eta_\mathsf{n} \,\big|\, \mathcal{X}_\mathsf{n}\big) \big\} \\
&= S_\mathsf{n} \;\; \text{P-a.s.}
\end{aligned} \tag{3.5}
$$

Thus, Algorithm 3.2 yields the random iteration process $x_{n+1} = \text{proj}_{C_k} x_n$ in which a single, randomly selected set is projected onto at iteration n. By contrast, projecting onto S_n would yield the barycentric projection method $x_{n+1} = p^{-1} \sum_{k=1}^{p} \text{proj}_{C_k} x_n$, which is deterministic and imposes the computation of all p projections at each iteration.

Another new feature of Algorithm 3.2 is that the relaxation parameters $(\lambda_n)_{n\in\mathbb{N}}$ are random. In addition, they need not be confined to the range $]0, 2[$ imposed in deterministic algorithms [5, 11, 14, 18, 25]. We call such relaxations *super relaxations*.

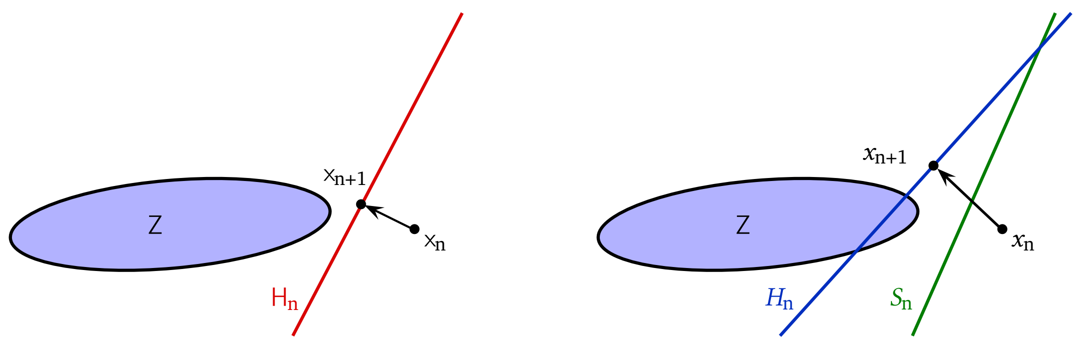


Figure 3.1 Geometry of algorithms for finding a point in Z with $\lambda_n = 1$. (a) Left: Iteration n of the deterministic Algorithm 3.1. (b) Right: Iteration n of the stochastic Algorithm 3.2 with $\varepsilon_n = 0$.

The deterministic setting of Algorithm 3.1 is known to capture a vast array of iterative methods in nonlinear analysis and optimization [18]. Our premise is that Algorithm 3.2 can serve the same purpose for their stochastic counterparts. Weak, strong, and linear convergence results will be established for Algorithm 3.2. In turn, these results will be applied to fixed point and feasibility problems, where they will be shown to provide new stochastic algorithms that go beyond the state of the art not only in terms of convergence guarantees but also of flexibility of implementation and scope.

The remainder of the paper is organized as follows. Notation and preliminary results are introduced in Section 3.2.2. The main theorems are presented in Section 3.2.3, where the asymptotic properties of Algorithm 3.2 are established. Section 3.2.4 is devoted to an application of the proposed theory to a randomly relaxed Krasnosel'skiĭ–Mann iteration process and includes new results on the convergence of the stochastic gradient method. Section 3.2.5 concerns an application to randomly activated and relaxed extrapolated fixed point methods for common fixed point problems in the presence of possibly uncountably many operators. These results significantly improve existing ones. Section 3.2.6 concludes the paper with applications to signal and image recovery. Applications of the results of Section 3.2.3 to the design and the analysis of stochastic splitting algorithms for monotone inclusion problems are addressed in the companion paper [19].

3.2.2 Notation and background

3.2.2.1 Notation

We use sans-serif letters to denote deterministic variables and italicized serif letters to denote random variables.

The Hilbert space H has identity operator Id, scalar product $\langle \cdot \mid \cdot \rangle_{\mathsf{H}}$, and associated norm $\|\cdot\|_{\mathsf{H}}$. The symbols $\rightharpoonup$ and $\to$ denote weak and strong convergence in H, respectively. The sets of strong and weak sequential cluster points of a sequence $(\mathsf{x}_n)_{n\in\mathbb{N}}$ in H are denoted by $\mathfrak{S}(\mathsf{x}_n)_{n\in\mathbb{N}}$ and $\mathfrak{W}(\mathsf{x}_n)_{n\in\mathbb{N}}$, respectively. The distance function of a set $\mathsf{C} \subset \mathsf{H}$ is denoted by $d_{\mathsf{C}} \colon \mathsf{x} \mapsto \inf_{\mathsf{y}\in\mathsf{C}} \|\mathsf{y}-\mathsf{x}\|_{\mathsf{H}}$ and the projection onto a nonempty closed convex set $\mathsf{C} \subset \mathsf{H}$ is denoted by $\mathrm{proj}_{\mathsf{C}}$. The fixed point set of an operator $\mathsf{T} \colon \mathsf{H} \to \mathsf{H}$ is $\mathrm{Fix}\,\mathsf{T} = \big\{\mathsf{x} \in \mathsf{H} \mid \mathsf{T}\mathsf{x} = \mathsf{x}\big\}$. The following notion will play an important role in Sections 3.2.4 and 3.2.5; see [2, Proposition 2.4] for examples of demiregular operators.

Definition 3.4 [2] $\mathsf{T} \colon \mathsf{H} \to \mathsf{H}$ is demiregular at $\mathsf{x} \in \mathsf{H}$ if, for every sequence $(\mathsf{x}_n)_{n\in\mathbb{N}}$ in H such that $\mathsf{x}_n \rightharpoonup \mathsf{x}$ and $\mathsf{T}\mathsf{x}_n \to \mathsf{T}\mathsf{x}$, we have $\mathsf{x}_n \to \mathsf{x}$.

Let $(\Xi, \mathcal{G})$ be a measurable space. A Ξ-valued random variable is a measurable mapping $x \colon (\Omega, \mathcal{F}, \mathsf{P}) \to (\Xi, \mathcal{G})$. Given $x \colon \Omega \to \Xi$ and $\mathsf{S} \in \mathcal{G}$, we set $[x \in \mathsf{S}] = \big\{\omega \in \Omega \mid x(\omega) \in \mathsf{S}\big\}$. Let x and y be random variables from $(\Omega, \mathcal{F}, \mathsf{P})$ to $(\Xi, \mathcal{G})$. Then y is a copy of x if, for every $\mathsf{S} \in \mathcal{G}$, $\mathsf{P}([x \in \mathsf{S}]) = \mathsf{P}([y \in \mathsf{S}])$. The Borel σ-algebra of H is denoted by $\mathcal{B}_{\mathsf{H}}$. An H-valued random variable is a measurable mapping $x \colon (\Omega, \mathcal{F}) \to (\mathsf{H}, \mathcal{B}_{\mathsf{H}})$. Let $\mathsf{p} \in [1, +\infty[$ and let $\mathcal{X}$ be a sub σ-algebra of $\mathcal{F}$. Then $L^{\mathsf{p}}(\Omega, \mathcal{X}, \mathsf{P}; \mathsf{H})$ denotes the space of equivalence classes of P-a.s. equal H-valued random variables $x \colon (\Omega, \mathcal{X}, \mathsf{P}) \to (\mathsf{H}, \mathcal{B}_{\mathsf{H}})$ such that $\mathsf{E}\|x\|_{\mathsf{H}}^{\mathsf{p}} < +\infty$. Endowed with the norm

$$\|\cdot\|_{L^{\mathsf{p}}(\Omega,\mathcal{X},\mathsf{P};\mathsf{H})} \colon x \mapsto \mathsf{E}^{1/\mathsf{p}}\|x\|_{\mathsf{H}}^{\mathsf{p}} = \left(\int_\Omega \|x(\omega)\|_{\mathsf{H}}^{\mathsf{p}} \mathsf{P}(d\omega)\right)^{1/\mathsf{p}}, \tag{3.6}$$

$L^{\mathsf{p}}(\Omega, \mathcal{X}, \mathsf{P}; \mathsf{H})$ is a real Banach space and $L^2(\Omega, \mathcal{X}, \mathsf{P}; \mathsf{H})$ is a real Hilbert space with scalar product

$$\langle \cdot \mid \cdot \rangle_{L^2(\Omega,\mathcal{X},\mathsf{P};\mathsf{H})} \colon (x, y) \mapsto \mathsf{E}\langle x \mid y\rangle_{\mathsf{H}} = \int_\Omega \big\langle x(\omega) \,\big|\, y(\omega)\big\rangle_{\mathsf{H}} \mathsf{P}(d\omega). \tag{3.7}$$

Further,

$$(\forall \mathsf{S} \in \mathcal{B}_{\mathsf{H}}) \quad L^{\mathsf{p}}(\Omega, \mathcal{X}, \mathsf{P}; \mathsf{S}) = \big\{x \in L^{\mathsf{p}}(\Omega, \mathcal{X}, \mathsf{P}; \mathsf{H}) \,\big|\, x \in \mathsf{S}\ \mathsf{P}\text{-a.s.}\big\}. \tag{3.8}$$

The σ-algebra generated by a family Φ of random variables is denoted by $\sigma(\Phi)$. Let $\mathfrak{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ be a sequence of sub σ-algebras of $\mathcal{F}$ such that $(\forall n \in \mathbb{N})$ $\mathcal{F}_n \subset \mathcal{F}_{n+1}$. Then $\ell_+(\mathfrak{F})$ is 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

$$(\forall \mathsf{p} \in \left]0, +\infty\right[) \quad \ell_+^{\mathsf{p}}(\mathfrak{F}) = \left\{(\xi_n)_{n\in\mathbb{N}} \in \ell_+(\mathfrak{F}) \,\middle|\, \sum_{n\in\mathbb{N}} \xi_n^{\mathsf{p}} < +\infty\ \mathsf{P}\text{-a.s.}\right\}. \tag{3.9}$$

We say that $\varphi\colon \Omega\times \mathsf{H}\to\mathbb{R}$ is a Carathéodory integrand if

$$\begin{cases}\text{for } \mathsf{P}\text{-almost every } \omega\in\Omega, & \varphi(\omega,\cdot) \text{ is continuous;}\\ \text{for every } \mathsf{x}\in\mathsf{H}, & \varphi(\cdot,\mathsf{x}) \text{ is } \mathcal{F}\text{-measurable.}\end{cases} \tag{3.10}$$

We denote by $\mathfrak{C}(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ the class of Carathéodory integrands $\varphi\colon\Omega\times\mathsf{H}\to[0,+\infty[$.

The reader is referred to [5] for background on convex analysis and fixed point theory, and to [31,36] for background on probability in Hilbert spaces.

3.2.2.2 Preliminary results

Definition 3.5 Let $\mathcal{X}$ be a sub σ-algebra of $\mathcal{F}$, $\mathsf{C}\in\mathcal{B}_{\mathsf{H}}$, and x be an H-valued random variable. Then x is a C-valued $\mathcal{X}$-simple mapping if there exist a finite family of disjoint sets $(\mathsf{F}_i)_{1\leqslant i\leqslant N}$ in $\mathcal{X}$ and a family of vectors $(\mathsf{z}_i)_{1\leqslant i\leqslant N}$ in C such that

$$\bigcup_{i=1}^{N}\mathsf{F}_i=\Omega \quad\text{and}\quad x=\sum_{i=1}^{N}1_{\mathsf{F}_i}\mathsf{z}_i \quad \mathsf{P}\text{-a.s.} \tag{3.11}$$

Remark 3.6 Let $\mathsf{p}\in[1,+\infty[$. Then every C-valued $\mathcal{X}$-simple mapping is in $L^{\mathsf{p}}(\Omega,\mathcal{X},\mathsf{P};\mathsf{C})$.

The following proposition is an adaptation of [31, Corollary 1.1.7].

Proposition 3.7 *Let C be a nonempty closed subset of H, $\mathcal{X}$ be a sub σ-algebra of $\mathcal{F}$, $\mathsf{p}\in[1,+\infty[$, and $x\in L^{\mathsf{p}}(\Omega,\mathcal{X},\mathsf{P};\mathsf{C})$. Then there exists a sequence $(x_n)_{n\in\mathbb{N}}$ of C-valued $\mathcal{X}$-simple mappings that converges strongly P-a.s. to x with $\sup_{n\in\mathbb{N}}\|x_n\|_{\mathsf{H}}^{\mathsf{p}}\leqslant\|x\|_{\mathsf{H}}^{\mathsf{p}}+1$ P-a.s.*

Proof. Let $\mathsf{z}\in\mathsf{C}$ be such that $\|\mathsf{z}\|_{\mathsf{H}}^{\mathsf{p}}\leqslant\inf_{\mathsf{y}\in\mathsf{C}}\|\mathsf{y}\|_{\mathsf{H}}^{\mathsf{p}}+1$ and let $\{\mathsf{z}_n\}_{n\in\mathbb{N}}$ be a countable dense subset of C with $\mathsf{z}_0=\mathsf{z}$. For every $n\in\mathbb{N}$ and every $\mathsf{y}\in\mathsf{C}$, define $I_{n,\mathsf{y}}=\big\{i\in\{0,\ldots,n\}\mid\|\mathsf{z}_i\|_{\mathsf{H}}^{\mathsf{p}}\leqslant\|\mathsf{y}\|_{\mathsf{H}}^{\mathsf{p}}+1\big\}$ and let $i_{n,\mathsf{y}}$ be the smallest integer $i\in I_{n,\mathsf{y}}$ such that $\|\mathsf{y}-\mathsf{z}_i\|_{\mathsf{H}}=\min_{j\in I_{n,\mathsf{y}}}\|\mathsf{y}-\mathsf{z}_j\|_{\mathsf{H}}$. Define, for every $n\in\mathbb{N}$, $\mathsf{T}_n\colon\mathsf{C}\to\mathsf{C}\colon\mathsf{y}\mapsto\mathsf{z}_{i_{n,\mathsf{y}}}$. It follows from the density of $\{\mathsf{z}_n\}_{n\in\mathbb{N}}$ in C that, for every $\mathsf{y}\in\mathsf{C}$, $\mathsf{T}_n\mathsf{y}\to\mathsf{y}$ and

$$(\forall n\in\mathbb{N})\quad\|\mathsf{T}_n\mathsf{y}\|_{\mathsf{H}}^{\mathsf{p}}\leqslant\|\mathsf{y}\|_{\mathsf{H}}^{\mathsf{p}}+1. \tag{3.12}$$

Set, for every $n\in\mathbb{N}$, $x_n=\mathsf{T}_n x$. Then $(x_n)_{n\in\mathbb{N}}$ converges strongly P-a.s. to x and

$$(\forall n\in\mathbb{N})\quad\|x_n\|_{\mathsf{H}}^{\mathsf{p}}\leqslant\|x\|_{\mathsf{H}}^{\mathsf{p}}+1\ \mathsf{P}\text{-a.s.} \tag{3.13}$$

It remains to show that $(x_n)_{n\in\mathbb{N}}$ is a sequence of C-valued $\mathcal{X}$-simple mappings. Fix $n\in\mathbb{N}$. Then

$$[x_n=\mathsf{z}_0]=\Big\{\omega\in\Omega\;\Big|\;\|x(\omega)-\mathsf{z}_0\|_{\mathsf{H}}=\min_{j\in I_{n,x(\omega)}}\|x(\omega)-\mathsf{z}_j\|_{\mathsf{H}}\Big\} \tag{3.14}$$

and, for every $i \in \{1, \ldots, n\}$,

$$[x_n = z_i] = \left\{ \omega \in \Omega \;\middle|\; i \in I_{n,x(\omega)} \text{ and } \|x(\omega) - z_i\|_H = \min_{j \in I_{n,x(\omega)}} \|x(\omega) - z_j\|_H < \min_{j \in I_{i-1,x(\omega)}} \|x(\omega) - z_j\|_H \right\}. \tag{3.15}$$

By construction, (3.14) and (3.15) define sets in $\mathcal{X}$. Further,

$$\bigcup_{i=0}^{n} [x_n = z_i] = \Omega \quad \text{and} \quad x_n = \sum_{i=0}^{n} 1_{[x_n = z_i]} z_i, \tag{3.16}$$

which confirms that x_n is a C-valued $\mathcal{X}$-simple mapping. ☐

Lemma 3.8 *Let $\mathfrak{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ be a sequence of sub σ-algebras of $\mathcal{F}$ such that $(\forall n \in \mathbb{N})\ \mathcal{F}_n \subset \mathcal{F}_{n+1}$. Let $(\alpha_n)_{n\in\mathbb{N}} \in \ell_+(\mathfrak{F})$, $(\theta_n)_{n\in\mathbb{N}} \in \ell_+(\mathfrak{F})$, and $(\eta_n)_{n\in\mathbb{N}} \in \ell_+(\mathfrak{F})$. Then the following hold:*

(i) *Suppose that $(\eta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathfrak{F})$ and there exists a sequence $(\chi_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathfrak{F})$ satisfying*

$$(\forall n \in \mathbb{N}) \quad \mathsf{E}(\alpha_{n+1} \,|\, \mathcal{F}_n) + \theta_n \leqslant (1 + \chi_n)\alpha_n + \eta_n \ \text{P-a.s.} \tag{3.17}$$

Then $(\theta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathfrak{F})$ and $(\alpha_n)_{n\in\mathbb{N}}$ converges P-a.s. to a $[0, +\infty[$-valued random variable.

(ii) *Suppose that $\mathsf{E}\alpha_0 < +\infty$, $\sum_{n\in\mathbb{N}} \mathsf{E}\eta_n < +\infty$, and there exists a sequence $(\chi_n)_{n\in\mathbb{N}}$ in $[0, +\infty[$ satisfying $\overline{\lim}\, \chi_n < 1$ and*

$$(\forall n \in \mathbb{N}) \quad \mathsf{E}(\alpha_{n+1} \,|\, \mathcal{F}_n) + \theta_n \leqslant \chi_n\alpha_n + \eta_n \ \text{P-a.s.} \tag{3.18}$$

Then $\sum_{n\in\mathbb{N}} \mathsf{E}\theta_n < +\infty$ and $\sum_{n\in\mathbb{N}} \mathsf{E}\alpha_n < +\infty$.

Proof. (i): This follows from [51, Theorem 1].

(ii): This follows from [22, Lemma 2.1(ii)]. ☐

Corollary 3.9 *Let $(\alpha_n)_{n\in\mathbb{N}}$, $(\theta_n)_{n\in\mathbb{N}}$, $(\eta_n)_{n\in\mathbb{N}}$, and $(\chi_n)_{n\in\mathbb{N}}$ be sequences in $[0, +\infty[$. Then the following hold:*

(i) *Suppose that $\sum_{n\in\mathbb{N}} \eta_n < +\infty$, $\sum_{n\in\mathbb{N}} \chi_n < +\infty$, and*

$$(\forall n \in \mathbb{N}) \quad \alpha_{n+1} + \theta_n \leqslant (1 + \chi_n)\alpha_n + \eta_n. \tag{3.19}$$

Then $\sum_{n\in\mathbb{N}} \theta_n < +\infty$ and $(\alpha_n)_{n\in\mathbb{N}}$ converges to a positive real number.

(ii) *Suppose that $\sum_{n\in\mathbb{N}} \eta_n < +\infty$, $\overline{\lim}\, \chi_n < 1$, and*

$$(\forall n \in \mathbb{N}) \quad \alpha_{n+1} + \theta_n \leqslant \chi_n\alpha_n + \eta_n \ \text{P-a.s.} \tag{3.20}$$

Then $\sum_{n\in\mathbb{N}} \theta_n < +\infty$ and $\sum_{n\in\mathbb{N}} \alpha_n < +\infty$.

Proof. An application of Lemma 3.8 with $(\forall n \in \mathbb{N})\ \mathcal{F}_n = \{\varnothing, \Omega\}$. ☐

Lemma 3.10 *Let* $\xi \in L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathbb{R})$*, let* Φ *be a family of random variables, set* $\mathcal{X} = \sigma(\Phi)$*, and let* $\eta \in L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathbb{R})$ *be independent of* $\sigma(\{\xi\} \cup \Phi)$*. Then* $\mathsf{E}(\eta\xi \,|\, \mathcal{X}) = \mathsf{E}\eta\mathsf{E}(\xi \,|\, \mathcal{X})$.

Proof. Note that $\mathcal{X} \subset \sigma(\{\xi\} \cup \Phi)$ and that ξ is $\sigma(\{\xi\} \cup \Phi)$-measurable. Hence, it follows from [54, Properties H*, K*, and J* in Section 2.7.4] that

$$\mathsf{E}(\eta\xi \,|\, \mathcal{X}) = \mathsf{E}\Big(\mathsf{E}\big(\eta\xi \,\big|\, \sigma(\{\xi\} \cup \Phi)\big) \,\Big|\, \mathcal{X}\Big) = \mathsf{E}\Big(\xi\mathsf{E}\big(\eta \,\big|\, \sigma(\{\xi\} \cup \Phi)\big) \,\Big|\, \mathcal{X}\Big) = \mathsf{E}\big(\xi\mathsf{E}\eta \,\big|\, \mathcal{X}\big) = \mathsf{E}\eta\mathsf{E}(\xi \,|\, \mathcal{X}), \quad (3.21)$$

which proves the identity. ☐

Lemma 3.11 *Let* $\boldsymbol{x} = (x_1, \ldots, x_\mathsf{N})$ *be an* H^N*-valued random variable, let* $(\mathsf{K}, \mathcal{K})$ *be a measurable space, and suppose that the random variable* $k\colon (\Omega, \mathcal{F}, \mathsf{P}) \to (\mathsf{K}, \mathcal{K})$ *is independent of* $\sigma(\boldsymbol{x})$*. Let* $\mathsf{f}\colon (\mathsf{K} \times \mathsf{H}, \mathcal{K} \otimes \mathcal{B}_\mathsf{H}) \to \mathbb{R}$ *be measurable and such that* $\mathsf{E}|\mathsf{f}(k, x_1)| < +\infty$*, and define* $\mathsf{g}\colon \mathsf{H} \to \mathbb{R}\colon \mathsf{x} \mapsto \mathsf{E}\mathsf{f}(k, \mathsf{x})$*. Then, for* P*-almost every* $\omega' \in \Omega$*,*

$$\mathsf{E}\big(\mathsf{f}(k, x_1) \,\big|\, \sigma(\boldsymbol{x})\big)(\omega') = \int_\Omega \mathsf{f}\big(k(\omega), x_1(\omega')\big)\mathsf{P}(d\omega) = \mathsf{g}\big(x_1(\omega')\big). \quad (3.22)$$

Proof. Define $\mathsf{f}\colon \mathsf{K} \times \mathsf{H}^\mathsf{N} \to \mathbb{R}\colon (\mathsf{k}, \mathsf{x}) \mapsto \mathsf{f}(\mathsf{k}, \mathsf{x}_1)$. Then f is an $\mathbb{R}$-valued measurable function. Let $\mathsf{S} \in \sigma(\boldsymbol{x})$. Then there exists $\mathsf{A} \in \bigotimes_{1 \leqslant \mathsf{i} \leqslant \mathsf{N}} \mathcal{B}_\mathsf{H}$ such that $\mathsf{S} = [\boldsymbol{x} \in \mathsf{A}]$. Thus, it follows from the image measure theorem [54, Theorem 2.6.7], the independence of k and $\sigma(\boldsymbol{x})$, and Fubini's theorem [54, Theorem 2.6.8] that

$$\begin{aligned}
\int_\mathsf{S} \mathsf{f}\big(k(\omega), x_1(\omega)\big)\mathsf{P}(d\omega) &= \int_\Omega \mathsf{f}\big(k(\omega), \boldsymbol{x}(\omega)\big)1_\mathsf{A}\big(\boldsymbol{x}(\omega)\big)\mathsf{P}(d\omega) \\
&= \int_{\mathsf{K}\times\mathsf{H}^\mathsf{N}} \mathsf{f}(\mathsf{k}, \mathsf{x})1_\mathsf{A}(\mathsf{x})\big(\mathsf{P} \circ (k, \boldsymbol{x})^{-1}\big)(d\mathsf{k}, d\mathsf{x}) \\
&= \int_{\mathsf{K}\times\mathsf{H}^\mathsf{N}} \mathsf{f}(\mathsf{k}, \mathsf{x})1_\mathsf{A}(\mathsf{x})\big((\mathsf{P} \circ k^{-1}) \otimes (\mathsf{P} \circ \boldsymbol{x}^{-1})\big)(d\mathsf{k}, d\mathsf{x}) \\
&= \int_{\mathsf{H}^\mathsf{N}} 1_\mathsf{A}(\mathsf{x})\left(\int_\mathsf{K} \mathsf{f}(\mathsf{k}, \mathsf{x})(\mathsf{P} \circ k^{-1})(d\mathsf{k})\right)(\mathsf{P} \circ \boldsymbol{x}^{-1})(d\mathsf{x}) \\
&= \int_{\mathsf{H}^\mathsf{N}} 1_\mathsf{A}(\mathsf{x})\left(\int_\mathsf{K} \mathsf{f}(\mathsf{k}, \mathsf{x}_1)(\mathsf{P} \circ k^{-1})(d\mathsf{k})\right)(\mathsf{P} \circ \boldsymbol{x}^{-1})(d\mathsf{x}) \\
&= \int_{\mathsf{H}^\mathsf{N}} 1_\mathsf{A}(\mathsf{x})\mathsf{g}(\mathsf{x}_1)(\mathsf{P} \circ \boldsymbol{x}^{-1})(d\mathsf{x}) \\
&= \int_\Omega 1_\mathsf{A}\big(\boldsymbol{x}(\omega)\big)\mathsf{g}\big(x_1(\omega)\big)\mathsf{P}(d\omega) \\
&= \int_\mathsf{S} \mathsf{g}\big(x_1(\omega)\big)\mathsf{P}(d\omega). \qquad (3.23)
\end{aligned}$$

Therefore $\mathsf{g}(x_1) = \mathsf{E}(\mathsf{f}(k, x_1) \,|\, \sigma(\boldsymbol{x}))$ P-a.s. ☐

Lemma 3.12 *Let* $\mathsf{p} \in \,]1, +\infty[$*, let* $(\xi_\mathsf{n})_{\mathsf{n}\in\mathbb{N}}$ *be a sequence in* $L^\mathsf{p}(\Omega, \mathcal{F}, \mathsf{P}; \mathbb{R})$ *such that* $\sup_{\mathsf{n}\in\mathbb{N}} \mathsf{E}|\xi_\mathsf{n}|^\mathsf{p} < +\infty$*, and let* $\xi\colon \Omega \to \mathbb{R}$ *be measurable. Suppose that* $\xi_\mathsf{n} \to \xi$ P-a.s. *Then* $\mathsf{E}|\xi| < +\infty$*,* $\xi_\mathsf{n} \to \xi$ *in* $L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathbb{R})$*, and* $\mathsf{E}\xi_\mathsf{n} \to \mathsf{E}\xi$*.*

Proof. Set $\mathsf{q} = (\mathsf{p}-1)/\mathsf{p}$. It follows from the Hölder and Markov inequalities that

$$\begin{aligned}
0 &\leqslant \lim_{\beta\to+\infty} \sup_{n\in\mathbb{N}} \int_{[|\xi_n|\geqslant\beta]} |\xi_n| d\mathsf{P} \\
&\leqslant \lim_{\beta\to+\infty} \sup_{n\in\mathbb{N}} \Big(\mathsf{E}^{1/p}|\xi_n|^p \mathsf{E}^{1/q} 1^q_{[|\xi_n|\geqslant\beta]} \Big) \\
&\leqslant \sup_{n\in\mathbb{N}} \mathsf{E}^{1/p}|\xi_n|^p \lim_{\beta\to+\infty} \sup_{n\in\mathbb{N}} \Big(\mathsf{P}\big([|\xi_n| \geqslant \beta]\big)\Big)^{1/q} \\
&\leqslant \sup_{n\in\mathbb{N}} \mathsf{E}^{1/p}|\xi_n|^p \lim_{\beta\to+\infty} \sup_{n\in\mathbb{N}} \frac{\mathsf{E}^{1/q}|\xi_n|^p}{\beta^{p/q}} \\
&= 0,
\end{aligned} \tag{3.24}$$

which shows that $(\xi_n)_{n\in\mathbb{N}}$ is uniformly integrable. We can therefore invoke [54, Theorem 2.6.4(b)], which asserts that $\xi \in L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathbb{R})$, $\mathsf{E}\xi_n \to \mathsf{E}\xi$, and $\xi_n \to \xi$ in $L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathbb{R})$. ☐

Lemma 3.13 [31, Proposition 2.6.31] *Let $x \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$, let $\mathcal{X}$ be a sub σ-algebra of $\mathcal{F}$, and let $y \in L^2(\Omega, \mathcal{X}, \mathsf{P}; \mathsf{H})$. Then $\mathsf{E}(\langle x \mid y\rangle_{\mathsf{H}} \,|\, \mathcal{X}) = \langle \mathsf{E}(x \,|\, \mathcal{X}) \,|\, y\rangle_{\mathsf{H}}$.*

Lemma 3.14 *Let C be a nonempty closed subset of H and let $(x_n)_{n\in\mathbb{N}}$ be a sequence of H-valued random variables. Define*

$$\mathfrak{X} = (\mathcal{X}_n)_{n\in\mathbb{N}}, \quad \textit{where} \quad (\forall n \in \mathbb{N}) \;\; \mathcal{X}_n = \sigma(x_0, \dots, x_n). \tag{3.25}$$

Let $\mathsf{p} \in [1, +\infty[$ and suppose that, for every $\mathsf{z} \in \mathsf{C}$, there exist $(\mu_n(\mathsf{z}))_{n\in\mathbb{N}} \in \ell^1_+(\mathfrak{X})$, $(\theta_n(\mathsf{z}))_{n\in\mathbb{N}} \in \ell_+(\mathfrak{X})$, and $(\nu_n(\mathsf{z}))_{n\in\mathbb{N}} \in \ell^1_+(\mathfrak{X})$ such that

$$(\forall n \in \mathbb{N}) \quad \mathsf{E}\big(\|x_{n+1} - \mathsf{z}\|^p_{\mathsf{H}} \,\big|\, \mathcal{X}_n\big) + \theta_n(\mathsf{z}) \leqslant \big(1 + \mu_n(\mathsf{z})\big)\|x_n - \mathsf{z}\|^p_{\mathsf{H}} + \nu_n(\mathsf{z}) \;\; \mathsf{P}\text{-a.s.} \tag{3.26}$$

Then the following hold:

(i) *Let $\mathsf{z} \in \mathsf{C}$. Then $\sum_{n\in\mathbb{N}} \theta_n(\mathsf{z}) < +\infty$ P-a.s.*

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

(iii) *$\mathfrak{W}(x_n)_{n\in\mathbb{N}} \neq \varnothing$ P-a.s.*

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

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

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

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

(viii) *Suppose that* $\mathsf{z} \in \mathsf{C}$ *and* $(\chi_n)_{n\in\mathbb{N}}$ *in* $[0, +\infty[$ *satisfy*

$$(\forall n \in \mathbb{N}) \quad \mathsf{E}\big(\|x_{n+1} - \mathsf{z}\|_{\mathsf{H}}^p \,\big|\, \mathcal{X}_n\big) \leqslant \chi_n \|x_n - \mathsf{z}\|_{\mathsf{H}}^p \ \mathsf{P}\text{-a.s.} \ \ \textit{and} \ \ \overline{\lim}\, \chi_n < 1. \tag{3.27}$$

Then the following hold:

(a) *Let* $n \in \mathbb{N}$. *Then* $\mathsf{E}(\|x_{n+1} - \mathsf{z}\|_{\mathsf{H}}^p \,|\, \mathcal{X}_0) \leqslant (\prod_{j=0}^{n} \chi_j) \|x_0 - \mathsf{z}\|_{\mathsf{H}}^p$ P-a.s.

(b) *Suppose that* $x_0 \in L^p(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges strongly in* $L^p(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ *and* P-a.s. *to* z.

Proof. (i)-(vii): Apply [20, Proposition 2.3] with $\phi = |\cdot|^p$. The measurability of the weak limit in (v) relies on [20, Proposition 2.3(iv)] which invokes [47, Corollary 1.13]. The applicability of the latter follows from the separability of H and the completeness of $(\Omega, \mathcal{F}, \mathsf{P})$; see [31, Sections 1.1a–b] for details.

(viii): Apply [22, Lemma 2.2] with $\phi = |\cdot|^p$. ∎

3.2.3 Main results

3.2.3.1 An abstract stochastic algorithm

The analysis of the asymptotic behavior of the following algorithm will serve as the backbone of subsequent convergence results. We recall from Section 3.2.1 that Z is the solution set of the problem under consideration and that it is assumed to be nonempty and closed.

Algorithm 3.15 Let $x_0 \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. Iterate

$$\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor \begin{array}{l}
\mathcal{X}_n = \sigma(x_0, \ldots, x_n) \\
\text{take } \lambda_n \in L^\infty(\Omega, \mathcal{F}, \mathsf{P}; \,]0, +\infty[), \ d_n \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H}), \text{ and } \delta_n \in \mathfrak{C}(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H}) \text{ such that} \\
\qquad \begin{cases}
\mathsf{E}\big(\lambda_n(2 - \lambda_n)\|d_n\|_{\mathsf{H}}^2 \,\big|\, \mathcal{X}_n\big) \geqslant 0 \ \mathsf{P}\text{-a.s.}; \\
(\forall \mathsf{z} \in \mathsf{Z}) \ \ \mathsf{E}\big(\lambda_n \langle \mathsf{z} + d_n - x_n \mid d_n \rangle_{\mathsf{H}} \,\big|\, \mathcal{X}_n\big) \leqslant \delta_n(\cdot, \mathsf{z})/2 \ \mathsf{P}\text{-a.s.}
\end{cases} \\
x_{n+1} = x_n - \lambda_n d_n.
\end{array} \right.
\end{array} \tag{3.28}$$

Let us outline the weak and strong convergence properties of Algorithm 3.15.

Theorem 3.16 *Let* $(x_n)_{n\in\mathbb{N}}$ *be the sequence generated by Algorithm* 3.15. *Then the following hold:*

(i) $(x_n)_{n\in\mathbb{N}}$ *is a well-defined sequence in* $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$.

(ii) *Let* $n \in \mathbb{N}$ *and* $\mathsf{z} \in \mathsf{Z}$. *Then*

$$\mathsf{E}\big(\|x_{n+1} - \mathsf{z}\|_{\mathsf{H}}^2 \,\big|\, \mathcal{X}_n\big) \leqslant \|x_n - \mathsf{z}\|_{\mathsf{H}}^2 - \mathsf{E}\big(\lambda_n(2 - \lambda_n)\|d_n\|_{\mathsf{H}}^2 \,\big|\, \mathcal{X}_n\big) + \delta_n(\cdot, \mathsf{z}) \ \ \mathsf{P}\text{-a.s.}$$

(iii) *Let* $\mathsf{n} \in \mathbb{N}$ *and* $z \in L^2(\Omega, \mathcal{X}_\mathsf{n}, \mathsf{P}; \mathsf{Z})$. *Then*

$$\mathsf{E}\big(\|x_{\mathsf{n}+1} - z\|_\mathsf{H}^2 \,\big|\, \mathcal{X}_\mathsf{n}\big) \leqslant \|x_\mathsf{n} - z\|_\mathsf{H}^2 - \mathsf{E}\big(\lambda_\mathsf{n}(2 - \lambda_\mathsf{n})\|d_\mathsf{n}\|_\mathsf{H}^2 \,\big|\, \mathcal{X}_\mathsf{n}\big) + \delta_\mathsf{n}(\cdot, z) \;\; \mathsf{P}\text{-a.s.}$$

(iv) *Let* $\mathsf{n} \in \mathbb{N}$ *and* $z \in L^2(\Omega, \mathcal{X}_\mathsf{n}, \mathsf{P}; \mathsf{Z})$. *Then*

$$\|x_{\mathsf{n}+1} - z\|^2_{L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})} \leqslant \|x_\mathsf{n} - z\|^2_{L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})} - \mathsf{E}\big(\lambda_\mathsf{n}(2 - \lambda_\mathsf{n})\|d_\mathsf{n}\|_\mathsf{H}^2\big) + \mathsf{E}\delta_\mathsf{n}(\cdot, z).$$

(v) *Suppose that, for every* $\mathsf{z} \in \mathsf{Z}$, $\sum_{\mathsf{n}\in\mathbb{N}} \delta_\mathsf{n}(\cdot, \mathsf{z}) < +\infty$ P-a.s. *Then the following hold:*

(a) $(\|x_\mathsf{n}\|_\mathsf{H})_{\mathsf{n}\in\mathbb{N}}$ *is bounded* P-a.s.

(b) *Let* z *be a* Z*-valued random variable. Then* $(\|x_\mathsf{n} - z\|_\mathsf{H})_{\mathsf{n}\in\mathbb{N}}$ *converges* P-a.s.

(c) $\sum_{\mathsf{n}\in\mathbb{N}} \mathsf{E}(\lambda_\mathsf{n}(2 - \lambda_\mathsf{n})\|d_\mathsf{n}\|_\mathsf{H}^2 \,|\, \mathcal{X}_\mathsf{n}) < +\infty$ P-a.s.

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

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

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

(vi) *Suppose that, for every* $\mathsf{z} \in \mathsf{Z}$, $\sum_{\mathsf{n}\in\mathbb{N}} \mathsf{E}\delta_\mathsf{n}(\cdot, \mathsf{z}) < +\infty$. *Then the following hold:*

(a) $(\|x_\mathsf{n}\|_{L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})})_{\mathsf{n}\in\mathbb{N}}$ *is bounded.*

(b) *Let* $z \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{Z})$. *Then* $(\|x_\mathsf{n} - z\|_{L^1(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})})_{\mathsf{n}\in\mathbb{N}}$ *converges.*

(c) $\sum_{\mathsf{n}\in\mathbb{N}} \mathsf{E}(\lambda_\mathsf{n}\,(2 - \lambda_\mathsf{n})\,\|d_\mathsf{n}\|_\mathsf{H}^2) < +\infty$.

(d) *Suppose that* $(x_\mathsf{n})_{\mathsf{n}\in\mathbb{N}}$ *converges weakly* P-a.s. *to an* H*-valued random variable* x. *Then* $x \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ *and* $(x_\mathsf{n})_{\mathsf{n}\in\mathbb{N}}$ *converges weakly in* $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ *to* x.

(e) *Let* x *be a* Z*-valued random variable. Then* $(x_\mathsf{n})_{\mathsf{n}\in\mathbb{N}}$ *converges strongly* P-a.s. *to* x *if and only if* $(x_\mathsf{n})_{\mathsf{n}\in\mathbb{N}}$ *converges strongly in* $L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ *to* x. *In this case,* $x \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{Z})$ *and* $(x_\mathsf{n})_{\mathsf{n}\in\mathbb{N}}$ *converges weakly in* $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ *to* x.

Proof. (i): By assumption, $x_0 \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. Now suppose that $x_\mathsf{n} \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. Then, since $d_\mathsf{n} \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ and $\lambda_\mathsf{n} \in L^\infty(\Omega, \mathcal{F}, \mathsf{P}; \left]0, +\infty\right[)$, $x_{\mathsf{n}+1} = x_\mathsf{n} - \lambda_\mathsf{n} d_\mathsf{n} \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. This establishes the claim by induction.

(ii): We derive from (3.28) that

$$\begin{aligned}
&\mathsf{E}\big(\|x_{\mathsf{n}+1} - \mathsf{z}\|_\mathsf{H}^2 \,\big|\, \mathcal{X}_\mathsf{n}\big)\\
&\quad= \mathsf{E}\big(\|x_\mathsf{n} - \mathsf{z}\|_\mathsf{H}^2 - 2\lambda_\mathsf{n}\langle x_\mathsf{n} - \mathsf{z} \,|\, d_\mathsf{n}\rangle_\mathsf{H} + \lambda_\mathsf{n}^2\|d_\mathsf{n}\|_\mathsf{H}^2 \,\big|\, \mathcal{X}_\mathsf{n}\big)\\
&\quad= \|x_\mathsf{n} - \mathsf{z}\|_\mathsf{H}^2 - \mathsf{E}\big(\lambda_\mathsf{n}(2 - \lambda_\mathsf{n})\|d_\mathsf{n}\|_\mathsf{H}^2 \,\big|\, \mathcal{X}_\mathsf{n}\big) + 2\mathsf{E}\big(\lambda_\mathsf{n}\langle \mathsf{z} + d_\mathsf{n} - x_\mathsf{n} \,|\, d_\mathsf{n}\rangle_\mathsf{H} \,\big|\, \mathcal{X}_\mathsf{n}\big)\\
&\quad\leqslant \|x_\mathsf{n} - \mathsf{z}\|_\mathsf{H}^2 - \mathsf{E}\big(\lambda_\mathsf{n}(2 - \lambda_\mathsf{n})\|d_\mathsf{n}\|_\mathsf{H}^2 \,\big|\, \mathcal{X}_\mathsf{n}\big) + \delta_\mathsf{n}(\cdot, \mathsf{z}) \;\; \mathsf{P}\text{-a.s.}
\end{aligned} \tag{3.29}$$

(iii): First, let s be a Z-valued $\mathcal{X}_n$-simple mapping. Then there exists a finite family of disjoint sets $(\mathsf{F}_i)_{i\in I}$ in $\mathcal{X}_n$ such that $\bigcup_{i\in I}\mathsf{F}_i = \Omega$, and a family $(\mathsf{z}_i)_{i\in I}$ in Z such that $s = \sum_{i\in I} 1_{\mathsf{F}_i}\mathsf{z}_i$. Then, by (ii),

$$
\begin{aligned}
\mathsf{E}\big(\|x_{n+1}-s\|_{\mathsf{H}}^2 \,|\, \mathcal{X}_n\big) &= \mathsf{E}\left(\left\|\sum_{i\in I} 1_{\mathsf{F}_i}(x_{n+1}-\mathsf{z}_i)\right\|_{\mathsf{H}}^2 \,\middle|\, \mathcal{X}_n\right)\\
&= \mathsf{E}\left(\sum_{i\in I}\|x_{n+1}-\mathsf{z}_i\|_{\mathsf{H}}^2 1_{\mathsf{F}_i} \,\middle|\, \mathcal{X}_n\right)\\
&= \sum_{i\in I}\mathsf{E}\big(\|x_{n+1}-\mathsf{z}_i\|_{\mathsf{H}}^2 \,|\, \mathcal{X}_n\big)1_{\mathsf{F}_i}\\
&\leqslant \sum_{i\in I}\|x_n-\mathsf{z}_i\|_{\mathsf{H}}^2 1_{\mathsf{F}_i} + \sum_{i\in I}\Big(-\mathsf{E}\big(\lambda_n(2-\lambda_n)\|d_n\|_{\mathsf{H}}^2 \,|\, \mathcal{X}_n\big) + \delta_n(\cdot,\mathsf{z}_i)\Big)1_{\mathsf{F}_i}\\
&= \left\|\sum_{i\in I}1_{\mathsf{F}_i}(x_n-\mathsf{z}_i)\right\|_{\mathsf{H}}^2 - \mathsf{E}\big(\lambda_n(2-\lambda_n)\|d_n\|_{\mathsf{H}}^2 \,|\, \mathcal{X}_n\big) + \sum_{i\in I}\delta_n(\cdot,\mathsf{z}_i)1_{\mathsf{F}_i}\\
&= \|x_n-s\|_{\mathsf{H}}^2 - \mathsf{E}\big(\lambda_n(2-\lambda_n)\|d_n\|_{\mathsf{H}}^2 \,|\, \mathcal{X}_n\big) + \delta_n(\cdot,s) \;\; \mathsf{P}\text{-a.s.}
\end{aligned}
\tag{3.30}
$$

Next, Proposition 3.7 guarantees the existence of a sequence of Z-valued $\mathcal{X}_n$-simple mappings $(s_j)_{j\in\mathbb{N}}$ such that $s_j \to z$ P-a.s. and $\sup_{j\in\mathbb{N}}\|s_j\|_{\mathsf{H}}^2 \leqslant \|z\|_{\mathsf{H}}^2+1$ P-a.s. Thus, we derive from (3.30) that

$$
(\forall j\in\mathbb{N})\quad \mathsf{E}\big(\|x_{n+1}-s_j\|_{\mathsf{H}}^2 \,|\, \mathcal{X}_n\big) \leqslant \|x_n-s_j\|_{\mathsf{H}}^2 - \mathsf{E}\big(\lambda_n(2-\lambda_n)\|d_n\|_{\mathsf{H}}^2 \,|\, \mathcal{X}_n\big) + \delta_n(\cdot,s_j) \;\; \mathsf{P}\text{-a.s.} \tag{3.31}
$$

Additionally,

$$
(\forall j\in\mathbb{N})\quad \|x_{n+1}-s_j\|_{\mathsf{H}}^2 \leqslant 2\|x_{n+1}\|_{\mathsf{H}}^2 + 2\|s_j\|_{\mathsf{H}}^2 \leqslant 2\|x_{n+1}\|_{\mathsf{H}}^2 + 2\|z\|_{\mathsf{H}}^2 + 2 \;\; \mathsf{P}\text{-a.s.} \tag{3.32}
$$

Note that the right-hand term in (3.32) is integrable and that $\|x_{n+1}-s_j\|_{\mathsf{H}}^2 \to \|x_{n+1}-z\|_{\mathsf{H}}^2$ P-a.s. as $j\to+\infty$. Therefore, by the conditional dominated convergence theorem [54, Theorem 2.7.2(a)],

$$
\mathsf{E}(\|x_{n+1}-s_j\|_{\mathsf{H}}^2 \,|\, \mathcal{X}_n) \to \mathsf{E}(\|x_{n+1}-z\|_{\mathsf{H}}^2 \,|\, \mathcal{X}_n) \;\; \mathsf{P}\text{-a.s. as } j\to+\infty. \tag{3.33}
$$

On the other hand, the continuity of δ_n with respect to the H-variable implies that $\delta_n(\cdot,s_j) \to \delta_n(\cdot,z)$ P-a.s. as $j\to+\infty$. Altogether, taking the limit as $j\to+\infty$ in (3.31) yields

$$
\mathsf{E}\big(\|x_{n+1}-z\|_{\mathsf{H}}^2 \,|\, \mathcal{X}_n\big) \leqslant \|x_n-z\|_{\mathsf{H}}^2 - \mathsf{E}\big(\lambda_n(2-\lambda_n)\|d_n\|_{\mathsf{H}}^2 \,|\, \mathcal{X}_n\big) + \delta_n(\cdot,z) \;\; \mathsf{P}\text{-a.s.} \tag{3.34}
$$

(iv): Take the expected value in (iii).

(v)(a): This follows from (ii) and Lemma 3.14(ii).

(v)(b): Let $\Omega'' \in \mathcal{F}$ be such that $\mathsf{P}(\Omega'') = 1$ and, for every $\omega\in\Omega''$, $z(\omega)\in\mathsf{Z}$. Further, let

$\Omega' \in \mathcal{F}$ be given as in Lemma 3.14(iv), which holds as a consequence of (ii). Then

$$(\forall \omega \in \Omega' \cap \Omega'') \quad \big(\|x_n(\omega) - z(\omega)\|_H\big)_{n\in\mathbb{N}} \text{ converges,} \tag{3.35}$$

which confirms that $(\|x_n - z\|_H)_{n\in\mathbb{N}}$ converges P-a.s. since $P(\Omega' \cap \Omega'') = 1$.

(v)(c): Let $z \in Z$. In view of (ii) and Lemma 3.14(i),

$$\sum_{n\in\mathbb{N}} E\big(\lambda_n(2-\lambda_n)\|d_n\|_H^2 \,\big|\, \mathcal{X}_n\big) < +\infty \text{ P-a.s.} \tag{3.36}$$

(v)(d)–(v)(f): These follow from (ii) and Lemma 3.14(v)–(vii).

(vi)(a): Note that $\{\varnothing, \Omega\} \subset \bigcap_{n\in\mathbb{N}} \mathcal{X}_n$. It follows from (iv) and the assumption that, for every $z \in Z$, $\sum_{n\in\mathbb{N}} E\delta_n(\cdot, z) < +\infty$, that $(x_n)_{n\in\mathbb{N}}$ is quasi-Fejér of Type III in $L^2(\Omega, \mathcal{F}, P; H)$ relative to $L^2(\Omega, \{\varnothing, \Omega\}, P; Z)$ [16, Definition 1.1]. Hence, [16, Proposition 3.3(i)] implies that $(x_n)_{n\in\mathbb{N}}$ is bounded in $L^2(\Omega, \mathcal{F}, P; H)$.

(vi)(b): It follows from (vi)(a) that $\sup_{n\in\mathbb{N}} E\|x_n - z\|_H^2 < +\infty$ and from (v)(b) that $(\|x_n - z\|_H)_{n\in\mathbb{N}}$ converges P-a.s. We then invoke Lemma 3.12 to deduce that $E\|x_n - z\|_H \to E\big(\lim \|x_n - z\|_H\big) < +\infty$.

(vi)(c): Let $z \in Z$. Then, in view of (iv) and Corollary 3.9(i),

$$\sum_{n\in\mathbb{N}} E\big(\lambda_n(2-\lambda_n)\|d_n\|_H^2\big) < +\infty. \tag{3.37}$$

(vi)(d): In view of (vi)(a), $(x_n)_{n\in\mathbb{N}}$ possesses a weak sequential cluster point $w \in L^2(\Omega, \mathcal{F}, P; H)$, i.e., there exists a strictly increasing sequence $(k_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$ such that $x_{k_n} \rightharpoonup w$ in $L^2(\Omega, \mathcal{F}, P; H)$. However, since H is separable, it contains a countable dense set $\{y_j\}_{j\in\mathbb{N}}$. Let us fix temporarily $j \in \mathbb{N}$ and identify y_j with a constant random variable in $L^2(\Omega, \mathcal{F}, P; H)$. Then $E\langle x_{k_n} - w \mid y_j\rangle_H \to 0$ and we can therefore extract a further subsequence $(x_{l_{k_n}})_{n\in\mathbb{N}}$ such that $\langle x_{l_{k_n}} - w \mid y_j\rangle_H \to 0$ P-a.s. On the other hand, the assumption yields $\langle x_{l_{k_n}} - x \mid y_j\rangle_H \to 0$ P-a.s. We deduce from the P-almost sure uniqueness of the limit that there exists $\Omega_j \in \mathcal{F}$ such that $P(\Omega_j) = 1$ and

$$(\forall \omega \in \Omega_j) \quad \langle x(\omega) \mid y_j\rangle_H = \langle w(\omega) \mid y_j\rangle_H. \tag{3.38}$$

Let us set $\Omega'' = \bigcap_{j\in\mathbb{N}} \Omega_j$ and note that $P(\Omega'') = 1$. Then (3.38) yields

$$(\forall j \in \mathbb{N})(\forall \omega \in \Omega'') \quad \langle x(\omega) - w(\omega) \mid y_j\rangle_H = 0. \tag{3.39}$$

By density, for every $\omega \in \Omega''$, there exists a strictly increasing sequence $(i_j)_{j\in\mathbb{N}}$ such that $y_{i_j} \to x(\omega) - w(\omega)$ and it results from (3.39) that

$$\|x(\omega) - w(\omega)\|_H^2 = \langle x(\omega) - w(\omega) \mid x(\omega) - w(\omega)\rangle_H = 0, \tag{3.40}$$

which shows that $x(\omega) = w(\omega)$. Thus, $x = w$ P-a.s. and it follows from [5, Lemma 2.46] that $x_n \rightharpoonup x$ in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$.

(vi)(e): Suppose that $x_n \to x$ P-a.s. Then it follows from (vi)(a) and Lemma 3.12 that $x \in L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{Z})$ and $x_n \to x$ in $L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. Conversely, suppose that $x \in L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{Z})$ and $x_n \to x$ in $L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$, i.e., $\mathsf{E}\|x_n - x\|_{\mathsf{H}} \to 0$. Then there exists a strictly increasing sequence $(k_n)_{n\in\mathbb{N}}$ in $\mathbb{N}$ such that $\|x_{k_n} - x\|_{\mathsf{H}} \to 0$ P-a.s. On the other hand, (v)(b) guarantees that $(\|x_n - x\|_{\mathsf{H}})_{n\in\mathbb{N}}$ converges P-a.s. Since the P-almost sure limit of any subsequence coincides with the P-almost sure limit of the sequence, we conclude that $\|x_n - x\|_{\mathsf{H}} \to 0$ P-a.s. Additionally, as P-almost sure strong convergence implies P-almost sure weak convergence, we deduce from (vi)(d) that $x \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ and $x_n \rightharpoonup x$ in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. ☐

We now assume that the tolerance variables $(\delta_n)_{n\in\mathbb{N}}$ are constant with respect to the H-variable and depend only on the Ω-variable.

Theorem 3.17 *Let $(x_n)_{n\in\mathbb{N}}$ be the sequence generated by Algorithm 3.15. For every $n \in \mathbb{N}$, assume that δ_n is constant with respect to the* H*-variable and set, for every $\omega \in \Omega$, $\vartheta_n(\omega) = \delta_n(\omega, 0)$. Then the following hold:*

(i) *Let $n \in \mathbb{N}$. Then* $\mathsf{E}(d_{\mathsf{Z}}^2(x_{n+1}) \,|\, \mathcal{X}_n) \leqslant d_{\mathsf{Z}}^2(x_n) + \vartheta_n$ P-a.s.

(ii) *Let $n \in \mathbb{N}$. Then* $\mathsf{E}d_{\mathsf{Z}}^2(x_{n+1}) \leqslant \mathsf{E}d_{\mathsf{Z}}^2(x_n) + \mathsf{E}\vartheta_n$.

(iii) *Suppose that* $\sum_{n\in\mathbb{N}} \vartheta_n < +\infty$ P-a.s. *Then* $(d_{\mathsf{Z}}(x_n))_{n\in\mathbb{N}}$ *converges* P-a.s.

(iv) *Suppose that* $\sum_{n\in\mathbb{N}} \mathsf{E}\vartheta_n < +\infty$. *Then the following hold:*

  (a) $(\mathsf{E}d_{\mathsf{Z}}^2(x_n))_{n\in\mathbb{N}}$ *converges.*

  (b) *Suppose that* Z *is convex and that* $\underline{\lim}\, \mathsf{E}d_{\mathsf{Z}}^2(x_n) = 0$. *Then $(x_n)_{n\in\mathbb{N}}$ converges strongly in* $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ *and* P-a.s. *to a* Z*-valued random variable.*

  (c) *Suppose that* Z *is convex and that there exists* $\chi \in\,]0, 1[$ *such that*

$$(\forall n \in \mathbb{N}) \quad \mathsf{E}\big(d_{\mathsf{Z}}^2(x_{n+1}) \,\big|\, \mathcal{X}_n\big) \leqslant \chi d_{\mathsf{Z}}^2(x_n) + \vartheta_n \;\; \text{P-a.s.} \tag{3.41}$$

  Then the following are satisfied:

  [A] *Let $n \in \mathbb{N}$. Then* $\mathsf{E}d_{\mathsf{Z}}^2(x_{n+1}) \leqslant \chi^{n+1}\mathsf{E}d_{\mathsf{Z}}^2(x_0) + \sum_{j=0}^{n} \chi^{n-j}\mathsf{E}\vartheta_j$.

  [B] *There exists* $x \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{Z})$ *such that $(x_n)_{n\in\mathbb{N}}$ converges strongly in* $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ *and* P-a.s. *to x, and*

$$(\forall n \in \mathbb{N}) \quad \mathsf{E}\|x_n - x\|_{\mathsf{H}}^2 \leqslant 4\chi^n \mathsf{E}d_{\mathsf{Z}}^2(x_0) + 4\sum_{j=0}^{n-1} \chi^{n-j-1}\mathsf{E}\vartheta_j + 2\sum_{j\geqslant n} \mathsf{E}\vartheta_j. \tag{3.42}$$

Proof. (i): Let $z \in \mathsf{Z}$. Then Theorem 3.16(ii) yields $\mathsf{E}(\|x_{n+1} - z\|_{\mathsf{H}}^2 \,|\, \mathcal{X}_n) \leqslant \|x_n - z\|_{\mathsf{H}}^2 + \vartheta_n$ P-a.s. On the other hand, $d_{\mathsf{Z}}(x_{n+1}) \leqslant \|x_{n+1} - z\|_{\mathsf{H}}$ P-a.s. Thus,

$$\mathsf{E}\big(d_{\mathsf{Z}}^2(x_{n+1}) \,\big|\, \mathcal{X}_n\big) \leqslant \mathsf{E}\big(\|x_{n+1} - z\|_{\mathsf{H}}^2 \,\big|\, \mathcal{X}_n\big) \leqslant \|x_n - z\|_{\mathsf{H}}^2 + \vartheta_n \;\; \text{P-a.s.} \tag{3.43}$$

Taking the infimum over $z \in Z$ yields the claim.

(ii): Take the expected value in (i).

(iii): This follows from (i) and Lemma 3.8(i).

(iv)(a): This follows from (ii) and Corollary 3.9(i).

(iv)(b): Let $n \in \mathbb{N}$, $m \in \mathbb{N} \smallsetminus \{0\}$, and $z \in L^2(\Omega, \mathcal{X}_n, P; Z)$. Then $z \in \bigcap_{1 \leqslant j \leqslant m} L^2(\Omega, \mathcal{X}_{n+j}, P; H)$ and we derive inductively from (3.28) and Theorem 3.16(iii) that

$$\begin{aligned} E(\|x_n - x_{n+m}\|_H^2 \,|\, \mathcal{X}_n) &\leqslant 2E\Big(\|x_n - z\|_H^2 + \|x_{n+m} - z\|_H^2 \,\Big|\, \mathcal{X}_n\Big) \\ &\leqslant 2\|x_n - z\|_H^2 + 2E\Big(E(\|x_{n+m} - z\|_H^2 \,|\, \mathcal{X}_{n+m-1}) \,\Big|\, \mathcal{X}_n\Big) \\ &\leqslant 4\|x_n - z\|_H^2 + 2\sum_{j=n}^{n+m-1} \vartheta_j \quad P\text{-a.s.} \end{aligned} \tag{3.44}$$

Now assume that $z = \operatorname{proj}_Z x_n$ and recall that proj_Z is nonexpansive [5, Proposition 4.16] while x_n is $(\mathcal{X}_n, \mathcal{B}_H)$-measurable. Consequently, z is $(\mathcal{X}_n, \mathcal{B}_H)$-measurable. Given $y \in L^2(\Omega, \mathcal{X}_n, P; Z)$,

$$\begin{aligned} \frac{1}{2}E\|z\|_H^2 = \frac{1}{2}E\|z - y + y\|_H^2 &\leqslant E\|\operatorname{proj}_Z x_n - \operatorname{proj}_Z y\|_H^2 + E\|y\|_H^2 \\ &\leqslant \|x_n - y\|^2_{L^2(\Omega,\mathcal{X}_n,P;Z)} + \|y\|^2_{L^2(\Omega,\mathcal{X}_n,P;Z)}, \end{aligned} \tag{3.45}$$

which shows that $z \in L^2(\Omega, \mathcal{X}_n, P; Z)$. Further, (3.44) yields

$$E(\|x_n - x_{n+m}\|_H^2 \,|\, \mathcal{X}_n) \leqslant 4d_Z^2(x_n) + 2\sum_{j=n}^{n+m-1} \vartheta_j \quad P\text{-a.s.} \tag{3.46}$$

Therefore, upon taking expectations, we get

$$E\|x_n - x_{n+m}\|_H^2 \leqslant 4Ed_Z^2(x_n) + 2\sum_{j=n}^{n+m-1} E\vartheta_j. \tag{3.47}$$

The assumption $\underline{\lim}\, Ed_Z^2(x_n) = 0$ and (iv)(a) yield $\lim Ed_Z^2(x_n) = 0$. In addition,

$$(\forall m \in \mathbb{N} \smallsetminus \{0\}) \quad 0 \leqslant \sum_{j=n}^{n+m-1} E\vartheta_j \leqslant \sum_{j \geqslant n} E\vartheta_j \to 0 \text{ as } n \to +\infty. \tag{3.48}$$

We thus infer from (3.47) that $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $L^2(\Omega, \mathcal{F}, P; H)$, which implies that there exists $x \in L^2(\Omega, \mathcal{F}, P; H)$ such that $x_n \to x$ in $L^2(\Omega, \mathcal{F}, P; H)$. Further, since $d_Z^2 \colon H \to [0, +\infty[$ is continuous, $d_Z^2(x_n) \to d_Z^2(x)$ P-a.s. In addition, it follows from Fatou's lemma that

$$0 \leqslant Ed_Z^2(x) \leqslant \underline{\lim}\, Ed_Z^2(x_n) = 0. \tag{3.49}$$

Hence $Ed_Z^2(x) = 0$, $d_Z^2(x) = 0$ P-a.s., and $x \in Z$ P-a.s. Finally, Theorem 3.16(vi)(e) yields

$x_n \to x$ P-a.s.

(iv)(c):

[A]: Taking expectations in (3.41) yields $\mathsf{E}d_Z^2(x_{n+1}) \leqslant \chi \mathsf{E}d_Z^2(x_n) + \mathsf{E}\vartheta_n$. The claim follows by induction.

[B]: It follows from Corollary 3.9(ii) that $\lim \mathsf{E}d_Z^2(x_n) = 0$. Therefore, (iv)(b) implies that $(x_n)_{n\in\mathbb{N}}$ converges strongly in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ and P-a.s. to a Z-valued random variable. Finally, arguing as in [16, Theorem 3.13(ii)], we obtain (3.42). $\square$

3.2.3.2 A stochastic algorithm with super relaxations

We study an implementation of Algorithm 3.2 in which the standard condition that the relaxations are deterministic and bounded above by 2 is not imposed. In Section 3.2.1 we called such relaxations super relaxations.

Algorithm 3.18 In Algorithm 3.2 assume that, for every $n \in \mathbb{N}$, $\varepsilon_n \in \mathfrak{C}(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$, λ_n is independent of $\sigma(\{x_0, \ldots, x_n, d_n\})$, and $\mathsf{E}(\lambda_n(2-\lambda_n)) \geqslant 0$.

Proposition 3.19 *Algorithm 3.18 is a special case of Algorithm 3.15 where, for every* $n \in \mathbb{N}$, $\delta_n = 2\varepsilon_n \mathsf{E}\lambda_n$.

Proof. Let $(x_n)_{n\in\mathbb{N}}$ be the sequence generated by Algorithm 3.18. Let us first show by induction that it is a well-defined sequence in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. By assumption, $x_0 \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. Fix $n \in \mathbb{N}$ and note that d_n is measurable as a combination of measurable functions. Additionally, (3.2) yields

$$\begin{aligned}
\frac{1}{2}\mathsf{E}\|d_n\|_{\mathsf{H}}^2 &= \frac{1}{2}\mathsf{E}\|\alpha_n t_n^*\|_{\mathsf{H}}^2 \\
&\leqslant \frac{1}{2}\mathsf{E}\left\|\frac{1_{[t_n^*\neq 0]}1_{[\langle x_n \mid t_n^*\rangle_{\mathsf{H}} > \eta_n]}\big(\langle x_n \mid t_n^*\rangle_{\mathsf{H}} - \eta_n\big)}{\|t_n^*\|_{\mathsf{H}}^2 + 1_{[t_n^*=0]}}\, t_n^*\right\|_{\mathsf{H}}^2 \\
&= \frac{1}{2}\mathsf{E}\left|\frac{1_{[t_n^*\neq 0]}1_{[\langle x_n \mid t_n^*\rangle_{\mathsf{H}} > \eta_n]}\big(\langle x_n \mid t_n^*\rangle_{\mathsf{H}} - \eta_n\big)}{\|t_n^*\|_{\mathsf{H}} + 1_{[t_n^*=0]}}\right|^2 \\
&\leqslant \mathsf{E}\left|\frac{1_{[t_n^*\neq 0]}1_{[\langle x_n \mid t_n^*\rangle_{\mathsf{H}} > \eta_n]}\langle x_n \mid t_n^*\rangle_{\mathsf{H}}}{\|t_n^*\|_{\mathsf{H}} + 1_{[t_n^*=0]}}\right|^2 + \mathsf{E}\left|\frac{1_{[t_n^*\neq 0]}1_{[\langle x_n \mid t_n^*\rangle_{\mathsf{H}} > \eta_n]}\eta_n}{\|t_n^*\|_{\mathsf{H}} + 1_{[t_n^*=0]}}\right|^2 \\
&\leqslant \mathsf{E}\left|\frac{\|x_n\|_{\mathsf{H}}\,\|t_n^*\|_{\mathsf{H}}}{\|t_n^*\|_{\mathsf{H}} + 1_{[t_n^*=0]}}\right|^2 + \mathsf{E}\left|\frac{1_{[t_n^*\neq 0]}1_{[\langle x_n \mid t_n^*\rangle_{\mathsf{H}} > \eta_n]}\eta_n}{\|t_n^*\|_{\mathsf{H}} + 1_{[t_n^*=0]}}\right|^2 \\
&\leqslant \mathsf{E}\|x_n\|_{\mathsf{H}}^2 + \mathsf{E}\left|\frac{1_{[t_n^*\neq 0]}1_{[\langle x_n \mid t_n^*\rangle_{\mathsf{H}} > \eta_n]}\eta_n}{\|t_n^*\|_{\mathsf{H}} + 1_{[t_n^*=0]}}\right|^2 \\
&< +\infty.
\end{aligned} \tag{3.50}$$

Thus, $d_n \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ and, since $\lambda_n \in L^\infty(\Omega, \mathcal{F}, \mathsf{P};]0, +\infty[)$, $x_{n+1} = x_n - \lambda_n d_n \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$, which completes the induction argument. The fact that $\lambda_n \in L^\infty(\Omega, \mathcal{F}, \mathsf{P};]0, +\infty[)$ also guarantees the integrability of λ_n and $\lambda_n(2 - \lambda_n)$. Further, since λ_n is independent of $\sigma(\{x_0, \ldots, x_n, d_n\})$ and $\mathsf{E}(\lambda_n(2 - \lambda_n)) \geqslant 0$, it follows from Lemma 3.10 that

$$\mathsf{E}\big(\lambda_n(2 - \lambda_n)\|d_n\|_{\mathsf{H}}^2 \,\big|\, \mathcal{X}_n\big) = \mathsf{E}\big(\lambda_n(2 - \lambda_n)\big)\mathsf{E}\big(\|d_n\|_{\mathsf{H}}^2 \,\big|\, \mathcal{X}_n\big) \geqslant 0 \ \ \mathsf{P}\text{-a.s.} \tag{3.51}$$

Next, we infer from (3.2) that

$$(\forall n \in \mathbb{N}) \quad \alpha_n \eta_n = \langle x_n \mid \alpha_n t_n^* \rangle_{\mathsf{H}} - \alpha_n^2 \|t_n^*\|_{\mathsf{H}}^2 = \langle x_n \mid d_n \rangle_{\mathsf{H}} - \|d_n\|_{\mathsf{H}}^2, \tag{3.52}$$

which shows that $\alpha_n \eta_n \in L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathbb{R})$. Now set $\delta_n = 2\varepsilon_n \mathsf{E}\lambda_n \in \mathfrak{C}(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ and let $\mathsf{z} \in \mathsf{Z}$. Then we deduce from (3.2), Lemma 3.13, and (3.52) that

$$\begin{aligned}(\forall n \in \mathbb{N}) \quad \mathsf{E}\big(\langle \mathsf{z} \mid d_n \rangle_{\mathsf{H}} \,\big|\, \mathcal{X}_n\big) &= \langle \mathsf{z} \mid \mathsf{E}(\alpha_n t_n^* \mid \mathcal{X}_n) \rangle_{\mathsf{H}} \\ &\leqslant \mathsf{E}(\alpha_n \eta_n \mid \mathcal{X}_n) + \varepsilon_n(\cdot, \mathsf{z}) \\ &= \mathsf{E}\big(\langle x_n \mid d_n \rangle_{\mathsf{H}} - \|d_n\|_{\mathsf{H}}^2 \,\big|\, \mathcal{X}_n\big) + \varepsilon_n(\cdot, \mathsf{z}) \ \ \mathsf{P}\text{-a.s.}\end{aligned} \tag{3.53}$$

Finally, we derive from (3.53) and Lemma 3.10 that

$$\begin{aligned}(\forall n \in \mathbb{N}) \quad \mathsf{E}\big(\lambda_n \langle \mathsf{z} + d_n - x_n \mid d_n \rangle_{\mathsf{H}} \,\big|\, \mathcal{X}_n\big) &= \mathsf{E}\big(\langle \mathsf{z} + d_n - x_n \mid d_n \rangle_{\mathsf{H}} \,\big|\, \mathcal{X}_n\big)\mathsf{E}\lambda_n \\ &= \mathsf{E}\big(\langle \mathsf{z} \mid d_n \rangle_{\mathsf{H}} + \|d_n\|_{\mathsf{H}}^2 - \langle x_n \mid d_n \rangle_{\mathsf{H}} \,\big|\, \mathcal{X}_n\big)\mathsf{E}\lambda_n \\ &\leqslant \varepsilon_n(\cdot, \mathsf{z})\mathsf{E}\lambda_n \\ &= \frac{\delta_n(\cdot, \mathsf{z})}{2} \ \ \mathsf{P}\text{-a.s.},\end{aligned} \tag{3.54}$$

which yields the claim. $\square$

The asymptotic behavior of Algorithm 3.18 is our next topic. We leverage Proposition 3.19 and Theorems 3.16 and 3.17 to obtain the following properties.

Theorem 3.20 *Let $(x_n)_{n \in \mathbb{N}}$ be the sequence generated by Algorithm 3.18.*

(i) *Suppose that, for every $\mathsf{z} \in \mathsf{Z}$, $\sum_{n \in \mathbb{N}} \varepsilon_n(\cdot, \mathsf{z})\mathsf{E}\lambda_n < +\infty$ P-a.s. Then the following hold:*

  (a) $\sum_{n \in \mathbb{N}} \mathsf{E}(\lambda_n (2 - \lambda_n))\mathsf{E}(\|d_n\|_{\mathsf{H}}^2 \mid \mathcal{X}_n) < +\infty$ P-a.s.

  (b) *Suppose that* $\inf_{n \in \mathbb{N}} \mathsf{E}(\lambda_n(2 - \lambda_n)) > 0$ *and there exists* $\rho \in [1, +\infty[$ *such that* $\sup_{n \in \mathbb{N}} \lambda_n < \rho$ P-a.s. *Then* $\sum_{n \in \mathbb{N}} \mathsf{E}(\|x_{n+1} - x_n\|_{\mathsf{H}}^2 \mid \mathcal{X}_n) < +\infty$ P-a.s. *and* $\sum_{n \in \mathbb{N}} \|x_{n+1} - x_n\|_{\mathsf{H}}^2 < +\infty$ P-a.s.

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

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

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

(ii) *Suppose that, for every* $z \in Z$, $\sum_{n\in\mathbb{N}} E\varepsilon_n(\cdot, z)E\lambda_n < +\infty$. *Then the following hold:*

(a) $\sum_{n\in\mathbb{N}} E(\lambda_n(2-\lambda_n))E\|d_n\|_H^2 < +\infty$.

(b) *Suppose that* $\inf_{n\in\mathbb{N}} E(\lambda_n(2-\lambda_n)) > 0$ *and there exists* $\rho \in [1, +\infty[$ *such that* $\sup_{n\in\mathbb{N}} \lambda_n < \rho$ P-a.s. *Then* $\sum_{n\in\mathbb{N}} E\|x_{n+1} - x_n\|_H^2 < +\infty$.

(c) *Suppose that* $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset Z$ P-a.s. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges weakly* P-a.s. *and weakly in* $L^2(\Omega, \mathcal{F}, P; H)$ *to a random variable* $x \in L^2(\Omega, \mathcal{F}, P; Z)$.

(d) *Suppose that* $\mathfrak{S}(x_n)_{n\in\mathbb{N}} \cap Z \neq \varnothing$ P-a.s. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges strongly* P-a.s. *and strongly in* $L^1(\Omega, \mathcal{F}, P; H)$ *to a random variable* $x \in L^2(\Omega, \mathcal{F}, P; Z)$. *Additionally,* $(x_n)_{n\in\mathbb{N}}$ *converges weakly in* $L^2(\Omega, \mathcal{F}, P; H)$ *to* x.

(e) *Suppose that* Z *is convex, that, for every* $n \in \mathbb{N}$, ε_n *is constant with respect to the* H*-variable, and that* $\underline{\lim}\, Ed_Z^2(x_n) = 0$. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges strongly in* $L^2(\Omega, \mathcal{F}, P; H)$ *and* P-a.s. *to a* Z*-valued random variable.*

(f) *Suppose that* Z *is convex, that, for every* $n \in \mathbb{N}$, ε_n *is constant with respect to the* H*-variable, and that there exists* $\chi \in]0, 1[$ *such that*

$$(\forall n \in \mathbb{N}) \quad E\big(d_Z^2(x_{n+1}) \,\big|\, \mathcal{X}_n\big) \leqslant \chi d_Z^2(x_n) + 2\varepsilon_n E\lambda_n \quad \text{P-a.s.} \tag{3.55}$$

Set, for every $n \in \mathbb{N}$ *and for every* $\omega \in \Omega$, $\vartheta_n(\omega) = \varepsilon_n(\omega, 0)$. *Then the following are satisfied:*

[A] *Let* $n \in \mathbb{N}$. *Then* $Ed_Z^2(x_{n+1}) \leqslant \chi^{n+1} Ed_Z^2(x_0) + 2\sum_{j=0}^{n} \chi^{n-j} E\vartheta_j E\lambda_j$.

[B] *There exists* $x \in L^2(\Omega, \mathcal{F}, P; Z)$ *such that* $(x_n)_{n\in\mathbb{N}}$ *converges strongly in* $L^2(\Omega, \mathcal{F}, P; H)$ *and* P-a.s. *to* x, *and*

$$(\forall n \in \mathbb{N}) \quad E\|x_n - x\|_H^2 \leqslant 4\chi^n Ed_Z^2(x_0) + 8\sum_{j=0}^{n-1} \chi^{n-j-1} E\vartheta_j E\lambda_j + 4\sum_{j\geqslant n} E\vartheta_j E\lambda_j. \tag{3.56}$$

Proof. In view of Proposition 3.19, we appeal to Theorems 3.16 and 3.17 to establish the claims.

(i)(a): It follows from Theorem 3.16(v)(c) and Lemma 3.10 that

$$\sum_{n\in\mathbb{N}} E(\lambda_n(2-\lambda_n))E\big(\|d_n\|_H^2 \,\big|\, \mathcal{X}_n\big) = \sum_{n\in\mathbb{N}} E\big(\lambda_n(2-\lambda_n)\|d_n\|_H^2 \,\big|\, \mathcal{X}_n\big) < +\infty \text{ P-a.s.} \tag{3.57}$$

(i)(a)⇒(i)(b): It follows from (3.2) that

$$\sum_{n\in\mathbb{N}} E(\lambda_n(2-\lambda_n))E\left(\frac{1}{\lambda_n^2}\|x_{n+1} - x_n\|_H^2 \,\middle|\, \mathcal{X}_n\right) = \sum_{n\in\mathbb{N}} E(\lambda_n(2-\lambda_n))E\big(\|d_n\|_H^2 \,\big|\, \mathcal{X}_n\big) < +\infty \text{ P-a.s.} \tag{3.58}$$

Hence, the assumption $\inf_{n\in\mathbb{N}} E(\lambda_n(2-\lambda_n)) > 0$ yields $\sum_{n\in\mathbb{N}} E\big(\|x_{n+1} - x_n\|_H^2/\lambda_n^2 \,\big|\, \mathcal{X}_n\big) < +\infty$ P-a.s. Further,

$$(\forall n \in \mathbb{N}) \quad 0 < \frac{1}{\rho^2} \leqslant \frac{1}{\lambda_n^2} \quad \text{P-a.s.} \tag{3.59}$$

Thus,

$$\sum_{n\in\mathbb{N}} \mathsf{E}\big(\|x_{n+1}-x_n\|_{\mathsf{H}}^2 \,\big|\, \mathcal{X}_n\big) < +\infty \ \mathsf{P}\text{-a.s.} \tag{3.60}$$

In addition,

$$(\forall n\in\mathbb{N})\quad \mathsf{E}\left(\sum_{k=0}^{n+1}\|x_{k+1}-x_k\|_{\mathsf{H}}^2 \,\middle|\, \mathcal{X}_{n+1}\right) = \sum_{k=0}^{n}\|x_{k+1}-x_k\|_{\mathsf{H}}^2 + \mathsf{E}\big(\|x_{n+2}-x_{n+1}\|_{\mathsf{H}}^2 \,\big|\, \mathcal{X}_{n+1}\big)\ \ \mathsf{P}\text{-a.s.} \tag{3.61}$$

It then follows from (3.60) and Lemma 3.8(i) that $(\sum_{k=0}^{n}\|x_{k+1}-x_k\|_{\mathsf{H}}^2)_{n\in\mathbb{N}}$ converges P-a.s. to a $[0,+\infty[$-valued random variable, hence $\sum_{n\in\mathbb{N}}\|x_{n+1}-x_n\|_{\mathsf{H}}^2 < +\infty$ P-a.s.

(i)(c)–(i)(e): These follow from Theorem 3.16(v)(d)–(v)(f).

(ii)(a): It follows from Theorem 3.16(vi)(c) and Lemma 3.10 that

$$\begin{aligned}
\sum_{n\in\mathbb{N}} \mathsf{E}(\lambda_n(2-\lambda_n))\mathsf{E}\|d_n\|_{\mathsf{H}}^2 &= \sum_{n\in\mathbb{N}} \mathsf{E}(\lambda_n(2-\lambda_n))\mathsf{E}\Big(\mathsf{E}\big(\|d_n\|_{\mathsf{H}}^2 \,\big|\, \mathcal{X}_n\big)\Big)\\
&= \sum_{n\in\mathbb{N}} \mathsf{E}\Big(\mathsf{E}(\lambda_n(2-\lambda_n))\mathsf{E}\big(\|d_n\|_{\mathsf{H}}^2 \,\big|\, \mathcal{X}_n\big)\Big)\\
&= \sum_{n\in\mathbb{N}} \mathsf{E}\Big(\mathsf{E}\Big(\lambda_n(2-\lambda_n)\|d_n\|_{\mathsf{H}}^2 \,\Big|\, \mathcal{X}_n\Big)\Big)\\
&= \sum_{n\in\mathbb{N}} \mathsf{E}\Big(\lambda_n(2-\lambda_n)\|d_n\|_{\mathsf{H}}^2\Big)\\
&< +\infty.
\end{aligned} \tag{3.62}$$

Hence $\sum_{n\in\mathbb{N}} \mathsf{E}(\lambda_n(2-\lambda_n))\mathsf{E}\|d_n\|_{\mathsf{H}}^2 < +\infty$.

(ii)(a)⇒(ii)(b): It follows from (3.28) that

$$\sum_{n\in\mathbb{N}} \mathsf{E}(\lambda_n(2-\lambda_n))\mathsf{E}\left(\frac{1}{\lambda_n^2}\|x_{n+1}-x_n\|_{\mathsf{H}}^2\right) = \sum_{n\in\mathbb{N}} \mathsf{E}(\lambda_n(2-\lambda_n))\mathsf{E}\|d_n\|_{\mathsf{H}}^2 < +\infty. \tag{3.63}$$

Thus, as in (i)(b), $\sum_{n\in\mathbb{N}} \mathsf{E}\|x_{n+1}-x_n\|_{\mathsf{H}}^2 < +\infty$.

(ii)(c)–(ii)(d): These follow from (i)(c)–(i)(d) and Theorem 3.16(vi)(d)–(vi)(e).

(ii)(e)–(ii)(f): These follow from Theorem 3.17(iv)(b)–(iv)(c). ☐

3.2.3.3 A stochastic algorithm with random relaxations bounded by 2

We present an implementation of Algorithm 3.2 with an alternative relaxation strategy.

Algorithm 3.21 In Algorithm 3.2, for every $n\in\mathbb{N}$, $\varepsilon_n \in \mathfrak{C}(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ and $\lambda_n \in L^\infty(\Omega,\mathcal{X}_n,\mathsf{P};]0,2[)$.

Proposition 3.22 *Algorithm 3.21 is a special case of Algorithm 3.15 where, for every* $n\in\mathbb{N}$, $\delta_n = 2\lambda_n\varepsilon_n$.

Proof. Set $(\forall n \in \mathbb{N})$ $\delta_n = 2\lambda_n\varepsilon_n$. Following the proof of Proposition 3.19, it is enough to show that

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \delta_n \in \mathfrak{C}(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H}); \\ \mathsf{E}\big(\lambda_n(2-\lambda_n)\|d_n\|_{\mathsf{H}}^2 \,\big|\, \mathcal{X}_n\big) \geqslant 0 \;\; \mathsf{P}\text{-a.s.}; \\ (\forall z \in \mathsf{Z}) \;\; \mathsf{E}\big(\lambda_n\langle z + d_n - x_n \mid d_n\rangle_{\mathsf{H}} \,\big|\, \mathcal{X}_n\big) \leqslant \delta_n(\cdot, z)/2 \;\; \mathsf{P}\text{-a.s.} \end{cases} \tag{3.64}$$

Let $n \in \mathbb{N}$. It follows from the positivity and measurability of λ_n, as well the fact that $\varepsilon_n \in \mathfrak{C}(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$, that $\delta_n \in \mathfrak{C}(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. Next, since $\lambda_n \in \,]0, 2[\,$ P-a.s., we have $\lambda_n(2 - \lambda_n) > 0$ P-a.s. and hence

$$\mathsf{E}\big(\lambda_n(2-\lambda_n)\|d_n\|_{\mathsf{H}}^2 \,\big|\, \mathcal{X}_n\big) \geqslant 0 \;\; \mathsf{P}\text{-a.s.} \tag{3.65}$$

Finally, let $z \in \mathsf{Z}$. It then follows from (3.53) and the fact that λ_n is positive and $\mathcal{X}_n$-measurable that

$$\begin{aligned} &\mathsf{E}\big(\lambda_n\langle z + d_n - x_n \mid d_n\rangle_{\mathsf{H}} \,\big|\, \mathcal{X}_n\big) \\ &\qquad = \lambda_n \mathsf{E}\big(\langle z \mid d_n\rangle_{\mathsf{H}} + \|d_n\|_{\mathsf{H}}^2 - \langle x_n \mid d_n\rangle_{\mathsf{H}} \,\big|\, \mathcal{X}_n\big) \leqslant \lambda_n\varepsilon_n(\cdot, z) = \frac{\delta_n(\cdot, z)}{2} \;\; \mathsf{P}\text{-a.s.}, \end{aligned} \tag{3.66}$$

which completes the proof. □

As in Section 3.2.3.2, we can derive weak, strong, and linear convergence results from Theorems 3.16 and 3.17. For brevity, we provide below only the weak convergence results but, as in Theorem 3.20, strong and linear convergence results can also be obtained.

Theorem 3.23 *Let $(x_n)_{n\in\mathbb{N}}$ be the sequence generated by Algorithm 3.21. Suppose that, for every $z \in \mathsf{Z}$, $\sum_{n\in\mathbb{N}} \lambda_n\varepsilon_n(\cdot, z) < +\infty$ P-a.s. and that $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \mathsf{Z}$ P-a.s. Then $(x_n)_{n\in\mathbb{N}}$ converges weakly P-a.s. to a Z-valued random variable x. If, in addition, for every $z \in \mathsf{Z}$, $\sum_{n\in\mathbb{N}} \mathsf{E}(\lambda_n\varepsilon_n(\cdot, z)) < +\infty$, then $x \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ and $(x_n)_{n\in\mathbb{N}}$ converges weakly in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ to x.*

Proof. In view of Proposition 3.22, the claim follows Theorem 3.16(v)(d) and 3.16(vi)(d). □

3.2.4 Randomly relaxed Krasnosel'skiĭ–Mann iterations

Let us first recall some definitions about an operator $\mathsf{T}\colon \mathsf{H} \to \mathsf{H}$ [5, Chapter 4]. First, $\mathsf{T}\colon \mathsf{H} \to \mathsf{H}$ is nonexpansive if it is 1-Lipschitzian and α-averaged for some $\alpha \in \,]0, 1[\,$ if $\mathrm{Id} + \alpha^{-1}(\mathsf{T} - \mathrm{Id})$ is nonexpansive [3]. On the other hand, T is β-cocoercive for some $\beta \in \,]0, +\infty[\,$ if

$$(\forall x \in \mathsf{H})(\forall y \in \mathsf{H}) \quad \langle x - y \mid \mathsf{T}x - \mathsf{T}y\rangle_{\mathsf{H}} \geqslant \beta\|\mathsf{T}x - \mathsf{T}y\|_{\mathsf{H}}^2 \tag{3.67}$$

and it is firmly nonexpansive if it is 1-cocoercive.

The Krasnosel'skiĭ–Mann iterative process is a basic algorithm to construct fixed points of nonexpansive operators [5, 25, 29, 35, 38, 49]. We propose a study of its asymptotic behavior in a novel environment featuring random relaxations and stochastic errors.

Theorem 3.24 *Let* $\mathsf{T}\colon\mathsf{H}\to\mathsf{H}$ *be a nonexpansive operator such that* $\operatorname{Fix}\mathsf{T}\neq\varnothing$ *and let* $x_0\in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$. *Iterate*

$$\begin{array}{l}\text{for } \mathsf{n}=0,1,\ldots\\ \left\lfloor\begin{array}{l}\text{take } e_{\mathsf{n}}\in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H}) \text{ and } \mu_{\mathsf{n}}\in L^\infty(\Omega,\mathcal{F},\mathsf{P};\,]0,1[)\\ x_{\mathsf{n}+1}=x_{\mathsf{n}}+\mu_{\mathsf{n}}\big(\mathsf{T}x_{\mathsf{n}}+e_{\mathsf{n}}-x_{\mathsf{n}}\big).\end{array}\right.\end{array}\tag{3.68}$$

Set $(\forall\mathsf{n}\in\mathbb{N})$ $\Phi_{\mathsf{n}}=\{x_0,\ldots,x_{\mathsf{n}}\}$ *and* $\mathcal{X}_{\mathsf{n}}=\sigma(\Phi_{\mathsf{n}})$. *Suppose that* $\sum_{\mathsf{n}\in\mathbb{N}}\mathsf{E}(\mu_{\mathsf{n}}(1-\mu_{\mathsf{n}}))=+\infty$ *and, for every* $\mathsf{n}\in\mathbb{N}$, μ_{n} *is independent of* $\sigma(\{e_{\mathsf{n}}\}\cup\Phi_{\mathsf{n}})$. *Then the following hold for some* $\operatorname{Fix}\mathsf{T}$*-valued random variable* x*:*

(i) *Suppose that* $\mathsf{E}(\|e_{\mathsf{n}}\|_{\mathsf{H}}^2\,|\,\mathcal{X}_{\mathsf{n}})\to 0$ P-a.s. *and* $\sum_{\mathsf{n}\in\mathbb{N}}\mathsf{E}\mu_{\mathsf{n}}\sqrt{\mathsf{E}(\|e_{\mathsf{n}}\|_{\mathsf{H}}^2\,|\,\mathcal{X}_{\mathsf{n}})}<+\infty$ P-a.s. *Then the following hold:*

  (a) $(\mathsf{T}x_{\mathsf{n}}-x_{\mathsf{n}})_{\mathsf{n}\in\mathbb{N}}$ *converges strongly* P-a.s. *to* 0.

  (b) $(x_{\mathsf{n}})_{\mathsf{n}\in\mathbb{N}}$ *converges weakly* P-a.s. *to* x.

  (c) *Suppose that* $\mathsf{T}-\mathrm{Id}$ *is demiregular at every point in* $\operatorname{Fix}\mathsf{T}$. *Then* $(x_{\mathsf{n}})_{\mathsf{n}\in\mathbb{N}}$ *converges strongly* P-a.s. *to* x.

(ii) *Suppose that* $\mathsf{E}\|e_{\mathsf{n}}\|_{\mathsf{H}}^2\to 0$ *and* $\sum_{\mathsf{n}\in\mathbb{N}}\sqrt{\mathsf{E}\mu_{\mathsf{n}}^2\mathsf{E}\|e_{\mathsf{n}}\|_{\mathsf{H}}^2}<+\infty$. *Then the following hold:*

  (a) $(\mathsf{T}x_{\mathsf{n}}-x_{\mathsf{n}})_{\mathsf{n}\in\mathbb{N}}$ *converges strongly in* $L^1(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ *and strongly* P-a.s. *to* 0.

  (b) $x\in L^2(\Omega,\mathcal{F},\mathsf{P};\operatorname{Fix}\mathsf{T})$ *and* $(x_{\mathsf{n}})_{\mathsf{n}\in\mathbb{N}}$ *converges weakly in* $L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ *and weakly* P-a.s. *to* x.

  (c) *Suppose that* $\mathsf{T}-\mathrm{Id}$ *is demiregular at every point in* $\operatorname{Fix}\mathsf{T}$. *Then* $(x_{\mathsf{n}})_{\mathsf{n}\in\mathbb{N}}$ *converges strongly in* $L^1(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ *and strongly* P-a.s. *to* x.

Proof. Let us show that the sequence $(x_{\mathsf{n}})_{\mathsf{n}\in\mathbb{N}}$ constructed by (3.68) corresponds to a sequence generated by Algorithm 3.18. To see this, set $\mathsf{Z}=\operatorname{Fix}\mathsf{T}$ and observe that, since T is nonexpansive,

$$\big(\forall\mathsf{n}\in\mathbb{N}\big)\big(\forall z\in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{Z})\big)\quad \mathsf{E}\|\mathsf{T}x_{\mathsf{n}}-z\|_{\mathsf{H}}^2\leqslant\mathsf{E}\|x_{\mathsf{n}}-z\|_{\mathsf{H}}^2.\tag{3.69}$$

Thus if, for some $\mathsf{n}\in\mathbb{N}$, $x_{\mathsf{n}}\in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$, then $\mathsf{T}x_{\mathsf{n}}\in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ and (3.68) yields $x_{\mathsf{n}+1}\in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$. This shows by induction that $(x_{\mathsf{n}})_{\mathsf{n}\in\mathbb{N}}$ and $(\mathsf{T}x_{\mathsf{n}})_{\mathsf{n}\in\mathbb{N}}$ lie in $L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$. Let us define

$$
(\forall \mathsf{n} \in \mathbb{N}) \quad \begin{cases} t_{\mathsf{n}}^* = \dfrac{x_{\mathsf{n}} - \mathsf{T}x_{\mathsf{n}} - e_{\mathsf{n}}}{2} \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H}); \\ \eta_{\mathsf{n}} = \dfrac{\|x_{\mathsf{n}}\|_{\mathsf{H}}^2 - \|\mathsf{T}x_{\mathsf{n}} + e_{\mathsf{n}}\|_{\mathsf{H}}^2}{4} \in L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathbb{R}); \\ \alpha_{\mathsf{n}} = 1_{[t_{\mathsf{n}}^* \neq 0]}; \\ (\forall \mathsf{z} \in \mathsf{Z}) \ \varepsilon_{\mathsf{n}}(\cdot, \mathsf{z}) = \dfrac{1}{4}\mathsf{E}\big(\|e_{\mathsf{n}}\|_{\mathsf{H}}^2 \,\big|\, \mathcal{X}_{\mathsf{n}}\big) + \dfrac{1}{2}\|x_{\mathsf{n}} - \mathsf{z}\|_{\mathsf{H}} \sqrt{\mathsf{E}\big(\|e_{\mathsf{n}}\|_{\mathsf{H}}^2 \,\big|\, \mathcal{X}_{\mathsf{n}}\big)}; \\ d_{\mathsf{n}} = t_{\mathsf{n}}^*; \\ \lambda_{\mathsf{n}} = 2\mu_{\mathsf{n}} \in \,]0, 2[\ \ \mathsf{P}\text{-a.s.} \end{cases} \tag{3.70}
$$

Now set $\mathsf{F} = (\mathsf{T} + \mathsf{Id})/2$. Since T is nonexpansive, F is firmly nonexpansive (see [5, Proposition 4.4] or [28, Proposition 1.11.2]. Hence, we deduce from Lemma 3.13 and (3.70) that, for every $\mathsf{z} \in \mathsf{Z}$ and every $\mathsf{n} \in \mathbb{N}$,

$$
\begin{aligned}
\big\langle \mathsf{z} \,\big|\, \mathsf{E}(\alpha_{\mathsf{n}} t_{\mathsf{n}}^* \,|\, \mathcal{X}_{\mathsf{n}}) \big\rangle_{\mathsf{H}} &= \mathsf{E}\bigg(\Big\langle \mathsf{z} \,\Big|\, x_{\mathsf{n}} - \mathsf{F}x_{\mathsf{n}} - \frac{1}{2}e_{\mathsf{n}} \Big\rangle_{\mathsf{H}} \,\bigg|\, \mathcal{X}_{\mathsf{n}}\bigg) \\
&= \big\langle \mathsf{z} \,\big|\, x_{\mathsf{n}} - \mathsf{F}x_{\mathsf{n}} \big\rangle_{\mathsf{H}} - \frac{1}{2}\mathsf{E}\big(\langle \mathsf{z} \,|\, e_{\mathsf{n}} \rangle_{\mathsf{H}} \,\big|\, \mathcal{X}_{\mathsf{n}}\big) \\
&\leqslant \big\langle \mathsf{F}x_{\mathsf{n}} \,\big|\, x_{\mathsf{n}} - \mathsf{F}x_{\mathsf{n}} \big\rangle_{\mathsf{H}} - \frac{1}{2}\mathsf{E}\big(\langle \mathsf{z} \,|\, e_{\mathsf{n}} \rangle_{\mathsf{H}} \,\big|\, \mathcal{X}_{\mathsf{n}}\big) \\
&= \mathsf{E}(\alpha_{\mathsf{n}}\eta_{\mathsf{n}} \,|\, \mathcal{X}_{\mathsf{n}}) + \frac{1}{4}\mathsf{E}\big(\|e_{\mathsf{n}}\|_{\mathsf{H}}^2 \,\big|\, \mathcal{X}_{\mathsf{n}}\big) + \frac{1}{2}\mathsf{E}\big(\langle \mathsf{T}x_{\mathsf{n}} - \mathsf{z} \,|\, e_{\mathsf{n}} \rangle_{\mathsf{H}} \,\big|\, \mathcal{X}_{\mathsf{n}}\big) \\
&\leqslant \mathsf{E}(\alpha_{\mathsf{n}}\eta_{\mathsf{n}} \,|\, \mathcal{X}_{\mathsf{n}}) + \frac{1}{4}\mathsf{E}\big(\|e_{\mathsf{n}}\|_{\mathsf{H}}^2 \,\big|\, \mathcal{X}_{\mathsf{n}}\big) + \frac{1}{2}\|\mathsf{T}x_{\mathsf{n}} - \mathsf{z}\|_{\mathsf{H}}\mathsf{E}\big(\|e_{\mathsf{n}}\|_{\mathsf{H}} \,\big|\, \mathcal{X}_{\mathsf{n}}\big) \\
&\leqslant \mathsf{E}(\alpha_{\mathsf{n}}\eta_{\mathsf{n}} \,|\, \mathcal{X}_{\mathsf{n}}) + \frac{1}{4}\mathsf{E}\big(\|e_{\mathsf{n}}\|_{\mathsf{H}}^2 \,\big|\, \mathcal{X}_{\mathsf{n}}\big) + \frac{1}{2}\|x_{\mathsf{n}} - \mathsf{z}\|_{\mathsf{H}}\mathsf{E}\big(\|e_{\mathsf{n}}\|_{\mathsf{H}} \,\big|\, \mathcal{X}_{\mathsf{n}}\big) \\
&\leqslant \mathsf{E}(\alpha_{\mathsf{n}}\eta_{\mathsf{n}} \,|\, \mathcal{X}_{\mathsf{n}}) + \varepsilon_{\mathsf{n}}(\cdot, \mathsf{z}) \ \ \mathsf{P}\text{-a.s.}
\end{aligned} \tag{3.71}
$$

Next, it is clear from (3.70) that $\varepsilon_{\mathsf{n}} \in \mathfrak{C}(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. Furthermore, in view of (3.70), (3.68) can be written as

$$
(\forall \mathsf{n} \in \mathbb{N}) \quad x_{\mathsf{n}+1} = x_{\mathsf{n}} - \lambda_{\mathsf{n}} d_{\mathsf{n}}. \tag{3.72}
$$

On the other hand, since $(\mu_{\mathsf{n}})_{\mathsf{n} \in \mathbb{N}}$ lies almost surely in $]0, 1[$, we have $\mathsf{E}(\lambda_{\mathsf{n}}(2 - \lambda_{\mathsf{n}})) \geqslant 0$. Additionally, for every $\mathsf{n} \in \mathbb{N}$, μ_{n} is independent of $\sigma(\{e_{\mathsf{n}}\} \cup \Phi_{\mathsf{n}})$ and, by (3.70), d_{n} is $\sigma(\{e_{\mathsf{n}}\} \cup \Phi_{\mathsf{n}})$-measurable. Hence, $\sigma(\{d_{\mathsf{n}}\} \cup \Phi_{\mathsf{n}}) \subset \sigma(\{e_{\mathsf{n}}\} \cup \Phi_{\mathsf{n}})$ and λ_{n} is independent of $\sigma(\{d_{\mathsf{n}}\} \cup \Phi_{\mathsf{n}})$. Altogether, $(x_{\mathsf{n}})_{\mathsf{n} \in \mathbb{N}}$ is a sequence generated by Algorithm 3.18. Finally, it follows from (3.68) and Lemma 3.10 that

$$
\begin{aligned}
(\forall \mathsf{n} \in \mathbb{N})(\forall \mathsf{z} \in \mathsf{Z}) \quad & \mathsf{E}(\|x_{\mathsf{n}+1} - \mathsf{z}\|_{\mathsf{H}} \,|\, \mathcal{X}_{\mathsf{n}}) \\
&\leqslant \mathsf{E}\big((1 - \mu_{\mathsf{n}})\|x_{\mathsf{n}} - \mathsf{z}\|_{\mathsf{H}} + \mu_{\mathsf{n}}\|\mathsf{T}x_{\mathsf{n}} - \mathsf{z}\|_{\mathsf{H}} + \mu_{\mathsf{n}}\|e_{\mathsf{n}}\|_{\mathsf{H}} \,\big|\, \mathcal{X}_{\mathsf{n}}\big) \\
&= (1 - \mathsf{E}\mu_{\mathsf{n}})\|x_{\mathsf{n}} - \mathsf{z}\|_{\mathsf{H}} + \mathsf{E}\mu_{\mathsf{n}}\|\mathsf{T}x_{\mathsf{n}} - \mathsf{z}\|_{\mathsf{H}} + \mathsf{E}\mu_{\mathsf{n}}\mathsf{E}(\|e_{\mathsf{n}}\|_{\mathsf{H}} \,|\, \mathcal{X}_{\mathsf{n}})
\end{aligned}
$$

$$\leqslant \|x_n - z\|_H + \mathsf{E}\mu_n \mathsf{E}(\|e_n\|_H \,|\, \mathcal{X}_n)$$
$$\leqslant \|x_n - z\|_H + \mathsf{E}\mu_n \sqrt{\mathsf{E}(\|e_n\|_H^2 \,|\, \mathcal{X}_n)} \ \ \mathsf{P}\text{-a.s.} \tag{3.73}$$

Further, by invoking the nonexpansiveness of T and the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned}
(\forall n \in \mathbb{N})(\forall z \in Z) \quad & \|x_{n+1} - z\|_{L^2(\Omega,\mathcal{F},\mathsf{P};H)} \\
&= \sqrt{\mathsf{E}\|x_{n+1} - z\|_H^2} \\
&= \sqrt{\mathsf{E}\|(1-\mu_n)(x_n - z) + \mu_n(\mathsf{T}x_n - \mathsf{T}z) + \mu_n e_n\|_H^2} \\
&\leqslant \sqrt{\mathsf{E}\big|\|x_n - z\|_H + \mu_n\|e_n\|_H\big|^2} \\
&= \sqrt{\mathsf{E}\|x_n - z\|_H^2 + 2\mathsf{E}\mu_n \mathsf{E}(\|x_n - z\|_H\|e_n\|_H) + \mathsf{E}\mu_n^2 \mathsf{E}\|e_n\|_H^2} \\
&\leqslant \sqrt{\mathsf{E}\|x_n - z\|_H^2 + 2\sqrt{\mathsf{E}\mu_n^2}\sqrt{\mathsf{E}\|x_n - z\|_H^2}\sqrt{\|e_n\|_H^2} + \mathsf{E}\mu_n^2 \mathsf{E}\|e_n\|_H^2} \\
&= \sqrt{\left|\sqrt{\mathsf{E}\|x_n - z\|_H^2} + \sqrt{\mathsf{E}\mu_n^2 \mathsf{E}\|e_n\|_H^2}\right|^2} \\
&= \|x_n - z\|_{L^2(\Omega,\mathcal{F},\mathsf{P};H)} + \sqrt{\mathsf{E}\mu_n^2 \mathsf{E}\|e_n\|_H^2}.
\end{aligned} \tag{3.74}$$

(i)(a): We derive from (3.73) and Lemma 3.14(ii) that $(\|x_n\|_H)_{n\in\mathbb{N}}$ is bounded P-a.s. Hence, for every $z \in Z$, $\sum_{n\in\mathbb{N}} \|x_n - z\|_H \mathsf{E}\mu_n \sqrt{\mathsf{E}(\|e_n\|_H^2 \,|\, \mathcal{X}_n)} < +\infty$ P-a.s. On the other hand, the assumptions $\lim \mathsf{E}(\|e_n\|_H^2 \,|\, \mathcal{X}_n) = 0$ and $\sum_{n\in\mathbb{N}} \mathsf{E}\mu_n \sqrt{\mathsf{E}(\|e_n\|_H^2 \,|\, \mathcal{X}_n)} < +\infty$ P-a.s. yield

$$\sum_{n\in\mathbb{N}} \mathsf{E}\mu_n \mathsf{E}(\|e_n\|_H^2 \,|\, \mathcal{X}_n) < +\infty \ \ \mathsf{P}\text{-a.s.} \tag{3.75}$$

Therefore, for every $z \in Z$, $\sum_{n\in\mathbb{N}} \varepsilon_n(\cdot, z)\mathsf{E}\lambda_n = 2\sum_{n\in\mathbb{N}} \varepsilon_n(\cdot, z)\mathsf{E}\mu_n < +\infty$ P-a.s. It then follows from Theorem 3.20(i)(a) and the assumption $\sum_{n\in\mathbb{N}} \mathsf{E}(\mu_n(1-\mu_n)) = +\infty$ that $\underline{\lim}\, \mathsf{E}(\|d_n\|_H^2 \,|\, \mathcal{X}_n) = 0$ P-a.s. Hence,

$$\begin{aligned}
0 &\leqslant \frac{1}{2}\underline{\lim}\|\mathsf{T}x_n - x_n\|_H^2 \\
&\leqslant \underline{\lim}\, \mathsf{E}\big(\|\mathsf{T}x_n + e_n - x_n\|_H^2 + \|e_n\|_H^2 \,\big|\, \mathcal{X}_n\big) \\
&= \underline{\lim}\, \mathsf{E}\big(\|\mathsf{T}x_n + e_n - x_n\|_H^2 \,\big|\, \mathcal{X}_n\big) + \lim \mathsf{E}(\|e_n\|_H^2 \,|\, \mathcal{X}_n) \\
&= 4\,\underline{\lim}\, \mathsf{E}\big(\|d_n\|_H^2 \,\big|\, \mathcal{X}_n\big) + \lim \mathsf{E}(\|e_n\|_H^2 \,|\, \mathcal{X}_n) \\
&= 0 \ \ \mathsf{P}\text{-a.s.}
\end{aligned} \tag{3.76}$$

Thus, Lemma 3.10 implies that, for every $n \in \mathbb{N}$,

$$\mathsf{E}\big(\|\mathsf{T}x_{n+1} - x_{n+1}\|_H \,\big|\, \mathcal{X}_n\big)$$

$$
\begin{aligned}
&= \mathsf{E}\big(\|\mathsf{T}x_{n+1} - \mathsf{T}x_n + (1-\mu_n)(\mathsf{T}x_n - x_n) - \mu_n e_n\|_{\mathsf{H}} \,\big|\, \mathcal{X}_n\big)\\
&\leqslant \mathsf{E}\big(\|\mathsf{T}x_{n+1} - \mathsf{T}x_n\|_{\mathsf{H}} \,\big|\, \mathcal{X}_n\big) + \mathsf{E}\big((1-\mu_n)\|\mathsf{T}x_n - x_n\|_{\mathsf{H}} \,\big|\, \mathcal{X}_n\big) + \mathsf{E}\big(\mu_n\|e_n\|_{\mathsf{H}} \,\big|\, \mathcal{X}_n\big)\\
&\leqslant \mathsf{E}\big(\|x_{n+1} - x_n\|_{\mathsf{H}} \,\big|\, \mathcal{X}_n\big) + (1-\mathsf{E}\mu_n)\|\mathsf{T}x_n - x_n\|_{\mathsf{H}} + \mathsf{E}\big(\mu_n\|e_n\|_{\mathsf{H}} \,\big|\, \mathcal{X}_n\big)\\
&= \mathsf{E}\big(\mu_n\|\mathsf{T}x_n + e_n - x_n\|_{\mathsf{H}} \,\big|\, \mathcal{X}_n\big) + (1-\mathsf{E}\mu_n)\|\mathsf{T}x_n - x_n\|_{\mathsf{H}} + \mathsf{E}\big(\mu_n\|e_n\|_{\mathsf{H}} \,\big|\, \mathcal{X}_n\big)\\
&= \mathsf{E}\big(\mu_n\|\mathsf{T}x_n - x_n\|_{\mathsf{H}} \,\big|\, \mathcal{X}_n\big) + (1-\mathsf{E}\mu_n)\|\mathsf{T}x_n - x_n\|_{\mathsf{H}} + 2\mathsf{E}\big(\mu_n\|e_n\|_{\mathsf{H}} \,\big|\, \mathcal{X}_n\big)\\
&= (\mathsf{E}\mu_n)\|\mathsf{T}x_n - x_n\|_{\mathsf{H}} + (1-\mathsf{E}\mu_n)\|\mathsf{T}x_n - x_n\|_{\mathsf{H}} + 2\mathsf{E}\mu_n\mathsf{E}(\|e_n\|_{\mathsf{H}} \,|\, \mathcal{X}_n)\\
&\leqslant \|\mathsf{T}x_n - x_n\|_{\mathsf{H}} + 2\mathsf{E}\mu_n\sqrt{\mathsf{E}(\|e_n\|_{\mathsf{H}}^2 \,|\, \mathcal{X}_n)} \ \ \mathsf{P}\text{-a.s.}
\end{aligned}
\tag{3.77}
$$

Consequently, Lemma 3.8(i) secures the convergence P-a.s. of the sequence $(\|\mathsf{T}x_n - x_n\|_{\mathsf{H}})_{n\in\mathbb{N}}$, which, in view of (3.76), forces

$$\lim \|\mathsf{T}x_n - x_n\|_{\mathsf{H}} = 0 \ \ \mathsf{P}\text{-a.s.} \tag{3.78}$$

(i)(b): Let us show that $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \mathsf{Z}$ P-a.s. Let $\omega \in \Omega$ be such that $\mathfrak{W}(x_n(\omega))_{n\in\mathbb{N}} \neq \varnothing$ and $\lim\|\mathsf{T}x_n(\omega) - x_n(\omega)\| = 0$. Let $\mathsf{x} \in \mathfrak{W}(x_n(\omega))_{n\in\mathbb{N}}$, say $x_{k_n}(\omega) \rightharpoonup \mathsf{x}$. The nonexpansiveness of T implies that $\mathrm{Id} - \mathsf{T}$ is demiclosed at 0 [5, Theorem 4.27]. In turn, $\mathsf{T}\mathsf{x} = \mathsf{x}$ and $\mathfrak{W}(x_n(\omega)) \subset \mathsf{Z}$. Since $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \neq \varnothing$ P-a.s. and $\lim\|\mathsf{T}x_n - x_n\| = 0$ P-a.s., we conclude that $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \mathsf{Z}$ P-a.s. Thus, the claim follows from Theorem 3.20(i)(c).

(i)(c): By (i)(a) and (i)(b), there exists $\Omega' \in \mathcal{F}$ such that $\mathsf{P}(\Omega') = 1$ and

$$(\forall \omega \in \Omega') \quad \mathsf{T}x_n(\omega) - x_n(\omega) \to 0 \ \text{ and } \ x_n(\omega) \rightharpoonup x(\omega). \tag{3.79}$$

It then follows from the demiregularity of $\mathsf{T} - \mathrm{Id}$ that, for every $\omega \in \Omega'$, $x_n(\omega) \to x(\omega)$. Hence, $(x_n)_{n\in\mathbb{N}}$ converges strongly P-a.s. to x.

(ii)(a): We derive from (3.74) and Corollary 3.9(i) that $(\|x_n\|_{L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})})_{n\in\mathbb{N}}$ is bounded. Therefore,

$$\big(\forall z \in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})\big) \quad \sup_{n\in\mathbb{N}} \mathsf{E}\|x_n - z\|_{\mathsf{H}}^2 < +\infty. \tag{3.80}$$

In turn, for every $z \in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$,

$$\sum_{n\in\mathbb{N}} \mathsf{E}\mu_n\sqrt{\mathsf{E}\|x_n - z\|_{\mathsf{H}}^2}\sqrt{\mathsf{E}\|e_n\|_{\mathsf{H}}^2} \leqslant \sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}\|x_n - z\|_{\mathsf{H}}^2}\sqrt{\mathsf{E}\mu_n^2\mathsf{E}\|e_n\|_{\mathsf{H}}^2} < +\infty. \tag{3.81}$$

On the other hand, since $\lim \mathsf{E}\|e_n\|_{\mathsf{H}}^2 = 0$ and $\sum_{n\in\mathbb{N}} \mathsf{E}\mu_n\sqrt{\mathsf{E}\|e_n\|_{\mathsf{H}}^2} < +\infty$, we have

$$\sum_{n\in\mathbb{N}} \mathsf{E}\mu_n\mathsf{E}\|e_n\|_{\mathsf{H}}^2 < +\infty. \tag{3.82}$$

Altogether, we deduce that

$$(\forall z \in \mathsf{Z}) \quad \sum_{n\in\mathbb{N}} \mathsf{E}\varepsilon_n(\cdot, z)\mathsf{E}\lambda_n = 2\sum_{n\in\mathbb{N}} \mathsf{E}\varepsilon_n(\cdot, z)\mathsf{E}\mu_n < +\infty, \tag{3.83}$$

which shows in particular that, for every $z \in \mathsf{Z}$, $\sum_{n\in\mathbb{N}} \varepsilon_n(\cdot, z)\mathsf{E}\lambda_n < +\infty$ P-a.s. Thus $\lim\|\mathsf{T}x_n - x_n\|_{\mathsf{H}} = 0$ P-a.s. On the other hand, it follows from Theorem 3.20(ii)(a) and the assumptions that $\underline{\lim}\, \mathsf{E}\|d_n\|^2_{\mathsf{H}} = 0$. Hence, proceeding as in (3.76), we obtain $\underline{\lim}\, \mathsf{E}\|\mathsf{T}x_n - x_n\|^2_{\mathsf{H}} = 0$. Moreover, taking expectations in (3.77) yields

$$\begin{aligned}(\forall n \in \mathbb{N}) \quad \mathsf{E}\|\mathsf{T}x_{n+1} - x_{n+1}\|_{\mathsf{H}} &\leqslant \mathsf{E}\|\mathsf{T}x_n - x_n\|_{\mathsf{H}} + 2\mathsf{E}\mu_n\sqrt{\mathsf{E}\|e_n\|^2_{\mathsf{H}}} \\ &\leqslant \mathsf{E}\|\mathsf{T}x_n - x_n\|_{\mathsf{H}} + 2\sqrt{\mathsf{E}\mu_n^2\mathsf{E}\|e_n\|^2_{\mathsf{H}}}.\end{aligned} \tag{3.84}$$

It then follows from Corollary 3.9(i) that $(\mathsf{E}\|\mathsf{T}x_n - x_n\|_{\mathsf{H}})_{n\in\mathbb{N}}$ converges. Since $\underline{\lim}\, \mathsf{E}\|\mathsf{T}x_n - x_n\|^2_{\mathsf{H}} = 0$, this implies that $\lim \mathsf{E}\|\mathsf{T}x_n - x_n\|^2_{\mathsf{H}} = 0$. Hence, appealing to (i)(a), $(\mathsf{T}x_n - x_n)_{n\in\mathbb{N}}$ converges strongly in $L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ and strongly P-a.s. to 0.

(ii)(b): We deduce weak convergence P-a.s. by arguing as in the proof of (i)(b), while weak convergence in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ follows from Theorem 3.20(ii)(c).

(ii)(c): As in the proof of (i)(c), it follows from (ii)(b) that $(x_n)_{n\in\mathbb{N}}$ converges strongly P-a.s. to x. Further, strong convergence in $L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ follows from Theorem 3.20(ii)(d). ☐

Remark 3.25 Theorem 3.24(ii) extends [20, Corollary 2.7], where the relaxations are only deterministic and the weak limit is not shown to be in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. Another connected result is [8, Theorem 2.8], which focuses on the finite-dimensional setting (which implies that demiregularity holds [2, Proposition 2.4]) with deterministic relaxations and the weaker summability condition $\sum_{n\in\mathbb{N}} \mu_n \mathsf{E}(\|e_n\|_{\mathsf{H}} \,|\, \mathcal{X}_n) < +\infty$ P-a.s. The case of deterministic relaxations and deterministic errors was considered in [16, Theorem 5.5(i)], as an extension of the classical result error-free result of [29, Corollary 3].

The following application of Theorem 3.24 concerns averaged operators.

Corollary 3.26 *Let $\alpha \in \left]0, 1\right[$, let $\mathsf{T}\colon \mathsf{H} \to \mathsf{H}$ be an α-averaged operator such that* $\operatorname{Fix}\mathsf{T} \neq \varnothing$*, and let $x_0 \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. Iterate*

$$\begin{array}{l}\text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} \text{take } e_n \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H}) \text{ and } \mu_n \in L^\infty(\Omega, \mathcal{F}, \mathsf{P}; \left]0, 1/\alpha\right[) \\ x_{n+1} = x_n + \mu_n\big(\mathsf{T}x_n + e_n - x_n\big). \end{array} \right.\end{array} \tag{3.85}$$

Set $(\forall n \in \mathbb{N})$ $\Phi_n = \{x_0, \ldots, x_n\}$ and $\mathcal{X}_n = \sigma(\Phi_n)$. Suppose that $\sum_{n\in\mathbb{N}} \mathsf{E}(\mu_n(1 - \alpha\mu_n)) = +\infty$ and, for every $n \in \mathbb{N}$, that μ_n is independent of $\sigma(\{e_n\} \cup \Phi_n)$. Then the following hold for some $\operatorname{Fix}\mathsf{T}$*-valued random variable x:*

(i) *Suppose that* $\mathsf{E}(\|e_n\|_{\mathsf{H}}^2 \,|\, \mathcal{X}_n) \to 0$ P-a.s. *and* $\sum_{n\in\mathbb{N}} \mathsf{E}\mu_n\sqrt{\mathsf{E}(\|e_n\|_{\mathsf{H}}^2 \,|\, \mathcal{X}_n)} < +\infty$ P-a.s. *Then the following hold:*

  (a) $(\mathsf{T}x_n - x_n)_{n\in\mathbb{N}}$ *converges strongly* P-a.s. *to* 0.

  (b) $(x_n)_{n\in\mathbb{N}}$ *converges weakly* P-a.s. *to* x.

  (c) *Suppose that* $\mathsf{T} - \mathrm{Id}$ *is demiregular at every point in* $\operatorname{Fix}\mathsf{T}$. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges strongly* P-a.s. *to* x.

(ii) *Suppose that* $\mathsf{E}\|e_n\|_{\mathsf{H}}^2 \to 0$ *and* $\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}\mu_n^2\mathsf{E}\|e_n\|_{\mathsf{H}}^2} < +\infty$. *Then the following hold:*

  (a) $(\mathsf{T}x_n - x_n)_{n\in\mathbb{N}}$ *converges strongly in* $L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ *and strongly* P-a.s. *to* 0.

  (b) $x \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \operatorname{Fix}\mathsf{T})$ *and* $(x_n)_{n\in\mathbb{N}}$ *converges weakly in* $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ *and weakly* P-a.s. *to* x.

  (c) *Suppose that* $\mathsf{T} - \mathrm{Id}$ *is demiregular at every point in* $\operatorname{Fix}\mathsf{T}$. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges strongly in* $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ *and strongly* P-a.s. *to* x.

Proof. Apply Theorem 3.24 to the nonexpansive operator $\mathrm{Id} + \alpha^{-1}(\mathsf{T} - \mathrm{Id})$ and observe that it has the same fixed points as T. ☐

Remark 3.27 As discussed in [17, 18], the Krasnosel'skiĭ–Mann iterative process for averaged operators is at the core of monotone operator splitting strategies such as the three operator splitting scheme of [24], the Douglas–Rachford algorithm [37], and the constant proximal parameter version of the forward-backward algorithm [39]. Stochastically relaxed and perturbed extensions of these algorithms can be derived from Corollary 3.26 with weaker assumptions than those of [21, Theorem 4.1].

We now consider a stochastic version of the (forward) Euler method to find a zero of a cocoercive operator. For simplicity, we adopt deterministic step-sizes $(\gamma_n)_{n\in\mathbb{N}}$. This result extends those of [20, 21, 52] by establishing, under weaker assumptions, weak convergence P-almost surely and, in addition, proving for the first time weak convergence in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$.

Corollary 3.28 *Let* $\beta \in \left]0, +\infty\right[$ *and let* $\mathsf{B}\colon \mathsf{H} \to \mathsf{H}$ *be* β*-cocoercive, with* $\operatorname{zer}\mathsf{B} = \left\{\mathsf{z} \in \mathsf{H} \mid \mathsf{B}\mathsf{z} = 0\right\} \neq \varnothing$. *Let* $(\mathsf{K}, \mathcal{K})$ *be a measurable space, let* $k\colon (\Omega, \mathcal{F}, \mathsf{P}) \to (\mathsf{K}, \mathcal{K})$ *be a random variable, let* $\xi \in \left]0, +\infty\right[$, *and let* $(\mathsf{B}_k)_{k\in\mathsf{K}}$ *be operators from* H *to* H *such that* $\mathsf{B}\colon (\mathsf{K}\times\mathsf{H}, \mathcal{K}\otimes\mathcal{B}_{\mathsf{H}}) \to (\mathsf{H}, \mathcal{B}_{\mathsf{H}})\colon (\mathsf{k}, \mathsf{x}) \mapsto \mathsf{B}_{\mathsf{k}}\mathsf{x}$ *is measurable and*

$$(\forall \mathsf{x} \in \mathsf{H}) \quad \mathsf{E}(\mathsf{B}_k\mathsf{x}) = \mathsf{B}\mathsf{x} \quad \textit{and} \quad \mathsf{E}\|\mathsf{B}_k\mathsf{x} - \mathsf{B}\mathsf{x}\|_{\mathsf{H}}^2 \leqslant \xi. \tag{3.86}$$

Let $\nu \in]2/3,1]$ *and* $x_0 \in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$. *Iterate*

$$\begin{aligned}
&\text{for } \mathsf{n} = 0,1,\ldots\\
&\left\lfloor
\begin{array}{l}
\mathcal{X}_\mathsf{n} = \sigma(x_0,\ldots,x_\mathsf{n})\\
k_\mathsf{n} \text{ is a copy of } k \text{ and is independent of } \mathcal{X}_\mathsf{n}\\
\gamma_\mathsf{n} = \dfrac{2\beta}{(\mathsf{n}+1)^\nu}\\
x_{\mathsf{n}+1} = x_\mathsf{n} - \gamma_\mathsf{n}\mathsf{B}_{k_\mathsf{n}}x_\mathsf{n}.
\end{array}
\right.
\end{aligned} \tag{3.87}$$

Then the following hold for some $x \in L^2(\Omega,\mathcal{F},\mathsf{P};\operatorname{zer}\mathsf{B})$*:*

(i) $(\mathsf{B}x_\mathsf{n})_{\mathsf{n}\in\mathbb{N}}$ *converges strongly in* $L^1(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ *and strongly* P-a.s. *to* 0.

(ii) $(x_\mathsf{n})_{\mathsf{n}\in\mathbb{N}}$ *converges weakly in* $L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ *and weakly* P-a.s. *to* x.

(iii) *Suppose that* B *is demiregular at every point in* $\operatorname{zer}\mathsf{B}$. *Then* $(x_\mathsf{n})_{\mathsf{n}\in\mathbb{N}}$ *converges strongly in* $L^1(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ *and strongly* P-a.s. *to* x.

Proof. We apply Corollary 3.26 to the reduction technique of [8]. Fix $\theta \in]0,2\beta[$ and set

$$(\forall \mathsf{n}\in\mathbb{N})\quad e_\mathsf{n} = \mathsf{B}x_\mathsf{n} - \mathsf{B}_{k_\mathsf{n}}x_\mathsf{n} = \mathsf{B}x_\mathsf{n} - \mathsf{B}\circ(k_\mathsf{n},x_\mathsf{n}) \;\text{ and }\; \mu_\mathsf{n} = \frac{\gamma_\mathsf{n}}{\theta} \in \left]0,\frac{2\beta}{\theta}\right[. \tag{3.88}$$

Then, for every $\mathsf{n}\in\mathbb{N}$, e_n is measurable with $\mathsf{E}(e_\mathsf{n}\,|\,\mathcal{X}_\mathsf{n}) = 0$ and $\mathsf{E}\mu_\mathsf{n} = \gamma_\mathsf{n}/\theta$. Let us also define $e'_0 = \gamma_0 e_0$ and

$$(\forall \mathsf{n}\in\mathbb{N})\quad e'_{\mathsf{n}+1} = (1-\gamma_{\mathsf{n}+1})e'_\mathsf{n} + \gamma_{\mathsf{n}+1}e_{\mathsf{n}+1}. \tag{3.89}$$

Set $\mathsf{T} = \mathsf{Id} - \theta\mathsf{B}$. Then $\operatorname{Fix}\mathsf{T} = \operatorname{zer}\mathsf{B}$ and T is $\theta/(2\beta)$-averaged [5, Proposition 4.39]. Finally, define, for every $\mathsf{n}\in\mathbb{N}$, $y_\mathsf{n} = x_\mathsf{n} - e'_\mathsf{n}$ and $\mathcal{Y}_\mathsf{n} = \sigma(y_0,\ldots,y_\mathsf{n})$. Then, we infer from (3.87) that

$$(\forall \mathsf{n}\in\mathbb{N})\quad y_{\mathsf{n}+1} = y_\mathsf{n} + \mu_\mathsf{n}\big(\mathsf{T}y_\mathsf{n} + e''_\mathsf{n} - y_\mathsf{n}\big), \tag{3.90}$$

where

$$(\forall \mathsf{n}\in\mathbb{N})\quad \begin{cases} e''_\mathsf{n} = \mathsf{T}x_\mathsf{n} - \mathsf{T}y_\mathsf{n} \in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H});\\ \|e''_\mathsf{n}\|_\mathsf{H} \leqslant \|x_\mathsf{n} - y_\mathsf{n}\|_\mathsf{H} = \|e'_\mathsf{n}\|_\mathsf{H} \;\; \mathsf{P}\text{-a.s.} \end{cases} \tag{3.91}$$

It follows from the choice of $(\gamma_\mathsf{n})_{\mathsf{n}\in\mathbb{N}}$, the uniformly bounded variance in (3.86), and [8, Example 2.7 and Theorem 2.5] that

$$\sum_{\mathsf{n}\in\mathbb{N}}\sqrt{\mathsf{E}\mu_\mathsf{n}^2\mathsf{E}\|e'_\mathsf{n}\|_\mathsf{H}^2} = \sum_{\mathsf{n}\in\mathbb{N}}\frac{\gamma_\mathsf{n}}{\theta}\sqrt{\mathsf{E}\|e'_\mathsf{n}\|_\mathsf{H}^2} < +\infty, \tag{3.92}$$

and

$$\sum_{\mathsf{n}\in\mathbb{N}}\mathsf{E}\bigg(\mu_\mathsf{n}\Big(1-\frac{\theta}{2\beta}\mu_\mathsf{n}\Big)\bigg) = \sum_{\mathsf{n}\in\mathbb{N}}\frac{\gamma_\mathsf{n}}{\theta}\Big(1-\frac{\gamma_\mathsf{n}}{2\beta}\Big) = +\infty. \tag{3.93}$$

We also deduce from the proofs of [8, Lemma 2.4 and Theorem 2.5] that $\mathsf{E}\|e'_\mathsf{n}\|_\mathsf{H}^2 \to 0$ and

$\|e'_n\|_H \to 0$ P-a.s. Consequently, by taking (3.91) into account, we obtain

$$\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}\mu_n^2 \mathsf{E}\|e''_n\|_H^2} < +\infty \quad \text{and} \quad \mathsf{E}\|e''_n\|_H^2 \to 0. \tag{3.94}$$

(i): It follows from Theorem 3.24(ii)(a) that $(By_n)_{n\in\mathbb{N}}$ converges strongly in $L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ and strongly P-a.s. to 0. On the other hand, (3.91) implies that

$$\begin{aligned}(\forall n \in \mathbb{N}) \quad \theta\|Bx_n - By_n\|_H &= \|\theta Bx_n - \theta By_n\|_H \\ &\leqslant \|x_n - \theta Bx_n - (y_n - \theta By_n)\|_H + \|x_n - y_n\|_H \\ &= \|e''_n\|_H + \|e'_n\|_H \\ &\leqslant 2\|e'_n\|_H \ \ \text{P-a.s.}\end{aligned} \tag{3.95}$$

Since $\mathsf{E}\|e'_n\|_H^2 \to 0$ and $\|e'_n\|_H \to 0$ P-a.s., we deduce that $(Bx_n - By_n)_{n\in\mathbb{N}}$ converges strongly in $L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ and strongly P-a.s. to 0 and, therefore, we obtain the convergence results for $(Bx_n)_{n\in\mathbb{N}}$.

(ii): We infer from Theorem 3.24(ii)(b) that $(y_n)_{n\in\mathbb{N}}$ converges weakly in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ and weakly P-a.s. to some $x \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \operatorname{zer} B)$. However, for every $n \in \mathbb{N}$, $x_n = y_n + e'_n$. Since $(e'_n)_{n\in\mathbb{N}}$ converges strongly P-a.s. and strongly in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ to 0, we conclude that $(x_n)_{n\in\mathbb{N}}$ converges weakly in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ and weakly P-a.s. to x.

(iii): This follows from Theorem 3.24(ii)(c) using the same arguments as in the proofs of (i) and (ii). ☐

The following special case of Corollary 3.28 concerns stochastic optimization and establishes new results on the convergence of the iterates generated by the standard stochastic gradient method, a method that goes back to the classical work of [7, 26, 50].

Corollary 3.29 *Let $\beta \in\,]0, +\infty[$ and let $f\colon \mathsf{H} \to \mathbb{R}$ be convex, differentiable, and such that ∇f is $1/\beta$-Lipschitzian, with $\operatorname{Argmin} f \neq \varnothing$. Let $(\mathsf{K}, \mathcal{K})$ be a measurable space, let $k\colon (\Omega, \mathcal{F}, \mathsf{P}) \to (\mathsf{K}, \mathcal{K})$ be a random variable, let $\xi \in\,]0, +\infty[$, and, for every $k \in \mathsf{K}$, let $g_k\colon \mathsf{H} \to \mathbb{R}$ be differentiable and such that $B\colon (\mathsf{K} \times \mathsf{H}, \mathcal{K} \otimes \mathcal{B}_H) \to (\mathsf{H}, \mathcal{B}_H)\colon (k, x) \mapsto \nabla g_k(x)$ is measurable and*

$$(\forall x \in \mathsf{H}) \quad \mathsf{E}\nabla g_k(x) = \nabla f(x) \quad \text{and} \quad \mathsf{E}\|\nabla g_k(x) - \nabla f(x)\|_H^2 \leqslant \xi. \tag{3.96}$$

Let $\nu \in\,]2/3, 1]$ and $x_0 \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. Iterate

$$\begin{array}{l}\text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} \mathcal{X}_n = \sigma(x_0, \ldots, x_n) \\ k_n \text{ is a copy of } k \text{ and is independent of } \mathcal{X}_n \\ \gamma_n = \dfrac{2\beta}{(n+1)^\nu} \\ x_{n+1} = x_n - \gamma_n \nabla g_{k_n}(x_n). \end{array}\right.\end{array} \tag{3.97}$$

Then the following hold for some $x \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \operatorname{Argmin} \mathsf{f})$*:*

(i) $(\nabla \mathsf{f}(x_n))_{n\in\mathbb{N}}$ *converges strongly in* $L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ *and strongly* P-a.s. *to* 0.

(ii) $(x_n)_{n\in\mathbb{N}}$ *converges weakly in* $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ *and weakly* P-a.s. *to* x.

(iii) *Suppose that* $\nabla \mathsf{f}$ *is demiregular at every point in* $\operatorname{Argmin} \mathsf{f}$. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges strongly in* $L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ *and strongly* P-a.s. *to* x.

Proof. Apply Corollary 3.28 to $\mathsf{B} = \nabla \mathsf{f}$, which is β-cocoercive [5, Corollary 18.17], and, for every $\mathsf{k} \in \mathsf{K}$, $\mathsf{B}_k = \nabla \mathsf{g}_k$. ☐

Remark 3.30 In Corollary 3.29(ii), the weak convergence P-a.s. and in L^2 results are new. In a finite-dimensional setting, we recover the P-a.s. convergence of [8, Corollary 4.5] with the novelty of the L^1 convergence. In the infinite-dimensional setting, we extend the result of [53] where the P-a.s. weak convergence is stated only for a subsequence of the iterates.

Remark 3.31 Variants of Corollary 3.28 can be explored by modifying the probabilistic assumptions in (3.86). In the context of Corollary 3.29, see for instance [13, 32, 46] and their bibliographies for possible candidates.

3.2.5 Application to common fixed point problems

The problem under consideration is a common fixed point problem involving an arbitrary family of firmly quasinonexpansive operators. Recall that $\mathsf{T} \colon \mathsf{H} \to \mathsf{H}$ is firmly quasinonexpansive [5, Definition 4.1(iv)] if

$$(\forall \mathsf{x} \in \mathsf{H})(\forall \mathsf{y} \in \operatorname{Fix} \mathsf{T}) \quad \|\mathsf{T}\mathsf{x} - \mathsf{y}\|_{\mathsf{H}}^2 + \|\mathsf{T}\mathsf{x} - \mathsf{x}\|_{\mathsf{H}}^2 \leqslant \|\mathsf{x} - \mathsf{y}\|_{\mathsf{H}}^2. \tag{3.98}$$

Example 3.32 ([4, Proposition 2.3]) Let $\mathsf{T} \colon \mathsf{H} \to \mathsf{H}$. Then T is firmly quasinonexpansive if one of the following holds:

(i) C is a nonempty closed convex subset of H and $\mathsf{T} = \operatorname{proj}_{\mathsf{C}}$ is the projector onto C. Here, $\operatorname{Fix} \mathsf{T} = \mathsf{C}$.

(ii) $\mathsf{f} \colon \mathsf{H} \to \left]-\infty, +\infty\right]$ is a proper lower semicontinuous convex function and

$$\mathsf{T} = \operatorname{prox}_{\mathsf{f}} \colon \mathsf{H} \to \mathsf{H} \colon \mathsf{x} \mapsto \underset{\mathsf{y}\in\mathsf{H}}{\operatorname{argmin}}\Big(\mathsf{f}(\mathsf{y}) + \frac{1}{2}\|\mathsf{x} - \mathsf{y}\|_{\mathsf{H}}^2\Big). \tag{3.99}$$

Here, $\operatorname{Fix} \mathsf{T} = \operatorname{Argmin} \mathsf{f}$.

(iii) $\mathsf{A} \colon \mathsf{H} \to 2^{\mathsf{H}}$ is maximally monotone and $\mathsf{T} = \mathsf{J}_{\mathsf{A}} = (\mathrm{Id} + \mathsf{A})^{-1}$. Here, $\operatorname{Fix} \mathsf{T} = \big\{\mathsf{z} \in \mathsf{H} \mid 0 \in \mathsf{A}\mathsf{z}\big\}$.

(iv) $\mathsf{f} \colon \mathsf{H} \to \mathbb{R}$ is a continuous convex function, $\mathsf{s} \colon \mathsf{H} \to \mathsf{H} \colon \mathsf{x} \mapsto \mathsf{s}(\mathsf{x}) \in \partial \mathsf{f}(\mathsf{x})$ is a selection of $\partial \mathsf{f}$, and

$$\mathsf{T} = \mathsf{G}_{\mathsf{f}} \colon \mathsf{H} \to \mathsf{H} \colon \mathsf{x} \mapsto \begin{cases} \mathsf{x} - \dfrac{\mathsf{f}(\mathsf{x})}{\|\mathsf{s}(\mathsf{x})\|_{\mathsf{H}}^2}\mathsf{s}(\mathsf{x}), & \text{if } \mathsf{f}(\mathsf{x}) > 0; \\ \mathsf{x}, & \text{if } \mathsf{f}(\mathsf{x}) \leqslant 0, \end{cases} \tag{3.100}$$

is the subgradient projector onto $\operatorname{Fix} T = \{x \in H \mid f(x) \leqslant 0\}$.

The following formulation covers a wide range of problems in mathematics and its applications [11,14,16].

Problem 3.33 Let $(K, \mathcal{K})$ be a measurable space and $(T_k)_{k \in K}$ a family of firmly quasinonexpansive operators such that $T\colon (K \times H, \mathcal{K} \otimes \mathcal{B}_H) \to (H, \mathcal{B}_H)\colon (k, x) \mapsto T_k x$ is measurable and, for every $k \in K$, $\mathrm{Id} - T_k$ is demiclosed at 0. Let $k\colon (\Omega, \mathcal{F}, P) \to (K, \mathcal{K})$ be a random variable. The task is to

$$\text{find} \;\; x \in Z = \big\{z \in H \mid z \in \operatorname{Fix} T_{k} \;\, P\text{-a.s.}\big\}, \tag{3.101}$$

under the assumption that $Z \neq \varnothing$.

Remark 3.34 Z is a closed convex subset of H. Indeed, let $(z_n)_{n \in \mathbb{N}}$ be a sequence in Z that converges to $z \in H$. For every $n \in \mathbb{N}$, let $\Omega_n \in \mathcal{F}$ be such that $P(\Omega_n) = 1$ and, for every $\omega \in \Omega_n$, let $z_n \in \operatorname{Fix} T_{k(\omega)}$. Set $\Omega' = \bigcap_{n \in \mathbb{N}} \Omega_n$. Then $P(\Omega') = 1$ and

$$(\forall \omega \in \Omega')(\forall n \in \mathbb{N}) \quad z_n \in \operatorname{Fix} T_{k(\omega)}. \tag{3.102}$$

Since each set of fixed points is closed [16, Proposition 2.3(v)], we deduce that, for every $\omega \in \Omega'$, $z \in \operatorname{Fix} T_{k(\omega)}$, i.e., $z \in Z$. So Z is closed. Likewise, let $z_1 \in Z$, $z_2 \in Z$, and $\alpha \in \,]0,1[$. Define almost sure events $\Omega_1 \in \mathcal{F}$ and $\Omega_2 \in \mathcal{F}$ as above. Then, it follows from the convexity of each set of fixed points [16, Proposition 2.3(v)] that

$$(\forall \omega \in \Omega_1 \cap \Omega_2) \quad \alpha z_1 + (1-\alpha) z_2 \in \operatorname{Fix} T_{k(\omega)}. \tag{3.103}$$

Since $P(\Omega_1 \cap \Omega_2) = 1$, we get $\alpha z_1 + (1-\alpha) z_2 \in Z$, which shows that Z is convex.

We propose the following stochastic variant of the extrapolated parallel block-iterative fixed point algorithm of [16]. It introduces stochasticity at four levels:

- The operators are indexed on a general measurable space rather than a countable set.
- The block of activated operators is randomly selected at each iteration.
- The evaluations of the operators at iteration n are averaged and extrapolated with random weights $(\beta_{i,n})_{1 \leqslant i \leqslant M}$.
- The relaxation parameter λ_n at iteration n is random and not confined to the interval $]0,2[$ as in traditional fixed point methods [5,16,25].

Theorem 3.35 *In the setting of Problem 3.33, let* $x_0 \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$, $0 < \mathsf{M} \in \mathbb{N}$, $\delta \in \,]0, 1/\mathsf{M}[$, *and* $\rho \in [2, +\infty[$. *Iterate*

$$
\begin{array}{l}
\text{for } \mathsf{n} = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\mathcal{X}_\mathsf{n} = \sigma(x_0, \ldots, x_\mathsf{n}) \\
\;\text{for } \mathsf{i} = 1, \ldots, \mathsf{M} \\
\;\left\lfloor
\begin{array}{l}
k_{\mathsf{i},\mathsf{n}} \text{ is a copy of } k \text{ and is independent of } \mathcal{X}_\mathsf{n} \\
p_{\mathsf{i},\mathsf{n}} = \mathsf{T}_{k_{\mathsf{i},\mathsf{n}}} x_\mathsf{n}
\end{array}
\right. \\
(\beta_{\mathsf{i},\mathsf{n}})_{1 \leqslant \mathsf{i} \leqslant \mathsf{M}} \textit{ are } [0,1]\text{-valued random variables such that} \\
\qquad \sum_{\mathsf{i}=1}^{\mathsf{M}} \beta_{\mathsf{i},\mathsf{n}} = 1 \ \mathsf{P}\text{-a.s. and } (\forall \mathsf{i} \in \{1, \ldots, \mathsf{M}\})\ \beta_{\mathsf{i},\mathsf{n}} \geqslant \delta 1_{[\|p_{\mathsf{i},\mathsf{n}} - x_\mathsf{n}\|_\mathsf{H} = \max_{1 \leqslant \mathsf{j} \leqslant \mathsf{M}} \|p_{\mathsf{j},\mathsf{n}} - x_\mathsf{n}\|_\mathsf{H}]} \qquad (3.104) \\
p_\mathsf{n} = \sum_{\mathsf{i}=1}^{\mathsf{M}} \beta_{\mathsf{i},\mathsf{n}} p_{\mathsf{i},\mathsf{n}} \\
L_\mathsf{n} = \dfrac{\sum_{\mathsf{i}=1}^{\mathsf{M}} \beta_{\mathsf{i},\mathsf{n}} \|p_{\mathsf{i},\mathsf{n}} - x_\mathsf{n}\|_\mathsf{H}^2 + 1_{[p_\mathsf{n} = x_\mathsf{n}]}}{\|p_\mathsf{n} - x_\mathsf{n}\|_\mathsf{H}^2 + 1_{[p_\mathsf{n} = x_\mathsf{n}]}} \\
a_\mathsf{n} = x_\mathsf{n} + L_\mathsf{n}(p_\mathsf{n} - x_\mathsf{n}) \\
\text{take } \lambda_\mathsf{n} \in L^\infty(\Omega, \mathcal{F}, \mathsf{P}; \,]0, \rho]) \\
x_{\mathsf{n}+1} = x_\mathsf{n} + \lambda_\mathsf{n}(a_\mathsf{n} - x_\mathsf{n}).
\end{array}
\right.
\end{array}
$$

Suppose that there exists $\mu \in \,]0, 1[$ *such that* $\inf_{\mathsf{n} \in \mathbb{N}} \mathsf{E}(\lambda_\mathsf{n}(2 - \lambda_\mathsf{n})) \geqslant \mu$ *and that, for every* $\mathsf{n} \in \mathbb{N}$, λ_n *is independent of* $\sigma(p_{1,\mathsf{n}}, \ldots, p_{\mathsf{M},\mathsf{n}}, \beta_{1,\mathsf{n}}, \ldots, \beta_{\mathsf{M},\mathsf{n}}, x_0, \ldots, x_\mathsf{n})$. *Then the following hold for some* $x \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{Z})$:

(i) $(x_\mathsf{n})_{\mathsf{n} \in \mathbb{N}}$ *converges weakly in* $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ *and weakly* P-a.s. *to* x.

(ii) *Suppose that there exists* $\mathsf{S} \in \mathcal{F}$ *such that*

$$\mathsf{S} \subset \left\{\omega \in \Omega \mid \mathsf{T}_{k(\omega)} - \mathrm{Id} \textit{ is demiregular at every point in } \operatorname{Fix} \mathsf{T}_{k(\omega)}\right\} \textit{ and } \mathsf{P}(\mathsf{S}) > 0. \tag{3.105}$$

Then $(x_\mathsf{n})_{\mathsf{n} \in \mathbb{N}}$ *converges strongly in* $L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ *and strongly* P-a.s. *to* x.

(iii) *Suppose that one of the following is satisfied:*

[A] *There exists* $\chi \in \,]0, 1[$ *such that*

$$(\forall \mathsf{n} \in \mathbb{N}) \quad \mathsf{E}\big(\mathsf{d}_\mathsf{Z}^2(x_{\mathsf{n}+1}) \,\big|\, \mathcal{X}_\mathsf{n}\big) \leqslant \chi \mathsf{d}_\mathsf{Z}^2(x_\mathsf{n}) \ \ \mathsf{P}\text{-a.s.} \tag{3.106}$$

[B] T *is linearly regular in the sense that there exists* $\nu \in [1, +\infty[$ *such that*

$$(\forall \mathsf{x} \in \mathsf{H}) \quad \mathsf{d}_\mathsf{Z}^2(\mathsf{x}) \leqslant \nu \mathsf{E}\|\mathsf{T}_k \mathsf{x} - \mathsf{x}\|_\mathsf{H}^2 = \nu \int_\Omega \big\|\mathsf{T}_{k(\omega)} \mathsf{x} - \mathsf{x}\big\|_\mathsf{H}^2 \mathsf{P}(d\omega), \tag{3.107}$$

in which case we set $\zeta = \inf_{\mathsf{j} \in \mathbb{N}} \mathsf{E}\lambda_\mathsf{j}^2$ *and* $\chi = 1 - \mu\delta\zeta/(\rho^2\nu)$.

Then $(x_n)_{n\in\mathbb{N}}$ *converges strongly in* $L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ *and strongly* P-a.s. *to* x, *and*

$$(\forall n\in\mathbb{N})\quad \mathsf{E}\|x_n-x\|_{\mathsf{H}}^2\leqslant 4\chi^n\mathsf{E}d_{\mathsf{Z}}^2(x_0). \tag{3.108}$$

Proof. We define

$$(\forall n\in\mathbb{N})\quad \begin{cases} t_n^*=x_n-p_n;\\ \eta_n=\displaystyle\sum_{i=1}^{M}\beta_{i,n}\langle p_{i,n}\mid x_n-p_{i,n}\rangle_{\mathsf{H}};\\ \alpha_n=\dfrac{1_{[t_n^*\neq 0]}1_{[\langle x_n\mid t_n^*\rangle_{\mathsf{H}}>\eta_n]}\big(\langle x_n\mid t_n^*\rangle_{\mathsf{H}}-\eta_n\big)}{\|t_n^*\|_{\mathsf{H}}^2+1_{[t_n^*=0]}};\\ \varepsilon_n=0\ \text{ P-a.s.};\\ d_n=x_n-a_n \end{cases} \tag{3.109}$$

and shall show that, in this setting, the sequence $(x_n)_{n\in\mathbb{N}}$ constructed by (3.104) corresponds to one generated by Algorithm 3.18. Let $n\in\mathbb{N}$. We first infer from (3.109) and (3.104) that

$$\begin{aligned} d_n&=x_n-a_n\\ &=L_n(x_n-p_n)\\ &=\frac{\sum_{i=1}^M\beta_{i,n}\|p_{i,n}-x_n\|_{\mathsf{H}}^2+1_{[p_n=x_n]}}{\|p_n-x_n\|_{\mathsf{H}}^2+1_{[p_n=x_n]}}(x_n-p_n)\\ &=\frac{\sum_{i=1}^M\beta_{i,n}\|x_n-p_{i,n}\|_{\mathsf{H}}^2+1_{[t_n^*=0]}}{\|t_n^*\|_{\mathsf{H}}^2+1_{[t_n^*=0]}}t_n^*\\ &=\frac{\sum_{i=1}^M\beta_{i,n}\big(\langle x_n\mid x_n-p_{i,n}\rangle_{\mathsf{H}}-\langle p_{i,n}\mid x_n-p_{i,n}\rangle_{\mathsf{H}}\big)}{\|t_n^*\|_{\mathsf{H}}^2+1_{[t_n^*=0]}}t_n^*\\ &=\frac{\langle x_n\mid t_n^*\rangle_{\mathsf{H}}-\sum_{i=1}^M\beta_{i,n}\langle p_{i,n}\mid x_n-p_{i,n}\rangle_{\mathsf{H}}}{\|t_n^*\|_{\mathsf{H}}^2+1_{[t_n^*=0]}}t_n^*\\ &=\alpha_n t_n^*\ \text{ P-a.s.} \end{aligned} \tag{3.110}$$

Next, let us show that

$$L_n\geqslant 1\ \text{ P-a.s.} \tag{3.111}$$

Fix $\mathsf{z}\in\mathsf{Z}$ and, for every $\mathsf{i}\in\{1,\ldots,\mathsf{M}\}$, let $\Omega_{\mathsf{i},n}\in\mathcal{F}$ be such that

$$\mathsf{P}(\Omega_{i,n})=1\quad\text{and}\quad(\forall\omega\in\Omega_{i,n})\quad \mathsf{z}\in\operatorname{Fix}\mathsf{T}_{k_{i,n}(\omega)}. \tag{3.112}$$

Thanks to (3.104), we then choose $\Omega_n\in\mathcal{F}$ such that

$$\mathsf{P}(\Omega_n)=1\quad\text{and}\quad(\forall\omega\in\Omega_n)\quad\bigcap_{1\leqslant i\leqslant M}\operatorname{Fix}\mathsf{T}_{k_{i,n}(\omega)}\neq\varnothing\quad\text{and}\quad\sum_{i=1}^{M}\beta_{i,n}(\omega)=1. \tag{3.113}$$

Given $\omega \in \Omega_n$, we consider the following two cases:

- Suppose that $p_n(\omega) = x_n(\omega)$. Then [16, Proposition 2.4] yields

$$x_n(\omega) \in \operatorname{Fix}\left(\sum_{i=1}^{M} \beta_{i,n}(\omega) \mathsf{T}_{k_{i,n}(\omega)}\right) = \bigcap_{1 \leqslant i \leqslant M} \operatorname{Fix} \mathsf{T}_{k_{i,n}(\omega)}, \tag{3.114}$$

  hence, $(\forall i \in \{1, \ldots, M\})$ $x_n(\omega) = p_{i,n}(\omega)$. Thus,

$$L_n(\omega) = \frac{\displaystyle\sum_{i=1}^{M} \beta_{i,n}(\omega) \|p_{i,n}(\omega) - x_n(\omega)\|_{\mathsf{H}}^2 + 1_{[p_n = x_n]}(\omega)}{\|p_n(\omega) - x_n(\omega)\|_{\mathsf{H}}^2 + 1_{[p_n = x_n]}(\omega)} = \frac{1_{[p_n = x_n]}(\omega)}{1_{[p_n = x_n]}(\omega)} = 1. \tag{3.115}$$

- Suppose that $p_n(\omega) \neq x_n(\omega)$. Then it follows from the convexity of $\|\cdot\|_{\mathsf{H}}^2$ that

$$0 < \|p_n(\omega) - x_n(\omega)\|_{\mathsf{H}}^2 = \left\|\sum_{i=1}^{M} \beta_{i,n}(\omega)\big(p_{i,n}(\omega) - x_n(\omega)\big)\right\|_{\mathsf{H}}^2 \leqslant \sum_{i=1}^{M} \beta_{i,n}(\omega)\big\|p_{i,n}(\omega) - x_n(\omega)\big\|_{\mathsf{H}}^2, \tag{3.116}$$

  which implies that $L_n(\omega) \geqslant 1$.

In view of (3.2), our next task is to show by induction that $(x_n)_{n\in\mathbb{N}}$ and $(t_n^*)_{n\in\mathbb{N}}$ are in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$, and that $(\eta_n)_{n\in\mathbb{N}}$ is in $L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathbb{R})$. To this end, let $n \in \mathbb{N}$ and $i \in \{1, \ldots, M\}$, and suppose that $x_n \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. Then $\mathsf{T}_{k_{i,n}} x_n = \mathsf{T} \circ (k_{i,n}, x_n)$ is measurable. On the other hand, for every $\omega \in \Omega_{i,n}$, $2\mathsf{T}_{k_{i,n}(\omega)} - \mathrm{Id}$ is quasinonexpansive with $\operatorname{Fix}(2\mathsf{T}_{k_{i,n}(\omega)} - \mathrm{Id}) = \operatorname{Fix} \mathsf{T}_{k_{i,n}(\omega)}$ [16, Proposition 2.2(v)] and hence

$$\begin{aligned} 2\big\|p_{i,n}(\omega)\big\|_{\mathsf{H}}^2 &= \frac{1}{2}\big\|2\mathsf{T}_{k_{i,n}(\omega)} x_n(\omega)\big\|_{\mathsf{H}}^2 \\ &\leqslant \big\|(2\mathsf{T}_{k_{i,n}(\omega)} - \mathrm{Id}) x_n(\omega) - \mathsf{z}\big\|_{\mathsf{H}}^2 + \big\|x_n(\omega) + \mathsf{z}\big\|_{\mathsf{H}}^2 \\ &\leqslant \big\|x_n(\omega) - \mathsf{z}\big\|_{\mathsf{H}}^2 + \big\|x_n(\omega) + \mathsf{z}\big\|_{\mathsf{H}}^2. \end{aligned} \tag{3.117}$$

Consequently, since $x_n \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ and $\mathsf{z} \in \mathsf{H}$, we have $p_{i,n} \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ and (3.104) therefore yields $p_n \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. Thus, $t_n^* = x_n - p_n \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. On the other hand, it follows from the Cauchy–Schwarz inequalities in H as well in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathbb{R})$ that

$$\mathsf{E}\Big|\langle p_{i,n} \mid x_n - p_{i,n}\rangle_{\mathsf{H}}\Big| \leqslant \mathsf{E}\Big(\big\|p_{i,n}\big\|_{\mathsf{H}}\big\|x_n - p_{i,n}\big\|_{\mathsf{H}}\Big) \leqslant \sqrt{\mathsf{E}\big\|p_{i,n}\big\|_{\mathsf{H}}^2 \mathsf{E}\big\|x_n - p_{i,n}\big\|_{\mathsf{H}}^2} < +\infty, \tag{3.118}$$

which shows that $\big\langle p_{i,n} \,\big|\, x_n - p_{i,n}\big\rangle_{\mathsf{H}} \in L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathbb{R})$. Since this is true for every $i \in \{1, \ldots, M\}$, we obtain $\eta_n \in L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathbb{R})$. Further, it follows from [5, Proposition 4.2(iv)] that

$$(\forall i \in \{1, \ldots, M\}) \quad \big\langle p_{i,n} - \mathsf{z} \,\big|\, x_n - p_{i,n}\big\rangle_{\mathsf{H}} = \big\langle \mathsf{T}_{k_{i,n}} x_n - \mathsf{z} \,\big|\, x_n - \mathsf{T}_{k_{i,n}} x_n\big\rangle_{\mathsf{H}} \geqslant 0 \;\; \mathsf{P}\text{-a.s.} \tag{3.119}$$

In turn, the concavity of $y \mapsto \langle y - z \,|\, x_n(\omega) - y\rangle_H$ yields

$$0 \leqslant \sum_{i=1}^{M} \beta_{i,n}\big\langle p_{i,n} - z \,\big|\, x_n - p_{i,n}\big\rangle_H \leqslant \big\langle p_n - z \,\big|\, x_n - p_n\big\rangle_H = \big\langle x_n - t_n^* - z \,\big|\, t_n^*\big\rangle_H \;\; \text{P-a.s.} \tag{3.120}$$

and therefore

$$\begin{aligned}
\frac{1}{2}\mathsf{E}\left|\frac{1_{[t_n^*\neq 0]}1_{[\langle x_n | t_n^*\rangle_H > \eta_n]}\eta_n}{\|t_n^*\|_H + 1_{[t_n^*=0]}}\right|^2 &\leqslant \mathsf{E}\left|\frac{\eta_n - \sum_{i=1}^M \beta_{i,n}\langle z \,|\, x_n - p_{i,n}\rangle_H}{\|t_n^*\|_H + 1_{[t_n^*=0]}}\right|^2 + \mathsf{E}\left|\frac{\sum_{i=1}^M \beta_{i,n}\langle z \,|\, x_n - p_{i,n}\rangle_H}{\|t_n^*\|_H + 1_{[t_n^*=0]}}\right|^2 \\
&= \mathsf{E}\left|\frac{\sum_{i=1}^M \beta_{i,n}\langle p_{i,n} - z \,|\, x_n - p_{i,n}\rangle_H}{\|t_n^*\|_H + 1_{[t_n^*=0]}}\right|^2 + \mathsf{E}\left|\frac{\langle z \,|\, t_n^*\rangle_H}{\|t_n^*\|_H + 1_{[t_n^*=0]}}\right|^2 \\
&\leqslant \mathsf{E}\left|\frac{\big\langle x_n - t_n^* - z \,\big|\, t_n^*\big\rangle_H}{\|t_n^*\|_H + 1_{[t_n^*=0]}}\right|^2 + \mathsf{E}\left|\frac{\|z\|_H\|t_n^*\|_H}{\|t_n^*\|_H + 1_{[t_n^*=0]}}\right|^2 \\
&\leqslant \mathsf{E}\left|\frac{\|x_n - t_n^* - z\|_H\|t_n^*\|_H}{\|t_n^*\|_H + 1_{[t_n^*=0]}}\right|^2 + \mathsf{E}\|z\|_H^2 \\
&\leqslant \mathsf{E}\|x_n - t_n^* - z\|_H^2 + \|z\|_H^2.
\end{aligned} \tag{3.121}$$

Since $\{x_n, t_n^*\} \subset L^2(\Omega, \mathcal{F}, \mathsf{P}; H)$ and $z \in H$, we thus obtain $1_{[t_n^*\neq 0]}\eta_n/(\|t_n^*\|_H + 1_{[t_n^*=0]}) \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathbb{R})$. Hence, arguing as in (3.50), we deduce from (3.110) that

$$x_n - a_n = \alpha_n t_n^* \in L^2(\Omega, \mathcal{F}, \mathsf{P}; H). \tag{3.122}$$

It therefore results from (3.104) that $x_{n+1} \in L^2(\Omega, \mathcal{F}, \mathsf{P}; H)$, which completes the induction argument. On the other hand, since $\alpha_n \in [0, +\infty[$ P-a.s., (3.120) yields

$$\big\langle z \,\big|\, \alpha_n t_n^*\big\rangle_H = \alpha_n \sum_{i=1}^{M} \beta_{i,n}\big\langle z \,\big|\, x_n - p_{i,n}\big\rangle_H \leqslant \alpha_n \sum_{i=1}^{M} \beta_{i,n}\big\langle p_{i,n} \,\big|\, x_n - p_{i,n}\big\rangle_H = \alpha_n\eta_n \;\; \text{P-a.s.} \tag{3.123}$$

Thus, appealing to Lemma 3.13 and (3.109), we obtain

$$\Big\langle z \,\Big|\, \mathsf{E}\big(\alpha_n t_n^* \,\big|\, \mathcal{X}_n\big)\Big\rangle_H = \mathsf{E}\Big(\big\langle z \,\big|\, \alpha_n t_n^*\big\rangle_H \,\Big|\, \mathcal{X}_n\Big) \leqslant \mathsf{E}\big(\alpha_n\eta_n \,\big|\, \mathcal{X}_n\big) + \varepsilon_n \;\; \text{P-a.s.} \tag{3.124}$$

Altogether, the sequence $(x_n)_{n\in\mathbb{N}}$ constructed by (3.104) corresponds to one generated by Algorithm 3.18. Now set $\zeta = \inf_{j\in\mathbb{N}} \mathsf{E}\lambda_n^2$ and note that $\zeta \geqslant \inf_{j\in\mathbb{N}} \mathsf{E}^2\lambda_j \geqslant \mu^2/4 > 0$. Hence, we infer from (3.104) and Lemma 3.10 that

$$\begin{aligned}
\mathsf{E}\big(\|x_{n+1} - x_n\|_H^2 \,\big|\, \mathcal{X}_n\big) &= \mathsf{E}\Big(\big\|\lambda_n(a_n - x_n)\big\|_H^2 \,\Big|\, \mathcal{X}_n\Big) \\
&= \mathsf{E}\Big(\big\|\lambda_n L_n(p_n - x_n)\big\|_H^2 \,\Big|\, \mathcal{X}_n\Big) \\
&= \mathsf{E}\Big(|\lambda_n L_n|^2\|p_n - x_n\|_H^2 \,\Big|\, \mathcal{X}_n\Big)
\end{aligned}$$

$$
\begin{aligned}
&= \mathsf{E}\left(\lambda_n^2 L_n \sum_{i=1}^{M} \beta_{i,n}\|p_{i,n} - x_n\|_H^2 \,\Big|\, \mathcal{X}_n\right)\\
&\geqslant \mathsf{E}\left(\lambda_n^2 \sum_{i=1}^{M} \beta_{i,n}\|p_{i,n} - x_n\|_H^2 \,\Big|\, \mathcal{X}_n\right)\\
&= (\mathsf{E}\lambda_n^2)\mathsf{E}\left(\sum_{i=1}^{M} \beta_{i,n}\|p_{i,n} - x_n\|_H^2 \,\Big|\, \mathcal{X}_n\right)\\
&\geqslant \zeta\mathsf{E}\left(\sum_{i=1}^{M} \beta_{i,n}\|p_{i,n} - x_n\|_H^2 \,\Big|\, \mathcal{X}_n\right)\\
&\geqslant \zeta\mathsf{E}\left(\delta \max_{1\leqslant j\leqslant M}\|p_{j,n} - x_n\|_H^2 \,\Big|\, \mathcal{X}_n\right)\\
&\geqslant \delta\zeta\mathsf{E}\left(\|p_{1,n} - x_n\|_H^2 \,\Big|\, \mathcal{X}_n\right)\\
&= \delta\zeta\mathsf{E}\left(\|\mathsf{T}_{k_{1,n}} x_n - x_n\|_H^2 \,\Big|\, \mathcal{X}_n\right).
\end{aligned} \tag{3.125}
$$

However, since $k_{1,n}$ is independent of $\mathcal{X}_n$, Lemma 3.11 implies that, for P-almost every $\omega' \in \Omega$,

$$
\begin{aligned}
\mathsf{E}\left(\|\mathsf{T}_{k_{1,n}} x_n - x_n\|_H^2 \,\Big|\, \mathcal{X}_n\right)(\omega') &= \int_\Omega \|\mathsf{T}_{k_{1,n}(\omega)} x_n(\omega') - x_n(\omega')\|_H^2 \mathsf{P}(d\omega)\\
&= \int_\Omega \|\mathsf{T}_{k(\omega)} x_n(\omega') - x_n(\omega')\|_H^2 \mathsf{P}(d\omega).
\end{aligned} \tag{3.126}
$$

Therefore, for P-almost every $\omega' \in \Omega$, (3.125) implies that

$$
\mathsf{E}(\|x_{n+1} - x_n\|_H^2 \,|\, \mathcal{X}_n)(\omega') \geqslant \delta\zeta \int_\Omega \|\mathsf{T}_{k(\omega)} x_n(\omega') - x_n(\omega')\|_H^2 \mathsf{P}(d\omega) \quad \mathsf{P}\text{-a.s.} \tag{3.127}
$$

Upon taking the expected value in (3.125), summing over $n \in \mathbb{N}$, and invoking Theorem 3.20(ii)(b), we obtain

$$
\mathsf{E}\left(\sum_{n\in\mathbb{N}} \int_\Omega \|\mathsf{T}_{k(\omega)} x_n - x_n\|_H^2 \mathsf{P}(d\omega)\right) = \sum_{n\in\mathbb{N}} \mathsf{E}\left(\int_\Omega \|\mathsf{T}_{k(\omega)} x_n - x_n\|_H^2 \mathsf{P}(d\omega)\right) < +\infty. \tag{3.128}
$$

Hence,

$$
\sum_{n\in\mathbb{N}} \int_\Omega \|\mathsf{T}_{k(\omega)} x_n - x_n\|_H^2 \mathsf{P}(d\omega) < +\infty \quad \mathsf{P}\text{-a.s.} \tag{3.129}
$$

Let $\Omega' \in \mathcal{F}$ such that $\mathsf{P}(\Omega') = 1$ and

$$
(\forall \omega' \in \Omega') \quad \sum_{n\in\mathbb{N}} \int_\Omega \|\mathsf{T}_{k(\omega)} x_n(\omega') - x_n(\omega')\|_H^2 \mathsf{P}(d\omega) < +\infty \quad \text{and} \quad \mathfrak{W}\big(x_n(\omega')\big)_{n\in\mathbb{N}} \neq \varnothing. \tag{3.130}
$$

The existence of such a set Ω' follows from (3.129) as well as Theorem 3.16(v)(a). Fix $\omega' \in \Omega'$ and let $x(\omega') \in \mathfrak{W}(x_n(\omega'))_{n\in\mathbb{N}}$, say $x_{j_n}(\omega') \rightharpoonup x(\omega')$. On the other hand, it follows from the monotone convergence theorem that

$$\int_{\Omega}\sum_{n\in\mathbb{N}}\big\|\mathsf{T}_{k(\omega)}x_n(\omega')-x_n(\omega')\big\|_{\mathsf{H}}^2\mathsf{P}(d\omega)=\sum_{n\in\mathbb{N}}\int_{\Omega}\big\|\mathsf{T}_{k(\omega)}x_n(\omega')-x_n(\omega')\big\|_{\mathsf{H}}^2\mathsf{P}(d\omega)<+\infty. \tag{3.131}$$

Hence, for P-almost every $\omega\in\Omega$, $\sum_{n\in\mathbb{N}}\|\mathsf{T}_{k(\omega)}x_n(\omega')-x_n(\omega')\|_{\mathsf{H}}^2<+\infty$. Therefore, there exists $\Omega''\in\mathcal{F}$ such that $\mathsf{P}(\Omega'')=1$ and

$$\big(\forall\omega\in\Omega''\big)\quad \mathsf{T}_{k(\omega)}x_n(\omega')-x_n(\omega')\to 0. \tag{3.132}$$

It then follows from the demiclosedness of the operators $(\mathsf{Id}-\mathsf{T}_k)_{k\in K}$ at 0 that

$$\big(\forall\omega\in\Omega''\big)\quad \mathsf{T}_{k(\omega)}x(\omega')=x(\omega'). \tag{3.133}$$

Therefore $x(\omega')\in\big\{\mathsf{z}\in\mathsf{H}\mid \mathsf{z}\in\operatorname{Fix}\mathsf{T}_k\ \mathsf{P}\text{-a.s.}\big\}=\mathsf{Z}$. Since ω' is arbitrarily taken in Ω', we conclude that

$$\mathfrak{W}(x_n)_{n\in\mathbb{N}}\subset\mathsf{Z}\ \mathsf{P}\text{-a.s.} \tag{3.134}$$

(i): This follows from (3.134) and Theorems 3.20(i)(c) and 3.20(ii)(c).

(ii): Let $\omega'\in\Omega'$. In view of (3.105), (3.133), and (3.132), there exists $\mathcal{F}\ni\Omega'''\subset\Omega''$ such that $\mathsf{P}(\Omega''')>0$ and

$$\big(\forall\omega\in\Omega'''\big)\quad \begin{cases}\mathsf{T}_{k(\omega)}-\mathsf{Id}\text{ is demiregular at }x(\omega');\\ \mathsf{T}_{k(\omega)}x_n(\omega')-x_n(\omega')\to 0.\end{cases} \tag{3.135}$$

However, (i) implies that, for P-almost every $\omega'\in\Omega$, $x_n(\omega')\rightharpoonup x(\omega')$. Therefore, by demiregularity, for P-almost every $\omega'\in\Omega$, we deduce from (3.135) that $x_n(\omega')\to x(\omega')$. Thus, $(x_n)_{n\in\mathbb{N}}$ converges strongly P-a.s. to x. Finally, the strong convergence in $L^1(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ follows from Theorem 3.20(ii)(d).

(iii): This follows from Theorem 3.20(ii)(f) when [A] holds. It remains to show that [B] implies [A]. Let us first show that

$$\chi\in\,]0,1[. \tag{3.136}$$

First, the concavity of $\xi\mapsto\xi(2-\xi)$ and Jensen's inequality yield

$$0<\mu\leqslant\inf_{n\in\mathbb{N}}\mathsf{E}\big(\lambda_n(2-\lambda_n)\big)\leqslant\inf_{n\in\mathbb{N}}\mathsf{E}\lambda_n(2-\mathsf{E}\lambda_n). \tag{3.137}$$

This quadratic inequality forces

$$0<1-\sqrt{1-\mu}\leqslant\inf_{n\in\mathbb{N}}\mathsf{E}\lambda_n, \tag{3.138}$$

and Jensen's inequality guarantees that $0<\inf_{n\in\mathbb{N}}\mathsf{E}\lambda_n^2=\zeta$. Next, since $\mu\in\,]0,1[$, $\delta\in\,]0,1[$,

$\nu \in [1,+\infty[$, $\rho \in [2,+\infty[$, and $\lambda_n \in]0,\rho]$ P-a.s., we have $\lambda_n^2/\rho^2 \in]0,1]$ P-a.s. and

$$\frac{\zeta}{\rho^2} = \frac{\inf_{n\in\mathbb{N}} \mathsf{E}\lambda_n^2}{\rho^2} \in]0,1]. \tag{3.139}$$

It follows then that $\mu\delta\zeta/(\rho^2\nu) \in]0,1[$ and therefore that $\chi = 1-\mu\delta\zeta/(\rho^2\nu) \in]0,1[$. Next, let $n \in \mathbb{N}$ and $z \in L^2(\Omega,\mathcal{X}_n,\mathsf{P};\mathsf{Z})$. Theorem 3.16(iii), the independence assumption for λ_n, and (3.2) imply that

$$\begin{aligned}\mathsf{E}\big(\|x_{n+1}-z\|_{\mathsf{H}}^2 \,\big|\, \mathcal{X}_n\big) &\leqslant \|x_n-z\|_{\mathsf{H}}^2 - \mathsf{E}\big(\lambda_n(2-\lambda_n)\big)\mathsf{E}\big(\|d_n\|_{\mathsf{H}}^2 \,\big|\, \mathcal{X}_n\big)\\ &= \|x_n-z\|_{\mathsf{H}}^2 - \mathsf{E}\big(\lambda_n(2-\lambda_n)\big)\mathsf{E}\bigg(\frac{1}{\lambda_n^2}\|x_{n+1}-x_n\|_{\mathsf{H}}^2 \,\bigg|\, \mathcal{X}_n\bigg) \quad \text{P-a.s.}\end{aligned} \tag{3.140}$$

Upon taking $z = \operatorname{proj}_{\mathsf{Z}} x_n$ in (3.140),

$$\begin{aligned}\mathsf{E}\big(\mathsf{d}_{\mathsf{Z}}^2(x_{n+1}) \,\big|\, \mathcal{X}_n\big) &\leqslant \mathsf{d}_{\mathsf{Z}}^2(x_n) - \mathsf{E}\big(\lambda_n(2-\lambda_n)\big)\mathsf{E}\bigg(\frac{1}{\lambda_n^2}\|x_{n+1}-x_n\|_{\mathsf{H}}^2 \,\bigg|\, \mathcal{X}_n\bigg)\\ &\leqslant \mathsf{d}_{\mathsf{Z}}^2(x_n) - \mu\mathsf{E}\bigg(\frac{1}{\lambda_n^2}\|x_{n+1}-x_n\|_{\mathsf{H}}^2 \,\bigg|\, \mathcal{X}_n\bigg)\\ &\leqslant \mathsf{d}_{\mathsf{Z}}^2(x_n) - \frac{\mu}{\rho^2}\mathsf{E}\big(\|x_{n+1}-x_n\|_{\mathsf{H}}^2 \,\big|\, \mathcal{X}_n\big).\end{aligned} \tag{3.141}$$

Thus, for P-almost every $\omega' \in \Omega$, we derive from (3.125) that

$$\begin{aligned}\mathsf{E}\big(\mathsf{d}_{\mathsf{Z}}^2(x_{n+1}) \,\big|\, \mathcal{X}_n\big)(\omega') &\leqslant \mathsf{d}_{\mathsf{Z}}^2(x_n)(\omega') - \frac{\mu\delta\zeta}{\rho^2}\int_\Omega \big\|\mathsf{T}_{k(\omega)}x_n(\omega') - x_n(\omega')\big\|_{\mathsf{H}}^2 \mathsf{P}(d\omega)\\ &\leqslant \chi\mathsf{d}_{\mathsf{Z}}^2(x_n)(\omega').\end{aligned} \tag{3.142}$$

Hence, $\mathsf{E}(\mathsf{d}_{\mathsf{Z}}^2(x_{n+1}) \,|\, \mathcal{X}_n) \leqslant \chi\mathsf{d}_{\mathsf{Z}}^2(x_n)$ P-a.s. and, in view of (3.136), [A] holds. The conclusion follows from Theorem 3.20(ii)(f). ☐

Remark 3.36

(i) In Algorithm (3.104), M is the batch size, i.e., the number of activated sets, p_n is the standard average of the selected operators, $L_n \geqslant 1$ is the extrapolation parameter, a_n is the extrapolated average, and λ_n is the relaxation parameter, which can exceed the standard bound 2 imposed by deterministic methods [16].

(ii) Problem 3.33 is studied in [27] for firmly nonexpansive operators with errors. A deterministic algorithm which activates all the operators at each iteration via a Bochner integral average is proposed. The weak convergence to a solution is established; see also [10] for a version in the context of projectors of Example 3.32(i). This result contrasts with Theorem 3.35 in which the convergence is guaranteed even when a finite number of operators are activated at each iteration.

(iii) In (3.104), we need not impose a lower bound on the weights $(\beta_{i,n})_{1\leqslant i\leqslant M}$ if we assume that, for every $\mathsf{i} \in \{1,\ldots,\mathsf{M}\}$, $\beta_{\mathsf{i},\mathsf{n}}$ is independent of $\sigma(p_{\mathsf{i},\mathsf{n}}, x_0, \ldots, x_\mathsf{n})$. Indeed, in such a case, Lemma 3.10 asserts that

$$\begin{aligned}\mathsf{E}\bigg(\sum_{\mathsf{i}=1}^{\mathsf{M}} \beta_{\mathsf{i},\mathsf{n}} \|p_{\mathsf{i},\mathsf{n}} - x_\mathsf{n}\|_{\mathsf{H}}^2 \,\bigg|\, \mathcal{X}_\mathsf{n}\bigg) &= \sum_{\mathsf{i}=1}^{\mathsf{M}} \mathsf{E}\big(\beta_{\mathsf{i},\mathsf{n}} \|p_{\mathsf{i},\mathsf{n}} - x_\mathsf{n}\|_{\mathsf{H}}^2 \,\big|\, \mathcal{X}_\mathsf{n}\big)\\ &= \sum_{\mathsf{i}=1}^{\mathsf{M}} (\mathsf{E}\beta_{\mathsf{i},\mathsf{n}})\mathsf{E}\big(\|p_{1,\mathsf{n}} - x_\mathsf{n}\|_{\mathsf{H}}^2 \,\big|\, \mathcal{X}_\mathsf{n}\big)\\ &= \mathsf{E}\big(\|p_{1,\mathsf{n}} - x_\mathsf{n}\|_{\mathsf{H}}^2 \,\big|\, \mathcal{X}_\mathsf{n}\big).\end{aligned} \tag{3.143}$$

(iv) Suppose that, for every $\mathsf{k} \in \mathsf{K}$, $\mathsf{T}_\mathsf{k} \colon \mathsf{H} \to \mathsf{H}$ is continuous. Then, to obtain the joint measurability of T, it is enough to suppose that, for every $\mathsf{x} \in \mathsf{H}$, $\mathsf{T}(\cdot, \mathsf{x}) \colon \mathsf{k} \mapsto \mathsf{T}_\mathsf{k}\mathsf{x}$ is measurable [1, Lemma 4.51].

Remark 3.37 In the literature, convergence to solutions has been established in specific instances of Problem 3.33 and algorithm (3.104).

(i) Several works have focused on the sequential unrelaxed case, that is, the scenario in which

$$\mathsf{M} = 1,\ \lambda_\mathsf{n} = 1,\ \text{and therefore } x_{\mathsf{n}+1} = a_\mathsf{n} = p_\mathsf{n} = p_{1,\mathsf{n}} = \mathsf{T}_{k_{1,\mathsf{n}}} x_\mathsf{n}. \tag{3.144}$$

In the context of the projectors of Example 3.32(i), [43] guarantees almost sure convergence to a solution when $\mathsf{H} = \mathbb{R}^\mathsf{N}$ and K is finite. This result is also found in [9] and in [34]. The setting of [34] involves a Euclidean space H and a general measurable space $(\mathsf{K}, \mathcal{K})$, and it also shows convergence in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. When the subsets are half-spaces or when the interior of Z is nonempty, [43] provides a rate for convergence in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. For general separable Hilbert spaces and under the assumption that the operators are averaged mappings, [30] shows weak almost sure convergence. In addition, a convergence rate is established in $L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ when (3.107) is satisfied. The paper [44] involves deterministic relaxations $\lambda_\mathsf{n} \in \,]0,2[$ in the context of subgradient projectors of Example 3.32(iv) in $\mathsf{H} = \mathbb{R}^\mathsf{N}$. Assuming that (3.107) holds and, additionally, that the subgradients are uniform bounded, almost sure convergence to a solution is established.

(ii) We now discuss works that have studied algorithms for $\mathsf{M} > 1$. In [33], K is countable, extrapolations are not allowed (hence $a_\mathsf{n} = p_\mathsf{n}$), λ_n is a deterministic parameter in $]0,2]$, and the condition $\operatorname{int} \mathsf{Z} \neq \varnothing$ is imposed. Finite convergence is established. In the context of projectors in $\mathsf{H} = \mathbb{R}^\mathsf{N}$, a similar approach to Algorithm 3.1 is studied in [40] and [42] with the following restrictions: deterministic relaxations $(\lambda_\mathsf{n})_{\mathsf{n}\in\mathbb{N}}$ in $]0,2[$ and iteration-independent fixed deterministic weights $\beta_{\mathsf{i},\mathsf{n}} \equiv 1/\mathsf{M}$. Mean-square rates of convergence are established by assuming that (3.107) holds, as well as ergodic convergence results. However, almost sure convergence is not proved. Similarly, [41] and [45] use a deterministic relaxation sequence $(\lambda_\mathsf{n})_{\mathsf{n}\in\mathbb{N}}$ in $]0,2[$ and iteration-independent fixed deterministic weights $\beta_{i,n} \equiv 1/M$ to solve Problem 3.33 in the context of subgradient projectors in $H = \mathbb{R}^N$. Under linear regularity assumptions and, additionally, uniform boundedness of the subgradients, rates of convergence in mean-square are provided. Nevertheless, almost sure convergence of the sequence of iterates is not guaranteed.

Remark 3.38 By combining the proofs to Theorem 3.24 and Theorem 3.35, it is possible to establish convergence results for the following error-tolerant algorithm for solving Problem 3.33: Let $x_0 \in L^2(\Omega, \mathcal{F}, P; H)$, $0 < M \in \mathbb{N}$, and $\delta \in \,]0, 1/M[$. Iterate

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\mathcal{X}_n = \sigma(x_0, \ldots, x_n) \\
\text{for } i = 1, \ldots, M \\
\left\lfloor
\begin{array}{l}
k_{i,n} \text{ is a copy of } k \text{ and is independent of } \mathcal{X}_n \\
\text{take } e_{i,n} \in L^2(\Omega, \mathcal{F}, P; H) \\
p_{i,n} = T_{k_{i,n}} x_n + e_{i,n}
\end{array}
\right. \\
(\beta_{i,n})_{1 \leqslant i \leqslant M} \text{ are } [0,1]\text{-valued random variables such that} \\
\qquad \sum_{i=1}^{M} \beta_{i,n} = 1 \text{ P-a.s. and } (\forall i \in \{1, \ldots, M\})\ \beta_{i,n} \geqslant \delta 1_{[\|p_{i,n} - x_n\|_H = \max\limits_{1 \leqslant j \leqslant M} \|p_{j,n} - x_n\|_H]} \\
p_n = \sum_{i=1}^{M} \beta_{i,n} p_{i,n} \\
\text{take } \lambda_n \in L^\infty(\Omega, \mathcal{F}, P;]0, 2[) \\
x_{n+1} = x_n + \lambda_n (p_n - x_n).
\end{array}
\right.
\end{array}
\tag{3.145}
$$

Suppose that $\inf_{n \in \mathbb{N}} E(\lambda_n(2 - \lambda_n)) > 0$, $\max_{1 \leqslant i \leqslant M} \sum_{n \in \mathbb{N}} \sqrt{E\|e_{i,n}\|_H^2} < +\infty$, and that, for every $n \in \mathbb{N}$, λ_n is independent of $\sigma(k_{1,n}, \ldots, k_{M,n}, e_{1,n}, \ldots, e_{M,n}, \beta_{1,n}, \ldots, \beta_{M,n}, x_0, \ldots, x_n)$. Then there exists $x \in L^2(\Omega, \mathcal{F}, P; Z)$ such that $(x_n)_{n \in \mathbb{N}}$ converges weakly in $L^2(\Omega, \mathcal{F}, P; H)$ and weakly P-a.s. to x.

3.2.6 Numerical experiments

We illustrate numerically our results in the context of Problem 3.33 with applications of the stochastic algorithm (3.104) with the deterministic relaxation strategies

$$(\forall n \in \mathbb{N}) \quad \lambda_n = 1.0 \tag{3.146}$$

and

$$(\forall n \in \mathbb{N}) \quad \lambda_n = 1.9. \tag{3.147}$$

We also consider the random relaxation strategies

$$(\forall n \in \mathbb{N}) \quad P([\lambda_n = 2.3]) = \frac{1}{2} \text{ and } P([\lambda_n = 1.5]) = \frac{1}{2} \tag{3.148}$$

and

$$(\forall n \in \mathbb{N}) \quad \lambda_n \sim \text{uniform}([1.5, 2.3]). \tag{3.149}$$

Note that (3.148) and (3.149) are super relaxation strategies that satisfy, for every $n \in \mathbb{N}$, $\mathsf{E}(\lambda_n(2-\lambda_n)) > 0$, $\mathsf{P}([\lambda_n > 2]) > 0$, and $\mathsf{E}\lambda_n = 1.9$. Problem 3.33 is specialized to the standard Euclidean space $\mathsf{H} = \mathbb{R}^N$ with $\|\cdot\|_{\mathsf{H}} = \|\cdot\|$, $\mathsf{K} = \{1,\dots,p\}$, and $k \sim \text{uniform}(\mathsf{K})$.

Problem 3.39 For every $k \in \{1,\dots,p\}$, $f_k \colon \mathbb{R}^N \to \mathbb{R}$ is a convex function and $C_k = \{x \in \mathbb{R}^N \mid f_k(x) \leqslant 0\}$. It is assumed that $Z = \bigcap_{1\leqslant k\leqslant p} C_k \neq \varnothing$. The task is to

$$\text{find } x \in \mathbb{R}^N \text{ such that } x \in Z. \tag{3.150}$$

Consider the setting of Problem 3.39. For every $k \in \{1,\dots,p\}$, let $T_k \colon \mathbb{R}^N \to \mathbb{R}^N$ be the subgradient projector onto C_k of Example 3.32(iv), so that, by [5, Propositions 16.20 and 29.41]

$$T_k \text{ is firmly quasinonexpansive},\ \operatorname{Fix} T_k = C_k,\ \text{and } \mathrm{Id} - T_k \text{ is demiclosed at } 0. \tag{3.151}$$

Subgradient projectors extend the classical projection operators in the following sense. Let C be a nonempty closed and convex subset of $\mathbb{R}^N$ and suppose that $f_k = d_C$. Then $C_k = C$ and $G_k = \operatorname{proj}_C$ [5, Example 29.44]. Their importance in solving Problem 3.39 stems from the fact that they are generally much easier to implement than exact ones.

3.2.6.1 Signal restoration

The goal is to recover the original signal $\overline{x} \in \mathbb{R}^N$ ($N = 1024$) shown in Fig. 3.2(a) from 20 noisy observations $(r_k)_{1\leqslant k\leqslant 20}$ given by

$$(\forall k \in \{1,\dots,20\}) \quad r_k = L_k\overline{x} + w_k \tag{3.152}$$

where $L_k \colon \mathbb{R}^N \to \mathbb{R}^N$ is a known linear operator, $\eta_k \in \left]0,+\infty\right[$, and $w_k \in [-\eta_k,\eta_k]^N$ is a bounded random noise vector. The parameters $(\eta_k)_{1\leqslant k\leqslant 20} \in \left]0,+\infty\right[^{20}$ are known. The operators $(L_k)_{1\leqslant k\leqslant 20}$ are Gaussian convolution filters with zero mean and standard deviation taken uniformly in $[10,30]$, $\eta_k = 0.15$, and w_k is taken uniformly in $[-\eta_k,\eta_k]^N$. Set, for every $k \in \{1,\dots,20\}$ and every $j \in \{1,\dots,N\}$,

$$C_{k,j} = \big\{x \in \mathbb{R}^N \mid -\eta_k \leqslant \langle L_k x - r_k \mid e_j\rangle \leqslant \eta_k\big\}. \tag{3.153}$$

Since the intersection of these sets is nonempty and their projectors are computable explicitly [5, Example 29.21], we solve the feasibility problem

$$\text{find } x \in \mathbb{R}^N \text{ such that } \quad (\forall k \in \{1,\dots,20\})(\forall j \in \{1,\dots,N\}) \;\; x \in C_{k,j} \tag{3.154}$$

by algorithm (3.104) implemented with exact projectors. We run two instances with $x_0 = 0$. In the first one, $M = 1$. Note that the relaxation scheme of (3.146) leads to the almost sure convergence result of [43] (see also [34]), while the relaxation schemes (3.147)–(3.149) are new even in this specialized context of randomly activated projection method. In the second instance $\mathsf{M} = 16$. Fig. 3.3 displays the normalized error versus execution time.

Fig. 3.3 (top) shows the benefits of large relaxations when $\mathsf{M} = 1$. Fig. 3.3 (bottom) shows the advantage of using $\mathsf{M} > 1$ random blocks, in which case the extrapolation parameter L_n is not equal to 1 and can attain large values. This behavior was previously observed for deterministic algorithms [6, 12, 15, 48]. Fig. 3.3 also suggests that, on a single run, the use of the proposed random super relaxation scheme can further improve the speed of convergence. It is worth noting that the execution time can naturally be reduced if Algorithm 3.1 is implemented on a multi-core architecture where, at each iteration, each (subgradient) projector is assigned to a dedicated core and all the cores work in parallel.

3.2.6.2 Image restoration

The goal is to recover the original image $\bar{x} \in \mathbb{R}^{N \times N}$ $(N = 256)$ shown in Fig. 3.4(a) from four observations $\{r_1, \ldots, r_4\}$ which are given by the degradation of $\bar{x}$ via a Gaussian kernel

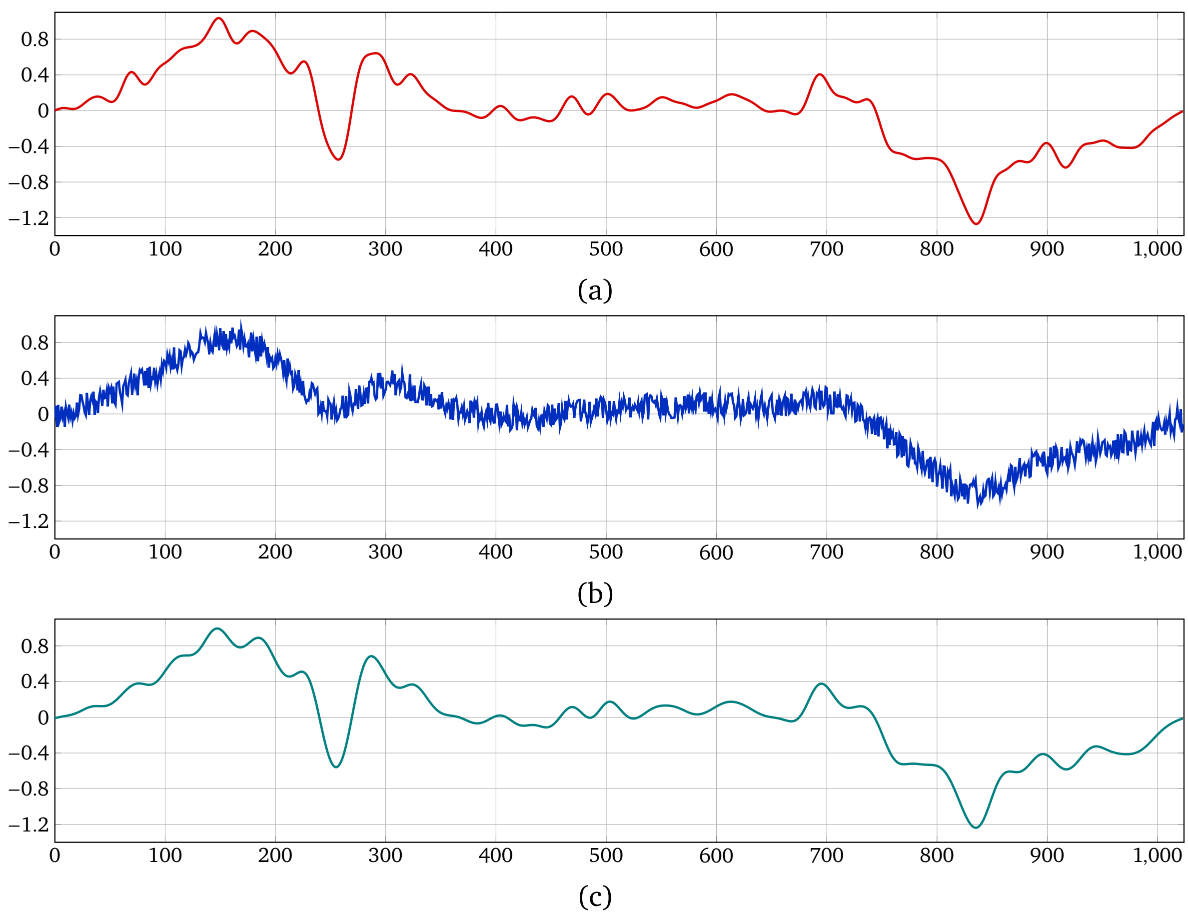


Figure 3.2 Experiment of Section 3.2.6.1. (a): Original signal $\bar{x}$. (b): Noisy observation r_1. (c): Solution produced by algorithm (3.104).

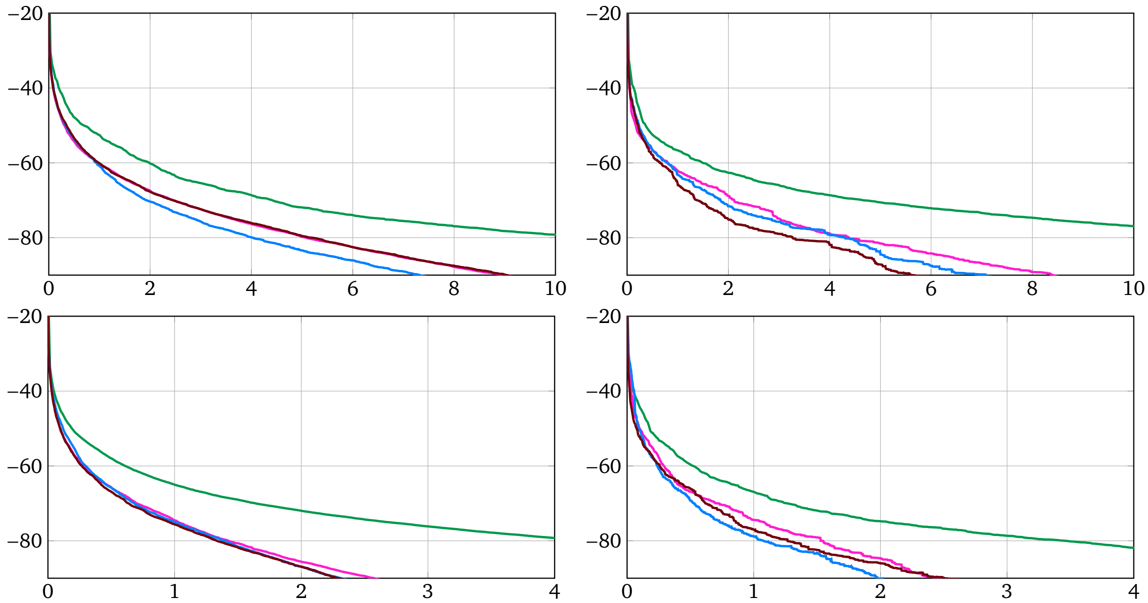


Figure 3.3 Experiment of Section 3.2.6.1. Normalized error $20\log(\|x_n - x_\infty\|/\|x_0 - x_\infty\|)$ (dB) versus execution time (s) on single processor machine for various relaxation strategies. Green: (3.146). Magenta: (3.147). Blue: (3.148). Brown: (3.149). Left: Average over ten runs. Right: A single run. Top: $M = 1$. Bottom: $M = 16$.

(a) (b) (c)

Figure 3.4 Experiment of Section 3.2.6.2. (a) Original image $\bar{x}$. (b) Noisy observation r_1. (c) Solution produced by algorithm (3.104).

with a standard deviation of 8 and the addition of random noise. The noise distribution is uniform$([0,5]^{N\times N})$. Let L be the block-Toeplitz matrix associated with the convolutional blur. Then

$$(\forall k \in \{1,2,3,4\}) \quad r_k = L\bar{x} + w_k, \quad \text{where} \quad w_k \sim \text{uniform}\big([0,5]^{N\times N}\big). \tag{3.155}$$

The entries of the random vectors $(w_k)_{1\leqslant k\leqslant 4}$ are i.i.d. Therefore, as shown in [23], for every

$k \in \{1, 2, 3, 4\}$, with a 95% confidence coefficient

$$\bar{x} \in C_k = \big\{x \in \mathbb{R}^{N\times N} \mid \|r_k - Lx\|^2 \leqslant \xi\big\}, \tag{3.156}$$

where $\xi = N^2 E|u|^2 + 1.96N\sqrt{E|u|^4 - E^2|u|^2}$ with $u \sim \text{uniform}([0,5])$. For every $k \in \{1, 2, 3, 4\}$, we compute the subgradient projector onto C_k in (3.100) via the function $f_k : x \mapsto \|r_k - Lx\|^2 - \xi$. In addition, the boundedness on pixel values is incorporated as the property set $C_5 = [0, 255]^{N\times N}$. Finally, it is assumed that the discrete Fourier transform $\mathfrak{F}(\bar{x})$ of $\bar{x}$ is known on a portion of its support for low frequencies in both directions. That is, let S be the set of frequency pairs $\{0, \ldots, N/8 - 1\}^2$ as well as those resulting from the symmetry properties of the 2D discrete Fourier transform of real images. The associated set is $C_6 = \big\{x \in \mathbb{R}^{N\times N} \mid \mathfrak{F}(x)1_S = \mathfrak{F}(\bar{x})1_S\big\}$ and its projection is given by $\text{proj}_{C_6} : x \mapsto \mathfrak{F}^{-1}(\mathfrak{F}(\bar{x})1_S + \mathfrak{F}(x)1_{\complement S})$. We run algorithm (3.104) with $x_0 = 0$ and $M = 2$. Fig. 3.5 displays the normalized error versus execution time. These results confirm the conclusions of Section 3.2.6.1.

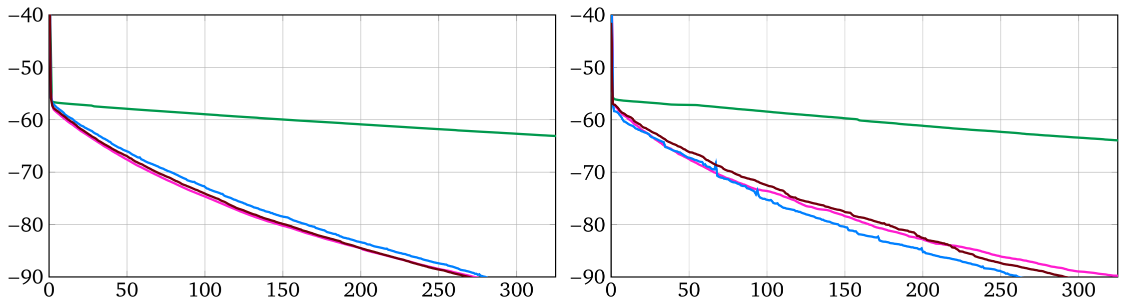


Figure 3.5 Experiment of Section 3.2.6.2 using $M = 2$. Normalized error $20\log(\|x_n - x_\infty\|/\|x_0 - x_\infty\|)$ (dB) versus execution time (s) on a single processor machine for various relaxation strategies. Green: (3.146). Magenta: (3.147). Blue: (3.148). Brown: (3.149). Left: Average over ten runs. Right: A single run.

Acknowledgement

The authors thank Dr. Minh N. Bùi for bringing to their attention a gap in the proof of an earlier version of Theorem 3.35.

Funding Information

This work was supported by the National Science Foundation under grant DMS-2513409.

References

[1] C. D. Aliprantis and K. C. Border, *Infinite Dimensional Analysis*, 3rd ed. Springer, New York, 2006.

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

[3] J.-B. Baillon, R. E. Bruck, and S. Reich, On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces, *Houston J. Math.*, vol. 4, pp. 1–9, 1978.

[4] H. H. Bauschke and P. L. Combettes, A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces, *Math. Oper. Res.*, vol. 26, pp. 248–264, 2001.

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

[6] H. H. Bauschke, P. L. Combettes, and S. G. Kruk, Extrapolation algorithm for affine-convex feasibility problems, *Numer. Algorithms*, vol. 41, pp. 239–274, 2006.

[7] J. R. Blum, Multidimensional stochastic approximation methods, *Ann. Math. Stat.*, vol. 25, pp. 737–744, 1954.

[8] M. Bravo and R. Cominetti, Stochastic fixed-point iterations for nonexpansive maps: Convergence and error bounds, *SIAM J. Control Optim.*, vol. 62, pp. 191–219, 2024.

[9] L. Briceño-Arias, J. Deride, and C. Vega, Random activations in primal-dual splittings for monotone inclusions with a priori information, *J. Optim. Theory Appl.*, vol. 192, pp. 56–81, 2022.

[10] D. Butnariu and S. D. Flåm, Strong convergence of expected-projection methods in Hilbert spaces, *Numer. Funct. Anal. Optim.*, vol. 16, pp. 601–636, 1995.

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

[12] Y. Censor, W. Chen, P. L. Combettes, R. Davidi, and G. T. Herman, On the effectiveness of projection methods for convex feasibility problems with linear inequality constraints, *Comput. Optim. Appl.*, vol. 51, pp. 1065–1088, 2012.

[13] Y.-L. Chen, S. Na, and M. Kolar, Convergence analysis of accelerated stochastic gradient descent under the growth condition, *Math. Oper. Res.*, vol. 49, pp. 2492–2526, 2024.

[14] P. L. Combettes, The convex feasibility problem in image recovery, in: *Advances in Imaging and Electron Physics* (P. Hawkes, Ed.), vol. 95, pp. 155–270. Academic Press, New York, 1996.

[15] P. L. Combettes, Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections, *IEEE Trans. Image Process.*, vol. 6, pp. 493–506, 1997.

[16] P. L. Combettes, Quasi-Fejérian analysis of some optimization algorithms, in: *Inherently Parallel Algorithms for Feasibility and Optimization*, (D. Butnariu, Y. Censor, and S. Reich, eds.), pp. 115–152. Elsevier, New York, 2001.

[17] P. L. Combettes, Solving monotone inclusions via compositions of nonexpansive averaged operators, *Optimization*, vol. 53, pp. 475–504, 2004.

[18] P. L. Combettes, The geometry of monotone operator splitting methods, *Acta Numer.*, vol. 33, pp. 487–632, 2024.

[19] P. L. Combettes and J. I. Madariaga, Asymptotic analysis of an abstract stochastic scheme for solving monotone inclusions, arxiv, 2025. `https://arxiv.org/pdf/2512.03023`

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

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

[22] P. L. Combettes and J.-C. Pesquet, Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping II: Mean-square and linear convergence, *Math. Program.*, vol. B174, pp. 433–451, 2019.

[23] P. L. Combettes and H. J. Trussell, The use of noise properties in set theoretic estimation, *IEEE Trans. Signal Process.*, vol. 39, pp. 1630–1641, 1991.

[24] D. Davis and W. Yin, A three-operator splitting scheme and its optimization applications, *Set-Valued Var. Anal.*, vol. 25, pp. 829–858, 2017.

[25] Q.-L. Dong, Y. J. Cho, S. He, P. M. Pardalos, and T. M. Rassias, *The Krasnosel'skiĭ–Mann Iterative Method – Recent Progress and Applications.* Springer, New York, 2022.

[26] Yu. M. Ermol'ev and Z. V. Nekrylova, Some methods of stochastic optimization, *Kibernetika (Kiev)*, vol. 1966, pp. 96–98, 1966.

[27] S. D. Flåm, Successive averages of firmly nonexpansive mappings, *Math. Oper. Res.*, vol. 20, pp. 497–512, 1995.

[28] K. Goebel and S. Reich, *Uniform Convexity, Hyperbolic Geometry, and Nonexpansive Mappings.* Marcel Dekker, New York, 1984.

[29] C. W. Groetsch, A note on segmenting Mann iterates, *J. Math. Anal. Appl.*, vol. 40, pp. 369–372, 1972.

[30] N. Hermer, D. R. Luke, and A. Sturm, Random function iterations for consistent stochastic feasibility, *Numer. Funct. Anal. Optim.*, vol. 40, pp. 386–420, 2019.

[31] T. Hytönen, J. van Neerven, M. Veraar, and L. Weis, *Analysis in Banach Spaces. Volume I: Martingales and Littlewood–Paley Theory.* Springer, New York, 2016.

[32] A. Jofré and P. Thompson, On variance reduction for stochastic smooth convex optimization with multiplicative noise, *Math. Program.*, vol. A174, pp. 253–292, 2019.

[33] V. I. Kolobov, S. Reich, and R. Zalas, Finitely convergent deterministic and stochastic iterative methods for solving convex feasibility problems, *Math. Program.*, vol. A194, pp. 1163–1183, 2022.

[34] V. R. Kostić and S. Salzo, The method of randomized Bregman projections for stochastic feasibility problems, *Numer. Algorithms*, vol. 93, pp. 1269–1307, 2023.

[35] M. A. Krasnosel'skiĭ, Two remarks on the method of successive approximations, *Uspekhi Mat. Nauk*, vol. 10, pp. 123–127, 1955.

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

[37] P.-L. Lions and B. Mercier, Splitting algorithms for the sum of two nonlinear operators, *SIAM J. Numer. Anal.*, vol. 16, pp. 964–979, 1979.

[38] W. R. Mann, Mean value methods in iteration, *Proc. Amer. Math. Soc.*, vol. 4, pp. 506–510, 1953.

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

[40] I. Necoara, Stochastic block projection algorithms with extrapolation for convex feasibility problems, *Optim. Methods Softw.*, vol. 37, pp. 1845–1875, 2022.

[41] I. Necoara and A. Nedić, Minibatch stochastic subgradient-based projection algorithms for feasibility problems with convex inequalities, *Comput. Optim. Appl.*, vol. 80, pp. 121–152, 2021.

[42] I. Necoara, P. Richtárik, and A. Patrascu, Randomized projection methods for convex feasibility: Conditioning and convergence rates, *SIAM J. Optim.*, vol. 29, pp. 2814–2852, 2019.

[43] A. Nedić, Random projection algorithms for convex set intersection problems, *Proc. 49th IEEE Conf. Decis. Control*, pp. 7655–7660, 2010.

[44] A. Nedić, Random algorithms for convex minimization problems, *Math. Program.*, vol. B129, pp. 225–253, 2011.

[45] A. Nedić and I. Necoara, Random minibatch subgradient algorithms for convex problems with functional constraints, *Appl. Math. Optim.*, vol. 80, pp. 801–833, 2019.

[46] L. M. Nguyen, P. H. Nguyen, P. Richtárik, K. Scheinberg, M. Takáč, and M. van Dijk, New convergence aspects of stochastic gradient algorithms, *J. Mach. Learn. Res.*, vol. 20, pp. 1–49, 2019.

[47] B. J. Pettis, On integration in vector spaces, *Trans. Amer. Math. Soc.*, vol. 44, pp. 277–304, 1938.

[48] G. Pierra, Decomposition through formalization in a product space, *Math. Program.*, vol. 28, pp. 96–115, 1984.

[49] S. Reich, Weak convergence theorems for nonexpansive mappings in Banach spaces, *J. Math. Anal. Appl.*, vol. 67, pp. 274–276, 1979.

[50] H. Robbins and S. Monro, A stochastic approximation method, *Ann. Math. Statist.*, vol. 22, pp. 400–407, 1951.

[51] H. Robbins and D. Siegmund, A convergence theorem for non negative almost supermartingales and some applications, in: *Optimizing Methods in Statistics*, (J. S. Rustagi, Ed.), pp. 233–257. Academic Press, New York, 1971.

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

[53] L. Rosasco, S. Villa, and B. C. Vũ, Convergence of stochastic proximal gradient algorithm, *Appl. Math. Optim.*, vol. 82, pp. 891–917, 2020.

[54] A. N. Shiryaev, *Probability–1*, 3rd ed. Springer, New York, 2016.

Chapter 4

ASYMPTOTIC ANALYSIS OF AN ABSTRACT STOCHASTIC SCHEME FOR SOLVING MONOTONE INCLUSIONS

4.1 Introduction and context

In this chapter, we address question **(Q3)** from Chapter 1. We specialize the framework of Chapter 3 to develop a stochastic scheme for solving the reduced inclusion problem with two operators.

This chapter presents the following journal article:

P. L. Combettes and J. I. Madariaga, Asymptotic analysis of an abstract stochastic scheme for solving monotone inclusions, submitted.

4.2 Article: Asymptotic analysis of an abstract stochastic scheme for solving monotone inclusions

Abstract. We propose an abstract stochastic scheme for solving a broad range of monotone operator inclusion problems in Hilbert spaces. This framework allows for the introduction of stochasticity at several levels in monotone operator splitting methods: approximation of operators, selection of coordinates and operators in block-iterative implementations, and relaxation parameters. The analysis involves an abstract reduced inclusion model with two operators. At each iteration of the proposed scheme, stochastic approximations to points in the graphs of these two operators are used to form the update. The results are applied to derive the almost sure and L^2 convergence of stochastic versions of the proximal point algorithm, as well as of

randomized block-iterative projective splitting methods for solving systems of coupled inclusions involving a mix of set-valued, cocoercive, and Lipschitzian monotone operators combined via various monotonicity-preserving operations.

4.2.1 Introduction

The object of the present paper is to study the asymptotic behavior of an abstract stochastic scheme for solving a broad class of monotone inclusion problems in Hilbert spaces. As in the deterministic methods unified in [18], our analysis is articulated around the following two-operator abstract model.

Problem 4.1 Let H be a separable real Hilbert space, let $\mathsf{W}\colon \mathsf{H} \to 2^{\mathsf{H}}$ be maximally monotone, let $\alpha \in \,]0,+\infty[$, and let $\mathsf{C}\colon \mathsf{H} \to \mathsf{H}$ be α-cocoercive and such that $\mathsf{Z} = \operatorname{zer}(\mathsf{W} + \mathsf{C}) \neq \varnothing$. The task is to

$$\text{find } \mathsf{x} \in \mathsf{H} \text{ such that } 0 \in \mathsf{W}\mathsf{x} + \mathsf{C}\mathsf{x}. \tag{4.1}$$

If the resolvent of W were numerically tractable, Problem 4.1 could be solved via the classical forward-backward algorithm [25, 37, 46]. However, in the general inclusion models to be considered, W is typically a composite operator defined on a product space, which makes such an assumption unrealistic. Instead, we merely assume the ability to pick points in the graph of W. This leads us to the following deterministic algorithmic template from [18, Section 4.4], which was first considered in [13, Proposition 3] in the context of saddle projective splitting methods.

Algorithm 4.2 In the setting of Problem 4.1, let $\mathsf{x}_0 \in \mathsf{H}$ and iterate

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\text{take } (\mathsf{w}_n, \mathsf{w}_n^*) \in \operatorname{gra} \mathsf{W} \text{ and } \mathsf{q}_n \in \mathsf{H} \\
\mathsf{t}_n^* = \mathsf{w}_n^* + \mathsf{C}\mathsf{q}_n \\
\Delta_n = \langle \mathsf{x}_n - \mathsf{w}_n \mid \mathsf{t}_n^* \rangle - (4\alpha)^{-1} \|\mathsf{w}_n - \mathsf{q}_n\|^2 \\
\theta_n = \begin{cases} \dfrac{\Delta_n}{\|\mathsf{t}_n^*\|^2}, & \text{if } \Delta_n > 0; \\ 0, & \text{otherwise} \end{cases} \\
\mathsf{d}_n = \theta_n \mathsf{t}_n^* \\
\text{take } \lambda_n \in \,]0,2[\\
\mathsf{x}_{n+1} = \mathsf{x}_n - \lambda_n \mathsf{d}_n.
\end{array}
\right.
\end{array}
\tag{4.2}
$$

As shown in [18], Algorithm 4.2 is at the core of a broad range of classical and block-iterative deterministic splitting methods, in particular those of [7, 11, 13, 15, 17, 19, 24, 26, 27, 31, 40, 45–48]. Stochasticity can be introduced in various components of these deterministic algorithms: stochastic approximation of operators, random selection of coordinates and operators in block-iterative implementations, and random relaxation parameters. To design and analyze such stochastic variants of existing models, we propose to transform Algorithm 4.2 into the following abstract stochastic scheme.

Algorithm 4.3 In the setting of Problem 4.1, let $\rho \in [2, +\infty[$, let $x_0 \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$, and iterate

$$
\begin{array}{l}
\text{for } \mathsf{n} = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\mathcal{X}_{\mathsf{n}} = \sigma(x_0, \ldots, x_{\mathsf{n}}) \\
\text{take } \{w_{\mathsf{n}}, w_{\mathsf{n}}^*, e_{\mathsf{n}}, e_{\mathsf{n}}^*\} \subset L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H}) \text{ such that } \big(w_{\mathsf{n}} + e_{\mathsf{n}}, w_{\mathsf{n}}^* + e_{\mathsf{n}}^*\big) \in \operatorname{gra} \mathsf{W} \ \mathsf{P}\text{-a.s.} \\
\text{take } \{q_{\mathsf{n}}, c_{\mathsf{n}}^*, f_{\mathsf{n}}^*\} \subset L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H}) \text{ such that } c_{\mathsf{n}}^* + f_{\mathsf{n}}^* = \mathsf{C} q_{\mathsf{n}} \ \mathsf{P}\text{-a.s.} \\
t_{\mathsf{n}}^* = w_{\mathsf{n}}^* + c_{\mathsf{n}}^* \\
\Delta_{\mathsf{n}} = \langle x_{\mathsf{n}} - w_{\mathsf{n}} \mid t_{\mathsf{n}}^* \rangle - (4\alpha)^{-1} \|w_{\mathsf{n}} - q_{\mathsf{n}}\|^2 \\
\theta_{\mathsf{n}} = \dfrac{1_{[t_{\mathsf{n}}^* \neq 0]} 1_{[\Delta_{\mathsf{n}} > 0]} \Delta_{\mathsf{n}}}{\|t_{\mathsf{n}}^*\|^2 + 1_{[t_{\mathsf{n}}^* = 0]}} \\
d_{\mathsf{n}} = \theta_{\mathsf{n}} t_{\mathsf{n}}^* \\
\text{take } \lambda_{\mathsf{n}} \in L^\infty(\Omega, \mathcal{F}, \mathsf{P};]0, \rho]) \\
x_{\mathsf{n}+1} = x_{\mathsf{n}} - \lambda_{\mathsf{n}} d_{\mathsf{n}}.
\end{array}
\right.
\end{array}
\tag{4.3}
$$

At iteration n of Algorithm 4.3, the variables e_{n}, e_{n}^*, and f_{n}^* model stochastic errors allowed in the activation of the operators W and C. Thus, the algorithm does not require an exact point in the graph of W but merely a stochastic approximation $(w_{\mathsf{n}}, w_{\mathsf{n}}^*)$ of such a point. Likewise, it does not require the exact evaluation of $\mathsf{C}q_{\mathsf{n}}$ but merely a stochastic approximation c_{n}^* of it. The broad reach of this algorithmic template stems from the flexibility it offers in choosing the triple $(w_{\mathsf{n}}, w_{\mathsf{n}}^*, q_{\mathsf{n}})$. Another notable new feature of (4.3) is the use of a random relaxation parameter λ_{n} which, furthermore, is not restricted to the usual interval $]0, 2[$.

Notation and preliminary results are presented in Section 4.2.2. The asymptotic behavior of Algorithm 4.3 is analyzed in Section 4.2.3, where we prove in particular weak almost sure convergence to a solution to Problem 4.1 under suitable assumptions. Just as the convergence analysis of Algorithm 4.2 provided a unifying framework to establish that of a wide array of classical and block-iterative methods in [18], those of Section 4.2.3 can be used to derive stochastic versions of these methods. Thus, in Section 4.2.4, we establish the almost-sure and L^2 weak convergence of the proximal point algorithm with stochastic approximations of the resolvents and random relaxations. To further illustrate the versatility of Algorithm 4.3, we consider in Section 4.2.5 a drastically different model, namely, a highly structured multivariate monotone inclusion problem involving a mix of set-valued, cocoercive, and Lipschitzian monotone operators, as well as linear operators, and various monotonicity-preserving operations among them. We design a stochastic version of the deterministic saddle projective splitting algorithm of [13] in which the blocks of variables and operators are now selected randomly over the course of the iterations, and the relaxations are random. Theorem 4.23 establishes for the first time the almost sure convergence of such a block-iterative algorithm. Likewise, Section 4.2.6 proposes a randomized version of the Kuhn–Tucker projective splitting method of [19] and analyzes its convergence as an instance of Algorithm 4.3.

4.2.2 Notation and preliminary results

4.2.2.1 General notation

We use sans-serif letters to denote deterministic variables and italicized serif letters to denote random variables. H is a separable real Hilbert space, with identity operator Id, power set 2^{H}, scalar product $\langle \cdot \mid \cdot \rangle$, and associated norm $\|\cdot\|$. The strong and weak convergence in H are denoted by the symbols $\to$ and $\rightharpoonup$, respectively. The sets of strong and weak sequential cluster points of a sequence $(\mathsf{x}_n)_{n\in\mathbb{N}}$ in H are denoted by $\mathfrak{S}(\mathsf{x}_n)_{n\in\mathbb{N}}$ and $\mathfrak{W}(\mathsf{x}_n)_{n\in\mathbb{N}}$, respectively. The reader is referred to [4] for background on convex analysis and fixed point theory, and to [34] for background on probability theory.

4.2.2.2 Operators

Let $\mathsf{M}\colon \mathsf{H} \to 2^{\mathsf{H}}$. The graph of M is $\operatorname{gra}\mathsf{M} = \big\{(\mathsf{x},\mathsf{x}^*) \in \mathsf{H}\times\mathsf{H} \mid \mathsf{x}^* \in \mathsf{M}\mathsf{x}\big\}$ and the set of zeros of M is $\operatorname{zer}\mathsf{M} = \big\{\mathsf{x} \in \mathsf{H} \mid 0 \in \mathsf{M}\mathsf{x}\big\}$. The inverse of M is the operator $\mathsf{M}^{-1}\colon \mathsf{H} \to 2^{\mathsf{H}}$ with graph $\operatorname{gra}\mathsf{M}^{-1} = \big\{(\mathsf{x}^*,\mathsf{x}) \in \mathsf{H}\times\mathsf{H} \mid \mathsf{x}^* \in \mathsf{M}\mathsf{x}\big\}$ and the resolvent of M is $\mathsf{J}_{\mathsf{M}} = (\mathrm{Id}+\mathsf{M})^{-1}$. We say that M is monotone if

$$\big(\forall(\mathsf{x},\mathsf{x}^*) \in \operatorname{gra}\mathsf{M}\big)\big(\forall(\mathsf{y},\mathsf{y}^*) \in \operatorname{gra}\mathsf{M}\big)\quad \langle \mathsf{x}-\mathsf{y} \mid \mathsf{x}^*-\mathsf{y}^*\rangle \geqslant 0, \tag{4.4}$$

and that it is maximally monotone if

$$\big(\forall(\mathsf{x},\mathsf{x}^*) \in \mathsf{H}\times\mathsf{H}\big)\quad \big[(\mathsf{x},\mathsf{x}^*) \in \operatorname{gra}\mathsf{M} \Leftrightarrow \big(\forall(\mathsf{y},\mathsf{y}^*) \in \operatorname{gra}\mathsf{M}\big)\ \ \langle \mathsf{x}-\mathsf{y} \mid \mathsf{x}^*-\mathsf{y}^*\rangle \geqslant 0\big]. \tag{4.5}$$

If M is maximally monotone, then J_{M} is a single-valued operator defined on H and which satisfies

$$\operatorname{Fix}\mathsf{J}_{\mathsf{M}} = \operatorname{zer}\mathsf{M} \quad\text{and}\quad (\forall\mathsf{x} \in \mathsf{H})(\forall\mathsf{y} \in \mathsf{H})\quad \|\mathsf{J}_{\mathsf{M}}\mathsf{x}-\mathsf{J}_{\mathsf{M}}\mathsf{y}\|^2 + \|(\mathrm{Id}-\mathsf{J}_{\mathsf{M}})\mathsf{x}-(\mathrm{Id}-\mathsf{J}_{\mathsf{M}})\mathsf{y}\|^2 \leqslant \|\mathsf{x}-\mathsf{y}\|^2. \tag{4.6}$$

Let $\beta \in \left]0,+\infty\right[$. Then M is β-strongly monotone if $\mathsf{M}-\beta\mathrm{Id}$ is monotone, i.e.,

$$\big(\forall(\mathsf{x},\mathsf{x}^*) \in \operatorname{gra}\mathsf{M}\big)\big(\forall(\mathsf{y},\mathsf{y}^*) \in \operatorname{gra}\mathsf{M}\big)\quad \langle \mathsf{x}-\mathsf{y} \mid \mathsf{x}^*-\mathsf{y}^*\rangle \geqslant \beta\|\mathsf{x}-\mathsf{y}\|^2. \tag{4.7}$$

The parallel sum of $\mathsf{B}\colon \mathsf{H} \to 2^{\mathsf{H}}$ and $\mathsf{D}\colon \mathsf{H} \to 2^{\mathsf{H}}$ is $\mathsf{B}\,\square\,\mathsf{D} = (\mathsf{B}^{-1}+\mathsf{D}^{-1})^{-1}$. An operator $\mathsf{C}\colon \mathsf{H} \to \mathsf{H}$ is cocoercive with constant $\alpha \in \left]0,+\infty\right[$ if

$$(\forall\mathsf{x} \in \mathsf{H})(\forall\mathsf{y} \in \mathsf{H})\quad \langle \mathsf{x}-\mathsf{y} \mid \mathsf{C}\mathsf{x}-\mathsf{C}\mathsf{y}\rangle \geqslant \alpha\|\mathsf{C}\mathsf{x}-\mathsf{C}\mathsf{y}\|^2. \tag{4.8}$$

We denote by $\Gamma_0(\mathsf{H})$ the class of lower semicontinuous convex functions $\mathsf{f}\colon \mathsf{H} \to \,]-\infty,+\infty]$ such that $\operatorname{dom}\mathsf{f} = \{\mathsf{x} \in \mathsf{H} \mid \mathsf{f}(\mathsf{x}) < +\infty\} \neq \varnothing$. The subdifferential of $\mathsf{f} \in \Gamma_0(\mathsf{H})$ is the maximally monotone operator $\partial\mathsf{f}\colon \mathsf{H} \to 2^{\mathsf{H}}\colon \mathsf{x} \mapsto \{\mathsf{x}^* \in \mathsf{H} \mid (\forall \mathsf{y} \in \mathsf{H})\ \langle \mathsf{y}-\mathsf{x} \mid \mathsf{x}^*\rangle + \mathsf{f}(\mathsf{x}) \leqslant \mathsf{f}(\mathsf{y})\}$ and the proximity operator of f is

$$\operatorname{prox}_{\mathsf{f}} = \mathsf{J}_{\partial\mathsf{f}}\colon \mathsf{H} \to \mathsf{H}\colon \mathsf{x} \mapsto \operatorname{argmin}_{\mathsf{z}\in\mathsf{H}}\left(\mathsf{f}(\mathsf{z}) + \frac{1}{2}\|\mathsf{x}-\mathsf{z}\|^2\right). \tag{4.9}$$

The infimal convolution of f and $\mathsf{h} \in \Gamma_0(\mathsf{H})$ is $\mathsf{f} \,\square\, \mathsf{h}\colon \mathsf{H} \to [-\infty,+\infty]\colon \mathsf{x} \mapsto \inf_{\mathsf{y}\in\mathsf{H}}(\mathsf{f}(\mathsf{y}) + \mathsf{h}(\mathsf{x}-\mathsf{y}))$.

4.2.2.3 Probabilistic setting

The underlying probability space $(\Omega, \mathcal{F}, \mathsf{P})$ is complete. Let $(\Xi, \mathcal{G})$ be a measurable space. A Ξ-valued random variable (random variable for short) is a measurable mapping $x\colon (\Omega, \mathcal{F}, \mathsf{P}) \to (\Xi, \mathcal{G})$. In particular, an H-valued random variable is a measurable mapping $x\colon (\Omega, \mathcal{F}, \mathsf{P}) \to (\mathsf{H}, \mathcal{B}_{\mathsf{H}})$, where $\mathcal{B}_{\mathsf{H}}$ denotes the Borel σ-algebra of H. Given $x\colon \Omega \to \Xi$ and $\mathsf{S} \in \mathcal{G}$, we set $[x \in \mathsf{S}] = \{\omega \in \Omega \mid x(\omega) \in \mathsf{S}\}$. Let $\mathsf{p} \in [1,+\infty[$ and let $\mathcal{X}$ be a sub σ-algebra of $\mathcal{F}$. Then $L^{\mathsf{p}}(\Omega, \mathcal{X}, \mathsf{P}; \mathsf{H})$ denotes the space of equivalence classes of P-a.s. equal H-valued random variables $x\colon (\Omega, \mathcal{X}, \mathsf{P}) \to (\mathsf{H}, \mathcal{B}_{\mathsf{H}})$ such that $\mathsf{E}\|x\|^{\mathsf{p}} < +\infty$. Endowed with the norm

$$\|\cdot\|_{L^{\mathsf{p}}(\Omega,\mathcal{X},\mathsf{P};\mathsf{H})}\colon x \mapsto \mathsf{E}^{1/\mathsf{p}}\|x\|^{\mathsf{p}} = \left(\int_\Omega \|x(\omega)\|^{\mathsf{p}}\mathsf{P}(d\omega)\right)^{1/\mathsf{p}}, \tag{4.10}$$

$L^{\mathsf{p}}(\Omega, \mathcal{X}, \mathsf{P}; \mathsf{H})$ is a real Banach space. Further,

$$(\forall \mathsf{S} \in \mathcal{B}_{\mathsf{H}}) \quad L^{\mathsf{p}}(\Omega, \mathcal{X}, \mathsf{P}; \mathsf{S}) = \left\{x \in L^{\mathsf{p}}(\Omega, \mathcal{X}, \mathsf{P}; \mathsf{H}) \;\middle|\; x \in \mathsf{S}\ \ \mathsf{P}\text{-a.s.}\right\}. \tag{4.11}$$

The σ-algebra generated by a family Φ of random variables is denoted by $\sigma(\Phi)$. Let $(x_{\mathsf{n}})_{\mathsf{n}\in\mathbb{N}}$ and x be H-valued random variables. We say that $(x_{\mathsf{n}})_{\mathsf{n}\in\mathbb{N}}$ converges in probability to x, denoted by $x_{\mathsf{n}} \xrightarrow{\mathsf{P}} x$, if $\|x_{\mathsf{n}} - x\|$ converges in probability to 0, i.e.,

$$(\forall \varepsilon \in \,]0,+\infty[) \quad \mathsf{P}\Big(\big[\|x_{\mathsf{n}} - x\| > \varepsilon\big]\Big) \to 0. \tag{4.12}$$

We say $\varphi\colon \Omega \times \mathsf{H} \to \mathbb{R}$ is a Carathéodory integrand if

$$\begin{cases} \text{for } \mathsf{P}\text{-almost every } \omega \in \Omega,\ \varphi(\omega,\cdot) \text{ is continuous;} \\ \text{for every } \mathsf{x} \in \mathsf{H},\ \varphi(\cdot,\mathsf{x}) \text{ is } \mathcal{F}\text{-measurable.} \end{cases} \tag{4.13}$$

We denote by $\mathfrak{C}(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ the class of Carathéodory integrands $\varphi\colon \Omega \times \mathsf{H} \to [0,+\infty[$.

4.2.2.4 Preliminary results

Our main results rest on several technical facts, which are presented below. The first two lemmas are direct consequences of the corresponding statements for $\mathbb{R}$-valued random variables;

see [44, Section 2.10].

Lemma 4.4 *Let $(x_n)_{n\in\mathbb{N}}$ and x be* H*-valued random variables and let* $p \in [1,+\infty[$ *be such that* $(x_n)_{n\in\mathbb{N}}$ *converges strongly in* $L^p(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ *to x. Then* $x_n \xrightarrow{\mathsf{P}} x$.

Lemma 4.5 *Let $(x_n)_{n\in\mathbb{N}}$ and x be* H*-valued random variables such that* $x_n \xrightarrow{\mathsf{P}} x$. *Then there exists a strictly increasing sequence* $(j_n)_{n\in\mathbb{N}}$ *in* $\mathbb{N}$ *such that* $(x_{j_n})_{n\in\mathbb{N}}$ *converges strongly* P-a.s. *to x.*

Lemma 4.6 *Let $(\xi_n)_{n\in\mathbb{N}}$, $(\Delta_n)_{n\in\mathbb{N}}$, and $(\chi_n)_{n\in\mathbb{N}}$ be sequences of $\mathbb{R}$-valued random variables such that*

$$\begin{cases}\overline{\lim}\,\Delta_n \leqslant 0 \text{ P-a.s.};\\ \chi_n \xrightarrow{\mathsf{P}} 0;\\ (\forall n\in\mathbb{N})\ \ \xi_n \geqslant 0 \text{ P-a.s. } \textit{and } \ \xi_n+\chi_n \leqslant \Delta_n \text{ P-a.s.}\end{cases} \tag{4.14}$$

Then $\xi_n \xrightarrow{\mathsf{P}} 0$.

Proof. Let $\varepsilon \in\,]0,+\infty[$ and $n\in\mathbb{N}$. Let $\omega\in\Omega$ and suppose that $\xi_n(\omega) > \varepsilon$. Then there are two cases:

- $\chi_n(\omega) < -\varepsilon/2$.
- $\chi_n(\omega) \geqslant -\varepsilon/2$, in which case $\varepsilon/2 = \varepsilon-\varepsilon/2 < \xi_n(\omega)+\chi_n(\omega) \leqslant \Delta_n(\omega)$. Therefore,

$$[\xi_n > \varepsilon] \subset [\chi_n < -\varepsilon/2] \cup [\Delta_n > \varepsilon/2]. \tag{4.15}$$

Note that $\mathsf{P}([\chi_n < -\varepsilon/2]) \to 0$ since $\chi_n \xrightarrow{\mathsf{P}} 0$. On the other hand, since $\overline{\lim}\,\Delta_n \leqslant 0$ P-a.s., we have

$$\begin{aligned}\overline{\lim}\,\mathsf{P}([\Delta_n > \varepsilon/2]) &\leqslant \mathsf{P}\Big(\overline{\lim}\,[\Delta_n > \varepsilon/2]\Big)\\ &= \mathsf{P}\Big(\big\{\omega\in\Omega \mid (\forall n\in\mathbb{N})(\exists\, k\in\{n,n+1,\ldots\})\ \Delta_k(\omega) > \varepsilon/2\big\}\Big)\\ &= 0.\end{aligned} \tag{4.16}$$

Altogether, $\mathsf{P}([|\xi_n| > \varepsilon]) = \mathsf{P}([\xi_n > \varepsilon]) \leqslant \mathsf{P}([\chi_n < -\varepsilon/2]) + \mathsf{P}([\Delta_n > \varepsilon/2]) \to 0$ and we conclude that $\xi_n \xrightarrow{\mathsf{P}} 0$. ☐

Lemma 4.7 *Let $x\in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ and let $\mathsf{T}\colon \mathsf{H}\to\mathsf{H}$ be Lipschitzian. Then $\mathsf{T}x \in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$.*

Proof. Let $\beta\in\,]0,+\infty[$ be the Lipschitz constant of T. Since T is continuous, the mapping $\omega\mapsto(\mathsf{T}\circ x)(\omega) = \mathsf{T}x(\omega)$ is measurable. Furthermore,

$$\frac{1}{2}\mathsf{E}\|\mathsf{T}x\|^2 \leqslant \mathsf{E}\|\mathsf{T}x-\mathsf{T}0\|^2 + \mathsf{E}\|\mathsf{T}0\|^2 \leqslant \beta\mathsf{E}\|x-0\|^2 + \mathsf{E}\|\mathsf{T}0\|^2 = \beta\mathsf{E}\|x\|^2 + \|\mathsf{T}0\|^2 < +\infty, \tag{4.17}$$

which confirms that $\mathsf{T}x\in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$. ☐

Lemma 4.8 *Let* $(x_n)_{n\in\mathbb{N}}$ *be a sequence in* $L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$*, let* $\mathsf{m}\in\mathbb{N}$*, and let* $\vartheta(\mathsf{m})$ *be a* $\{0,\ldots,\mathsf{m}\}$*-valued random variable. Then the function* $x_{\vartheta(\mathsf{m})}\colon\omega\mapsto x_{\vartheta(\mathsf{m})(\omega)}(\omega)$ *is in* $L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$.

Proof. We note that

$$x_{\vartheta(\mathsf{m})}=\sum_{\mathsf{j}=0}^{\mathsf{m}}1_{[\vartheta(\mathsf{m})=\mathsf{j}]}x_{\mathsf{j}}\quad\mathsf{P}\text{-a.s.},\tag{4.18}$$

which shows that $x_{\vartheta(\mathsf{m})}$ is measurable, as $(\Omega,\mathcal{F},\mathsf{P})$ is complete, and that

$$\mathsf{E}\|x_{\vartheta(\mathsf{m})}\|^2\leqslant\mathsf{m}\max_{1\leqslant\mathsf{j}\leqslant\mathsf{m}}\mathsf{E}\|x_{\mathsf{j}}\|^2<+\infty.\tag{4.19}$$

Thus, $x_{\vartheta(\mathsf{m})}\in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$. ☐

The following theorem is a straightforward consequence of [21, Theorems 3.2 and 3.6].

Theorem 4.9 *Let* Z *be a nonempty closed convex subset of* H*, let* $x_0\in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$*, and let* $\rho\in[2,+\infty[$*. Iterate*

$$\begin{array}{l}
\text{for } \mathsf{n}=0,1,\ldots\\
\left\lfloor
\begin{array}{l}
\mathcal{X}_{\mathsf{n}}=\sigma(x_0,\ldots,x_{\mathsf{n}})\\
t_{\mathsf{n}}^*\in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})\text{ and }\eta_{\mathsf{n}}\in L^1(\Omega,\mathcal{F},\mathsf{P};\mathbb{R})\text{ satisfy}\\
\quad\left\{
\begin{array}{l}
\dfrac{1_{[t_{\mathsf{n}}^*\neq 0]}1_{[\langle x_{\mathsf{n}}\mid t_{\mathsf{n}}^*\rangle>\eta_{\mathsf{n}}]}\eta_{\mathsf{n}}}{\|t_{\mathsf{n}}^*\|+1_{[t_{\mathsf{n}}^*=0]}}\in L^2(\Omega,\mathcal{F},\mathsf{P};\mathbb{R});\\
\theta_{\mathsf{n}}=\dfrac{1_{[t_{\mathsf{n}}^*\neq 0]}1_{[\langle x_{\mathsf{n}}\mid t_{\mathsf{n}}^*\rangle>\eta_{\mathsf{n}}]}\big(\langle x_{\mathsf{n}}\mid t_{\mathsf{n}}^*\rangle-\eta_{\mathsf{n}}\big)}{\|t_{\mathsf{n}}^*\|^2+1_{[t_{\mathsf{n}}^*=0]}};\\
(\forall\mathsf{z}\in\mathsf{Z})\quad\langle\mathsf{z}\mid\mathsf{E}(\theta_{\mathsf{n}}t_{\mathsf{n}}^*\mid\mathcal{X}_{\mathsf{n}})\rangle\leqslant\mathsf{E}(\theta_{\mathsf{n}}\eta_{\mathsf{n}}\mid\mathcal{X}_{\mathsf{n}})+\varepsilon_{\mathsf{n}}(\cdot,\mathsf{z})\ \mathsf{P}\text{-a.s.},\\
\hfill\text{where }\varepsilon_{\mathsf{n}}\in\mathfrak{C}(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})
\end{array}
\right.\\
d_{\mathsf{n}}=\theta_{\mathsf{n}}t_{\mathsf{n}}^*\\
\lambda_{\mathsf{n}}\in L^\infty(\Omega,\mathcal{F},\mathsf{P};\,]0,\rho])\\
x_{\mathsf{n}+1}=x_{\mathsf{n}}-\lambda_{\mathsf{n}}d_{\mathsf{n}}.
\end{array}
\right.
\end{array}\tag{4.20}$$

Suppose that, for every $\mathsf{n}\in\mathbb{N}$*,* λ_{n} *is independent of* $\sigma(\{x_0,\ldots,x_{\mathsf{n}},d_{\mathsf{n}}\})$*, and* $\mathsf{E}(\lambda_{\mathsf{n}}(2-\lambda_{\mathsf{n}}))\geqslant 0$*. Then the following hold:*

(i) $(x_{\mathsf{n}})_{\mathsf{n}\in\mathbb{N}}$ *is a well-defined sequence in* $L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$.

(ii) *Suppose that, for every* $\mathsf{z}\in\mathsf{Z}$*,* $\sum_{\mathsf{n}\in\mathbb{N}}\mathsf{E}\varepsilon_{\mathsf{n}}(\cdot,\mathsf{z})\mathsf{E}\lambda_{\mathsf{n}}<+\infty$*. Then the following are satisfied:*

  (a) $(\|x_{\mathsf{n}}\|)_{\mathsf{n}\in\mathbb{N}}$ *is bounded* P-a.s. *and* $(\mathsf{E}\|x_{\mathsf{n}}\|^2)_{\mathsf{n}\in\mathbb{N}}$ *is bounded.*

  (b) $\sum_{\mathsf{n}\in\mathbb{N}}\mathsf{E}(\lambda_{\mathsf{n}}(2-\lambda_{\mathsf{n}}))\mathsf{E}\|d_{\mathsf{n}}\|^2<+\infty$.

  (c) *Suppose that* $\inf_{\mathsf{n}\in\mathbb{N}}\mathsf{E}(\lambda_{\mathsf{n}}(2-\lambda_{\mathsf{n}}))>0$*. Then* $\sum_{\mathsf{n}\in\mathbb{N}}\mathsf{E}\|x_{\mathsf{n}+1}-x_{\mathsf{n}}\|^2<+\infty$.

  (d) *Suppose that* $\mathfrak{W}(x_{\mathsf{n}})_{\mathsf{n}\in\mathbb{N}}\subset\mathsf{Z}$ P-a.s. *Then* $(x_{\mathsf{n}})_{\mathsf{n}\in\mathbb{N}}$ *converges weakly* P-a.s. *and weakly in* $L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ *to a random variable* $x\in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{Z})$.

  (e) *Suppose that* $\mathfrak{S}(x_{\mathsf{n}})_{\mathsf{n}\in\mathbb{N}}\cap\mathsf{Z}\neq\varnothing$ P-a.s. *Then* $(x_{\mathsf{n}})_{\mathsf{n}\in\mathbb{N}}$ *converges strongly* P-a.s. *and* *strongly in $L^1(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ to a random variable $x \in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{Z})$. Additionally, $(x_n)_{n\in\mathbb{N}}$ converges weakly in $L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ to x.*

Lemma 4.10 ([13, Lemma A.2]) *Let $\alpha \in [0,+\infty[$, let $\mathsf{A}\colon \mathsf{H} \to \mathsf{H}$ be α-Lipschitzian, let $\sigma \in]0,+\infty[$, and let $\gamma \in]0,1/(\alpha+\sigma)]$. Then $\gamma^{-1}\mathrm{Id} - \mathsf{A}$ is σ-strongly monotone.*

4.2.3 Convergence analysis

This section is dedicated to establishing the weak convergence to solutions to Problem 4.1, in the almost sure and $L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ modes, of the sequence $(x_n)_{n\in\mathbb{N}}$ generated by the stochastic Algorithm 4.3.

Theorem 4.11 *In the context of Problem 4.1, let $(x_n)_{n\in\mathbb{N}}$ be the sequence generated by Algorithm 4.3. For every $n \in \mathbb{N}$ and every $z \in \mathsf{Z}$, set*

$$\varepsilon_n(\cdot,z) = \max\Big\{0, \mathsf{E}\Big(\theta_n\Big(\big\langle w_n - z \,\big|\, e_n^* + f_n^*\big\rangle\Big) + \big\langle e_n \,\big|\, w_n^* + \mathsf{C}z\big\rangle + \big\langle e_n \,\big|\, e_n^*\big\rangle \,\Big|\, \mathcal{X}_n\Big)\Big\}, \tag{4.21}$$

and suppose that λ_n is independent of $\sigma(\{x_0,\ldots,x_n,d_n\})$ and that $\mathsf{E}(\lambda_n(2-\lambda_n)) \geqslant 0$. Then the following hold:

(i) *Let $n \in \mathbb{N}$ and $z \in \mathsf{Z}$. Then*

$$\big\langle z \,\big|\, \mathsf{E}(\theta_n t_n^* \,|\, \mathcal{X}_n)\big\rangle \leqslant \mathsf{E}\Big(\theta_n \langle w_n \,|\, t_n^*\rangle \,\Big|\, \mathcal{X}_n\Big) + \frac{1}{4\alpha}\mathsf{E}\big(\theta_n\|w_n - q_n\|^2 \,\big|\, \mathcal{X}_n\big) + \varepsilon_n(\cdot,z) \ \ \mathsf{P}\text{-a.s.} \tag{4.22}$$

(ii) *$(x_n)_{n\in\mathbb{N}}$ lies in $L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$.*

(iii) *Suppose that, for every $z \in \mathsf{Z}$, $\sum_{n\in\mathbb{N}} \mathsf{E}\varepsilon_n(\cdot,z)\mathsf{E}\lambda_n < +\infty$. Then the following are satisfied:*

  (a) *$(\|x_n\|)_{n\in\mathbb{N}}$ is bounded P-a.s. and $(\mathsf{E}\|x_n\|^2)_{n\in\mathbb{N}}$ is bounded.*

  (b) *$\sum_{n\in\mathbb{N}} \mathsf{E}(\lambda_n(2-\lambda_n))\mathsf{E}\|d_n\|^2 < +\infty$.*

  (c) *Suppose that $\inf_{n\in\mathbb{N}} \mathsf{E}(\lambda_n(2-\lambda_n)) > 0$. Then $\sum_{n\in\mathbb{N}} \mathsf{E}\|x_{n+1} - x_n\|^2 < +\infty$.*

  (d) *Suppose that $\inf_{n\in\mathbb{N}} \lambda_n > 0$ P-a.s. and that $(t_n^*)_{n\in\mathbb{N}}$ is bounded P-a.s. Then $\overline{\lim}\,\Delta_n \leqslant 0$ P-a.s.*

  (e) *Suppose that $x_n - w_n - e_n \rightharpoonup 0$ P-a.s., $w_n + e_n - q_n \to 0$ P-a.s., and $w_n^* + e_n^* + \mathsf{C}q_n \to 0$ P-a.s. Then $(x_n)_{n\in\mathbb{N}}$ converges weakly P-a.s. and weakly in $L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ to a Z-valued random variable.*

  (f) *Suppose that $\dim \mathsf{H} < +\infty$, $x_n - w_n - e_n \xrightarrow{\mathsf{P}} 0$, $w_n + e_n - q_n \xrightarrow{\mathsf{P}} 0$, and $w_n^* + e_n^* + \mathsf{C}q_n \xrightarrow{\mathsf{P}} 0$. Then $(x_n)_{n\in\mathbb{N}}$ converges P-a.s. and in $L^1(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ to a Z-valued random variable.*

Proof. (i): Note that $(z,-\mathsf{C}z) \in \operatorname{gra}\mathsf{W}$. Hence, (4.3) and the monotonicity of W yield

$$\begin{aligned}&\big\langle z - w_n - e_n \,\big|\, w_n^* + e_n^* + c_n^*\big\rangle\\ &\quad= \langle z - w_n - e_n \,|\, w_n^* + e_n^* + \mathsf{C}q_n\rangle - \langle z - w_n - e_n \,|\, f_n^*\rangle\end{aligned}$$

$$
\begin{aligned}
&= \langle z - w_n - e_n \mid w_n^* + e_n^* + Cz\rangle + \langle z - w_n - e_n \mid Cq_n - Cz\rangle - \langle z - w_n - e_n \mid f_n^*\rangle \\
&\leqslant \langle z - w_n - e_n \mid Cq_n - Cz\rangle - \langle z - w_n - e_n \mid f_n^*\rangle \\
&= -\langle z - q_n \mid Cz - Cq_n\rangle + \langle w_n - q_n \mid Cz - Cq_n\rangle + \langle e_n \mid Cz - Cq_n\rangle - \langle z - w_n - e_n \mid f_n^*\rangle \\
&\leqslant -\alpha\|Cz - Cq_n\|^2 + \|w_n - q_n\|\|Cz - Cq_n\| + \langle e_n \mid Cz - Cq_n\rangle - \langle z - w_n - e_n \mid f_n^*\rangle \\
&= \frac{\|w_n - q_n\|^2}{4\alpha} - \Big|(2\sqrt{\alpha})^{-1}\|w_n - q_n\| - \sqrt{\alpha}\|Cz - Cq_n\|\Big|^2 \\
&\quad + \langle e_n \mid Cz - Cq_n\rangle - \langle z - w_n - e_n \mid f_n^*\rangle \\
&\leqslant \frac{\|w_n - q_n\|^2}{4\alpha} + \langle w_n - z \mid f_n^*\rangle + \langle e_n \mid Cz - Cq_n\rangle + \langle e_n \mid f_n^*\rangle \ \ \text{P-a.s.}
\end{aligned}
\tag{4.23}
$$

Therefore, since $t_n^* = w_n^* + c_n^*$,

$$
\big\langle z \,\big|\, t_n^*\big\rangle \leqslant \big\langle w_n \,\big|\, t_n^*\big\rangle + \frac{\|w_n - q_n\|^2}{4\alpha} + \big\langle w_n - z \,\big|\, e_n^* + f_n^*\big\rangle + \big\langle e_n \,\big|\, w_n^* + Cz\big\rangle + \big\langle e_n \,\big|\, e_n^*\big\rangle \ \ \text{P-a.s.} \tag{4.24}
$$

On the other hand, because $\theta_n \geqslant 0$ P-a.s., it follows from scaling by θ_n and taking the conditional expectation with respect to $\mathcal{X}_n$ in (4.24) that (4.22) holds.

(ii): Let $n \in \mathbb{N}$ and set $\eta_n = \langle w_n \mid t_n^*\rangle + (4\alpha)^{-1}\|w_n - q_n\|^2$. Then $\eta_n \in L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathbb{R})$ and $\varepsilon_n \in \mathfrak{C}(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. Furthermore, by the Cauchy–Schwarz inequality,

$$
\begin{aligned}
\frac{1}{2}\mathsf{E}\left|\frac{1_{[t_n^*\neq 0]}1_{[\langle x_n \mid t_n^*\rangle > \eta_n]}\eta_n}{\|t_n^*\| + 1_{[t_n^*=0]}}\right|^2 &\leqslant \mathsf{E}\left|\frac{1_{[t_n^*\neq 0]}1_{[\langle x_n \mid t_n^*\rangle > \eta_n]}\big(\langle x_n \mid t_n^*\rangle - \eta_n\big)}{\|t_n^*\| + 1_{[t_n^*=0]}}\right|^2 + \mathsf{E}\left|\frac{1_{[t_n^*\neq 0]}1_{[\langle x_n \mid t_n^*\rangle > \eta_n]}\langle x_n \mid t_n^*\rangle}{\|t_n^*\| + 1_{[t_n^*=0]}}\right|^2 \\
&\leqslant \mathsf{E}\left|\frac{1_{[t_n^*\neq 0]}1_{[\langle x_n \mid t_n^*\rangle > \eta_n]}\big(\langle x_n \mid t_n^*\rangle - \eta_n\big)}{\|t_n^*\| + 1_{[t_n^*=0]}}\right|^2 + \mathsf{E}\|x_n\|^2 \\
&= \mathsf{E}\left|\frac{1_{[t_n^*\neq 0]}1_{[\langle x_n \mid t_n^*\rangle > \eta_n]}\big(\langle x_n - w_n \mid t_n^*\rangle - (4\alpha)^{-1}\|w_n - q_n\|^2\big)}{\|t_n^*\| + 1_{[t_n^*=0]}}\right|^2 + \mathsf{E}\|x_n\|^2 \\
&\leqslant \mathsf{E}\left|\frac{1_{[t_n^*\neq 0]}1_{[\langle x_n \mid t_n^*\rangle > \eta_n]}\langle x_n - w_n \mid t_n^*\rangle}{\|t_n^*\| + 1_{[t_n^*=0]}}\right|^2 + \mathsf{E}\|x_n\|^2 \\
&\leqslant \mathsf{E}\|x_n - w_n\|^2 + \mathsf{E}\|x_n\|^2 \\
&< +\infty.
\end{aligned}
\tag{4.25}
$$

Altogether, in view of (i), we deduce that (4.3) is a realization of (4.20). Hence, the claim follows from Theorem 4.9(i).

(iii)(a)–(iii)(c): These follow from Theorem 4.9(ii)(a)–(ii)(c).

(iii)(d): Since $\inf_{n\in\mathbb{N}} \lambda_n > 0$ P-a.s., we proceed, for P-almost every $\omega \in \Omega$, as in the proof of [13, Proposition 3(iii)] to get the result using (iii)(c).

(iii)(e): In view of (iii)(a), we fix $\Omega' \in \mathcal{F}$ such that

$$\mathsf{P}(\Omega') = 1 \quad \text{and} \quad (\forall \omega \in \Omega') \quad \begin{cases} x_n(\omega) - w_n(\omega) - e_n(\omega) \rightharpoonup 0; \\ w_n(\omega) + e_n(\omega) - q_n(\omega) \to 0; \\ w_n^*(\omega) + e_n^*(\omega) + \mathsf{C}q_n(\omega) \to 0; \\ (\|x_n(\omega)\|)_{n\in\mathbb{N}} \text{ is bounded.} \end{cases} \tag{4.26}$$

Now let $\omega \in \Omega'$ and $\mathsf{x} \in \mathfrak{W}(x_n(\omega))_{n\in\mathbb{N}}$. Then there exists a strictly increasing sequence in $\mathbb{N}$, say $(k_n)_{n\in\mathbb{N}}$, such that $x_{k_n}(\omega) \rightharpoonup \mathsf{x}$. Furthermore,

$$w_{k_n}(\omega) + e_{k_n}(\omega) = x_{k_n}(\omega) - \big(x_{k_n}(\omega) - w_{k_n}(\omega) - e_{k_n}(\omega)\big) \rightharpoonup \mathsf{x} \tag{4.27}$$

and, since C is α^{-1}-Lipschitzian,

$$\begin{aligned} &\big\|w_{k_n}^*(\omega) + e_{k_n}^*(\omega) + \mathsf{C}\big(w_{k_n}(\omega) + e_{k_n}(\omega)\big)\big\| \\ &\quad\leqslant \big\|w_{k_n}^*(\omega) + e_{k_n}^*(\omega) + \mathsf{C}q_{k_n}(\omega)\big\| + \big\|\mathsf{C}\big(w_{k_n}(\omega) + e_{k_n}(\omega)\big) - \mathsf{C}q_{k_n}(\omega)\big\| \\ &\quad\leqslant \big\|w_{k_n}^*(\omega) + e_{k_n}^*(\omega) + \mathsf{C}q_{k_n}(\omega)\big\| + \frac{\big\|w_{k_n}(\omega) + e_{k_n}(\omega) - q_{k_n}(\omega)\big\|}{\alpha} \\ &\quad\to 0. \end{aligned} \tag{4.28}$$

On the other hand, (4.3) yields

$$(\forall n \in \mathbb{N}) \quad \Big(w_{k_n}(\omega) + e_{k_n}(\omega), w_{k_n}^*(\omega) + e_{k_n}^*(\omega) + \mathsf{C}\big(w_{k_n}(\omega) + e_{k_n}(\omega)\big)\Big) \in \operatorname{gra}(\mathsf{W} + \mathsf{C}). \tag{4.29}$$

Since, by [4, Corollary 25.5(i)], $\mathsf{W} + \mathsf{C}$ is maximally monotone, (4.27), (4.28), (4.29), and [4, Proposition 20.38(ii)] imply that $\mathsf{x} \in \mathsf{Z}$. Since x is arbitrarily chosen in $\mathfrak{W}(x_n(\omega))_{n\in\mathbb{N}}$, we deduce that $\mathfrak{W}(x_n(\omega))_{n\in\mathbb{N}} \subset \mathsf{Z}$ and, since $\mathsf{P}(\Omega') = 1$, that $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \mathsf{Z}$ P-a.s. Therefore, it follows from Theorems 4.9(ii)(d) that $(x_n)_{n\in\mathbb{N}}$ converges weakly P-a.s. and weakly in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ to a Z-valued random variable.

(iii)(f): Lemma 4.5 guarantees the existence of a strictly increasing sequence in $\mathbb{N}$, say $(l_n)_{n\in\mathbb{N}}$, such that $x_{l_n} - w_{l_n} - e_{l_n} \to 0$ P-a.s., $w_{l_n} + e_{l_n} - q_{l_n} \to 0$ P-a.s., and $w_{l_n}^* + e_{l_n}^* + \mathsf{C}q_{l_n} \to 0$ P-a.s. Additionally, it follows from (iii)(a) that $(\|x_{l_n}\|)_{n\in\mathbb{N}}$ is bounded P-a.s. Let $\Omega' \in \mathcal{F}$ be such that

$$\mathsf{P}(\Omega') = 1 \quad \text{and} \quad (\forall \omega \in \Omega') \quad \begin{cases} x_{l_n}(\omega) - w_{l_n}(\omega) - e_{l_n}(\omega) \to 0; \\ w_{l_n}(\omega) + e_{l_n}(\omega) - q_{l_n}(\omega) \to 0; \\ w_{l_n}^*(\omega) + e_{l_n}^*(\omega) + \mathsf{C}q_{l_n}(\omega) \to 0; \\ (\|x_{l_n}(\omega)\|)_{n\in\mathbb{N}} \text{ is bounded.} \end{cases} \tag{4.30}$$

Let $\omega \in \Omega'$. We derive from (4.30) and the fact that H is finite-dimensional that there exists

$x \in H$ and a further subsequence $(k_{l_n})_{n\in\mathbb{N}}$ such that $x_{k_{l_n}}(\omega) \to x$,

$$w_{k_{l_n}}(\omega) + e_{k_{l_n}}(\omega) = x_{k_{l_n}}(\omega) - \big(x_{k_{l_n}}(\omega) - w_{k_{l_n}}(\omega) - e_{k_{l_n}}(\omega)\big) \to x, \tag{4.31}$$

and, as in (4.28),

$$w^*_{k_{l_n}}(\omega) + e^*_{k_{l_n}}(\omega) + C\big(w_{k_{l_n}}(\omega) + e_{k_{l_n}}(\omega)\big) \to 0. \tag{4.32}$$

However, as in (4.29),

$$(\forall n \in \mathbb{N}) \quad \Big(w_{k_{l_n}}(\omega) + e_{k_{l_n}}(\omega), w^*_{k_{l_n}}(\omega) + e^*_{k_{l_n}}(\omega) + C\big(w_{k_{l_n}}(\omega) + e_{k_{l_n}}(\omega)\big)\Big) \in \operatorname{gra}(W + C), \tag{4.33}$$

and the maximal monotonicity of $W + C$ yields $x \in Z$. Thus, $(x_n(\omega))_{n\in\mathbb{N}}$ has a cluster point in Z and we conclude that $\mathfrak{S}(x_n)_{n\in\mathbb{N}} \cap Z \neq \varnothing$ P-a.s. Therefore, it follows from Theorem 4.9(ii)(e) that $(x_n)_{n\in\mathbb{N}}$ converges P-a.s. and in $L^1(\Omega, \mathcal{F}, P; H)$ to a Z-valued random variable. □

Remark 4.12 The random relaxations parameters $(\lambda_n)_{n\in\mathbb{N}}$ satisfy $\inf_{n\in\mathbb{N}} E(\lambda_n(2-\lambda_n)) \geqslant 0$. When the relaxation parameters are deterministic, this condition imposes that, for every $n \in \mathbb{N}$, $\lambda_n \in \,]0, 2[$, which is the standard range found in deterministic methods in the literature [13, 18, 29, 32]. However, Theorem 4.11 allows for the use of so-called super relaxation parameters [21] which may exceed 2 by satisfying

$$\inf_{n\in\mathbb{N}} E\big(\lambda_n(2 - \lambda_n)\big) > 0 \quad \text{and} \quad \inf_{n\in\mathbb{N}} P([\lambda_n > 2]) > 0. \tag{4.34}$$

Note that the use of super relaxation parameters leads to novel results and faster convergence; see [21, Section 6] for examples of super relaxation strategies.

4.2.4 Stochastic proximal point algorithm

The proximal point algorithm is a classical method for finding a zero of a maximal monotone operator $A\colon H \to 2^H$ [5, 35, 36, 41]. In this section, we propose a stochastic version of it which involves stochastic approximations of the resolvents together with random relaxations.

Theorem 4.13 *Let $A\colon H \to 2^H$ be a maximally monotone operator such that $\operatorname{zer} A \neq \varnothing$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $]0, +\infty[$, and let $x_0 \in L^2(\Omega, \mathcal{F}, P; H)$. Iterate*

$$\begin{array}{l} \text{for } n = 0, 1, \dots \\ \left\lfloor \begin{array}{l} \text{take } e_n \in L^2(\Omega, \mathcal{F}, P; H) \text{ and } \lambda_n \in L^\infty(\Omega, \mathcal{F}, P; \,]0, 2[) \\ x_{n+1} = x_n + \lambda_n\big(J_{\gamma_n A} x_n - e_n - x_n\big). \end{array} \right. \end{array} \tag{4.35}$$

Suppose that, for every $n \in \mathbb{N}$, λ_n is independent of $\sigma(x_0, \dots, x_n, e_n)$, and that one of the following holds:

(i) $\sum_{n\in\mathbb{N}} E(\lambda_n(2 - \lambda_n)) = +\infty$, $\sum_{n\in\mathbb{N}} \sqrt{E|\lambda_n|^2 E\|e_n\|^2} < +\infty$, $(E\|e_n\|^2)_{n\in\mathbb{N}}$ *is bounded, and* $(\forall n \in \mathbb{N})$ $\gamma_n = 1$.

(ii) $\inf_{n\in\mathbb{N}} \mathsf{E}(\lambda_n(2-\lambda_n)) > 0$, $\inf_{n\in\mathbb{N}} \gamma_n > 0$, *and* $\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}\|e_n\|^2} < +\infty$.

(iii) $\sum_{n\in\mathbb{N}} \gamma_n^2 = +\infty$, $\sum_{n\in\mathbb{N}} \sqrt{\mathsf{E}\|e_n\|^2} < +\infty$, *and* $(\forall n \in \mathbb{N})$ $\lambda_n = 1$ P-a.s.

Then $(x_n)_{n\in\mathbb{N}}$ *converges weakly* P-a.s. *and weakly in* $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ *to a* $(\operatorname{zer} \mathsf{A})$*-valued random variable.*

Proof. We apply Theorem 4.11 with $\mathsf{W} = \mathsf{A}$, $\mathsf{C} = 0$ (hence $\mathsf{Z} = \operatorname{zer} \mathsf{A}$) and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} w_n = \mathsf{J}_{\gamma_n \mathsf{A}} x_n - e_n; \\ w_n^* = \gamma_n^{-1}(x_n - w_n); \\ q_n = w_n; \\ c_n^* = f_n^* = 0; \\ e_n^* = -\gamma_n^{-1} e_n. \end{cases} \tag{4.36}$$

In this setting, it follows from [4, Proposition 23.22] that

$$(\forall n \in \mathbb{N}) \quad \big(w_n + e_n, w_n^* + e_n^*\big) = \big(\mathsf{J}_{\gamma_n \mathsf{A}} x_n, \gamma_n^{-1}(x_n - \mathsf{J}_{\gamma_n \mathsf{A}} x_n)\big) \in \operatorname{gra} \mathsf{A} \ \text{ P-a.s.} \tag{4.37}$$

and that algorithm (4.35) is an instantiation of Algorithm 4.3 with

$$(\forall n \in \mathbb{N}) \quad t_n^* = \gamma_n^{-1}(x_n - w_n) \quad \text{and} \quad \theta_n = \gamma_n. \tag{4.38}$$

We therefore deduce from Theorem 4.11(ii) that the sequence $(x_n)_{n\in\mathbb{N}}$ lies in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. Next, let us define a family of auxiliary sequences as follows. For every $k \in \mathbb{N}$, set

$$y_{0,k} = x_k \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad y_{n+1,k} = y_{n,k} + \lambda_{n+k}\big(\mathsf{J}_{\gamma_{n+k}\mathsf{A}} y_{n,k} - y_{n,k}\big). \tag{4.39}$$

Let $k \in \mathbb{N}$. Then, as above, $(y_{n,k})_{n\in\mathbb{N}}$ is a sequence generated by an instantiation of Algorithm 4.3 now initialized at x_k with, for every $n \in \mathbb{N}$, $e_n = 0$ and $q_n = \mathsf{J}_{\gamma_{n+k}\mathsf{A}} y_{n,k}$. Consequently, Theorem 4.11(iii)(a) asserts that $(\|y_{n,k}\|)_{n\in\mathbb{N}}$ is bounded P-a.s. and that $(\mathsf{E}\|y_{n,k}\|^2)_{n\in\mathbb{N}}$ is bounded. Additionally, we deduce from Theorem 4.11(iii)(b) that

$$\sum_{n\in\mathbb{N}} \mathsf{E}\big(\lambda_n(2-\lambda_n)\big)\mathsf{E}\big\|y_{n,k} - \mathsf{J}_{\gamma_{n+k}\mathsf{A}} y_{n,k}\big\|^2 < +\infty. \tag{4.40}$$

Next, let us show that, under any of scenarios (i)–(iii),

$$\|y_{n,k} - \mathsf{J}_{\gamma_{n+k}\mathsf{A}} y_{n,k}\| \to 0 \ \text{ P-a.s. as } \ n \to +\infty. \tag{4.41}$$

- Suppose that (i) holds. Then we deduce from (4.40) that $\underline{\lim}\, \mathsf{E}\|y_{n,k} - \mathsf{J}_{\mathsf{A}} y_{n,k}\|^2 = 0$. In turn, Fatou's lemma yields $\underline{\lim} \|y_{n,k} - \mathsf{J}_{\mathsf{A}} y_{n,k}\| = 0$ P-a.s. Now set $\mathsf{T} = 2\mathsf{J}_{\mathsf{A}} - \mathrm{Id}$ and recall that it is nonexpansive [4, Corollary 23.11(ii)]. Therefore,

$$(\forall n \in \mathbb{N}) \quad 2\big\|y_{n+1,k} - J_A y_{n+1,k}\big\| = \big\|Ty_{n+1,k} - y_{n+1,k}\big\|$$
$$= \big\|Ty_{n+1,k} - Ty_{n,k} + (1-\lambda_n/2)(Ty_{n,k} - y_{n,k})\big\|$$
$$\leqslant \|y_{n+1,k} - y_{n,k}\| + (1-\lambda_n/2)\big\|Ty_{n,k} - y_{n,k}\big\|$$
$$= (\lambda_n/2)\big\|Ty_{n,k} - y_{n,k}\big\| + (1-\lambda_n/2)\big\|Ty_{n,k} - y_{n,k}\big\|$$
$$= 2\big\|y_{n,k} - J_A y_{n,k}\big\| \ \text{ P-a.s.,} \tag{4.42}$$

which shows that $(\|y_{n,k} - J_A y_{n,k}\|)_{n\in\mathbb{N}}$ decreases P-a.s. Hence, $\|y_{n,k} - J_A y_{n,k}\| \to 0$ P-a.s. as $n \to +\infty$.

- Suppose that (ii) or (iii) holds. Then it follows from (4.40) that

$$\mathsf{E}\sum_{n\in\mathbb{N}}\big\|y_{n,k} - J_{\gamma_{n+k}A}y_{n,k}\big\|^2 = \sum_{n\in\mathbb{N}}\mathsf{E}\big\|y_{n,k} - J_{\gamma_{n+k}A}y_{n,k}\big\|^2 < +\infty. \tag{4.43}$$

Thus $\sum_{n\in\mathbb{N}}\|y_{n,k} - J_{\gamma_{n+k}A}y_{n,k}\|^2 < +\infty$ P-a.s. and hence $\|y_{n,k} - J_{\gamma_{n+k}A}y_{n,k}\| \to 0$ P-a.s. as $n \to +\infty$.

This establishes (4.41). On the other hand, let us note that, under any of scenarios (i)–(iii),

$$\mathsf{E}\sum_{n\in\mathbb{N}}|\lambda_n|\,\|e_n\| = \sum_{n\in\mathbb{N}}\mathsf{E}\big(|\lambda_n|\,\|e_n\|\big) = \sum_{n\in\mathbb{N}}\mathsf{E}|\lambda_n|\,\mathsf{E}\|e_n\| \leqslant \sum_{n\in\mathbb{N}}\sqrt{\mathsf{E}|\lambda_n|^2\mathsf{E}\|e_n\|^2} < +\infty. \tag{4.44}$$

Hence $\sum_{n\in\mathbb{N}}|\lambda_n|\,\|e_n\| < +\infty$ P-a.s. Consequently, taking into account (4.35), (4.39), and (4.6), we infer that, for every $n \in \mathbb{N}\smallsetminus\{0\}$,

$$\big\|x_{n+k} - y_{n,k}\big\| \leqslant \sum_{j=k}^{n+k-1}\big\|\lambda_j e_j\big\| \leqslant \sum_{j=k}^{+\infty}|\lambda_j|\,\|e_j\| < +\infty \ \text{ P-a.s.} \tag{4.45}$$

and

$$\big\|x_{n+k} - y_{n,k}\big\|_{L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})} \leqslant \sum_{j=k}^{n+k-1}\big\|\lambda_j e_j\big\|_{L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})}$$
$$= \sum_{j=k}^{n+k-1}\sqrt{\mathsf{E}\big|\lambda_j\|e_j\|\big|^2}$$
$$= \sum_{j=k}^{n+k-1}\sqrt{\mathsf{E}|\lambda_j|^2\mathsf{E}\|e_j\|^2}$$
$$\leqslant \sum_{j=k}^{+\infty}\sqrt{\mathsf{E}|\lambda_j|^2\mathsf{E}\|e_j\|^2}$$
$$< +\infty \tag{4.46}$$

In turn, since $(\mathsf{E}\|y_{n,k}\|^2)_{n\in\mathbb{N}}$ is bounded, so is $(\mathsf{E}\|x_n\|^2)_{n\in\mathbb{N}}$. Next, fix $z \in \operatorname{zer} A$. We derive from (4.21), (4.36), (4.38), the Cauchy–Schwarz inequality, and (4.6) that

$$
\begin{aligned}
\sum_{n\in\mathbb{N}} \mathsf{E}\varepsilon_n(\cdot,z)\mathsf{E}\lambda_n &= \sum_{n\in\mathbb{N}} \mathsf{E}\max\Big\{0, \mathsf{E}\Big(\langle z - J_{\gamma_n A}x_n \mid e_n\rangle + \langle e_n \mid x_n - J_{\gamma_n A}x_n\rangle + \|e_n\|^2 \,\Big|\, \mathcal{X}_n\Big)\Big\}\mathsf{E}\lambda_n \\
&\leqslant \sum_{n\in\mathbb{N}}\Big(\sqrt{\mathsf{E}\|z - J_{\gamma_n A}x_n\|^2} + \sqrt{\mathsf{E}\|x_n - J_{\gamma_n A}x_n\|^2} + \sqrt{\mathsf{E}\|e_n\|^2}\Big)\sqrt{\mathsf{E}\|e_n\|^2}\;\mathsf{E}\lambda_n \\
&\leqslant \sum_{n\in\mathbb{N}}\Big(\sqrt{\mathsf{E}\|z - x_n\|^2} + \sqrt{\mathsf{E}\|(\mathrm{Id} - J_{\gamma_n A})x_n - (\mathrm{Id} - J_{\gamma_n A})z\|^2} + \sqrt{\mathsf{E}\|e_n\|^2}\Big)\sqrt{\mathsf{E}|\lambda_n|^2\mathsf{E}\|e_n\|^2} \\
&\leqslant \sum_{n\in\mathbb{N}}\Big(2\sqrt{\mathsf{E}\|x_n - z\|^2} + \sqrt{\mathsf{E}\|e_n\|^2}\Big)\sqrt{\mathsf{E}|\lambda_n|^2\mathsf{E}\|e_n\|^2} \\
&< +\infty.
\end{aligned}
\tag{4.47}
$$

We conclude the proof using Theorem 4.11(iii)(e).

- Convergence under assumption (i) or (ii): In view of (4.36), let us show that

$$
\begin{cases}
x_n - w_n - e_n = x_n - J_{\gamma_n A}x_n \rightharpoonup 0 \;\; \mathsf{P}\text{-a.s.};\\
w_n + e_n - q_n = J_{\gamma_n A}x_n - x_n \to 0 \;\; \mathsf{P}\text{-a.s.};\\
w_n^* + e_n^* + Cq_n = \gamma_n^{-1}\big(x_n - J_{\gamma_n A}x_n\big) \to 0 \;\; \mathsf{P}\text{-a.s.}
\end{cases}
\tag{4.48}
$$

  By invoking (4.6), (4.45), and (4.41), we obtain

$$
\begin{aligned}
\varlimsup_{m\to+\infty}\big\|x_m - J_{\gamma_m A}x_m\big\| &= \varlimsup_{n\to+\infty}\big\|x_{n+k} - J_{\gamma_{n+k}A}x_{n+k}\big\| \\
&\leqslant \varlimsup_{n\to+\infty}\Big(\|x_{n+k} - y_{n,k}\| + \big\|J_{\gamma_{n+k}A}x_{n+k} - J_{\gamma_{n+k}A}y_{n,k}\big\| + \big\|y_{n,k} - J_{\gamma_{n+k}A}y_{n,k}\big\|\Big) \\
&\leqslant \varlimsup_{n\to+\infty}\Big(2\|x_{n+k} - y_{n,k}\| + \big\|y_{n,k} - J_{\gamma_{n+k}A}y_{n,k}\big\|\Big) \\
&\leqslant \varlimsup_{n\to+\infty}\Big(2\sum_{j=k}^{+\infty}\big\|\lambda_j e_j\big\| + \big\|y_{n,k} - J_{\gamma_{n+k}A}y_{n,k}\big\|\Big) \\
&= 2\sum_{j=k}^{+\infty}|\lambda_j|\,\|e_j\| + \lim_{n\to+\infty}\big\|y_{n,k} - J_{\gamma_{n+k}A}y_{n,k}\big\| \\
&= 2\sum_{j=k}^{+\infty}|\lambda_j|\,\|e_j\| \;\; \mathsf{P}\text{-a.s.}
\end{aligned}
\tag{4.49}
$$

  Thus, upon taking the limit as $k \to +\infty$ in (4.49), we obtain $\lim_{m\to+\infty}\|x_m - J_{\gamma_m A}x_m\| = 0$ P-a.s. Hence, since $(\gamma_n)_{n\in\mathbb{N}}$ is bounded away from 0, (4.48) holds.

- Convergence under assumption (iii): Note that (4.40) yields

$$
\sum_{n\in\mathbb{N}} \gamma_{n+k}^2\mathsf{E}\|\gamma_{n+k}^{-1}(y_{n,k} - J_{\gamma_{n+k}A}y_{n,k})\|^2 < +\infty,
\tag{4.50}
$$

which forces $\underline{\lim}\|\gamma_{n+k}^{-1}(y_{n,k} - J_{\gamma_{n+k}A}y_{n,k})\| = 0$ P-a.s. Upon invoking [4, Proposition 23.22] and the Cauchy–Schwarz inequality, we obtain, for every $n \in \mathbb{N}$,

$$\begin{aligned}
0 &\leqslant \frac{1}{\gamma_{n+1+k}}\left\langle J_{\gamma_{n+k}A}y_{n,k} - J_{\gamma_{n+1+k}A}y_{n+1,k} \,\middle|\, \frac{y_{n,k} - J_{\gamma_{n+k}A}y_{n,k}}{\gamma_{n+k}} - \frac{y_{n+1,k} - J_{\gamma_{n+k+1}A}y_{n+1,k}}{\gamma_{n+1+k}}\right\rangle \\
&= \left\langle \frac{J_{\gamma_{n+k}A}y_{n,k} - J_{\gamma_{n+1+k}A}y_{n+1,k}}{\gamma_{n+1+k}} \,\middle|\, \frac{y_{n,k} - J_{\gamma_{n+k}A}y_{n,k}}{\gamma_{n+k}}\right\rangle - \left\|\frac{y_{n+1,k} - J_{\gamma_{n+k+1}A}y_{n+1,k}}{\gamma_{n+1+k}}\right\|^2 \\
&\leqslant \left\|\frac{y_{n+1,k} - J_{\gamma_{n+k}A}y_{n+1,k}}{\gamma_{n+1+k}}\right\|\left(\left\|\frac{y_{n,k} - J_{\gamma_{n+k}A}y_{n,k}}{\gamma_{n+k}}\right\| - \left\|\frac{y_{n+1,k} - J_{\gamma_{n+k+1}A}y_{n+1,k}}{\gamma_{n+1+k}}\right\|\right) \text{ P-a.s.} \qquad (4.51)
\end{aligned}$$

Hence, $(\|\gamma_{n+k}^{-1}(y_{n,k} - J_{\gamma_{n+k}A}y_{n,k})\|)_{n\in\mathbb{N}}$ decreases P-a.s., which implies that

$$\gamma_{n+k}^{-1}(y_{n,k} - J_{\gamma_{n+k}A}y_{n,k}) \to 0 \text{ P-a.s.} \quad \text{as} \quad n \to +\infty. \qquad (4.52)$$

Consequently, we deduce then from Theorem (iii)(e) that, for every $k \in \mathbb{N}$, $(y_{n,k})_{n\in\mathbb{N}}$ converges weakly P-a.s. and weakly in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ to some $(\operatorname{zer} A)$-valued random variable which we denote by y_k. In addition, we deduce from (4.35), (4.39), and (4.6) that

$$(\forall k \in \mathbb{N})(\forall n \in \mathbb{N}) \quad \begin{cases} \|y_{n,k+1} - y_{n+1,k}\| \leqslant \|e_k\| \text{ P-a.s.;} \\ \|y_{n,k+1} - y_{n+1,k}\|_{L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})} \leqslant \|e_k\|_{L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})}. \end{cases} \qquad (4.53)$$

In turn, the weak lower semicontinuity of the norm and Fatou's lemma imply that

$$(\forall k \in \mathbb{N}) \quad \begin{cases} \|y_{k+1} - y_k\| \leqslant \underline{\lim}\,\|y_{n,k+1} - y_{n+1,k}\| \leqslant \|e_k\| \text{ P-a.s.;} \\ \mathsf{E}\|y_{k+1} - y_k\|^2 \leqslant \underline{\lim}\,\mathsf{E}\|y_{n,k+1} - y_{n+1,k}\|^2 \leqslant \mathsf{E}\|e_k\|^2. \end{cases} \qquad (4.54)$$

Since $\sum_{n\in\mathbb{N}}\|e_n\| < +\infty$ P-a.s. and $\sum_{n\in\mathbb{N}}\sqrt{\mathsf{E}\|e_n\|^2} < +\infty$, (4.54) shows that $(y_k)_{k\in\mathbb{N}}$ is a Cauchy sequence both P-a.s. and in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. Hence, we deduce from (4.45), (4.46), and (4.54) that there exists a $(\operatorname{zer} A)$-valued random variable y such that

$$\begin{cases} x_{n+k} - y_{n,k} \to 0 \text{ P-a.s. and in } L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H}) \text{ as } n \to +\infty \text{ and } k \to +\infty; \\ \text{for every } k \in \mathbb{N},\ y_{n,k} - y_k \rightharpoonup 0 \text{ P-a.s. and in } L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H}) \text{ as } n \to +\infty; \\ y_k - y \to 0 \text{ P-a.s. and in } L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H}) \text{ as } k \to +\infty. \end{cases} \qquad (4.55)$$

Thus, $x_{n+k} - y = x_{n+k} - y_{n,k} + y_{n,k} - y_k + y_k - y \rightharpoonup 0$ P-a.s. and in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ as $n \to +\infty$ and $k \to +\infty$. This confirms that $(x_n)_{n\in\mathbb{N}}$ converges weakly P-a.s. and weakly in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ to y.

□

Remark 4.14 Here are a few commentaries on Theorem 4.13.

(i) In the deterministic setting with $(\lambda_n)_{n\in\mathbb{N}}$ in $]0,2[$, Theorem 4.13(i) follows from [16, Theorem 2.1(i)(a)], Theorem 4.13(ii) was established in [29, Theorem 3], and Theorem 4.13(iii) was established in [5, Remarque 14(a)].

(ii) In the case of deterministic relaxations $(\lambda_n)_{n\in\mathbb{N}}$ in $]0,2[$ and constant proximal parameters $(\gamma_n)_{n\in\mathbb{N}}$, the almost sure weak convergence result in Theorem 4.13(ii) follows from [22, Proposition 5.1].

(iii) As discussed in [18, Section 5], the deterministic proximal point algorithm can be employed to solve equilibrium problems beyond the simple inclusion $0 \in \mathsf{A}\mathsf{x}$. It captures in particular the method of partial inverses to split multi-operator inclusions, problems involving resolvent compositions, and the Chambolle–Pock algorithm. Stochasticity can be introduced in these methods via Theorem 4.13.

4.2.5 Randomized block-iterative saddle projective splitting

4.2.5.1 Problem setting

We consider a highly structured composite multivariate primal-dual inclusion problem introduced in [13] and further studied in [18, Section 10]. This model includes a mix of set-valued, cocoercive, and Lipschitzian monotone operators, as well as linear operators and various monotonicity-preserving operations among them. Its multivariate structure captures problems in areas such as domain decomposition methods [1,2], game theory [12,38], mean field games [8], machine learning [3,6], network flow problems [9,42], neural networks [23], and stochastic programming [10,30].

Problem 4.15 Let $(\mathsf{H}_i)_{i\in I}$ and $(\mathsf{G}_k)_{k\in K}$ be finite families of Euclidean spaces with respective direct sums $\mathsf{H} = \bigoplus_{i\in I} \mathsf{H}_i$ and $\mathsf{G} = \bigoplus_{k\in K} \mathsf{G}_k$. Denote by $\mathsf{x} = (\mathsf{x}_i)_{i\in I}$ a generic element in H. For every $i \in I$ and every $k \in K$, let $\mathsf{s}_i^* \in \mathsf{H}_i$, let $\mathsf{r}_k \in \mathsf{G}_k$, and suppose that the following are satisfied:

[a] $\mathsf{A}_i \colon \mathsf{H}_i \to 2^{\mathsf{H}_i}$ is maximally monotone, $\mathsf{C}_i \colon \mathsf{H}_i \to \mathsf{H}_i$ is cocoercive with constant $\alpha_i^{c} \in\,]0,+\infty[$, $\mathsf{Q}_i \colon \mathsf{H}_i \to \mathsf{H}_i$ is monotone and Lipschitzian with constant $\alpha_i^{\ell} \in [0,+\infty[$, and $\mathsf{R}_i \colon \mathsf{H} \to \mathsf{H}_i$.

[b] $\mathsf{B}_k^{m} \colon \mathsf{G}_k \to 2^{\mathsf{G}_k}$ is maximally monotone, $\mathsf{B}_k^{c} \colon \mathsf{G}_k \to \mathsf{G}_k$ is cocoercive with constant $\beta_k^{c} \in\,]0,+\infty[$, and $\mathsf{B}_k^{\ell} \colon \mathsf{G}_k \to \mathsf{G}_k$ is monotone and Lipschitzian with constant $\beta_k^{\ell} \in [0,+\infty[$.

[c] $\mathsf{D}_k^{m} \colon \mathsf{G}_k \to 2^{\mathsf{G}_k}$ is maximally monotone, $\mathsf{D}_k^{c} \colon \mathsf{G}_k \to \mathsf{G}_k$ is cocoercive with constant $\delta_k^{c} \in\,]0,+\infty[$, and $\mathsf{D}_k^{\ell} \colon \mathsf{G}_k \to \mathsf{G}_k$ is monotone and Lipschitzian with constant $\delta_k^{\ell} \in [0,+\infty[$.

[d] $\mathsf{L}_{ki} \colon \mathsf{H}_i \to \mathsf{G}_k$ is linear.

In addition, it is assumed that

[e] $\mathsf{R} \colon \mathsf{H} \to \mathsf{H} \colon \mathsf{x} \mapsto (\mathsf{R}_i\mathsf{x})_{i\in I}$ is monotone and Lipschitzian with constant $\chi \in [0,+\infty[$.

The objective is to solve the primal problem

$$\text{find } \overline{x} \in \mathsf{H} \text{ such that } (\forall i \in I)\;\; s_i^* \in A_i\overline{x}_i + C_i\overline{x}_i + Q_i\overline{x}_i + R_i\overline{x} + \sum_{k\in K} L_{ki}^*\Bigg(\Big(\big(B_k^{m} + B_k^{c} + B_k^{\ell}\big) \,\square\, \big(D_k^{m} + D_k^{c} + D_k^{\ell}\big)\Big)\Bigg(\sum_{j\in I} L_{kj}\overline{x}_j - r_k\Bigg)\Bigg) \tag{4.56}$$

and the associated dual problem

$$\text{find } \overline{v}^* \in \mathsf{G} \text{ such that } (\exists x \in \mathsf{H})(\forall i \in I)(\forall k \in K)\;\; \begin{cases} s_i^* - \sum_{j\in K} L_{ji}^*\overline{v}_j^* \in A_i x_i + C_i x_i + Q_i x_i + R_i x;\\ \overline{v}_k^* \in \Big(\big(B_k^{m} + B_k^{c} + B_k^{\ell}\big) \,\square\, \big(D_k^{m} + D_k^{c} + D_k^{\ell}\big)\Big)\Bigg(\sum_{j\in I} L_{kj}x_j - r_k\Bigg). \end{cases} \tag{4.57}$$

Finally, $\mathscr{P}$ denotes the set of solutions to (4.56), $\mathscr{D}$ denotes the set of solutions to (4.57), and we set $\underline{\mathsf{X}} = \mathsf{H} \oplus \mathsf{G} \oplus \mathsf{G} \oplus \mathsf{G}$.

To deal with large size problems in which I and/or K is sizable, the deterministic block-iterative algorithm proposed in [13] has the ability to activate only subgroups of coordinates and operators at each iteration instead of all of them as in classical methods. We propose a stochastic version of this block-iterative algorithm with almost sure convergence to a solution of Problem 4.15. The convergence analysis will rely on an application of Theorem 4.11 in $\underline{\mathsf{X}}$ using the following saddle formalism.

Definition 4.16 ([13, Definition 1]) The *saddle operator* associated with Problem 4.15 is

$$\begin{aligned} \boldsymbol{\mathcal{S}} \colon \underline{\mathsf{X}} \to 2^{\underline{\mathsf{X}}} \colon (x, y, z, v^*) \mapsto & \Bigg(\bigtimes_{i\in I}\Bigg(-s_i^* + A_i x_i + C_i x_i + Q_i x_i + R_i x + \sum_{k\in K} L_{ki}^* v_k^*\Bigg), \bigtimes_{k\in K}\big(B_k^{m} y_k + B_k^{c} y_k + B_k^{\ell} y_k - v_k^*\big),\\ & \bigtimes_{k\in K}\big(D_k^{m} z_k + D_k^{c} z_k + D_k^{\ell} z_k - v_k^*\big), \bigtimes_{k\in K}\Bigg\{r_k + y_k + z_k - \sum_{i\in I} L_{ki}x_i\Bigg\} \Bigg), \end{aligned} \tag{4.58}$$

and the *saddle form* of Problem 4.15 is to

$$\text{find } \underline{\overline{x}} \in \underline{\mathsf{X}} \text{ such that } \underline{\mathbf{0}} \in \boldsymbol{\mathcal{S}}\underline{\overline{x}}. \tag{4.59}$$

Item (ii) below asserts that finding a saddle point, i.e., solving (4.59), provides a solution to Problem 4.15.

Proposition 4.17 ([13, Proposition 1]) *Consider the setting of Problem 4.15 and Definition 4.16. Then the following hold:*

(i) *$\boldsymbol{\mathcal{S}}$ is maximally monotone.*

(ii) *Suppose that $\underline{\overline{x}} = (\overline{x}, \overline{y}, \overline{z}, \overline{v}^*) \in \operatorname{zer} \boldsymbol{\mathcal{S}}$. Then $(\overline{x}, \overline{v}^*) \in \mathscr{P} \times \mathscr{D}$.*

(iii) $\mathscr{D} \neq \varnothing \Leftrightarrow \operatorname{zer} \mathcal{S} \neq \varnothing \Rightarrow \mathscr{P} \neq \varnothing$.

To use Theorem 4.11, we decompose the saddle operator $\mathcal{S}$ of (4.58) as the sum of

$$\begin{aligned} \underline{\mathsf{W}} \colon \underline{\mathsf{X}} \to 2^{\underline{\mathsf{X}}} \colon (\mathsf{x}, \mathsf{y}, \mathsf{z}, \mathsf{v}^*) \mapsto \Bigg(& \bigtimes_{i \in I} \Big(-\mathsf{s}_i^* + \mathsf{A}_i \mathsf{x}_i + \mathsf{Q}_i \mathsf{x}_i + \mathsf{R}_i \mathsf{x} + \sum_{k \in K} \mathsf{L}_{ki}^* \mathsf{v}_k^* \Big), \bigtimes_{k \in K} \big(\mathsf{B}_k^{m} \mathsf{y}_k + \mathsf{B}_k^{\ell} \mathsf{y}_k - \mathsf{v}_k^* \big), \\ & \bigtimes_{k \in K} \big(\mathsf{D}_k^{m} \mathsf{z}_k + \mathsf{D}_k^{\ell} \mathsf{z}_k - \mathsf{v}_k^* \big), \bigtimes_{k \in K} \Big\{ \mathsf{r}_k + \mathsf{y}_k + \mathsf{z}_k - \sum_{i \in I} \mathsf{L}_{ki} \mathsf{x}_i \Big\} \Bigg) \end{aligned} \tag{4.60}$$

and

$$\underline{\mathsf{C}} \colon \underline{\mathsf{X}} \to \underline{\mathsf{X}} \colon (\mathsf{x}, \mathsf{y}, \mathsf{z}, \mathsf{v}^*) \mapsto \Big((\mathsf{C}_i \mathsf{x}_i)_{i \in I}, (\mathsf{B}_k^{c} \mathsf{y}_k)_{k \in K}, (\mathsf{D}_k^{c} \mathsf{z}_k)_{k \in K}, \mathbf{0} \Big). \tag{4.61}$$

As seen in [13, Proposition 2(ii)–(iii)], $\underline{\mathsf{W}}$ is maximally monotone and $\underline{\mathsf{C}}$ is α-cocoercive with $\alpha = \min\{\alpha_i^{c}, \beta_k^{c}, \delta_k^{c}\}_{i \in I, k \in K}$. This confirms that (4.59) fits the framework described in Problem 4.1.

4.2.5.2 Algorithm and convergence

The following assumptions regulate the way in which the coordinates and the sets are randomly activated over the course of the iterations.

Assumption 4.18 *I and K are nonempty finite sets, $(\pi_i)_{i \in I}$ and $(\zeta_k)_{k \in K}$ are in $]0,1]$, and $N \in \mathbb{N} \smallsetminus \{0\}$. $(I_n)_{n \in \mathbb{N}}$ are nonempty sets composed of elements randomly taken in I and $(K_n)_{n \in \mathbb{N}}$ are nonempty sets composed of elements randomly taken in K. Further, for every finite collection of positive integers $n_1, \ldots, n_m$,*

$$\begin{cases} (\forall i \in I) \quad \mathsf{P}\Big(\bigcap_{j=1}^{m} [i \in I_{n_j}] \Big) = \prod_{j=1}^{m} \mathsf{P}\big([i \in I_{n_j}] \big); \\ (\forall k \in K) \quad \mathsf{P}\Big(\bigcap_{j=1}^{m} [k \in K_{n_j}] \Big) = \prod_{j=1}^{m} \mathsf{P}\big([k \in K_{n_j}] \big). \end{cases} \tag{4.62}$$

Moreover, $I_0 = I$, $K_0 = K$, and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} (\forall i \in I) \;\; \mathsf{P}\Bigg(\Bigg[i \in \bigcup_{j=n}^{n+N-1} I_j \Bigg] \Bigg) \geqslant \pi_i; \\ (\forall k \in K) \;\; \mathsf{P}\Bigg(\Bigg[k \in \bigcup_{j=n}^{n+N-1} K_j \Bigg] \Bigg) \geqslant \zeta_k. \end{cases} \tag{4.63}$$

Example 4.19

(i) The (deterministic) rule of [13, Assumption 2] satisfies Assumption 4.18 by setting, for every $i \in I$ and every $k \in K$, $\pi_i = 1$ and $\zeta_k = 1$.

(ii) Set, for every $n \in \mathbb{N}$, $I_n = \{i_n\}$ and $K_n = \{k_n\}$, where $(i_n)_{n \in \mathbb{N}}$ are i.i.d. random variables uniformly distributed on I and $(k_n)_{n \in \mathbb{N}}$ are i.i.d. random variables uniformly distributed on K. This rule satisfies Assumption 4.18 for $\mathsf{N} = 1$, $\pi_i \equiv 1/\operatorname{card} \mathtt{I}$, and $\zeta_k \equiv 1/\operatorname{card} \mathsf{K}$.

Proposition 4.20 *Let* $\mathtt{I}$ *be a nonempty finite set and let* $(I_n)_{n\in\mathbb{N}}$ *be nonempty sets composed of elements randomly taken in* $\mathtt{I}$. *Suppose that* $I_0 = \mathtt{I}$, *and that* $i \in \mathtt{I}$ *is such that* $([i \in I_n])_{n\in\mathbb{N}}$ *is an independent sequence in* $\mathcal{F}$ *that satisfies*

$$(\exists\, \mathsf{N} \in \mathbb{N} \smallsetminus \{0\})(\exists\, \pi_i \in \,]0,1])(\forall n \in \mathbb{N}) \quad \mathsf{P}\left(\left[i \in \bigcup_{j=n}^{n+\mathsf{N}-1} I_j\right]\right) \geqslant \pi_i. \tag{4.64}$$

Set, for every $n \in \mathbb{N}$, $\vartheta_i(n) = \max\big\{j \in \mathbb{N} \mid j \leqslant n \text{ and } i \in I_j\big\}$. *Further, let* $(x_n)_{n\in\mathbb{N}}$ *be a sequence in* $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ *such that* $\sum_{n\in\mathbb{N}} \mathsf{E}\|x_{n+1} - x_n\|^2 < +\infty$ P-a.s. *Then* $x_{\vartheta_i(n)} - x_n \to 0$ *in* $L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$.

Proof. Note that $(\forall n \in \mathbb{N})\ \vartheta_i(n) \in \{0, \ldots, n\}$ P-a.s. Hence, Lemma 4.8 ensures that, for every $n \in \mathbb{N}$, $x_{\vartheta_i(n)} \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. On the other hand, it follows from the independence condition and (4.64) that

$$\begin{aligned}(\forall n \in \mathbb{N}) \quad \mathsf{P}\left(\left[i \notin \bigcup_{j=n}^{+\infty} I_j\right]\right) &= \mathsf{P}\left(\bigcap_{j=n}^{+\infty}\left[i \in \complement I_j\right]\right)\\ &= \mathsf{P}\left(\lim_{0<m\to+\infty} \bigcap_{j=n}^{n+m\mathsf{N}-1}\left[i \in \complement I_j\right]\right)\\ &= \lim_{0<m\to+\infty} \mathsf{P}\left(\bigcap_{j=n}^{n+m\mathsf{N}-1}\left[i \in \complement I_j\right]\right)\\ &= \lim_{0<m\to+\infty} \prod_{k=0}^{m-1} \mathsf{P}\left(\bigcap_{j=n+k\mathsf{N}}^{n+(k+1)\mathsf{N}-1}\left[i \in \complement I_j\right]\right)\\ &= \lim_{0<m\to+\infty} \prod_{k=0}^{m-1} \mathsf{P}\left(\complement\left[i \in \bigcup_{j=n+k\mathsf{N}}^{(n+k\mathsf{N})+\mathsf{N}-1} I_j\right]\right)\\ &\leqslant \lim_{0<m\to+\infty} (1-\pi_i)^m\\ &= 0. \end{aligned} \tag{4.65}$$

Therefore $\vartheta_i(n) \to +\infty$ P-a.s. as $n \to +\infty$ and, since $\sum_{n\in\mathbb{N}} \|x_{n+1} - x_n\|^2 < +\infty$ P-a.s., we have $\sum_{j\geqslant\vartheta_i(n)} \|x_{j+1} - x_j\|^2 \downarrow 0$ P-a.s. as $n \to +\infty$. Thus,

$$(\forall n \in \mathbb{N}) \quad 0 \leqslant \sum_{j\geqslant\vartheta_i(n)} \|x_{j+1} - x_j\|^2 \leqslant \sum_{j\in\mathbb{N}} \|x_{j+1} - x_j\|^2 \in L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathbb{R}), \tag{4.66}$$

from which we deduce via [44, Theorem 2.6.1(b)] that $\mathsf{E}\sum_{j\geqslant\vartheta_i(n)} \|x_{j+1} - x_j\|^2 \to 0$ as $n \to +\infty$. On the other hand, let $n \in \mathbb{N}$ and $m \in \mathbb{N}$ be such that $m\mathsf{N} \leqslant n < (m+1)\mathsf{N}$. Then

$$\mathsf{E}\big(n - \vartheta_i(n)\big)$$

$$
\begin{aligned}
&= \sum_{l=0}^{n-1}(n-l)P\big([i \in I_l]\big)P\left(\left[i \notin \bigcup_{j=l+1}^{n} I_j\right]\right)\\
&\leqslant \sum_{l=0}^{n-1}(n-l)P\left(\left[i \notin \bigcup_{j=l+1}^{n} I_j\right]\right)\\
&\leqslant N(n-mN) + \sum_{k=0}^{m-1}\sum_{l=kN}^{(k+1)N-1}(n-l)P\left(\left[i \notin \bigcup_{j=l+1}^{n} I_j\right]\right)\\
&\leqslant N^2 + \sum_{k=0}^{m-1}\sum_{l=kN}^{(k+1)N-1}(n-l)P\left(\left[i \notin \bigcup_{j=(k+1)N}^{mN-1} I_j\right]\right)\\
&\leqslant N^2 + \sum_{k=0}^{m-1}(n-kN)\sum_{l=kN}^{(k+1)N-1}(1-\pi_i)^{m-k-1}\\
&= N^2 + \sum_{k=0}^{m-1}(n-kN)N(1-\pi_i)^{m-k-1}\\
&\leqslant N^2 + \sum_{k=0}^{m-1}\big((m+1)N-kN\big)N(1-\pi_i)^{m-k-1}\\
&= N^2\left(1 + \sum_{k=0}^{m-1}\big(m+1-k\big)(1-\pi_i)^{m-k-1}\right)\\
&= N^2\left(1 + \sum_{l=0}^{m-1}\big(l+2\big)(1-\pi_i)^{l}\right)\\
&= N^2\left(1 + \sum_{l=0}^{m-1}l(1-\pi_i)^{l} + \sum_{l=0}^{m-1}2(1-\pi_i)^{l}\right)\\
&= N^2\left(1 + (1-\pi_i)\frac{1-m(1-\pi_i)^{m-1}+(m-1)(1-\pi_i)^m}{\pi_i^2} + 2\frac{1-(1-\pi_i)^m}{\pi_i}\right)\\
&= N^2\left(1 + \frac{1-\pi_i-m(1-\pi_i)^{m}+(m-1)(1-\pi_i)^{m+1}}{\pi_i^2} + \frac{2\pi_i+2(1-\pi_i)^{m+1}-2(1-\pi_i)^m}{\pi_i^2}\right)\\
&= N^2\left(1 + \frac{(m+1)(1-\pi_i)^{m+1}-(m+2)(1-\pi_i)^{m}+1+\pi_i}{\pi_i^2}\right), \qquad (4.67)
\end{aligned}
$$

which shows that $\overline{\lim}\, E(n-\vartheta_i(n)) \leqslant N^2(1+(1+\pi_i)/\pi_i^2) < +\infty$. Thus,

$$
\begin{aligned}
E\|x_n - x_{\vartheta_i(n)}\| &\leqslant E\sum_{j=\vartheta_i(n)}^{n}\|x_{j+1}-x_j\|\\
&\leqslant E\left(\sqrt{n+1-\vartheta_i(n)}\sqrt{\sum_{j=\vartheta_i(n)}^{n}\|x_{j+1}-x_j\|^2}\right)
\end{aligned}
$$

$$
\begin{aligned}
&\leqslant \sqrt{1+\mathsf{E}(\mathsf{n}-\vartheta_{\mathsf{i}}(\mathsf{n}))}\sqrt{\mathsf{E}\sum_{\mathsf{j}=\vartheta_{\mathsf{i}}(\mathsf{n})}^{+\infty}\|x_{\mathsf{j}+1}-x_{\mathsf{j}}\|^2}\\
&\to 0. \qquad (4.68)
\end{aligned}
$$

This confirms that $x_{\vartheta_{\mathsf{i}}(\mathsf{n})}-x_{\mathsf{n}}\to 0$ in $L^1(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$. $\square$

Assumption 4.21 *In the setting of Problem 4.15, set* $\alpha=\min\{\alpha_{\mathsf{i}}^{c},\beta_{\mathsf{k}}^{c},\delta_{\mathsf{k}}^{c}\}_{\mathsf{i}\in\mathsf{I},\mathsf{k}\in\mathsf{K}}$*, and let* $\sigma\in\,]0,+\infty[$ *and* $\varepsilon\in\,]0,1[$ *be such that* $\sigma>1/(4\alpha)$ *and* $1/\varepsilon>\max\{\alpha_{\mathsf{i}}^{\ell}+\chi+\sigma,\beta_{\mathsf{k}}^{\ell}+\sigma,\delta_{\mathsf{k}}^{\ell}+\sigma\}_{\mathsf{i}\in\mathsf{I},\mathsf{k}\in\mathsf{K}}$*, and suppose that the following are satisfied:*

[a] *For every* $\mathsf{i}\in\mathsf{I}$ *and every* $\mathsf{n}\in\mathbb{N}$*,* $\gamma_{\mathsf{i},\mathsf{n}}\in\big[\varepsilon,1/(\alpha_{\mathsf{i}}^{\ell}+\chi+\sigma)\big]$.

[b] *For every* $\mathsf{k}\in\mathsf{K}$ *and every* $\mathsf{n}\in\mathbb{N}$*,* $\mu_{\mathsf{k},\mathsf{n}}\in\big[\varepsilon,1/(\beta_{\mathsf{k}}^{\ell}+\sigma)\big]$*,* $\nu_{\mathsf{k},\mathsf{n}}\in\big[\varepsilon,1/(\delta_{\mathsf{k}}^{\ell}+\sigma)\big]$*, and* $\sigma_{\mathsf{k},\mathsf{n}}\in[\varepsilon,1/\varepsilon]$.

[c] *For every* $\mathsf{i}\in\mathsf{I}$*,* $x_{\mathsf{i},0}\in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H}_{\mathsf{i}})$ *and, for every* $\mathsf{k}\in\mathsf{K}$*,* $\{y_{\mathsf{k},0},z_{\mathsf{k},0},v_{\mathsf{k},0}^*\}\subset L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{G}_{\mathsf{k}})$.

We now introduce our stochastic block-iterative algorithm. It differs from that of [13] in that the selection of the blocks of variables and operators to be activated at each iteration is random, and so is the relaxation strategy. In addition, the relaxation parameters need not be bounded by 2.

Algorithm 4.22 Consider the setting of Problem 4.15 and suppose that Assumptions 4.18 and 4.21 are in force. Let $\rho\in[2,+\infty[$ and iterate

$$
\begin{array}{l}
\text{for } \mathsf{n}=0,1,\ldots\\
\left\lfloor
\begin{array}{l}
\text{for every } \mathsf{i}\in I_{\mathsf{n}}\\
\left\lfloor
\begin{array}{l}
l_{\mathsf{i},\mathsf{n}}^*=\mathsf{Q}_{\mathsf{i}}x_{\mathsf{i},\mathsf{n}}+\mathsf{R}_{\mathsf{i}}\boldsymbol{x}_{\mathsf{n}}+\sum_{\mathsf{k}\in\mathsf{K}}\mathsf{L}_{\mathsf{ki}}^*v_{\mathsf{k},\mathsf{n}}^*;\\
a_{\mathsf{i},\mathsf{n}}=\mathsf{J}_{\gamma_{\mathsf{i},\mathsf{n}}\mathsf{A}_{\mathsf{i}}}\big(x_{\mathsf{i},\mathsf{n}}+\gamma_{\mathsf{i},\mathsf{n}}(\mathsf{s}_{\mathsf{i}}^*-l_{\mathsf{i},\mathsf{n}}^*-\mathsf{C}_{\mathsf{i}}x_{\mathsf{i},\mathsf{n}})\big);\\
a_{\mathsf{i},\mathsf{n}}^*=\gamma_{\mathsf{i},\mathsf{n}}^{-1}(x_{\mathsf{i},\mathsf{n}}-a_{\mathsf{i},\mathsf{n}})-l_{\mathsf{i},\mathsf{n}}^*+\mathsf{Q}_{\mathsf{i}}a_{\mathsf{i},\mathsf{n}};\\
\xi_{\mathsf{i},\mathsf{n}}=\|a_{\mathsf{i},\mathsf{n}}-x_{\mathsf{i},\mathsf{n}}\|^2;
\end{array}
\right.\\
\text{for every } \mathsf{i}\in\mathsf{I}\smallsetminus I_{\mathsf{n}}\\
\left\lfloor
\begin{array}{l}
a_{\mathsf{i},\mathsf{n}}=a_{\mathsf{i},\mathsf{n}-1};\ a_{\mathsf{i},\mathsf{n}}^*=a_{\mathsf{i},\mathsf{n}-1}^*;\ \xi_{\mathsf{i},\mathsf{n}}=\xi_{\mathsf{i},\mathsf{n}-1};
\end{array}
\right.
\end{array}
\right.
\end{array}
$$

$$
\left\lfloor
\begin{array}{l}
\text{for every } k\in K_n\\
\left\lfloor
\begin{array}{l}
u^*_{k,n} = v^*_{k,n} - B^{\ell}_k y_{k,n};\\
w^*_{k,n} = v^*_{k,n} - D^{\ell}_k z_{k,n};\\
b_{k,n} = J_{\mu_{k,n}B^{m}_k}\big(y_{k,n} + \mu_{k,n}(u^*_{k,n} - B^{c}_k y_{k,n})\big);\\
d_{k,n} = J_{\nu_{k,n}D^{m}_k}\big(z_{k,n} + \nu_{k,n}(w^*_{k,n} - D^{c}_k z_{k,n})\big);\\
e^*_{k,n} = \sigma_{k,n}\big(\sum_{i\in I} L_{ki}x_{i,n} - y_{k,n} - z_{k,n} - r_k\big) + v^*_{k,n};\\
q^*_{k,n} = \mu^{-1}_{k,n}(y_{k,n} - b_{k,n}) + u^*_{k,n} + B^{\ell}_k b_{k,n} - e^*_{k,n};\\
t^*_{k,n} = \nu^{-1}_{k,n}(z_{k,n} - d_{k,n}) + w^*_{k,n} + D^{\ell}_k d_{k,n} - e^*_{k,n};\\
\eta_{k,n} = \|b_{k,n} - y_{k,n}\|^2 + \|d_{k,n} - z_{k,n}\|^2;\\
e_{k,n} = r_k + b_{k,n} + d_{k,n} - \sum_{i\in I} L_{ki}a_{i,n};
\end{array}\right.\\
\text{for every } k\in K\smallsetminus K_n\\
\left\lfloor
\begin{array}{l}
b_{k,n} = b_{k,n-1};\ d_{k,n} = d_{k,n-1};\ e^*_{k,n} = e^*_{k,n-1};\ q^*_{k,n} = q^*_{k,n-1};\ t^*_{k,n} = t^*_{k,n-1};\\
\eta_{k,n} = \eta_{k,n-1};\ e_{k,n} = r_k + b_{k,n} + d_{k,n} - \sum_{i\in I} L_{ki}a_{i,n};
\end{array}\right.\\
\text{for every } i\in I\\
\left\lfloor
\begin{array}{l}
p^*_{i,n} = a^*_{i,n} + R_i\boldsymbol{a}_n + \sum_{k\in K} L^*_{ki}e^*_{k,n};
\end{array}\right.\\
\Delta_n = -(4\alpha)^{-1}\big(\sum_{i\in I}\xi_{i,n} + \sum_{k\in K}\eta_{k,n}\big) + \sum_{i\in I}\langle x_{i,n} - a_{i,n} \mid p^*_{i,n}\rangle\\
\qquad + \sum_{k\in K}\big(\langle y_{k,n} - b_{k,n} \mid q^*_{k,n}\rangle + \langle z_{k,n} - d_{k,n} \mid t^*_{k,n}\rangle + \langle e_{k,n} \mid v^*_{k,n} - e^*_{k,n}\rangle\big);\\
\theta_n = \dfrac{1_{[\Delta_n>0]}\Delta_n}{\sum_{i\in I}\|p^*_{i,n}\|^2 + \sum_{k\in K}\big(\|q^*_{k,n}\|^2 + \|t^*_{k,n}\|^2 + \|e_{k,n}\|^2\big) + 1_{[\Delta_n\leqslant 0]}};\\
\text{take } \lambda_n\in L^\infty(\Omega,\mathcal{F},P;[\varepsilon,\rho])\\
\text{for every } i\in I\\
\left\lfloor
\begin{array}{l}
x_{i,n+1} = x_{i,n} - \lambda_n\theta_n p^*_{i,n};
\end{array}\right.\\
\text{for every } k\in K\\
\left\lfloor
\begin{array}{l}
y_{k,n+1} = y_{k,n} - \lambda_n\theta_n q^*_{k,n};\ z_{k,n+1} = z_{k,n} - \lambda_n\theta_n t^*_{k,n};\ v^*_{k,n+1} = v^*_{k,n} - \lambda_n\theta_n e_{k,n};
\end{array}\right.
\end{array}\right.
\tag{4.69}
$$

The convergence properties of Algorithm 4.22 are established in the following theorem.

Theorem 4.23 *Consider the setting of Algorithm 4.22. Suppose that* $\inf_{n\in\mathbb{N}} E(\lambda_n(2-\lambda_n)) > 0$ *and that* $\mathscr{D}\neq\varnothing$. *Then the following hold:*

(i) *Let* $i\in I$. *Then* $(x_{i,n})_{n\in\mathbb{N}}$ *lies in* $L^2(\Omega,\mathcal{F},P;H_i)$ *and* $\sum_{n\in\mathbb{N}} E\|x_{i,n+1} - x_{i,n}\|^2 < +\infty$.

(ii) *Let* $k\in K$. *Then* $(y_{k,n})_{n\in\mathbb{N}}$, $(z_{k,n})_{n\in\mathbb{N}}$, *and* $(v^*_{k,n})_{n\in\mathbb{N}}$ *are sequences in* $L^2(\Omega,\mathcal{F},P;G_k)$. *Further,* $\sum_{n\in\mathbb{N}} E\|y_{k,n+1} - y_{k,n}\|^2 < +\infty$, $\sum_{n\in\mathbb{N}} E\|z_{k,n+1} - z_{k,n}\|^2 < +\infty$, *and* $\sum_{n\in\mathbb{N}} E\|v^*_{k,n+1} - v^*_{k,n}\|^2 < +\infty$.

(iii) *Let* $i\in I$ *and* $k\in K$. *Then* $x_{i,n} - a_{i,n} \xrightarrow{P} 0$, $y_{k,n} - b_{k,n} \xrightarrow{P} 0$, $z_{k,n} - d_{k,n} \xrightarrow{P} 0$, *and* $v^*_{k,n} - e^*_{k,n} \xrightarrow{P} 0$.

(iv) *There exist a* $\mathscr{P}$*-valued random variable* $\overline{\boldsymbol{x}}$ *and a* $\mathscr{D}$*-valued random variable* $\overline{\boldsymbol{v}}^*$ *such that, for every* $i\in I$ *and every* $k\in K$, $x_{i,n}\to\overline{x}_i$ P-a.s., $a_{i,n}\to\overline{x}_i$ P-a.s., *and* $v^*_{k,n}\to\overline{v}^*_k$ P-a.s.

Proof. The results will be derived from Theorem 4.11 applied to $\underline{Z} = \operatorname{zer}\mathcal{S}$ in $\underline{X}$, following the general pattern of the deterministic proof of [13, Theorem 1]. We use the notation of Definition 4.16, as well as (4.60) and (4.61). Note that, since $\mathscr{D} \neq \varnothing$, Proposition 4.17(iii) asserts that $\operatorname{zer} \boldsymbol{\mathcal{S}} \neq \varnothing$. Let us show that (4.69) is a special case of (4.3). We define the random indices

$$(\forall i \in I)(\forall n \in \mathbb{N}) \quad \vartheta_i(n) = \max\big\{j \in \mathbb{N} \mid j \leqslant n \text{ and } i \in I_j\big\} \tag{4.70}$$

and

$$(\forall k \in K)(\forall n \in \mathbb{N}) \quad \varrho_k(n) = \max\big\{j \in \mathbb{N} \mid j \leqslant n \text{ and } k \in K_j\big\}. \tag{4.71}$$

It then follows from (4.69) that

$$(\forall i \in I)(\forall n \in \mathbb{N}) \quad a_{i,n} = a_{i,\vartheta_i(n)} \text{ P-a.s.}, \quad a^*_{i,n} = a^*_{i,\vartheta_i(n)} \text{ P-a.s.}, \quad \xi_{i,n} = \xi_{i,\vartheta_i(n)} \text{ P-a.s.}, \tag{4.72}$$

and

$$(\forall k \in K)(\forall n \in \mathbb{N}) \quad \begin{cases} b_{k,n} = b_{k,\varrho_k(n)} \text{ P-a.s.}; \;\; d_{k,n} = d_{k,\varrho_k(n)} \text{ P-a.s.}; \;\; \eta_{k,n} = \eta_{k,\varrho_k(n)} \text{ P-a.s.};\\ e^*_{k,n} = e^*_{k,\varrho_k(n)} \text{ P-a.s.}; \;\; q^*_{k,n} = q^*_{k,\varrho_k(n)} \text{ P-a.s.}; \;\; t^*_{k,n} = t^*_{k,\varrho_k(n)} \text{ P-a.s.} \end{cases} \tag{4.73}$$

To match the notation of Theorem 4.11, set

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \underline{\boldsymbol{x}}_n = (\boldsymbol{x}_n, \boldsymbol{y}_n, \boldsymbol{z}_n, \boldsymbol{v}^*_n);\\ \widetilde{\underline{\boldsymbol{q}}}_n = (\boldsymbol{x}_n, \boldsymbol{y}_n, \boldsymbol{z}_n, \boldsymbol{e}^*_n);\\ \underline{\boldsymbol{w}}_n = (\boldsymbol{a}_n, \boldsymbol{b}_n, \boldsymbol{d}_n, \boldsymbol{e}^*_n);\\ \underline{\boldsymbol{w}}^*_n = \big(\boldsymbol{p}^*_n - (C_i x_{i,\vartheta_i(n)})_{i\in I}, \boldsymbol{q}^*_n - (B^{\ell}_k y_{k,\varrho_k(n)})_{k\in K}, \boldsymbol{t}^*_n - (D^{\ell}_k z_{k,\varrho_k(n)})_{k\in K}, \boldsymbol{e}_n\big);\\ \underline{\boldsymbol{q}}_n = \big((x_{i,\vartheta_i(n)})_{i\in I}, (y_{k,\varrho_k(n)})_{k\in K}, (z_{k,\varrho_k(n)})_{k\in K}, (e^*_{k,n})_{k\in K}\big);\\ \underline{\boldsymbol{t}}^*_n = (\boldsymbol{p}^*_n, \boldsymbol{q}^*_n, \boldsymbol{t}^*_n, \boldsymbol{e}_n);\\ (\underline{\boldsymbol{e}}_n, \underline{\boldsymbol{e}}^*_n, \underline{\boldsymbol{f}}^*_n) = (\underline{\boldsymbol{0}}, \underline{\boldsymbol{0}}, \underline{\boldsymbol{0}}). \end{cases} \tag{4.74}$$

Then it follows from (4.21) that, for every $n \in \mathbb{N}$ and every $\underline{z} \in \operatorname{zer} \boldsymbol{\mathcal{S}}$, $\varepsilon_n(\cdot, \underline{z}) = 0$ P-a.s. Next, we observe that, for every $i \in I$ and every $n \in \mathbb{N}$, (4.72), (4.70), (4.69), and [4, Proposition 23.2(ii)] imply that

$$\begin{aligned} a^*_{i,n} - C_i x_{i,\vartheta_i(n)} &= a^*_{i,\vartheta_i(n)} - C_i x_{i,\vartheta_i(n)}\\ &= \gamma^{-1}_{i,\vartheta_i(n)}\big(x_{i,\vartheta_i(n)} - a_{i,\vartheta_i(n)}\big) - l^*_{i,\vartheta_i(n)} - C_i x_{i,\vartheta_i(n)} + Q_i a_{i,\vartheta_i(n)}\\ &\in -s^*_i + A_i a_{i,\vartheta_i(n)} + Q_i a_{i,\vartheta_i(n)}\\ &= -s^*_i + A_i a_{i,n} + Q_i a_{i,n} \quad \text{P-a.s.} \end{aligned} \tag{4.75}$$

and, therefore, that

$$
\begin{aligned}
p^*_{i,n} - \mathsf{C}_i x_{i,\vartheta_i(n)} &= a^*_{i,n} - \mathsf{C}_i x_{i,\vartheta_i(n)} + \mathsf{R}_i \boldsymbol{a}_n + \sum_{k\in K} \mathsf{L}^*_{ki} e^*_{k,n}\\
&\in -\mathsf{s}^*_i + \mathsf{A}_i a_{i,n} + \mathsf{Q}_i a_{i,n} + \mathsf{R}_i \boldsymbol{a}_n + \sum_{k\in K} \mathsf{L}^*_{ki} e^*_{k,n} \quad \text{P-a.s.}
\end{aligned}
\tag{4.76}
$$

Likewise, we derive from (4.73), (4.71), (4.69), and [4, Proposition 23.2(ii)] that

$$
(\forall k \in K)(\forall n \in \mathbb{N}) \quad
\begin{cases}
q^*_{k,n} - \mathsf{B}^c_k y_{k,\varrho_k(n)} \in \mathsf{B}^m_k b_{k,n} + \mathsf{B}^\ell_k b_{k,n} - e^*_{k,n} \quad \text{P-a.s.};\\
t^*_{k,n} - \mathsf{D}^c_k z_{k,\varrho_k(n)} \in \mathsf{D}^m_k d_{k,n} + \mathsf{D}^\ell_k d_{k,n} - e^*_{k,n} \quad \text{P-a.s.};\\
e_{k,n} = \mathsf{r}_k + b_{k,n} + d_{k,n} - \sum_{i\in I} \mathsf{L}_{ki} a_{i,n} \quad \text{P-a.s.}
\end{cases}
\tag{4.77}
$$

In turn, we derive from (4.74) and (4.60) that the sequence $(\underline{\boldsymbol{w}}_n, \underline{\boldsymbol{w}}^*_n)_{n\in\mathbb{N}}$ lies in $\operatorname{gra} \underline{\mathsf{W}}$ P-a.s. Next, using (4.74) and (4.61), we obtain, for every $n \in \mathbb{N}$, $\underline{\boldsymbol{t}}^*_n = \underline{\boldsymbol{w}}^*_n + \underline{\mathsf{C}}\underline{\boldsymbol{q}}_n$ P-a.s. Additionally, (4.69) and (4.72)–(4.74) yield

$$
(\forall n \in \mathbb{N}) \quad \sum_{i\in I} \xi_{i,n} + \sum_{k\in K} \eta_{k,n} = \|\underline{\boldsymbol{w}}_n - \underline{\boldsymbol{q}}_n\|^2 \quad \text{P-a.s.}
\tag{4.78}
$$

Hence, in view of (4.69),

$$
(\forall n \in \mathbb{N}) \quad \Delta_n = \langle \underline{\boldsymbol{x}}_n - \underline{\boldsymbol{w}}_n \mid \underline{\boldsymbol{t}}^*_n \rangle - (4\alpha)^{-1} \|\underline{\boldsymbol{w}}_n - \underline{\boldsymbol{q}}_n\|^2 \quad \text{P-a.s.}
\tag{4.79}
$$

On the other hand,

$$
(\forall i \in I)(\forall k \in K)(\forall n \in \mathbb{N}) \quad
\begin{cases}
\mathsf{R}_i, \mathsf{Q}_i, \mathsf{B}^\ell_k, \mathsf{D}^\ell_k \text{ and } \mathsf{L}_{ki} \text{ are Lipschitzian;}\\
\mathsf{C}_i, \mathsf{B}^c_k, \text{ and } \mathsf{D}^c_k \text{ are cocoercive, hence Lipschitzian;}\\
\mathsf{J}_{\gamma_{i,n}\mathsf{A}_i}, \mathsf{J}_{\mu_{k,n}\mathsf{B}^m_k}, \text{ and } \mathsf{J}_{\nu_{k,n}\mathsf{D}^m_k} \text{ are 1-Lipschitzian.}
\end{cases}
\tag{4.80}
$$

It therefore follows from Assumption 4.21[c], Lemmas 4.7 and 4.8, and an inductive argument that the variables defined in (4.74) belong to $L^2(\Omega, \mathcal{F}, \mathsf{P}; \underline{\mathsf{X}})$. Altogether, taking into account the assumptions, we have shown that (4.69) is a realization of (4.3). In turn, Theorem 4.11(iii)(c) asserts that

$$
\sum_{n\in\mathbb{N}} \mathsf{E}\|\underline{\boldsymbol{x}}_{n+1} - \underline{\boldsymbol{x}}_n\|^2 < +\infty.
\tag{4.81}
$$

(i)–(ii): These follow from Theorem 4.11(ii), (4.81), and (4.74).

(iii)–(iv): Theorem 4.11(iii)(a) implies that $(\underline{\boldsymbol{x}}_n)_{n\in\mathbb{N}}$ is bounded P-a.s. Therefore, arguing as in the proof of [13, Theorem 1],

$$
\big(\widetilde{\underline{\boldsymbol{q}}}_n\big)_{n\in\mathbb{N}}, \quad (\underline{\boldsymbol{w}}_n)_{n\in\mathbb{N}}, \quad \text{and} \quad (\underline{\boldsymbol{t}}^*_n)_{n\in\mathbb{N}} \quad \text{are bounded P-a.s.}
\tag{4.82}
$$

As a result, (4.79) and Theorem 4.11(iii)(d) yield

$$\overline{\lim}\,\big(\langle \underline{\boldsymbol{x}}_n - \underline{\boldsymbol{w}}_n \mid \underline{\boldsymbol{t}}_n^*\rangle - (4\alpha)^{-1}\|\underline{\boldsymbol{w}}_n - \underline{\boldsymbol{q}}_n\|^2\big) = \overline{\lim}\,\Delta_n \leqslant 0 \ \ \mathsf{P}\text{-a.s.} \tag{4.83}$$

Now define

$$\mathsf{L}\colon \mathsf{H} \to \mathsf{G}\colon \mathsf{x} \mapsto \Bigg(\sum_{i\in I} \mathsf{L}_{ki}\mathsf{x}_i\Bigg)_{k\in K}, \quad \text{with adjoint } \ \mathsf{L}^*\colon \mathsf{G} \to \mathsf{H}\colon \mathsf{v}^* \mapsto \Bigg(\sum_{k\in K} \mathsf{L}_{ki}^*\mathsf{v}_k^*\Bigg)_{i\in I}, \tag{4.84}$$

and

$$\underline{\mathsf{U}}\colon \underline{\mathsf{X}} \to \underline{\mathsf{X}}\colon (\mathsf{x},\mathsf{y},\mathsf{z},\mathsf{v}^*) \mapsto \big(\mathsf{L}^*\mathsf{v}^*, -\mathsf{v}^*, -\mathsf{v}^*, -\mathsf{L}\mathsf{x}+\mathsf{y}+\mathsf{z}\big). \tag{4.85}$$

Further, for every $n \in \mathbb{N}$, set

$$\begin{cases} (\forall i \in I) \ \ \mathsf{F}_{i,n} = \gamma_{i,\vartheta_i(n)}^{-1}\mathrm{Id} - \mathsf{Q}_i;\\ (\forall k \in K) \ \ \mathsf{S}_{k,n} = \mu_{k,\varrho_k(n)}^{-1}\mathrm{Id} - \mathsf{B}_k^{\ell}; \ \ \mathsf{T}_{k,n} = \nu_{k,\varrho_k(n)}^{-1}\mathrm{Id} - \mathsf{D}_k^{\ell};\\ \underline{\mathsf{F}}_n\colon \underline{\mathsf{X}} \to \underline{\mathsf{X}}\colon (\mathsf{x},\mathsf{y},\mathsf{z},\mathsf{v}^*) \mapsto \big((\mathsf{F}_{i,n}\mathsf{x}_i)_{i\in I}, (\mathsf{S}_{k,n}\mathsf{y}_k)_{k\in K}, (\mathsf{T}_{k,n}\mathsf{z}_k)_{k\in K}, (\sigma_{k,\varrho_k(n)}^{-1}\mathsf{v}_k^*)_{k\in K}\big) \end{cases} \tag{4.86}$$

and

$$\begin{cases} \widetilde{\underline{\boldsymbol{x}}}_n = \Big((x_{i,\vartheta_i(n)})_{i\in I}, (y_{k,\varrho_k(n)})_{k\in K}, (z_{k,\varrho_k(n)})_{k\in K}, (v_{k,\varrho_k(n)}^*)_{k\in K}\Big);\\ \underline{\boldsymbol{v}}_n^* = \underline{\mathsf{F}}_n\underline{\boldsymbol{x}}_n - \underline{\mathsf{F}}_n\underline{\boldsymbol{w}}_n; \ \ \underline{\boldsymbol{u}}_n^* = \underline{\mathsf{U}}\underline{\boldsymbol{w}}_n - \underline{\mathsf{U}}\underline{\boldsymbol{x}}_n;\\ \underline{\boldsymbol{r}}_n^* = \big((\mathsf{R}_i\boldsymbol{a}_n - \mathsf{R}_i\boldsymbol{x}_n)_{i\in I}, \boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0}\big); \ \ \widetilde{\underline{\boldsymbol{r}}}_n^* = \big((\mathsf{R}_i\boldsymbol{a}_n - \mathsf{R}_i\boldsymbol{x}_{\vartheta_i(n)})_{i\in I}, \boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0}\big);\\ \underline{\boldsymbol{l}}_n^* = \Big(\big(-\sum_{k\in K}\mathsf{L}_{ki}^*v_{k,\vartheta_i(n)}^*\big)_{i\in I}, \big(v_{k,\varrho_k(n)}^*\big)_{k\in K}, \big(v_{k,\varrho_k(n)}^*\big)_{k\in K}, \big(\sum_{i\in I}\mathsf{L}_{ki}x_{i,\varrho_k(n)} - y_{k,\varrho_k(n)} - z_{k,\varrho_k(n)}\big)_{k\in K}\Big). \end{cases} \tag{4.87}$$

Assumptions [a]–[c] in Problem 4.15 and 4.21[a]&[b], together with Lemma 4.10, imply that

$$(\forall n \in \mathbb{N}) \quad \text{the operators} \ \begin{cases} (\mathsf{F}_{i,n})_{i\in I} \ \text{are } (\chi+\sigma)\text{-strongly monotone;}\\ (\mathsf{S}_{k,n})_{k\in K} \ \text{and} \ (\mathsf{T}_{k,n})_{k\in K} \ \text{are } \sigma\text{-strongly monotone.} \end{cases} \tag{4.88}$$

Consequently, in view of (4.86), there exists $\kappa \in \left]0,+\infty\right[$ such that

$$\text{the operators } \ (\underline{\mathsf{F}}_n)_{n\in\mathbb{N}} \ \text{ are } \kappa\text{-Lipschitzian.} \tag{4.89}$$

Next, using the same arguments as in the proof of [13, Theorem 1], we obtain

$$(\forall n \in \mathbb{N}) \quad \underline{\boldsymbol{t}}_n^* = \underline{\mathsf{F}}_n\widetilde{\underline{\boldsymbol{x}}}_n - \underline{\mathsf{F}}_n\underline{\boldsymbol{w}}_n + \widetilde{\underline{\boldsymbol{r}}}_n^* + \underline{\boldsymbol{l}}_n^* + \underline{\mathsf{U}}\underline{\boldsymbol{w}}_n \ \mathsf{P}\text{-a.s.} \tag{4.90}$$

We also observe that, in view of (4.81), (4.70), (4.71), and Assumption 4.18, Proposition 4.20

and Lemma 4.4 imply that

$$(\forall i \in I)(\forall k \in K) \quad \begin{cases} \boldsymbol{x}_{\vartheta_i(n)} - \boldsymbol{x}_n \xrightarrow{\mathsf{P}} \mathbf{0}; & \boldsymbol{x}_{\varrho_k(n)} - \boldsymbol{x}_n \xrightarrow{\mathsf{P}} \mathbf{0}; \\ \boldsymbol{v}^*_{\vartheta_i(n)} - \boldsymbol{v}^*_n \xrightarrow{\mathsf{P}} \mathbf{0}; & \boldsymbol{y}_{\varrho_k(n)} - \boldsymbol{y}_n \xrightarrow{\mathsf{P}} \mathbf{0}; \\ \boldsymbol{z}_{\varrho_k(n)} - \boldsymbol{z}_n \xrightarrow{\mathsf{P}} \mathbf{0}; & \boldsymbol{v}^*_{\varrho_k(n)} - \boldsymbol{v}^*_n \xrightarrow{\mathsf{P}} \mathbf{0}. \end{cases} \tag{4.91}$$

Thus, (4.87), (4.84), and (4.85) yield

$$\underline{\boldsymbol{l}}^*_n + \underline{\mathsf{U}}\underline{\boldsymbol{x}}_n \xrightarrow{\mathsf{P}} \underline{\mathbf{0}}, \tag{4.92}$$

while assumption [e] in Problem 4.15 gives

$$(\forall i \in I) \quad \|\mathsf{R}_i \boldsymbol{x}_{\vartheta_i(n)} - \mathsf{R}_i \boldsymbol{x}_n\| \leqslant \chi \|\boldsymbol{x}_{\vartheta_i(n)} - \boldsymbol{x}_n\| \xrightarrow{\mathsf{P}} 0. \tag{4.93}$$

On the other hand, (4.89), (4.87), and (4.91) yield

$$\|\underline{\mathsf{E}}_n \underline{\widetilde{\boldsymbol{x}}}_n - \underline{\mathsf{E}}_n \underline{\boldsymbol{x}}_n\| \leqslant \kappa \|\underline{\widetilde{\boldsymbol{x}}}_n - \underline{\boldsymbol{x}}_n\| \xrightarrow{\mathsf{P}} 0 \tag{4.94}$$

which, combined with (4.90), (4.87), (4.92), and (4.93) leads to

$$\underline{\boldsymbol{t}}^*_n - (\underline{\boldsymbol{v}}^*_n + \underline{\boldsymbol{r}}^*_n + \underline{\boldsymbol{u}}^*_n) = \underline{\boldsymbol{l}}^*_n + \underline{\mathsf{U}}\underline{\boldsymbol{x}}_n + \underline{\mathsf{E}}_n \underline{\widetilde{\boldsymbol{x}}}_n - \underline{\mathsf{E}}_n \underline{\boldsymbol{x}}_n + \underline{\widetilde{\boldsymbol{r}}}^*_n - \underline{\boldsymbol{r}}^*_n \xrightarrow{\mathsf{P}} \underline{\mathbf{0}}. \tag{4.95}$$

Additionally, (4.74) and (4.91) yield

$$\underline{\widetilde{\boldsymbol{q}}}_n - \underline{\boldsymbol{q}}_n \xrightarrow{\mathsf{P}} \underline{\mathbf{0}}. \tag{4.96}$$

Therefore, by Cauchy–Schwarz and (4.82),

$$\left|\langle \underline{\boldsymbol{w}}_n - \underline{\widetilde{\boldsymbol{q}}}_n \mid \underline{\widetilde{\boldsymbol{q}}}_n - \underline{\boldsymbol{q}}_n \rangle\right| \leqslant \left(\sup_{m \in \mathbb{N}} \|\underline{\boldsymbol{w}}_m\| + \sup_{m \in \mathbb{N}} \|\underline{\widetilde{\boldsymbol{q}}}_m\| \right) \|\underline{\widetilde{\boldsymbol{q}}}_n - \underline{\boldsymbol{q}}_n\| \xrightarrow{\mathsf{P}} 0 \tag{4.97}$$

while, by (4.95),

$$\left|\langle \underline{\boldsymbol{x}}_n - \underline{\boldsymbol{w}}_n \mid \underline{\boldsymbol{t}}^*_n - (\underline{\boldsymbol{v}}^*_n + \underline{\boldsymbol{r}}^*_n + \underline{\boldsymbol{u}}^*_n) \rangle\right| \leqslant \left(\sup_{m \in \mathbb{N}} \|\underline{\boldsymbol{x}}_m\| + \sup_{m \in \mathbb{N}} \|\underline{\boldsymbol{w}}_m\| \right) \|\underline{\boldsymbol{t}}^*_n - (\underline{\boldsymbol{v}}^*_n + \underline{\boldsymbol{r}}^*_n + \underline{\boldsymbol{u}}^*_n)\| \xrightarrow{\mathsf{P}} 0. \tag{4.98}$$

However, it follows from (4.85) and assumption [d] in Problem 4.15 that $\underline{\mathsf{U}}$ is linear and bounded, with $\underline{\mathsf{U}}^* = -\underline{\mathsf{U}}$. It then results from (4.87) that, for every $n \in \mathbb{N}$, $\langle \underline{\boldsymbol{x}}_n - \underline{\boldsymbol{w}}_n \mid \underline{\boldsymbol{u}}^*_n \rangle = 0$ P-a.s. On the other hand, note that, for every $n \in \mathbb{N}$,

$$\begin{aligned} &\langle \underline{\boldsymbol{x}}_n - \underline{\boldsymbol{w}}_n \mid \underline{\boldsymbol{t}}^*_n \rangle - (4\alpha)^{-1} \|\underline{\boldsymbol{w}}_n - \underline{\boldsymbol{q}}_n\|^2 \\ &\quad = \langle \underline{\boldsymbol{x}}_n - \underline{\boldsymbol{w}}_n \mid \underline{\boldsymbol{v}}^*_n + \underline{\boldsymbol{r}}^*_n + \underline{\boldsymbol{u}}^*_n \rangle + \langle \underline{\boldsymbol{x}}_n - \underline{\boldsymbol{w}}_n \mid \underline{\boldsymbol{t}}^*_n - (\underline{\boldsymbol{v}}^*_n + \underline{\boldsymbol{r}}^*_n + \underline{\boldsymbol{u}}^*_n) \rangle - (4\alpha)^{-1} \|\underline{\boldsymbol{w}}_n - \underline{\boldsymbol{q}}_n\|^2 \\ &\quad = \langle \underline{\boldsymbol{x}}_n - \underline{\boldsymbol{w}}_n \mid \underline{\boldsymbol{v}}^*_n + \underline{\boldsymbol{r}}^*_n \rangle + \langle \underline{\boldsymbol{x}}_n - \underline{\boldsymbol{w}}_n \mid \underline{\boldsymbol{t}}^*_n - (\underline{\boldsymbol{v}}^*_n + \underline{\boldsymbol{r}}^*_n + \underline{\boldsymbol{u}}^*_n) \rangle \end{aligned}$$

$$-(4\alpha)^{-1}\Big(\big\|\underline{w}_n-\underline{\widetilde{q}}_n\big\|^2+2\big\langle\underline{w}_n-\underline{\widetilde{q}}_n\,\big|\,\underline{\widetilde{q}}_n-\underline{q}_n\big\rangle+\big\|\underline{\widetilde{q}}_n-\underline{q}_n\big\|^2\Big)\ \text{P-a.s.}\tag{4.99}$$

Moreover, as in [13, Equation (95)], it follows from (4.87), (4.74), (4.86), (4.88), Assumption 4.21[b], and assumption [e] in Problem 4.15 that, for every $n\in\mathbb{N}$,

$$\begin{aligned}&\langle\underline{x}_n-\underline{w}_n\,|\,\underline{v}_n^*+\underline{r}_n^*\rangle-(4\alpha)^{-1}\|\underline{w}_n-\underline{\widetilde{q}}_n\|^2\\&\qquad\geqslant\big(\sigma-(4\alpha)^{-1}\big)\big(\|\boldsymbol{x}_n-\boldsymbol{a}_n\|^2+\|\boldsymbol{y}_n-\boldsymbol{b}_n\|^2+\|\boldsymbol{z}_n-\boldsymbol{d}_n\|^2\big)+\varepsilon\|\boldsymbol{v}_n^*-\boldsymbol{e}_n^*\|^2\ \text{P-a.s.}\end{aligned}\tag{4.100}$$

For every $n\in\mathbb{N}$, let us define

$$\begin{cases}\xi_n=\big(\sigma-(4\alpha)^{-1}\big)\big(\|\boldsymbol{x}_n-\boldsymbol{a}_n\|^2+\|\boldsymbol{y}_n-\boldsymbol{b}_n\|^2+\|\boldsymbol{z}_n-\boldsymbol{d}_n\|^2\big)+\varepsilon\|\boldsymbol{v}_n^*-\boldsymbol{e}_n^*\|^2;\\ \chi_n=\langle\underline{x}_n-\underline{w}_n\,|\,\underline{t}_n^*-(\underline{v}_n^*+\underline{r}_n^*+\underline{u}_n^*)\rangle-(4\alpha)^{-1}\big(2\langle\underline{w}_n-\underline{\widetilde{q}}_n\,|\,\underline{\widetilde{q}}_n-\underline{q}_n\rangle+\|\underline{\widetilde{q}}_n-\underline{q}_n\|^2\big).\end{cases}\tag{4.101}$$

Then $\inf_{n\in\mathbb{N}}\xi_n\geqslant 0$ P-a.s. Moreover, (4.99) and (4.100) imply that, for every $n\in\mathbb{N}$, $\xi_n+\chi_n\leqslant\Delta_n$ P-a.s. In addition, $\varlimsup\Delta_n\leqslant 0$ P-a.s. by (4.83) and $\chi_n\overset{\mathrm{P}}{\to}0$ by (4.96)–(4.98). Therefore, in view of Lemma 4.6, $\xi_n\overset{\mathrm{P}}{\to}0$ and therefore

$$\boldsymbol{x}_n-\boldsymbol{a}_n\overset{\mathrm{P}}{\to}\boldsymbol{0},\quad \boldsymbol{y}_n-\boldsymbol{b}_n\overset{\mathrm{P}}{\to}\boldsymbol{0},\quad \boldsymbol{z}_n-\boldsymbol{d}_n\overset{\mathrm{P}}{\to}\boldsymbol{0},\quad \boldsymbol{v}_n^*-\boldsymbol{e}_n^*\overset{\mathrm{P}}{\to}\boldsymbol{0},\tag{4.102}$$

which establishes (iii). In turn, (4.74) and (4.89) force

$$\underline{x}_n-\underline{w}_n\overset{\mathrm{P}}{\to}\underline{\boldsymbol{0}}\quad\text{and}\quad(\forall n\in\mathbb{N})\quad\|\underline{E}_n\underline{x}_n-\underline{E}_n\underline{w}_n\|\leqslant\kappa\|\underline{x}_n-\underline{w}_n\|.\tag{4.103}$$

Hence,

$$\underline{E}_n\underline{x}_n-\underline{E}_n\underline{w}_n\overset{\mathrm{P}}{\to}\underline{\boldsymbol{0}}.\tag{4.104}$$

Likewise, (4.91) yields $\underline{w}_n-\underline{q}_n\overset{\mathrm{P}}{\to}\underline{\boldsymbol{0}}$. Further, we infer from (4.87), (4.102), and Problem 4.15[e] that

$$\|\underline{r}_n^*\|^2=\|\mathrm{R}\boldsymbol{a}_n-\mathrm{R}\boldsymbol{x}_n\|^2\leqslant\chi^2\|\boldsymbol{a}_n-\boldsymbol{x}_n\|^2\overset{\mathrm{P}}{\to}0.\tag{4.105}$$

As a result, it follows from (4.87), (4.95), (4.104), and (4.105) that

$$\underline{t}_n^*=\Big(\underline{t}_n^*-\big(\underline{v}_n^*+\underline{r}_n^*+\underline{u}_n^*\big)\Big)+\big(\underline{E}_n\underline{x}_n-\underline{E}_n\underline{w}_n\big)+\underline{U}(\underline{w}_n-\underline{x}_n)+\underline{r}_n^*\overset{\mathrm{P}}{\to}\underline{\boldsymbol{0}}.\tag{4.106}$$

Altogether,

$$\underline{x}_n-\underline{w}_n-\underline{e}_n\overset{\mathrm{P}}{\to}\underline{\boldsymbol{0}},\ \underline{w}_n+\underline{e}_n-\underline{q}_n\overset{\mathrm{P}}{\to}\underline{\boldsymbol{0}},\ \text{ and }\ \underline{w}_n^*+\underline{e}_n^*+\underline{C}\underline{q}_n\overset{\mathrm{P}}{\to}\underline{\boldsymbol{0}}\tag{4.107}$$

and Theorem 4.11(iii)(f) therefore guarantees that there exists a $\operatorname{zer}\boldsymbol{\mathcal{S}}$-valued random variable $\underline{\overline{x}}=(\overline{\boldsymbol{x}},\overline{\boldsymbol{y}},\overline{\boldsymbol{z}},\overline{\boldsymbol{v}}^*)$ such that $\underline{x}_n\to\underline{\overline{x}}$ P-a.s. This and (4.102) imply that, for every $i\in I$ and every $k\in K$, $x_{i,n}\to\overline{x}_i$ P-a.s., $a_{i,n}\to\overline{x}_i$ P-a.s., and $v_{k,n}^*\to\overline{v}_k^*$ P-a.s. Finally, Proposition 4.17(ii) asserts that $\overline{\boldsymbol{x}}$ solves (4.56) P-a.s. and that $\overline{\boldsymbol{v}}^*$ solves (4.57) P-a.s. ☐

Remark 4.24 Here are some observations pertaining to Theorem 4.23.

(i) There does not exist any result on stochastic algorithms for solving Problem 4.15 with random block selection or random relaxations. In the case of deterministic relaxations $(\lambda_n)_{n\in\mathbb{N}}$ in $]0,2[$ and deterministic blocks selection (see Example 4.19(i)), Theorem 4.23 appears in [13, Theorem 1(iv)].

(ii) For notational simplicity, we have not considered stochastic errors in the evaluations of the single-valued operators and the resolvents, as is done in the simpler settings of Theorem 4.13 and [20, 22, 33, 39, 43]. For this reason, we have implemented Theorem 4.11 with $(\forall n \in \mathbb{N})(\forall \underline{z} \in \operatorname{zer} \boldsymbol{\mathcal{S}})$ $\varepsilon_n(\cdot, \underline{z}) = 0$ P-a.s. Such stochastic errors can be introduced in Algorithm 4.22 under suitable summability conditions to guarantee that $(\forall \underline{z} \in \operatorname{zer} \boldsymbol{\mathcal{S}})$ $\sum_{n\in\mathbb{N}} \mathsf{E}\varepsilon_n(\cdot, \underline{z})\mathsf{E}\lambda_n < +\infty$.

(iii) The convergence results invoke Theorem 4.11(iii)(f), which requires Euclidean spaces. Note that we cannot use Theorem 4.11(iii)(e), which would provide weak convergence in general Hilbert spaces, because the convergences in (4.107) are only in probability and not almost sure.

4.2.5.3 Application to multivariate minimization

We consider a multivariate composite minimization problem.

Problem 4.25 Let $(\mathsf{H}_i)_{i\in I}$ and $(\mathsf{G}_k)_{k\in K}$ be finite families of Euclidean spaces with respective direct sums $\mathsf{H} = \bigoplus_{i\in I} \mathsf{H}_i$ and $\mathsf{G} = \bigoplus_{k\in K} \mathsf{G}_k$. Denote by $\mathsf{x} = (\mathsf{x}_i)_{i\in I}$ a generic element in H. For every $i \in I$ and every $k \in K$, let $\mathsf{f}_i \in \Gamma_0(\mathsf{H}_i)$, let $\alpha_i \in]0, +\infty[$, let $\varphi_i \colon \mathsf{H}_i \to \mathbb{R}$ be convex and differentiable with a $(1/\alpha_i)$-Lipschitzian gradient, let $\mathsf{g}_k \in \Gamma_0(\mathsf{G}_k)$, let $\mathsf{h}_k \in \Gamma_0(\mathsf{G}_k)$, let $\beta_k \in]0, +\infty[$, let $\psi_k \colon \mathsf{G}_k \to \mathbb{R}$ be convex and differentiable with a $(1/\beta_k)$-Lipschitzian gradient, and suppose that $\mathsf{L}_{ki} \colon \mathsf{H}_i \to \mathsf{G}_k$ is linear. In addition, let $\chi \in [0, +\infty[$ and let $\Theta \colon \mathsf{H} \to \mathbb{R}$ be convex and differentiable with a χ-Lipschitzian gradient. The objective is to

$$\underset{\mathsf{x}\in\mathsf{H}}{\text{minimize}} \;\; \Theta(\mathsf{x}) + \sum_{i\in I}\big(\mathsf{f}_i(\mathsf{x}_i) + \varphi_i(\mathsf{x}_i)\big) + \sum_{k\in K}\big((\mathsf{g}_k + \psi_k) \,\square\, \mathsf{h}_k\big)\Bigg(\sum_{i\in I} \mathsf{L}_{ki}\mathsf{x}_i\Bigg). \tag{4.108}$$

We denote by $\mathscr{P}$ the set of solutions to (4.108).

Algorithm 4.26 Consider the setting of Problem 4.25 and suppose that Assumptions 4.18 and 4.21 are in force with, for every $i \in I$ and every $k \in K$, $\alpha_i^{c} = \alpha_i$, $\beta_k^{c} = \beta_k$, $\alpha_i^{\ell} = \beta_k^{\ell} = \delta_k^{c} = \delta_k^{\ell} = 0$, and $\nabla_i\Theta$ denotes the partial derivative of Θ relative to H_i. Iterate as in (4.69), where the following adjustments are made

$$\begin{cases} J_{\gamma_{i,n}A_i} = \operatorname{prox}_{\gamma_{i,n}\mathsf{f}_i};\;\; C_i = \nabla\varphi_i;\;\; Q_i = 0;\;\; R_i = \nabla_i\Theta;\;\; s_i^* = 0;\\ J_{\mu_{k,n}B_k^{m}} = \operatorname{prox}_{\mu_{k,n}\mathsf{g}_k};\;\; B_k^{c} = \nabla\psi_k;\;\; J_{\nu_{k,n}D_k^{m}} = \operatorname{prox}_{\nu_{k,n}\mathsf{h}_k};\;\; B_k^{\ell} = D_k^{c} = D_k^{\ell} = 0;\;\; r_k = 0. \end{cases} \tag{4.109}$$

Corollary 4.27 *Consider the setting of Algorithm 4.26. Suppose that* $\inf_{n\in\mathbb{N}} \mathsf{E}(\lambda_n(2-\lambda_n)) > 0$ *and that a Kuhn–Tucker point* $(\widetilde{x}, \widetilde{v}^*) \in \mathsf{H} \times \mathsf{G}$ *exists, that is,*

$$(\forall i \in I)(\forall k \in K) \quad \begin{cases} -\sum_{j\in K} \mathsf{L}_{ji}^*\widetilde{\mathsf{v}}_j^* \in \partial \mathsf{f}_i(\widetilde{\mathsf{x}}_i) + \nabla \varphi_i(\widetilde{\mathsf{x}}_i) + \nabla_i \Theta(\widetilde{\mathsf{x}}); \\ \sum_{j\in I} \mathsf{L}_{kj}\widetilde{\mathsf{x}}_j \in \partial\big(\mathsf{g}_k^* \,\square\, \psi_k^*\big)(\widetilde{\mathsf{v}}_k^*) + \partial \mathsf{h}_k^*(\widetilde{\mathsf{v}}_k^*). \end{cases} \tag{4.110}$$

Then there exists a $\mathscr{P}$*-valued random variable* $\overline{\boldsymbol{x}}$ *such that, for every* $i \in I$, $x_{i,n} \to \overline{x}_i$ P-a.s.

4.2.6 Randomized block-iterative Kuhn–Tucker projective splitting

We revisit a multivariate primal-dual inclusion problem studied in [19] and randomize the algorithm proposed there to solve it. See also [18, Section 9] and [28] for further discussions on the deterministic setting.

Problem 4.28 Let $(\mathsf{H}_i)_{i\in I}$ and $(\mathsf{G}_k)_{k\in K}$ be finite families of Euclidean spaces with respective direct sums $\mathsf{H} = \bigoplus_{i\in I} \mathsf{H}_i$ and $\mathsf{G} = \bigoplus_{k\in K} \mathsf{G}_k$. Denote by $\mathsf{x} = (\mathsf{x}_i)_{i\in I}$ a generic element in H. For every $i \in I$ and every $k \in K$, $\mathsf{A}_i \colon \mathsf{H}_i \to 2^{\mathsf{H}_i}$ is maximally monotone, $\mathsf{B}_k \colon \mathsf{G}_k \to 2^{\mathsf{G}_k}$ is maximally monotone, and $\mathsf{L}_{ki} \colon \mathsf{H}_i \to \mathsf{G}_k$ is linear. The objective is to solve the primal problem

$$\text{find } \overline{\mathsf{x}} \in \mathsf{H} \text{ such that } (\forall i \in I)\ 0 \in \mathsf{A}_i\overline{\mathsf{x}}_i + \sum_{k\in K} \mathsf{L}_{ki}^*\left(\mathsf{B}_k\left(\sum_{j\in I} \mathsf{L}_{kj}\overline{\mathsf{x}}_j\right)\right) \tag{4.111}$$

and the associated dual problem

$$\text{find } \overline{\mathsf{v}}^* \in \mathsf{G} \text{ such that } (\exists\, \mathsf{x} \in \mathsf{H}) \quad \begin{cases} (\forall i \in I) & \mathsf{x}_i \in \mathsf{A}_i^{-1}\left(-\sum_{k\in K} \mathsf{L}_{ki}^*\overline{\mathsf{v}}_k^*\right); \\ (\forall k \in K) & \sum_{i\in I} \mathsf{L}_{ki}\mathsf{x}_i \in \mathsf{B}_k^{-1}\overline{\mathsf{v}}_k^*. \end{cases} \tag{4.112}$$

Finally, $\mathscr{P}$ denotes the set of solutions to (4.111) and $\mathscr{D}$ the set of solutions to (4.112).

The Kuhn–Tucker operator associated with Problem 4.28 is [18, Equation (9.18)]

$$\mathsf{W} \colon \mathsf{H} \oplus \mathsf{G} \to 2^{\mathsf{H}\oplus\mathsf{G}} \colon (\mathsf{x}, \mathsf{v}^*) \mapsto \left(\bigtimes_{i\in I} \left(\mathsf{A}_i\mathsf{x}_i + \sum_{k\in K} \mathsf{L}_{ki}^*\mathsf{v}_k^*\right), \bigtimes_{k\in K} \left(\mathsf{B}_k^{-1}\mathsf{v}_k^* - \sum_{i\in I} \mathsf{L}_{ki}\mathsf{x}_i\right) \right) \tag{4.113}$$

As shown in [18, Lemma 9.7(ii)], $\operatorname{zer} \mathsf{W} \subset \mathscr{P} \times \mathscr{D}$. We can therefore approach Problem 4.28 as an instance of Problem 4.1 with $\mathsf{C} = \mathbf{0}$ and then α can be selected arbitrarily large. By applying Theorem 4.11 in this context, we obtain a randomized version of the deterministic algorithm of [19], which relied on Algorithm 4.2. To this end, let us make the following assumption.

Assumption 4.29 *In the setting of Problem 4.28, set* $\varepsilon \in \,]0,1[$ *and suppose that for every* $i \in I$, *every* $k \in K$, *and every* $n \in \mathbb{N}$, $\gamma_{i,n} \in [\varepsilon, 1/\varepsilon]$, $\mu_{k,n} \in [\varepsilon, 1/\varepsilon]$, $x_{i,0} \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H}_i)$, *and* $v^*_{k,0} \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{G}_k)$.

Algorithm 4.30 Consider the setting of Problem 4.28 and suppose that Assumptions 4.18 and 4.29 are in force. Let $\rho \in [2, +\infty[$ and iterate

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\text{for every } i \in I_n \\
\left\lfloor
\begin{array}{l}
l^*_{i,n} = \sum_{k\in K} L^*_{ki} v^*_{k,n}; \\
a_{i,n} = J_{\gamma_{i,n} A_i}\big(x_{i,n} - \gamma_{i,n} l^*_{i,n}\big);\;\; a^*_{i,n} = \gamma_{i,n}^{-1}(x_{i,n} - a_{i,n}) - l^*_{i,n};
\end{array}
\right. \\
\text{for every } i \in I \smallsetminus I_n \\
\left\lfloor
\begin{array}{l}
a_{i,n} = a_{i,n-1};\;\; a^*_{i,n} = a^*_{i,n-1};
\end{array}
\right. \\
\text{for every } k \in K_n \\
\left\lfloor
\begin{array}{l}
l_{k,n} = \sum_{i\in I} L_{ki} x_{i,n}; \\
b_{k,n} = J_{\mu_{k,n} B_k}\big(l_{k,n} + \mu_{k,n} v^*_{k,n}\big);\;\; b^*_{k,n} = v^*_{k,n} + \mu_{k,n}^{-1}(l_{k,n} - b_{k,n});
\end{array}
\right. \\
\text{for every } k \in K \smallsetminus K_n \\
\left\lfloor
\begin{array}{l}
b_{k,n} = b_{k,n-1};\;\; b^*_{k,n} = b^*_{k,n-1};
\end{array}
\right. \\
\text{for every } i \in I \\
\left\lfloor
\begin{array}{l}
t^*_{i,n} = a^*_{i,n} + \sum_{k\in K} L^*_{ki} b^*_{k,n};
\end{array}
\right. \\
\text{for every } k \in K \\
\left\lfloor
\begin{array}{l}
t_{k,n} = b^*_{k,n} + \sum_{i\in I} L_{ki} a^*_{i,n};
\end{array}
\right. \\
\Delta_n = \sum_{i\in I}\big(\langle x_{i,n} \mid t^*_{i,n}\rangle - \langle a_{i,n} \mid a^*_{i,n}\rangle\big) + \sum_{k\in K}\big(\langle t_{k,n} \mid v^*_{k,n}\rangle + \langle b_{k,n} \mid b^*_{k,n}\rangle\big); \\
\theta_n = \dfrac{1_{[\Delta_n>0]}\Delta_n}{\sum_{i\in I}\|t^*_{i,n}\|^2 + \sum_{k\in K}\|t_{k,n}\|^2 + 1_{[\Delta_n\leqslant 0]}}; \\
\text{take } \lambda_n \in L^\infty(\Omega, \mathcal{F}, \mathsf{P}; [\varepsilon, \rho]) \\
\text{for every } i \in I \\
\left\lfloor
\begin{array}{l}
x_{i,n+1} = x_{i,n} - \lambda_n \theta_n t^*_{i,n};
\end{array}
\right. \\
\text{for every } k \in K \\
\left\lfloor
\begin{array}{l}
v^*_{k,n+1} = v^*_{k,n} - \lambda_n \theta_n t_{k,n}.
\end{array}
\right.
\end{array}
\right.
\end{array}
\tag{4.114}
$$

The convergence properties of Algorithm 4.30 are established in the following theorem.

Theorem 4.31 *Consider the setting of Algorithm 4.30. Suppose that* $\mathscr{D} \neq \varnothing$ *and* $\inf_{n\in\mathbb{N}} \mathsf{E}(\lambda_n(2 - \lambda_n)) > 0$. *Then there exist a* $\mathscr{P}$*-valued random variable* $\overline{\boldsymbol{x}}$ *and a* $\mathscr{D}$*-valued random variable* $\overline{\boldsymbol{v}}^*$ *such that, for every* $i \in I$ *and every* $k \in K$, $x_{i,n} \to \overline{x}_i$ P-a.s. *and* $v^*_{k,n} \to \overline{v}^*_k$ P-a.s.

Proof. (Sketch) We apply Theorem 4.11 to find a zero (x, v^*) of W following the deterministic pattern of the proof of [19, Theorem 13] and using probabilistic arguments made in the proof of Theorem 4.23, which shares the same Assumption 4.18 and involves a more sophisticated version of Assumption 4.29. ☐

Remark 4.32 We complement Theorem 4.31 with the following observations.

(i) In the case of deterministic relaxations $(\lambda_n)_{n\in\mathbb{N}}$ in $]0,2[$ and deterministic blocks selection, Theorem 4.31 appears in [19, Theorem 13].

(ii) A stochastic block-iterative algorithm for solving Problem 4.28 was proposed in [22, Corollary 5.3], with almost sure convergence of its iterates. This algorithm involves deterministic relaxations in $]0,2[$ and necessitates inversions to handle the linear operators. In the case when $\mathtt{I}$ is a singleton, further algorithms with the same features were proposed in [20]. The algorithm of [39, Proposition 4.6] also guarantees almost sure convergence of the iterates but it requires knowledge of the norms of linear operators. The same comments apply to the algorithm of [14, Theorem 2.1 and Algorithm 3.1], which considers the minimization case with $\mathtt{I}$ as a singleton. Additionally, none of these prior works show convergence in L^2, nor can they benefit from adaptive strategies as in Assumption 4.18 since their block-selection distributions remain constant throughout the iterations.

(iii) As in Remark 4.24, stochastic errors can be introduced in the evaluations of the resolvents in Algorithm 4.30.

Acknowledgement

This work was supported by the National Science Foundation under grant DMS-2513409.

References

[1] G. Alduncin, Multidomain optimal control of variational subpotential mixed evolution inclusions, *Appl. Math. Optim.*, vol. 88, art. 35, 2023.

[2] H. Attouch, L. M. Briceño–Arias, and P. L. Combettes, A strongly convergent primal-dual method for nonoverlapping domain decomposition, *Numer. Math.*, vol. 133, pp. 433–470, 2016.

[3] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, Optimization with sparsity-inducing penalties, *Found. Trends Machine Learn.*, vol. 4, pp. 1–106, 2012.

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

[5] H. Brézis and P. L. Lions, Produits infinis de résolvantes, *Israel J. Math.*, vol. 29, pp. 329–345, 1978.

[6] L. M. Briceño–Arias, G. Chierchia, E. Chouzenoux, and J.-C. Pesquet, A random block-coordinate Douglas–Rachford splitting method with low computational complexity for binary logistic regression, *Comput. Optim. Appl.*, vol. 72, pp. 707–726, 2019.

[7] L. M. Briceño-Arias and D. Davis, Forward-backward-half forward algorithm for solving monotone inclusions, *SIAM J. Optim.*, vol. 28, pp. 2839–2871, 2018.

[8] L. M. Briceño–Arias, D. Kalise, and F. J. Silva, Proximal methods for stationary mean field games with local couplings, *SIAM J. Control Optim.*, vol. 56, pp. 801–836, 2018.

[9] M. N. Bùi, A decomposition method for solving multicommodity network equilibria, *Oper. Res. Lett.*, vol. 50, pp. 40–44, 2022.

[10] M. N. Bùi, A block-activated decomposition algorithm for multi-stage stochastic variational inequalities, arxiv, 2025. `https://arxiv.org/abs/2509.26198`

[11] M. N. Bùi and P. L. Combettes, Warped proximal iterations for monotone inclusions, *J. Math. Anal. Appl.*, vol. 491, art. 124315, 2020.

[12] M. N. Bùi and P. L. Combettes, Analysis and numerical solution of a modular convex Nash equilibrium problem, *J. Convex Anal.*, vol. 29, pp. 1007–1021, 2022.

[13] M. N. Bùi and P. L. Combettes, Multivariate monotone inclusions in saddle form, *Math. Oper. Res.*, vol. 47, pp. 1082–1109, 2022.

[14] A. Chambolle, C. Delplancke, M. J. Ehrhardt, C.-B. Schönlieb, and J. Tang, Stochastic primal-dual hybrid gradient algorithm with adaptive step sizes, *J. Math. Imaging Vision*, vol. 66, pp. 294–313, 2024.

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

[16] P. L. Combettes, Iterative construction of the resolvent of a sum of maximal monotone operators, *J. Convex Anal.*, vol. 16, pp. 727–748, 2009.

[17] P. L. Combettes, Systems of structured monotone inclusions: Duality, algorithms, and applications, *SIAM J. Optim.*, vol. 23, pp. 2420–2447, 2013.

[18] P. L. Combettes, The geometry of monotone operator splitting methods, *Acta Numer.*, vol. 33, pp. 487–632, 2024.

[19] P. L. Combettes and J. Eckstein, Asynchronous block-iterative primal-dual decomposition methods for monotone inclusions, *Math. Program.*, vol. B168, pp. 645–672, 2018.

[20] P. L. Combettes and J. I. Madariaga, Almost-surely convergent randomly activated monotone operator splitting methods, *SIAM J. Imaging Sci.*, vol. 18, pp. 2177–2205, 2025.

[21] P. L. Combettes and J. I. Madariaga, A geometric framework for stochastic iterations, *Math. Comp.*, to appear.

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

[23] P. L. Combettes and J.-C. Pesquet, Deep neural network structures solving variational inequalities, *Set-Valued Var. Anal.*, vol. 28, pp. 491–518, 2020.

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

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

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

[27] D. Davis and W. Yin, A three-operator splitting scheme and its optimization applications, *Set-Valued Var. Anal.*, vol. 25, pp. 829–858, 2017.

[28] J. Eckstein, A simplified form of block-iterative operator splitting and an asynchronous algorithm resembling the multi-block alternating direction method of multipliers, *J. Optim. Theory Appl.*, vol. 173, pp. 155–182, 2017.

[29] J. Eckstein and D. P. Bertsekas, On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators, *Math. Program.*, vol. 55, pp. 293–318, 1992.

[30] J. Eckstein, J.-P. Watson, and D. L. Woodruff, Projective hedging algorithms for multistage stochastic programming, supporting distributed and asynchronous implementation, *Oper. Res.*, vol. 73, pp. 311–324, 2025.

[31] P. Giselsson, Nonlinear forward-backward splitting with projection correction, *SIAM J. Optim.*, vol. 31, pp. 2199–2226, 2021.

[32] E. G. Gol'shtein and N. V. Tret'yakov, Modified Lagrangians in convex programming and their generalizations, *Math. Program. Studies*, vol. 10, pp. 86–97, 1979.

[33] P. R. Johnstone, J. Eckstein, T. Flynn, and S. Yoo, Stochastic projective splitting, *Comput. Optim. Appl.*, vol. 87, pp. 397–437, 2024.

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

[35] J. Lieutaud, Approximations d'opérateurs monotones par des méthodes de splitting, in *Theory and Applications of Monotone Operators* (A. Ghizzetti, ed.), Edizioni Oderisi, pp. 259–264, 1969.

[36] B. Martinet, Régularisation d'inéquations variationnelles par approximations successives, *Rev. Fr. Inform. Rech. Oper.*, vol. 4, pp. 154–158, 1970.

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

[38] L. Pavel, On operator theory and applications in game theory, *Annu. Rev. Control Robot. Autonom. Syst.*, published online 2025-10-08.

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

[40] H. Raguet, A note on the forward-Douglas-Rachford splitting for monotone inclusion and convex optimization, *Optim. Lett.*, vol. 13, pp. 717–740, 2019.

[41] R. T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM J. Control Optim.*, vol. 14, pp. 877–898, 1976.

[42] R. T. Rockafellar, Monotone relations and network equilibrium, in: *Variational Inequalities and Network Equilibrium Problems*, (F. Giannessi and A. Maugeri, eds.), pp. 271–288. Plenum Press, New York, 1995.

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

[44] A. N. Shiryaev, *Probability–1*, 3rd ed. Springer, New York, 2016.

[45] J. E. Spingarn, Partial inverse of a monotone operator, *Appl. Math. Optim.*, vol. 10, pp. 247–265, 1983.

[46] P. Tseng, Applications of a splitting algorithm to decomposition in convex programming and variational inequalities, *SIAM J. Control Optim.*, vol. 29, pp. 119–138, 1991.

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

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

Chapter 5

CONVERGENCE OF THE ITERATES OF THE STOCHASTIC PROXIMAL GRADIENT METHOD

5.1 Introduction and context

This chapter is dedicated to answering question **(Q4)** of Chapter 1. We derive almost sure convergence of the sequence generated by the stochastic proximal gradient method, under suitable assumptions.

This chapter presents the following journal article:

J. I. Madariaga, Convergence of the iterates of the stochastic proximal gradient method, submitted.

5.2 Article: Convergence of the iterates of the stochastic proximal gradient method

Abstract. We propose a novel study of the stochastic proximal gradient method for minimizing the sum of two convex functions, one of which is smooth. Under suitable assumptions and without requiring any boundedness or control of the variance of the random variables, we derive the almost sure convergence and the convergence in the mean of the iterates to a solution of the minimization problem. The results are applied to classification and convex feasibility problems.

5.2.1 Introduction

Let H be a Euclidean space and let $(\Omega, \mathcal{F}, \mathsf{P})$ be a complete probability space. We consider the minimization of the sum of two functions in $\Gamma_0(\mathsf{H})$, one of which is smooth. As shown in [10], this framework models many problems in applied mathematics and engineering.

Problem 5.1 Let $\beta \in \left]0, +\infty\right[$, let $\mathsf{f} \in \Gamma_0(\mathsf{H})$, and let $\mathsf{g} \colon \mathsf{H} \to \mathbb{R}$ be convex, differentiable, and such that $\nabla \mathsf{g}$ is β-Lipschitzian, with $\operatorname{Argmin}(\mathsf{f} + \mathsf{g}) \neq \varnothing$. The task is to

$$\underset{\mathsf{x} \in \mathsf{H}}{\text{minimize}} \;\; \mathsf{f}(\mathsf{x}) + \mathsf{g}(\mathsf{x}). \tag{5.1}$$

A standard approach to solve this problem is to use the forward-backward algorithm, also called the proximal-gradient algorithm in this context: Given $\mathsf{x}_0 \in \mathsf{H}$ and a sequence $(\gamma_n)_{n\in\mathbb{N}}$ in $\left]0, +\infty\right[$ such that $0 < \inf \gamma_n \leqslant \sup \gamma_n < 2/\beta$, iterate

$$(\forall n \in \mathbb{N}) \quad \mathsf{x}_{n+1} = \operatorname{prox}_{\gamma_n \mathsf{f}}\big(\mathsf{x}_n - \gamma_n \nabla \mathsf{g}(\mathsf{x}_n)\big). \tag{5.2}$$

Then $(\mathsf{x}_n)_{n\in\mathbb{N}}$ is guaranteed to converge to a solution to Problem 5.1 [16, 27]. The implementation of this method requires both the proximity operator of f and the gradient of g to be numerically tractable. However, evaluating these operators could be computationally expensive or even impossible. This paper investigates the following version of Problem 5.1.

Problem 5.2 In the context of Problem 5.1, let $(\mathsf{K}, \mathcal{K})$ be a measurable space. For every $\mathsf{k} \in \mathsf{K}$, let $\mathsf{f}_\mathsf{k} \in \Gamma_0(\mathsf{H})$ and let $\mathsf{g}_\mathsf{k} \colon \mathsf{H} \to \mathbb{R}$ be convex and differentiable. Further, let $k \colon (\Omega, \mathcal{F}, \mathsf{P}) \to (\mathsf{K}, \mathcal{K})$ be a random variable such that

$$\begin{cases} (\forall \mathsf{x} \in \operatorname{dom} \mathsf{f}) & \mathsf{f}(\mathsf{x}) = \displaystyle\int_\Omega \mathsf{f}_{k(\omega)}(\mathsf{x}) \mathsf{P}(d\omega); \\ (\forall \mathsf{x} \in \mathsf{H}) & \mathsf{g}(\mathsf{x}) = \displaystyle\int_\Omega \mathsf{g}_{k(\omega)}(\mathsf{x}) \mathsf{P}(d\omega). \end{cases} \tag{5.3}$$

The contribution of this paper is to provide new results on the asymptotic behavior of the following stochastic version of the proximal-gradient method for solving Problem 5.2.

Algorithm 5.3 In the setting of Problem 5.2, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $\left]0, +\infty\right[$ and let $x_0 \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. Iterate

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} k_n \text{ is a copy of } k \text{ and is independent of } \sigma(x_0, \ldots, x_n) \\ x_{n+1} = \operatorname{prox}_{\gamma_n \mathsf{f}_{k_n}}\big(x_n - \gamma_n \nabla \mathsf{g}_{k_n}(x_n)\big). \end{array} \right. \end{array} \tag{5.4}$$

Algorithm 5.3 can be interpreted as the inexact version

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \operatorname{prox}_{\gamma_n \mathsf{f}}\big(x_n - \gamma_n(\nabla \mathsf{g}(x_n) + b_n)\big) + a_n, \tag{5.5}$$

of (5.2), in which

$$(\forall n \in \mathbb{N}) \quad \begin{cases} a_n &= \operatorname{prox}_{\gamma_n f_{k_n}}\big(x_n - \gamma_n \nabla g_{k_n}(x_n)\big) - \operatorname{prox}_{\gamma_n f}\big(x_n - \gamma_n \nabla g_{k_n}(x_n)\big); \\ b_n &= \nabla g_{k_n}(x_n) - \nabla g(x_n). \end{cases} \tag{5.6}$$

Thus, we can establish the asymptotic behavior of Algorithm 5.3 through some stochastic inexact version of the forward-backward algorithm; see, e.g., [12, 14]. However, the general frameworks established in these works do not make use of the structure of the functions in (5.3) and, instead, they rely on strong conditions on the sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ such as convergence to zero and boundedness of their variance.

Algorithm 5.3 has been studied in the case of a deterministic function f, i.e., when only ∇g is randomly approximated [2, 6, 17, 19, 24, 28], and in the case when $g = 0$ [1, 18, 21, 26]. However, Algorithm 5.3 in its full generality has been less explored. Existing analyses either do not prove almost sure convergence [20, 22] or rely on restrictive assumptions, such as the uniform boundedness of all gradients and subgradients [4] or the existence of solutions with subgradients in L^{2p} [5]. In the present paper, we establish almost-sure and L^1 convergence of the sequence generated by Algorithm 5.3 under much weaker assumptions.

The rest of the paper is organized as follows. Section 5.2.2 introduces the general notation and the preliminary results used throughout the manuscript. Section 5.2.3 establishes the convergence of Algorithm 5.3 to a solution of Problem 5.2 under mild conditions. Section 5.2.4 proposes an application to mixed-loss classification problems and Section 5.2.5 to inconsistent convex feasibility problems.

5.2.2 Notation

Throughout, H is a Euclidean space with identity operator Id, scalar product $\langle\cdot\,|\,\cdot\rangle$, and associated norm $\|\cdot\|$. Let C be a nonempty closed convex subset of H. Then ι_C denotes the indicator function of C, proj_C the projection operator onto C, N_C the normal cone to C, and $d_C\colon x \mapsto \inf_{y\in C}\|x-y\|$ the distance function of C. The class of lower semicontinuous convex functions $h\colon \mathsf{H} \to \left]-\infty,+\infty\right]$ such that $\operatorname{dom} h = \big\{x \in \mathsf{H} \mid h(x) < +\infty\big\} \neq \varnothing$ is denoted by $\Gamma_0(\mathsf{H})$. Let $h \in \Gamma_0(\mathsf{H})$. The subdifferential of h at $x \in \mathsf{H}$ is the set $\partial h(x) = \big\{u \in \mathsf{H} \mid (\forall z \in \mathsf{H})\ \langle z-x \mid u\rangle + h(x) \leqslant h(z)\big\}$, the element of minimal norm in $\partial h(x)$ is ${}^0\partial h(x)$, and the proximity operator of h is

$$\operatorname{prox}_h\colon \mathsf{H} \to \mathsf{H}\colon x \mapsto \underset{z\in\mathsf{H}}{\operatorname{argmin}}\Big(h(z) + \frac{1}{2}\|x-z\|^2\Big). \tag{5.7}$$

Let $(\Xi, \mathcal{G})$ be a measurable space. A Ξ-valued random variable is a measurable mapping $x\colon (\Omega, \mathcal{F}, \mathsf{P}) \to (\Xi, \mathcal{G})$. Given $x\colon \Omega \to \Xi$ and $S \in \mathcal{G}$, we set $[x \in S] = \big\{\omega \in \Omega \mid x(\omega) \in S\big\}$. Let x and y be random variables from $(\Omega, \mathcal{F}, \mathsf{P})$ to $(\Xi, \mathcal{G})$. Then y is a copy of x if, for every $S \in \mathcal{G}$, $\mathsf{P}([x \in S]) = \mathsf{P}([y \in S])$. The Borel σ-algebra of H is denoted by $\mathcal{B}$. An H-valued random variable is a measurable mapping $x\colon (\Omega, \mathcal{F}) \to (\mathsf{H}, \mathcal{B})$. Let $p \in [1, +\infty[$. Then $L^p(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ denotes the space of equivalence classes of P-a.s. equal H-valued random variables $x\colon (\Omega,\mathcal{F},\mathsf{P}) \to (\mathsf{H},\mathcal{B})$ such that $\mathsf{E}\|x\|^{\mathsf{p}} < +\infty$. Endowed with the norm

$$\|\cdot\|_{L^{\mathsf{p}}(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})}\colon x \mapsto \mathsf{E}^{1/\mathsf{p}}\|x\|^{\mathsf{p}} = \left(\int_\Omega \|x(\omega)\|^{\mathsf{p}}\mathsf{P}(d\omega)\right)^{1/\mathsf{p}}, \tag{5.8}$$

$L^{\mathsf{p}}(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ is a real Banach space. Further,

$$(\forall \mathsf{S} \in \mathcal{B}) \quad L^{\mathsf{p}}(\Omega,\mathcal{F},\mathsf{P};\mathsf{S}) = \big\{x \in L^{\mathsf{p}}(\Omega,\mathcal{F},\mathsf{P};\mathsf{H}) \;\big|\; x \in \mathsf{S}\ \mathsf{P}\text{-a.s.}\big\}. \tag{5.9}$$

The σ-algebra generated by a family Φ of random variables is denoted by $\sigma(\Phi)$.

The reader is referred to [3] for background on convex analysis, and to [25] for background on probability.

5.2.3 Convergence analysis

We propose to study the convergence of Algorithm 5.3 to solutions to Problem 5.2 under the following assumptions.

Assumption 5.4 *In the context of Problem 5.2, there exists* $\bar{\mathsf{z}} \in \operatorname{Argmin}(\mathsf{f}+\mathsf{g})$ *and, for every* $\mathsf{k} \in \mathsf{K}$, $\mathsf{s}_\mathsf{k} \in \partial\mathsf{f}_\mathsf{k}(\bar{\mathsf{z}})$*, such that*

$$\int_\Omega \big(\mathsf{s}_{k(\omega)} + \nabla\mathsf{g}_{k(\omega)}(\bar{\mathsf{z}})\big)\mathsf{P}(d\omega) = 0 \;\; \textit{and} \;\; \int_\Omega \big\|\mathsf{s}_{k(\omega)} + \nabla\mathsf{g}_{k(\omega)}(\bar{\mathsf{z}})\big\|^2\mathsf{P}(d\omega) < +\infty. \tag{5.10}$$

Assumption 5.5 *In the context of Problem 5.2,*

$$(\forall \mathsf{k} \in \mathsf{K}) \quad \operatorname{dom}\partial\mathsf{f}_\mathsf{k} = \operatorname{dom}\partial\mathsf{f}. \tag{5.11}$$

In addition, there exists a coercive function $\psi\colon [0,+\infty[\, \to [0,+\infty[$ *such that*

$$(\forall \mathsf{x} \in \operatorname{dom}\partial\mathsf{f}) \quad \int_\Omega \big\|{}^0\partial\mathsf{f}_{k(\omega)}(\mathsf{x}) + \nabla\mathsf{g}_{k(\omega)}(\mathsf{x})\big\|^2\mathsf{P}(d\omega) \leqslant \psi\big(\|\mathsf{x}\|\big). \tag{5.12}$$

Assumption 5.4 is significantly weaker than the standard assumptions in the literature, which typically require uniform boundedness of all measurable selections of subgradients at every solution to Problem 5.2 [1,4], or the existence of subgradients in $L^{2\mathsf{p}}$ [5]. Assumption 5.5 ensures that the sequence generated by Algorithm 5.3 remains within $\operatorname{dom}\mathsf{f}$. In addition, it allows for arbitrary subgradient growth, since ψ can be any coercive function. Assumption 5.5 is weaker than those in the literature, which require controlling every measurable selection of subgradients [1] or restricting the function ψ to a particular form [5].

We present two technical lemmas.

Lemma 5.6 *Let* $\boldsymbol{x} = (x_1, \dots, x_N)$ *be an* H^N*-valued random variable, let* $(\mathsf{K}, \mathcal{K})$ *be a measurable space, and suppose that the random variable* $k\colon (\Omega, \mathcal{F}, \mathsf{P}) \to (\mathsf{K}, \mathcal{K})$ *is independent of* $\sigma(\boldsymbol{x})$*. Let* $\mathsf{h}\colon (\mathsf{K} \times \mathsf{H}, \mathcal{K} \otimes \mathcal{B}) \to [0, +\infty[$ *be measurable and define* $\phi\colon \mathsf{H} \to [0, +\infty]\colon \mathsf{x} \mapsto \mathsf{E}\mathsf{h}(k, \mathsf{x})$*. Then, for* P*-almost every* $\omega' \in \Omega$*,*

$$\mathsf{E}\big(\mathsf{h}(k, x_1) \,\big|\, \sigma(\boldsymbol{x})\big)(\omega') = \int_\Omega \mathsf{h}\big(k(\omega), x_1(\omega')\big)\mathsf{P}(d\omega) = \phi\big(x_1(\omega')\big). \tag{5.13}$$

Proof. The proof is analogous to that of [11, Lemma 2.8], replacing Fubini's theorem with Tonelli's theorem. ☐

The following fact is used in several papers without proof.

Lemma 5.7 *In the context of Algorithm 5.3, suppose that Assumptions 5.4 and 5.5 are in force. Let* $\mathsf{n} \in \mathbb{N} \smallsetminus \{0\}$*, let* $\mathsf{z} \in \operatorname{dom} \mathsf{f}$*, and set* $\mathcal{X}_\mathsf{n} = \sigma(x_0, \dots, x_\mathsf{n})$*. Then, with probability 1,*

(i) $\mathsf{E}\big(\mathsf{f}_{k_\mathsf{n}}(\mathsf{z}) \,\big|\, \mathcal{X}_\mathsf{n}\big) = \mathsf{f}(\mathsf{z})$*, and* $\mathsf{E}\big(\mathsf{g}_{k_\mathsf{n}}(\mathsf{z}) \,\big|\, \mathcal{X}_\mathsf{n}\big) = \mathsf{g}(\mathsf{z})$.

(ii) $\mathsf{E}\big(\mathsf{f}_{k_\mathsf{n}}(x_\mathsf{n}) \,\big|\, \mathcal{X}_\mathsf{n}\big) = \mathsf{f}(x_\mathsf{n})$ *and* $\mathsf{E}\big(\mathsf{g}_{k_\mathsf{n}}(x_\mathsf{n}) \,\big|\, \mathcal{X}_\mathsf{n}\big) = \mathsf{g}(x_\mathsf{n})$.

(iii) $\mathsf{E}\big(\nabla \mathsf{g}_{k_\mathsf{n}}(\mathsf{z}) \,\big|\, \mathcal{X}_\mathsf{n}\big) = \nabla \mathsf{g}(\mathsf{z})$ *and* $\mathsf{E}\big(\nabla \mathsf{g}_{k_\mathsf{n}}(x_\mathsf{n}) \,\big|\, \mathcal{X}_\mathsf{n}\big) = \nabla \mathsf{g}(x_\mathsf{n})$.

Proof. (i): This follows from (5.3) and the fact that k_n is a copy of k.

(ii): We note that

$$\mathsf{f}_{k_\mathsf{n}}(x_\mathsf{n}) = \mathsf{f}_{k_\mathsf{n}}(x_\mathsf{n}) - \mathsf{f}_{k_\mathsf{n}}(\bar{\mathsf{z}}) + \langle \bar{\mathsf{z}} - x_\mathsf{n} \mid \mathsf{s}_{k_\mathsf{n}} \rangle + \mathsf{f}_{k_\mathsf{n}}(\bar{\mathsf{z}}) - \langle \bar{\mathsf{z}} - x_\mathsf{n} \mid \mathsf{s}_{k_\mathsf{n}} \rangle \ \ \mathsf{P}\text{-a.s.} \tag{5.14}$$

It follows from the definition of the subdifferential that $\mathsf{f}_{k_\mathsf{n}}(x_\mathsf{n}) - \mathsf{f}_{k_\mathsf{n}}(\bar{\mathsf{z}}) + \langle \bar{\mathsf{z}} - x_\mathsf{n} \mid \mathsf{s}_{k_\mathsf{n}} \rangle \geqslant 0$ P-a.s. Therefore, by Lemma 5.6, (ii), and Assumption 5.4, we get

$$\mathsf{E}\big(\mathsf{f}_{k_\mathsf{n}}(x_\mathsf{n}) \,\big|\, \mathcal{X}_\mathsf{n}\big) = \mathsf{f}(x_\mathsf{n}) - \mathsf{f}(\bar{\mathsf{z}}) + \big\langle \bar{\mathsf{z}} - x_\mathsf{n} \,\big|\, -\nabla \mathsf{g}(\bar{\mathsf{z}}) \big\rangle + \mathsf{f}(\bar{\mathsf{z}}) - \big\langle \bar{\mathsf{z}} - x_\mathsf{n} \,\big|\, -\nabla \mathsf{g}(\bar{\mathsf{z}}) \big\rangle = \mathsf{f}(x_\mathsf{n}) \ \ \mathsf{P}\text{-a.s.} \tag{5.15}$$

Similarly, we deduce $\mathsf{E}\big(\mathsf{g}_{k_\mathsf{n}}(x_\mathsf{n}) \,\big|\, \mathcal{X}_\mathsf{n}\big) = \mathsf{g}(x_\mathsf{n})$ P-a.s.

(iii): From Assumption 5.5, $\mathsf{E}\|\nabla \mathsf{g}_k(\cdot)\|^2$ is locally bounded. Thus, the conclusion follows from (ii) and the dominated convergence theorem. ☐

We now show the almost sure convergence and L^1 of the iterates of Algorithm 5.3.

Theorem 5.8 *In the setting of Problem 5.2, suppose that Assumptions 5.4 and 5.5 are in force, and let* $(x_\mathsf{n})_{\mathsf{n}\in\mathbb{N}}$ *be the sequence generated by Algorithm 5.3. In addition, suppose that* $\sum_{\mathsf{n}\in\mathbb{N}} \gamma_\mathsf{n} = +\infty$ *and* $\sum_{\mathsf{n}\in\mathbb{N}} \gamma_\mathsf{n}^2 < +\infty$*. Then the following hold:*

(i) $(x_\mathsf{n})_{\mathsf{n}\in\mathbb{N}}$ *is bounded* P-a.s. *in* H *and bounded in* $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$.

(ii) $\underline{\lim}(\mathsf{f} + \mathsf{g})(x_\mathsf{n}) = \inf(\mathsf{f} + \mathsf{g})(\mathsf{H})$ P-a.s.

(iii) $(x_\mathsf{n})_{\mathsf{n}\in\mathbb{N}}$ *converges* P-a.s. *to a random variable* $x \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \operatorname{Argmin}(\mathsf{f} + \mathsf{g}))$.

(iv) $\nabla g(x_n) \to \nabla g(\bar{z})$ P-a.s.

(v) *Suppose that there exists* $\xi \in]0,+\infty[$ *such that* $\psi = \xi(1+|\cdot|^2)$. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges in* $L^1(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ *to* x.

Proof. (i): Let $n \in \mathbb{N}$ and set $\mathcal{X}_n = \sigma = (x_0,\ldots,x_n)$. It follows from [3, Proposition 12.28] that, for P-almost every $\omega \in \Omega$, $\operatorname{prox}_{\gamma_n f_{k_n(\omega)}}$ is firmly nonexpansive, hence nonexpansive. On the other hand, we deduce from Assumption 5.4 and the characterization of the proximity operator [3, Proposition 16.44] that, for P-almost every $\omega \in \Omega$,

$$s_{k_n(\omega)} \in \partial f_{k_n(\omega)}(\bar{z}) \Leftrightarrow \bar{z} + \gamma_n s_{k_n(\omega)} - \bar{z} \in \gamma_n \partial f_{k_n(\omega)}(\bar{z}) \Leftrightarrow \bar{z} = \operatorname{prox}_{\gamma_n f_{k_n(\omega)}}\big(\bar{z} + \gamma_n s_{k_n(\omega)}\big) \tag{5.16}$$

Then it follows from (5.4), (5.3), and (5.16) that

$$\begin{aligned}
&\|x_{n+1} - \bar{z}\|^2\\
&= \big\|\operatorname{prox}_{\gamma_n f_{k_n}}(x_n - \gamma_n \nabla g_{k_n}(x_n)) - \operatorname{prox}_{\gamma_n f_{k_n}}(\bar{z} + \gamma_n s_{k_n})\big\|^2\\
&\leqslant \big\|x_n - \gamma_n \nabla g_{k_n}(x_n) - (\bar{z} + \gamma_n s_{k_n})\big\|^2\\
&= \big\|(x_n - \bar{z}) - \gamma_n(\nabla g_{k_n}(x_n) + s_{k_n})\big\|^2\\
&= \|x_n - \bar{z}\|^2 - 2\gamma_n \big\langle x_n - \bar{z} \,\big|\, \nabla g_{k_n}(x_n) + s_{k_n}\big\rangle + \gamma_n^2 \big\|s_{k_n} + \nabla g_{k_n}(x_n)\big\|^2\\
&\leqslant \|x_n - \bar{z}\|^2 - 2\gamma_n \big\langle x_n - \bar{z} \,\big|\, \nabla g_{k_n}(x_n) + s_{k_n}\big\rangle + 2\gamma_n^2 \big\|\nabla g_{k_n}(x_n) - \nabla g_{k_n}(\bar{z})\big\|^2 + 2\gamma_n^2 \big\|s_{k_n} + \nabla g_{k_n}(\bar{z})\big\|^2\\
&\leqslant \|x_n - \bar{z}\|^2 - 2\gamma_n \big\langle x_n - \bar{z} \,\big|\, \nabla g_{k_n}(x_n) + s_{k_n}\big\rangle + 2\beta\gamma_n^2 \|x_n - \bar{z}\|^2 + 2\gamma_n^2 \big\|s_{k_n} + \nabla g_{k_n}(\bar{z})\big\|^2\\
&= (1 + 2\beta\gamma_n^2)\|x_n - \bar{z}\|^2 - 2\gamma_n \big\langle x_n - \bar{z} \,\big|\, \nabla g_{k_n}(x_n) + s_{k_n}\big\rangle + 2\gamma_n^2 \big\|s_{k_n} + \nabla g_{k_n}(\bar{z})\big\|^2 \quad \text{P-a.s.}
\end{aligned} \tag{5.17}$$

Therefore, since k_n is independent of $\mathcal{X}_n$, $x_n - \bar{z}$ is $\mathcal{X}_n$-measurable, ∇g is $(1/\beta)$-cocoercive [3, Corollary 18.17], and by Lemmas 5.6 and 5.7, we get

$$\begin{aligned}
&\mathsf{E}(\|x_{n+1} - \bar{z}\|^2 \,|\, \mathcal{X}_n)\\
&\leqslant (1 + 2\beta\gamma_n^2)\|x_n - \bar{z}\|^2 - 2\gamma_n \Big\langle x_n - \bar{z} \,\Big|\, \mathsf{E}(\nabla g_{k_n}(x_n) + s_{k_n} \,|\, \mathcal{X}_n)\Big\rangle + 2\gamma_n^2 \mathsf{E}(\|s_{k_n} + \nabla g_{k_n}(\bar{z})\|^2 \,|\, \mathcal{X}_n)\\
&= (1 + 2\beta\gamma_n^2)\|x_n - \bar{z}\|^2 - 2\gamma_n \big\langle x_n - \bar{z} \,\big|\, \mathsf{E}\nabla g_k(x_n) + \mathsf{E}s_k\big\rangle + 2\gamma_n^2 \mathsf{E}\|s_k + \nabla g_k(\bar{z})\|^2\\
&= (1 + 2\beta\gamma_n^2)\|x_n - \bar{z}\|^2 - 2\gamma_n \big\langle x_n - \bar{z} \,\big|\, \nabla g(x_n) - \nabla g(\bar{z})\big\rangle + 2\gamma_n^2 \mathsf{E}\|s_k + \nabla g_k(\bar{z})\|^2\\
&\leqslant (1 + 2\beta\gamma_n^2)\|x_n - \bar{z}\|^2 - 2\beta^{-1}\gamma_n \big\|\nabla g(x_n) - \nabla g(\bar{z})\big\|^2 + 2\gamma_n^2 \mathsf{E}\|s_k + \nabla g_k(\bar{z})\|^2 \quad \text{P-a.s.}
\end{aligned} \tag{5.18}$$

Since $\sum_{n\in\mathbb{N}} \gamma_n^2 < +\infty$ and $\mathsf{E}\|s_k + \nabla g_k(\bar{z})\|^2 < +\infty$, we deduce from (5.18) that $(x_n)_{n\in\mathbb{N}}$ is stochastic quasi-Fejérian relative to the set $\{\bar{z}\}$ in the sense of [13, Proposition 2.3]. It then follows from [13, Proposition 2.3(i) and (ii)] that $(x_n)_{n\in\mathbb{N}}$ is bounded P-a.s. and

$$\sum_{n\in\mathbb{N}} \gamma_n \big\|\nabla g(x_n) - \nabla g(\bar{z})\big\|^2 < +\infty \quad \text{P-a.s.} \tag{5.19}$$

Hence, the assumption $\sum_{n\in\mathbb{N}} \gamma_n = +\infty$ yields

$$\underline{\lim} \left\|\nabla g(x_n) - \nabla g(\bar{z})\right\|^2 = 0 \ \text{ P-a.s.} \tag{5.20}$$

Similarly, by taking the expected value in (5.18) we get

$$(\forall n \in \mathbb{N}) \quad \mathsf{E}\|x_{n+1} - \bar{z}\|^2 \leqslant \left(1 + 2\gamma_n^2\beta\right)\mathsf{E}\|x_n - \bar{z}\|^2 + 2\gamma_n^2\mathsf{E}\|s_k + \nabla g_k(\bar{z})\|^2, \tag{5.21}$$

which shows that $(x_n)_{n\in\mathbb{N}}$ is quasi-Fejérian in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ relative to the set $\{\bar{z}\}$ [15, Definition 3.1]. Hence, it follows from [15, Proposition 3.2(ii)] that $(x_n)_{n\in\mathbb{N}}$ is bounded in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$.

(ii): Denote $\varphi = f + g$ and, for every $k \in K$, $\varphi_k = f_k + g_k$. Let $z \in \operatorname{Argmin} \varphi$, and let $n \in \mathbb{N} \smallsetminus \{0\}$. We infer from Assumption 5.5 and (5.4) that $x_n \in \operatorname{dom} \partial f_{k_{n-1}} \subset \operatorname{dom} f$ P-a.s. and, similarly to (5.16) and (5.17), we deduce that

$$\begin{aligned}
\|x_{n+1} - x_n\| &= \left\|\operatorname{prox}_{\gamma_n f_{k_n}}\big(x_n - \gamma_n \nabla g_{k_n}(x_n)\big) - \operatorname{prox}_{\gamma_n f_{k_n}}\big(x_n + \gamma_n {}^0\partial f_{k_n}(x_n)\big)\right\| \\
&\leqslant \left\|\gamma_n {}^0\partial f_{k_n}(x_n) + \gamma_n \nabla g_{k_n}(x_n)\right\| \\
&= \gamma_n\left\|{}^0\partial f_{k_n}(x_n) + \nabla g_{k_n}(x_n)\right\| \ \text{ P-a.s.}
\end{aligned} \tag{5.22}$$

On the other hand, it follows from (5.4) and [3, Proposition 12.26 and Theorem 18.15] that

$$\begin{cases}
\gamma_n f_{k_n}(x_{n+1}) \leqslant \gamma_n f_{k_n}(z) + \left\langle x_{n+1} - z \,\middle|\, x_n - x_{n+1} - \gamma_n \nabla g_{k_n}(x_n)\right\rangle \ \text{ P-a.s.} \\
\gamma_n g_{k_n}(x_{n+1}) \leqslant \gamma_n g_{k_n}(z) + \left\langle x_{n+1} - z \,\middle|\, \gamma_n \nabla g_{k_n}(x_n)\right\rangle + \dfrac{\gamma_n\beta}{2}\|x_{n+1} - x_n\|^2 \ \text{ P-a.s.}
\end{cases} \tag{5.23}$$

Thus, after adding both inequalities and rearranging the terms, we obtain

$$\left\langle x_{n+1} - z \,\middle|\, x_{n+1} - x_n\right\rangle \leqslant \gamma_n\big(\varphi_{k_n}(z) - \varphi_{k_n}(x_{n+1})\big) + \frac{\gamma_n\beta}{2}\|x_{n+1} - x_n\|^2 \ \text{ P-a.s.} \tag{5.24}$$

Then (5.24), the definition of the subdifferential, and (5.22) yield

$$\begin{aligned}
&\left\langle x_{n+1} - z \,\middle|\, x_{n+1} - x_n\right\rangle \\
&\leqslant \gamma_n\big(\varphi_{k_n}(z) - \varphi_{k_n}(x_n) + \varphi_{k_n}(x_n) - \varphi_{k_n}(x_{n+1})\big) + \frac{\gamma_n\beta}{2}\|x_{n+1} - x_n\|^2 \\
&\leqslant \gamma_n\Big(\varphi_{k_n}(z) - \varphi_{k_n}(x_n) + \left\langle x_n - x_{n+1} \,\middle|\, {}^0\partial f_{k_n}(x_n) + \nabla g_{k_n}(x_n)\right\rangle\Big) + \frac{\gamma_n\beta}{2}\|x_{n+1} - x_n\|^2 \\
&\leqslant \gamma_n\big(\varphi_{k_n}(z) - \varphi_{k_n}(x_n)\big) + \gamma_n\|x_n - x_{n+1}\|\left\|{}^0\partial f_{k_n}(x_n) + \nabla g_{k_n}(x_n)\right\| + \frac{\gamma_n\beta}{2}\|x_{n+1} - x_n\|^2 \\
&\leqslant \gamma_n\big(\varphi_{k_n}(z) - \varphi_{k_n}(x_n)\big) + \gamma_n^2\left\|{}^0\partial f_{k_n}(x_n) + \nabla g_{k_n}(x_n)\right\|^2 + \frac{\gamma_n\beta}{2}\|x_{n+1} - x_n\|^2 \ \text{ P-a.s.}
\end{aligned} \tag{5.25}$$

Thus, we deduce from (5.25), Lemmas 5.6 and 5.7, and Assumption 5.5 that, with probability 1,

$$\begin{aligned}
&\mathsf{E}\big(\|x_{n+1}-\mathsf{z}\|^2 \,\big|\, \mathcal{X}_n\big)\\
&\quad= \|x_n-\mathsf{z}\|^2 + 2\mathsf{E}\big(\langle x_n-\mathsf{z} \mid x_{n+1}-x_n\rangle \,\big|\, \mathcal{X}_n\big) + \mathsf{E}\big(\|x_{n+1}-x_n\|^2 \,\big|\, \mathcal{X}_n\big)\\
&\quad= \|x_n-\mathsf{z}\|^2 + 2\mathsf{E}\big(\langle x_{n+1}-\mathsf{z} \mid x_{n+1}-x_n\rangle \,\big|\, \mathcal{X}_n\big) - \mathsf{E}\big(\|x_{n+1}-x_n\|^2 \,\big|\, \mathcal{X}_n\big)\\
&\quad\leqslant \|x_n-\mathsf{z}\|^2 + 2\gamma_n\mathsf{E}\big(\varphi_{k_n}(\mathsf{z})-\varphi_{k_n}(x_n) \,\big|\, \mathcal{X}_n\big) + 2\gamma_n^2\mathsf{E}\big(\big\|{}^0\partial\mathsf{f}_{k_n}(x_n)+\nabla\mathsf{g}_{k_n}(x_n)\big\|^2 \,\big|\, \mathcal{X}_n\big)\\
&\qquad\quad + \gamma_n\beta\mathsf{E}\big(\|x_{n+1}-x_n\|^2 \,\big|\, \mathcal{X}_n\big) - \mathsf{E}\big(\|x_{n+1}-x_n\|^2 \,\big|\, \mathcal{X}_n\big)\\
&\quad= \|x_n-\mathsf{z}\|^2 + 2\gamma_n\big(\varphi(\mathsf{z})-\varphi(x_n)\big) + 2\gamma_n^2\mathsf{E}\big(\big\|{}^0\partial\mathsf{f}_{k_n}(x_n)+\nabla\mathsf{g}_{k_n}(x_n)\big\|^2 \,\big|\, \mathcal{X}_n\big)\\
&\qquad\quad + \big(\gamma_n\beta-1\big)\mathsf{E}\big(\|x_{n+1}-x_n\|^2 \,\big|\, \mathcal{X}_n\big)\\
&\quad\leqslant \|x_n-\mathsf{z}\|^2 + 2\gamma_n\big(\varphi(\mathsf{z})-\varphi(x_n)\big) + 2\gamma_n^2\psi\big(\|x_n\|\big) + \big(\gamma_n\beta-1\big)\mathsf{E}\big(\|x_{n+1}-x_n\|^2 \,\big|\, \mathcal{X}_n\big)\\
&\quad\leqslant \|x_n-\mathsf{z}\|^2 + 2\gamma_n\big(\varphi(\mathsf{z})-\varphi(x_n)\big) + 2\gamma_n^2\psi\big(\|x_n\|\big) + \max\big\{0,\gamma_n\beta-1\big\}\mathsf{E}\big(\|x_{n+1}-x_n\|^2 \,\big|\, \mathcal{X}_n\big).
\end{aligned} \tag{5.26}$$

We infer from (i) and Assumption 5.5 that $\psi(\|x_n\|)$ is bounded P-a.s. In addition, $\sum_{n\in\mathbb{N}}\gamma_n^2 < +\infty$ yields $\gamma_n\beta - 1 < 0$ for n large enough. Altogether, (5.26) yields that $(x_n)_{n\in\mathbb{N}}$ is stochastic quasi-Fejérian relative to $\operatorname{Argmin}\varphi$ in the sense of [13, Proposition 2.3]. It then follows from [13, Proposition 2.3(ii)] that

$$\sum_{n\in\mathbb{N}} \gamma_n\big(\varphi(x_n)-\varphi(\mathsf{z})\big) < +\infty \quad \text{P-a.s.} \tag{5.27}$$

Hence, since $\sum_{n\in\mathbb{N}}\gamma_n = +\infty$, we have $\underline{\lim}\ \varphi(x_n) = \inf\varphi(\mathsf{H})$ P-a.s.

(iii): Let us show that $(x_n)_{n\in\mathbb{N}}$ corresponds to a sequence generated by [11, Algorithm 3.4]. To this end, set $\mathsf{Z} = \operatorname{Argmin}\varphi$ and

$$(\forall n\in\mathbb{N})\quad \begin{cases} t_n^* = 2(x_n-x_{n+1}) \in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H});\\ \eta_n = \big\langle x_{n+1}+x_n \,\big|\, x_n-x_{n+1}\big\rangle \in L^1(\Omega,\mathcal{F},\mathsf{P};\mathbb{R});\\ \alpha_n = 1;\\ \varepsilon_n = 2\gamma_n^2\psi\big(\|x_n\|\big) + \max\big\{0,\gamma_n\beta-1\big\}\mathsf{E}\big(\|x_{n+1}-x_n\|^2 \,\big|\, \mathcal{X}_n\big) \in \left[0,+\infty\right[\ \text{P-a.s.};\\ \lambda_n = \dfrac{1}{2}. \end{cases} \tag{5.28}$$

The Cauchy–Schwarz inequality shows that

$$(\forall n\in\mathbb{N})\quad \frac{1_{[t_n^*\neq 0]}1_{[\langle x_n\mid t_n^*\rangle>\eta_n]}\eta_n}{\|t_n^*\|+1_{[t_n^*=0]}} \in L^2(\Omega,\mathcal{F},\mathsf{P};\mathbb{R}). \tag{5.29}$$

In addition, we can show analogously to (5.25) and (5.26) that, for every $n\in\mathbb{N}$ and every $\mathsf{z}\in\mathsf{Z}$,

$$\big\langle \mathsf{z} \,\big|\, \mathsf{E}(\alpha_n t_n^* \mid \mathcal{X}_n)\big\rangle \leqslant \mathsf{E}(\alpha_n\eta_n \mid \mathcal{X}_n) + \varepsilon_n \quad \text{P-a.s.} \tag{5.30}$$

Finally, we derive that

$$(\forall n\in\mathbb{N})\quad x_{n+1} = x_n - \lambda_n\alpha_n t_n^*. \tag{5.31}$$

Altogether, we confirm that $(x_n)_{n\in\mathbb{N}}$ is a sequence constructed by [11, Algorithm 3.4]. On the other hand, it follows from (i) and (ii) that there exists $\Omega' \in \mathcal{F}$ such that

$$\mathsf{P}(\Omega') = 1 \text{ and } (\forall \omega \in \Omega') \ \left[(x_n(\omega))_{n\in\mathbb{N}} \text{ is bounded and } \underline{\lim}\, \varphi(x_n(\omega)) = \inf \varphi(\mathsf{H}) \right]. \tag{5.32}$$

Let $\omega \in \Omega'$ and let $(\mathsf{j}_n)_{n\in\mathbb{N}}$ be a strictly increasing sequence in $\mathbb{N}$ such that $\lim \varphi(x_{\mathsf{j}_n}(\omega)) = \inf \varphi(\mathsf{H})$. Since the sequence is bounded, there exists a point $\mathsf{x} \in \mathsf{H}$ and a further subsequence, say $x_{\mathsf{l}_{\mathsf{j}_n}}(\omega)$, such that $x_{\mathsf{l}_{\mathsf{j}_n}}(\omega) \to \mathsf{x}$. Note that the lower semicontinuity of f and the continuity of g lead to the lower semicontinuity of φ, which yields

$$\inf \varphi(\mathsf{H}) \leqslant \varphi(\mathsf{x}) \leqslant \lim \varphi\big(x_{\mathsf{l}_{\mathsf{j}_n}}(\omega)\big) = \inf \varphi(\mathsf{H}). \tag{5.33}$$

Hence $\mathsf{x} \in \operatorname{Argmin} \varphi$. Since ω is arbitrarily taken in Ω', we deduce that, for P-almost every $\omega \in \Omega$, there exists a cluster point of $(x_n(\omega))_{n\in\mathbb{N}}$ that belongs to $\operatorname{Argmin}\varphi$. Therefore, it follows from [11, Theorem 3.6(i)(d)] and the fact that $\sum_{n\in\mathbb{N}} \varepsilon_n < +\infty$ P-a.s. that there exists an $(\operatorname{Argmin}\varphi)$-valued random variable x such that $x_n \to x$ P-a.s. Furthermore, (i) and Fatou's lemma guarantee that

$$0 \leqslant \mathsf{E}\|x\|^2 \leqslant \mathsf{E}\big(\underline{\lim}\|x_n\|^2\big) \leqslant \underline{\lim}\, \mathsf{E}\|x_n\|^2 \leqslant \sup \mathsf{E}\|x_n\|^2 < +\infty, \tag{5.34}$$

which shows that $x \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \operatorname{Argmin}\varphi)$.

(iv): The continuity of $\nabla \mathsf{g}$ and (iii) yield $\nabla \mathsf{g}(x_n) \to \nabla \mathsf{g}(x)$ P-a.s. On the other hand, (5.20) shows that $\underline{\lim}\, \|\nabla \mathsf{g}(x_n) - \nabla \mathsf{g}(\bar{\mathsf{z}})\| = 0$ P-a.s. Then

$$\big\|\nabla \mathsf{g}(x) - \nabla \mathsf{g}(\bar{\mathsf{z}})\big\| \leqslant \underline{\lim}\Big(\big\|\nabla \mathsf{g}(x_n) - \nabla \mathsf{g}(x)\big\| + \big\|\nabla \mathsf{g}(x_n) - \nabla \mathsf{g}(\bar{\mathsf{z}})\big\|\Big) = 0 \ \ \mathsf{P}\text{-a.s.}, \tag{5.35}$$

which shows that $\nabla \mathsf{g}(x) = \nabla \mathsf{g}(\bar{\mathsf{z}})$ P-a.s. Therefore, $\nabla \mathsf{g}(x_n) \to \nabla \mathsf{g}(\bar{\mathsf{z}})$ P-a.s.

(v): It follows from (i) that

$$\sup_{n\in\mathbb{N}} \mathsf{E}\psi(\|x_n\|) = \sup_{n\in\mathbb{N}} \xi\big(1 + \mathsf{E}\|x_n\|^2\big) < +\infty. \tag{5.36}$$

Hence $\sum_{n\in\mathbb{N}} \mathsf{E}\varepsilon_n < +\infty$ and the convergence of $(x_n)_{n\in\mathbb{N}}$ to x in $L^1(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ follows from [11, Theorem 3.6(ii)(d)]. ☐

We present two corollaries of Theorem 5.8 that introduce novel almost surely convergent results for the stochastic proximal point algorithm and the stochastic gradient method.

Corollary 5.9 *Let* $\mathsf{f} \in \Gamma_0(\mathsf{H})$ *and let* $(\mathsf{K}, \mathcal{K})$ *be a measurable space. For every* $\mathsf{k} \in \mathsf{K}$*, let* $\mathsf{f}_\mathsf{k} \in \Gamma_0(\mathsf{H})$ *such that* $\operatorname{dom} \partial \mathsf{f}_\mathsf{k} = \operatorname{dom} \partial \mathsf{f}$*. Further, let* $k\colon (\Omega, \mathcal{F}, \mathsf{P}) \to (\mathsf{K}, \mathcal{K})$ *be a random variable such that*

$$(\forall \mathsf{x} \in \operatorname{dom} \partial \mathsf{f}) \quad \mathsf{f}(\mathsf{x}) = \int_\Omega \mathsf{f}_{k(\omega)}(\mathsf{x}) \mathsf{P}(d\omega). \tag{5.37}$$

Assume that there exists $\bar{z} \in \operatorname{Argmin} f$ *such that, for every* $k \in K$, $s_k \in \partial f_k(\bar{z})$, *and*

$$\int_\Omega s_{k(\omega)} P(d\omega) = 0 \ \ \text{and} \ \ \int_\Omega \|s_{k(\omega)}\|^2 P(d\omega) < +\infty. \tag{5.38}$$

Further, assume that there exists a coercive function $\psi\colon [0,+\infty[\, \to [0,+\infty[$ *such that*

$$(\forall x \in H) \quad \int_\Omega \left\|{}^0\partial f_{k(\omega)}(x)\right\|^2 P(d\omega) \leqslant \psi(\|x\|). \tag{5.39}$$

Let $x_0 \in L^2(\Omega,\mathcal{F},P;H)$ *and let* $(\gamma_n)_{n\in\mathbb{N}}$ *be a sequence in* $]0,+\infty[$ *such that* $\sum_{n\in\mathbb{N}} \gamma_n = +\infty$ *and* $\sum_{n\in\mathbb{N}} \gamma_n^2 < +\infty$. *Iterate*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \operatorname{prox}_{\gamma_n f_{k_n}}(x_n). \tag{5.40}$$

Then $(x_n)_{n\in\mathbb{N}}$ *converges* P-a.s. *to a random variable* $x \in L^2(\Omega,\mathcal{F},P;\operatorname{Argmin} f)$.

Corollary 5.10 *Let* $\beta \in \,]0,+\infty[$ *and let* $g\colon H \to \mathbb{R}$ *be a convex differentiable function such that* ∇g *is* β*-Lipschitzian. Let* $(K,\mathcal{K})$ *be a measurable space. For every* $k \in K$, *let* $g_k\colon H \to \mathbb{R}$ *be a convex differentiable function such that* ∇g_k *is* β*-Lipschitzian. Further, let* $k\colon (\Omega,\mathcal{F},P) \to (K,\mathcal{K})$ *be a random variable such that*

$$(\forall x \in H) \quad g(x) = \int_\Omega g_{k(\omega)}(x) P(d\omega). \tag{5.41}$$

Assume that there exists a coercive function $\psi\colon [0,+\infty[\, \to [0,+\infty[$ *such that*

$$(\forall x \in H) \quad \int_\Omega \left\|\nabla g_{k(\omega)}(x)\right\|^2 P(d\omega) \leqslant \psi(\|x\|). \tag{5.42}$$

Let $x_0 \in L^2(\Omega,\mathcal{F},P;H)$ *and let* $(\gamma_n)_{n\in\mathbb{N}}$ *be a sequence in* $]0,+\infty[$ *such that* $\sum_{n\in\mathbb{N}} \gamma_n = +\infty$ *and* $\sum_{n\in\mathbb{N}} \gamma_n^2 < +\infty$. *Iterate*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n - \gamma_n \nabla g_{k_n}(x_n). \tag{5.43}$$

Then $(x_n)_{n\in\mathbb{N}}$ *converges* P-a.s. *to a random variable* $x \in L^2(\Omega,\mathcal{F},P;\operatorname{Argmin} g)$.

5.2.4 Application to mixed-loss classification problems

We address a binary classification problem which is modeled via the combination of two loss functions.

Problem 5.11 The training data samples are split into two finite collections in $\mathbb{R}^N \times \{-1,1\}$: $(u_k,\xi_k)_{k\in K_1}$ and $(u_k,\xi_k)_{k\in K_2}$. Let $\alpha \in \,]0,1[$. The task is to

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \ \ \frac{1}{\operatorname{card} K_1}\sum_{k\in K_1} f_k(x) + \frac{1}{\operatorname{card} K_2}\sum_{k\in K_2} g_k(x), \tag{5.44}$$

where

$$\begin{cases}(\forall k\in K_1) & f_k(x)=\alpha\max\{0,1-\xi_k\langle x\mid u_k\rangle\};\\(\forall k\in K_2) & g_k(x)=(1-\alpha)\ln\big(1+\exp(-\xi_k\langle x\mid u_k\rangle)\big).\end{cases}\tag{5.45}$$

Mixed-loss problems, in particular Problem 5.11, are commonly used to train multi-task learning models; see, e.g., [9]. The goal of Problem 5.11 is to learn a linear classifier $x\in H=\mathbb{R}^N$ by minimizing the mixed-loss function. In our model, K_1 represents a set of noisy data, which we handle using the hinge loss, whereas K_2 represents accurate data, for which we use the logistic loss. To solve Problem 5.11 using Algorithm 5.3, let us first provide the proximity operator of the hinge loss and the gradient of the logistic loss. As shown in [3, Example 24.37], for every $x\in H$, $\gamma\in\,]0,+\infty[$, $i\in K_1$, and $j\in K_2$,

$$\begin{cases}\operatorname{prox}_{\gamma f_i}(x)=\begin{cases}x, & \text{if } \xi_i\langle u_i\mid x\rangle>1;\\ x+\dfrac{1-\xi_i\langle u_i\mid x\rangle}{\|u_i\|^2}, & \text{if } 1\geqslant\xi_i\langle u_i\mid x\rangle\geqslant 1-\alpha\gamma\|u_i\|^2;\\ x+\alpha\gamma\xi_i\langle u_i\mid x\rangle, & \text{if } 1-\alpha\gamma\|u_i\|^2>\xi_i\langle u_i\mid x\rangle;\end{cases}\\ \nabla g_j(x)=-\dfrac{(1-\alpha)}{1+\exp\big(\xi_j\langle x\mid u_j\rangle\big)}\xi_j u_j.\end{cases}\tag{5.46}$$

Proposition 5.12 *In the context of Problem 5.11, let $x_0\in L^2(\Omega,\mathcal{F},\mathsf{P};H)$ and let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $]0,+\infty[$ such that $\sum_{n\in\mathbb{N}}\gamma_n=+\infty$ and $\sum_{n\in\mathbb{N}}\gamma_n^2<+\infty$. Iterate*

$$\begin{array}{l}\textit{for } n=0,1,\ldots\\ \left\lfloor\begin{array}{l}\text{take } (i_n,j_n) \text{ uniformly in } K_1\times K_2 \text{ and independent of } \sigma(x_0,\ldots,x_n)\\ x_{n+1}=\operatorname{prox}_{\gamma_n f_{i_n}}\big(x_n-\gamma_n\nabla g_{j_n}(x_n)\big).\end{array}\right.\end{array}\tag{5.47}$$

Denote by Z the set of solutions to Problem 5.11 and assume that $Z\neq\varnothing$. Then $(x_n)_{n\in\mathbb{N}}$ converges P-a.s. and in $L^1(\Omega,\mathcal{F},\mathsf{P};H)$ to a Z-valued random variable.

Proof. We deduce from (5.46) that, for every $x\in H$ and $j\in K_2$,

$$\|\nabla^2 g_j(x)\|=\frac{(1-\alpha)\exp\big(\xi_j\langle x\mid u_j\rangle\big)}{\Big(1+\exp\big(\xi_j\langle x\mid u_j\rangle\big)\Big)^2}\|u_j\|^2\leqslant\frac{1-\alpha}{4}\|u_j\|^2.\tag{5.48}$$

Set

$$\beta=\frac{1-\alpha}{4}\max_{j\in K_2}\|u_j\|^2.\tag{5.49}$$

Hence, for every $k\in K_2$, ∇g_k is β-Lipschitzian. Thus, we confirm that Problem 5.11 is an instance of Problem 5.2, and (5.47) is an instance of Algorithm 5.3. It follows from Fermat's rule [3, Theorem 16.3] and [23, Theorems 1.37 and 3.8] that there exists $\bar{z}\in Z$ and, for every $k\in K_1$, $s_k\in\partial f_k(\bar{z})$ such that

$$0=\int_\Omega\big(s_{k(\omega)}+\nabla g_{k(\omega)}(\bar{z})\big)\mathsf{P}(d\omega).\tag{5.50}$$

Furthermore, we infer from (5.45) and the fact that the sets K_1 and K_2 are finite that the subgradients of $(\mathsf{f}_k)_{k\in\mathsf{K}_1}$ and the gradients of $(\mathsf{g}_k)_{k\in\mathsf{K}_2}$ are uniformly bounded. Let $\rho \in\,]0,+\infty[$ be the bound. Hence Assumptions 5.4 and 5.5 hold with $\psi = \rho + |\cdot|$. Thus the conclusion follows from Theorems 5.8(iii) and 5.8(v). ☐

5.2.5 Application to inconsistent convex feasibility problems

We apply the stochastic proximal gradient method to the inconsistent convex feasibility problem.

Problem 5.13 Let $(\mathsf{K},\mathcal{K})$ be a measurable space and let $k\colon(\Omega,\mathcal{F},\mathsf{P})\to(\mathsf{K},\mathcal{K})$ be a random variable. Let C be a nonempty closed convex subset of H, and, for every $\mathsf{k}\in\mathsf{K}$, let Z_k be a nonempty closed convex subset of H. It is assumed that the mapping

$$\mathsf{T}\colon(\mathsf{K}\times\mathsf{H},\mathcal{K}\otimes\mathcal{B}_\mathsf{H})\to(\mathsf{H},\mathcal{B}_\mathsf{H})\colon(\mathsf{k},\mathsf{x})\mapsto d^2_{\mathsf{Z}_\mathsf{k}}\mathsf{x} \tag{5.51}$$

is measurable, $\mathsf{E}\mathsf{T}(k,0)<+\infty$, and $\operatorname{Argmin}\mathsf{E}\mathsf{T}(k,\cdot)\neq\varnothing$. The task is to

$$\underset{\mathsf{x}\in\mathsf{C}}{\text{minimize}}\quad\int_\Omega\frac{1}{2}d^2_{\mathsf{Z}_{k(\omega)}}(\mathsf{x})\mathsf{P}(d\omega). \tag{5.52}$$

Minimizing the integral of the squared distances dates back to the expected-projection method [7, 8]. However, this method requires the activation of every set at every iteration via a Bochner integral average. For the consistent case, random iterative methods have been proposed; see [11, Remark 5.6]. These methods activate only a finite number of sets at each iteration and guarantee convergence to a solution. For the inconsistent case, the random iterative method of [20] guarantees convergence in distribution to an invariant measure by randomly selecting one set at every iteration. Stronger modes of convergence have not been shown for the inconsistent case. As an application of Theorem 5.8, we introduce a randomized single-set activation algorithm for solving Problem 5.13 that converges both almost surely and in $L^1(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$.

Proposition 5.14 *In the context of Problem 5.13, let $x_0\in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ and let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $]0,+\infty[$ such that $\sum_{n\in\mathbb{N}}\gamma_n=+\infty$ and $\sum_{n\in\mathbb{N}}\gamma_n^2<+\infty$. Iterate*

$$\begin{array}{l}\textit{for } n=0,1,\ldots\\ \left\lfloor\begin{array}{l}k_n \text{ is a copy of } k \text{ and is independent of } \sigma(x_0,\ldots,x_n)\\ x_{n+1}=\operatorname{proj}_\mathsf{C}\big((1-\gamma_n)x_n+\gamma_n\operatorname{proj}_{\mathsf{Z}_{k_n}}(x_n)\big).\end{array}\right.\end{array} \tag{5.53}$$

Denote by Z the set of solutions to Problem 5.13. Then $(x_n)_{n\in\mathbb{N}}$ converges P-a.s. and in $L^1(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ to a Z-valued random variable.

Proof. Let us define

$$(\forall k \in K) \quad \begin{cases} f_k = \iota_C \in \Gamma_0(H); \\ g_k = \dfrac{1}{2} d_{Z_k}^2. \end{cases} \tag{5.54}$$

We deduce from [3, Example 12.25 and Corollary 12.31] that

$$(\forall k \in K) \quad \begin{cases} (\forall n \in \mathbb{N}) \quad \mathrm{prox}_{\gamma_n f_k} = \mathrm{proj}_C; \\ \nabla g_k = \mathrm{Id} - \mathrm{proj}_{Z_k}. \end{cases} \tag{5.55}$$

It then follows from [3, Corollary 4.18] that, for every $k \in K$, ∇g_k is firmly nonexpansive, hence 1-Lipschitzian. This confirms that Problem 5.13 is an instance of Problem 5.2 with $\beta = 1$, and (5.53) is an instance of Algorithm 5.3. We deduce from Fermat's rule [3, Theorem 16.3] and [23, Theorems 1.37 and 3.8] that there exists $\bar{z} \in Z$ and $s \in N_C(\bar{z})$ such that

$$0 = s + \int_\Omega \nabla g_{k(\omega)}(\bar{z}) P(d\omega). \tag{5.56}$$

Moreover, for every $x \in H$,

$$E\|\nabla g_k(x)\|^2 = E d_{Z_k}^2(x) \leqslant 2E d_{Z_k}^2(0) + 2\|x\|^2 = 2ET(k,0) + 2\|x\|^2. \tag{5.57}$$

Combining (5.56) and (5.57) with $x = \bar{z}$, we deduce that Assumption 5.4 holds. On the other hand, for every $k \in K$ and every $x \in C$, ${}^0\partial f_k(x) = {}^0 N_C(x) = 0$. Hence Assumption 5.5 also holds by setting $\psi = \max\{2, 2ET(k,0)\}(1+|\cdot|^2)$ in (5.57). Therefore the conclusion follows from Theorems 5.8(iii) and 5.8(v). ☐

Acknowledgement

This work is part of the author's Ph.D. dissertation. The author gratefully acknowledges the guidance of his Ph.D. advisor P. L. Combettes throughout this work. This work was supported by the National Science Foundation under grant DMS-2513409.

References

[1] H. Asi and J. C. Duchi, Stochastic (approximate) proximal point methods: Convergence, optimality, and adaptivity, *SIAM J. Optim.*, vol. 29, pp. 2257–2290, 2019.

[2] Y. F. Atchadé, G. Fort, and E. Moulines, On perturbed proximal gradient algorithms, *J. Mach. Learn. Res.*, vol. 18, pp. 1–33, 2017.

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

[4] D. P. Bertsekas, Incremental proximal methods for large scale convex optimization, *Math. Program.*, vol. B129, pp. 163–196, 2011.

[5] P. Bianchi and W. Hachem, Dynamical behavior of a stochastic forward-backward algorithm using random monotone operators, *J. Optim. Theory Appl.*, vol. 171, pp. 90–120, 2016.

[6] M. Bravo and R. Cominetti, Stochastic fixed-point iterations for nonexpansive maps: Convergence and error bounds, *SIAM J. Control Optim.*, vol. 62, pp. 191–219, 2024.

[7] D. Butnariu, The expected-projection method: Its behavior and applications to linear operator equations and convex optimization, *J. Appl. Anal.*, vol. 1, pp. 93–108, 1995.

[8] D. Butnariu and S. D. Flåm, Strong convergence of expected-projection methods in Hilbert spaces, *Numer. Funct. Anal. Optim.*, vol. 16, pp. 601–636, 1995.

[9] S. Chen, Y. Zhang, and Q. Yang, Multi-task learning in natural language processing: An overview, *ACM Comput. Surv.*, vol. 56, pp. 1–32, 2024.

[10] P. L. Combettes, The geometry of monotone operator splitting methods, *Acta Numer.*, vol. 33, pp. 487–632, 2024.

[11] P. L. Combettes and J. I. Madariaga, A geometric framework for stochastic iterations, *Math. Comput.*, to appear.

[12] P. L. Combettes and J. I. Madariaga, Asymptotic analysis of an abstract stochastic scheme for solving monotone inclusions, arxiv, 2025. https://arxiv.org/pdf/2512.03023

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

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

[15] P. L. Combettes and B. C. Vũ, Variable metric quasi-Fejér monotonicity, *Nonlinear Anal.*, vol. 78, pp. 17–31, 2013.

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

[17] S. Cui, U. Shanbhag, M. Staudigl, and P. Vuong, Stochastic relaxed inertial forward-backward-forward splitting for monotone inclusions in Hilbert spaces, *Comput. Optim. Appl.*, vol. 83, pp. 465–524, 2022.

[18] M. Eisenmann, T. Stillfjord, and M. Williamson, Sub-linear convergence of a stochastic proximal iteration method in Hilbert space, *Comput. Optim. Appl.*, vol. 83, pp. 181–210, 2022.

[19] G. Fort, E. Ollier, and A. Samson, Stochastic proximal-gradient algorithms for penalized mixed models, *Stat. Comput.*, vol. 29, pp. 231–253, 2019.

[20] N. Hermer, D. R. Luke, and A. Sturm, Nonexpansive Markov operators and random function iterations for stochastic fixed point problems, *J. Convex Anal.*, vol. 30, pp. 1073–1114, 2023.

[21] H. Iiduka, Almost sure convergence of random projected proximal and subgradient algorithms for distributed nonsmooth convex optimization, *Optimization*, vol. 66, pp. 35–59, 2017.

[22] A. Patrascu and P. Irofti, Stochastic proximal splitting algorithm for composite minimization, *Optim. Lett.*, vol. 15, pp. 2255–2273, 2021.

[23] T. Pennanen and A.-P. Perkkiö, *Convex Stochastic Optimization.* Springer, Berlin, 2024.

[24] L. Rosasco, S. Villa, and B. C. Vũ, Convergence of stochastic proximal gradient algorithm, *Appl. Math. Optim.*, vol. 82, pp. 891–917, 2020.

[25] A. N. Shiryaev, *Probability–1*, 3rd ed. Springer, New York, 2016.

[26] C. Traoré, V. Apidopoulos, S. Salzo, and S. Villa, Variance reduction techniques for stochastic proximal point algorithms, *J. Optim. Theory Appl.*, vol. 203, pp. 1910–1939, 2024.

[27] P. Tseng, Applications of a splitting algorithm to decomposition in convex programming and variational inequalities, *SIAM J. Control Optim.*, vol. 29, pp. 119–138, 1991.

[28] L. Xiao, and T. Zhang, A proximal stochastic gradient method with progressive variance reduction, *SIAM J. Optim.*, vol. 24, pp. 2057–2075, 2014.

Chapter 6

AN ABSTRACT STOCHASTIC HAUGAZEAU METHOD FOR BEST APPROXIMATION

6.1 Introduction and context

We dedicate this chapter to question **(Q5)** of Chapter 1. We propose a novel abstract stochastic Haugazeau method for computing the best approximation from a closed convex set. We establish the strong convergence of the generated sequence of random variables under general hypotheses.

This chapter presents the following journal article:

> J. I. Madariaga, An abstract stochastic Haugazeau method for best approximation, submitted.

6.2 Article: An abstract stochastic Haugazeau method for best approximation

Abstract. The Haugazeau method was originally designed to compute the best approximation from an intersection of closed convex sets in Hilbert spaces using the projection operators onto the individual sets iteratively. We propose an abstract stochastic version of it to compute the best approximation from a closed convex set by successive projections onto randomly generated stochastic outer approximations of that set. Strong convergence in the mean square and the almost sure modes is derived under general hypotheses on the outer approximations. The results are applied to the development of stochastic algorithms to construct the best approximation from an arbitrary intersection of fixed point sets by random activation of blocks of

operators. A numerical application to the computation of Chebyshev centers is provided.

6.2.1 Introduction

Throughout this paper, H is a separable real Hilbert space with identity operator Id, scalar product $\langle\cdot\,|\,\cdot\rangle_{\mathsf{H}}$, and associated norm $\|\cdot\|_{\mathsf{H}}$. $(\Omega, \mathcal{F}, \mathsf{P})$ is a complete probability space.

In his unpublished thesis, Yves Haugazeau presented a geometric strategy for finding the projection onto the intersection Z of a finite collection $(\mathsf{Z}_k)_{1\leqslant k\leqslant p}$ of closed convex subsets of H by periodic projections onto these sets individually. To describe it, let $(\mathsf{x},\mathsf{y},\mathsf{z}) \in \mathsf{H}^3$ and define

$$\begin{cases} \mathsf{H}(\mathsf{x},\mathsf{y}) & = \big\{\mathsf{z}\in\mathsf{H} \mid \langle \mathsf{z}-\mathsf{y} \mid \mathsf{x}-\mathsf{y}\rangle_{\mathsf{H}} \leqslant 0\big\};\\ \mathsf{O}(\mathsf{x},\mathsf{y},\mathsf{z}) & = \begin{cases} \mathsf{H}(\mathsf{x},\mathsf{y})\cap\mathsf{H}(\mathsf{y},\mathsf{z}), & \text{if } \mathsf{H}(\mathsf{x},\mathsf{y})\cap\mathsf{H}(\mathsf{y},\mathsf{z})\neq\varnothing;\\ \{\mathsf{y}\}, & \text{if } \mathsf{H}(\mathsf{x},\mathsf{y})\cap\mathsf{H}(\mathsf{y},\mathsf{z})=\varnothing; \end{cases}\\ \mathsf{Q}(\mathsf{x},\mathsf{y},\mathsf{z}) & = \mathrm{proj}_{\mathsf{O}(\mathsf{x},\mathsf{y},\mathsf{z})}\mathsf{x}. \end{cases} \tag{6.1}$$

The half-space $\mathsf{H}(\mathsf{x},\mathsf{y})$ is defined so that the projection of x onto $\mathsf{H}(\mathsf{x},\mathsf{y})$ coincides with y, that is, $\mathsf{y} = \mathrm{proj}_{\mathsf{H}(\mathsf{x},\mathsf{y})}\mathsf{x}$. Additionally, $\mathsf{O}(\mathsf{x},\mathsf{y},\mathsf{z})$ is a nonempty closed convex subset of H, which implies that $\mathsf{Q}(\mathsf{x},\mathsf{y},\mathsf{z})$ is well-defined. Given a starting point $\mathsf{x}_0 \in \mathsf{H}$, it is shown in [18, Théorème 3-1] that the sequence $(\mathsf{x}_n)_{n\in\mathbb{N}}$ generated by the iterative algorithm

$$(\forall n\in\mathbb{N})\quad \mathsf{x}_{n+1} = \mathsf{Q}\big(\mathsf{x}_0,\mathsf{x}_n,\mathrm{proj}_{\mathsf{Z}_{n(\mathrm{mod}\,p)+1}}\mathsf{x}_n\big) \tag{6.2}$$

converges strongly to the best approximation $\mathrm{proj}_{\mathsf{Z}}\mathsf{x}_0$ to x_0 from Z. In this process, x_0 is projected onto the set $\mathsf{O}\big(\mathsf{x}_0,\mathsf{x}_n,\mathrm{proj}_{\mathsf{Z}_{n(\mathrm{mod}\,p)+1}}\mathsf{x}_n\big)$, which is an outer approximation to Z. These ideas have also been used to design parallel projection methods to project x_0 onto Z [28]. Note that

$$\mathsf{Z} = \bigcap_{1\leqslant k\leqslant p}\mathsf{Z}_k \subset \mathsf{Z}_{n(\mathrm{mod}\,p)+1} \subset \mathsf{H}\big(\mathsf{x}_n,\mathsf{a}_n\big),\ \ \text{where}\ \ \mathsf{a}_n = \mathrm{proj}_{\mathsf{Z}_{n(\mathrm{mod}\,p)+1}}\mathsf{x}_n. \tag{6.3}$$

This observation led to the development in [6] of the following abstract version of Haugazeau's method to find the best approximation to $\mathsf{x}_0 \in \mathsf{H}$ from a nonempty closed convex subset Z of H.

$$\begin{array}{l} \text{for } n = 0,1,\ldots\\ \left\lfloor \begin{array}{l} \text{take } \mathsf{a}_n\in\mathsf{H} \text{ such that } \mathsf{Z}\subset\mathsf{H}(\mathsf{x}_n,\mathsf{a}_n)\\ \text{take } \lambda_n\in\,]0,1]\\ \mathsf{r}_n = \mathsf{x}_n+\lambda_n(\mathsf{a}_n-\mathsf{x}_n)\\ \mathsf{x}_{n+1} = \mathsf{Q}(\mathsf{x}_0,\mathsf{x}_n,\mathsf{r}_n). \end{array}\right. \end{array} \tag{6.4}$$

This algorithm covers the original Haugazeau method, which is obtained by setting $\mathsf{Z} = \bigcap_{1\leqslant k\leqslant p}\mathsf{Z}_k$, and, for every $n\in\mathbb{N}$, $\lambda_n = 1$ and $\mathsf{a}_n = \mathrm{proj}_{\mathsf{Z}_{n(\mathrm{mod}\,p)+1}}\mathsf{x}_n$. In the general case, the strong convergence of the sequence $(x_n)_{n\in\mathbb{N}}$ in (6.4) to the solution of the best approximation problem is guaranteed as long as each weak cluster point of $(x_n)_{n\in\mathbb{N}}$ belongs to Z [3,6]. Applications of this framework can be found for instance in [2,5,6,9,10,25,26,29].

In optimization theory, many stochastic methods have been proposed to handle large-scale problems and high-dimensional settings; see, e.g., [11,13,15,19,21,24] and their bibliographies for discussions of the modeling and computational benefits of stochastic methods. However, it remains an open question whether stochastic versions of (6.4) can be developed and if so, with which convergence properties. Closely related to the best approximation problem is the convex feasibility problem, which consists of finding an arbitrary point in the intersection of a collection of closed convex sets. Stochastic methods have been proposed for this problem in which, at each iteration, only a finite randomly selected subcollection of sets is activated [12,19,22,27]. This feature is especially valuable when the collection of sets is uncountably infinite since, in that setting, deterministic methods cannot guarantee convergence of the iterates by activating finite blocks of sets. In this spirit, we propose to study the convergence of a stochastic counterpart to (6.4) for finding the best approximation to $x_0 \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ from an arbitrary collection of sets. This abstract stochastic Haugazeau method operates as follows.

Algorithm 6.1 Let Z be a nonempty closed convex subset of H and let $x_0 \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. Iterate

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\text{take } a_n \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H}) \text{ such that } \mathsf{Z} \subset \mathsf{H}(x_n, a_n)\ \mathsf{P}\text{-a.s.} \\
\text{take } \lambda_n \in L^\infty(\Omega, \mathcal{F}, \mathsf{P};]0, 1]) \\
r_n = x_n + \lambda_n(a_n - x_n) \\
x_{n+1} = \mathsf{Q}(x_0, x_n, r_n).
\end{array}
\right.
\end{array}
\tag{6.5}
$$

In contrast to the deterministic setting, where the weak-to-strong convergence principle discovered in [3] enables the transformation of a broad range of weakly convergent methods in nonlinear analysis into strongly convergent best approximation methods [9], there is no stochastic weak-to-strong convergence principle for stochastic methods. This is because they are constructed over random outer approximations that are not deterministic and do not act as exact cuts in the sense that Z is not almost surely contained in the random outer approximation, as illustrated in [12, Figure 1]. Thus, a dedicated analysis of Algorithm 6.1 is required to establish convergence guarantees in this stochastic framework.

The paper is organized as follows. In Section 6.2.2, we introduce the notation and preliminary results. The study of Algorithm 6.1 is presented in Section 6.2.3, where we show the strong convergence of the sequences of iterates almost surely and in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. In Section 6.2.4, we apply the method to develop a randomly activated block-iterative algorithm for solving a best approximation problem in which Z is described as the common fixed points of an arbitrary family of operators, possibly uncountably infinite. In addition, we compare Haugazeau's original cyclic method with a specialized version of our method for finding the projection onto the finite intersection of closed convex sets. Section 6.2.5 concludes the paper with a numerical application to the computation of Chebyshev centers of nonempty and bounded sets in $\mathbb{R}^N$.

6.2.2 Notation and background

6.2.2.1 Notation

Random variables are denoted by italicized serif letters and deterministic variables by sans-serif letters.

The symbols $\rightharpoonup$ and $\to$ denote weak and strong convergence in H, respectively. The set of weak sequential cluster points of a sequence $(\mathsf{x}_n)_{n\in\mathbb{N}}$ in H is denoted by $\mathfrak{W}(\mathsf{x}_n)_{n\in\mathbb{N}}$. The projection onto a nonempty closed convex set $\mathsf{C} \subset \mathsf{H}$ is denoted by $\mathrm{proj}_{\mathsf{C}}$. The fixed point set of an operator $\mathsf{T}\colon \mathsf{H} \to \mathsf{H}$ is $\operatorname{Fix} \mathsf{T} = \{\mathsf{x} \in \mathsf{H} \mid \mathsf{T}\mathsf{x} = \mathsf{x}\}$, T is firmly quasinonexpansive [4, Definition 4.1(iv)] if

$$(\forall \mathsf{x} \in \mathsf{H})(\forall \mathsf{y} \in \operatorname{Fix} \mathsf{T}) \quad \|\mathsf{T}\mathsf{x} - \mathsf{y}\|_{\mathsf{H}}^2 + \|\mathsf{T}\mathsf{x} - \mathsf{x}\|_{\mathsf{H}}^2 \leqslant \|\mathsf{x} - \mathsf{y}\|_{\mathsf{H}}^2, \tag{6.6}$$

and T is demiclosed at $\mathsf{y} \in \mathsf{H}$ if for every $\mathsf{x} \in \mathsf{H}$ and every sequence $(\mathsf{x}_n)_{n\in\mathbb{N}}$ in H such that $\mathsf{x}_n \rightharpoonup \mathsf{x}$ and $\mathsf{T}\mathsf{x}_n \to \mathsf{y}$, we have $\mathsf{T}\mathsf{x} = \mathsf{y}$. The reader is referred to [4] for background on convex analysis and fixed point theory.

Let $(\Xi, \mathcal{G})$ be a measurable space. We say that $x\colon (\Omega, \mathcal{F}, \mathsf{P}) \to (\Xi, \mathcal{G})$ is a Ξ-valued random variable (random variable for short) if it is measurable. In particular, an H-valued random variable is a measurable mapping $x\colon (\Omega, \mathcal{F}, \mathsf{P}) \to (\mathsf{H}, \mathcal{B}_{\mathsf{H}})$, where $\mathcal{B}_{\mathsf{H}}$ denotes the Borel σ-algebra of H. The sub σ-algebra of $\mathcal{F}$ generated by a family Φ of random variables is denoted by $\sigma(\Phi)$. Given $x\colon \Omega \to \Xi$ and $\mathsf{S} \in \mathcal{G}$, we set $[x \in \mathsf{S}] = \{\omega \in \Omega \mid x(\omega) \in \mathsf{S}\}$. Let x and y be Ξ-valued random variables. We say that y is a copy of x if, for every $\mathsf{S} \in \mathcal{G}$, $\mathsf{P}([x \in \mathsf{S}]) = \mathsf{P}([y \in \mathsf{S}])$. Let $\mathcal{X}$ be a sub σ-algebra of $\mathcal{F}$. Then $L^2(\Omega, \mathcal{X}, \mathsf{P}; \mathsf{H})$ denotes the space of equivalence classes of P-a.s. equal H-valued random variables $x\colon (\Omega, \mathcal{X}, \mathsf{P}) \to (\mathsf{H}, \mathcal{B}_{\mathsf{H}})$ such that $\mathsf{E}\|x\|_{\mathsf{H}}^2 < +\infty$. Endowed with the scalar product

$$\langle \cdot \mid \cdot \rangle_{L^2(\Omega,\mathcal{X},\mathsf{P};\mathsf{H})}\colon (x, y) \mapsto \mathsf{E}\langle x \mid y \rangle_{\mathsf{H}} = \int_{\Omega} \big\langle x(\omega) \,\big|\, y(\omega) \big\rangle_{\mathsf{H}} \mathsf{P}(d\omega), \tag{6.7}$$

$L^2(\Omega, \mathcal{X}, \mathsf{P}; \mathsf{H})$ is a real Hilbert space. The reader is referred to [20, 23] for background on probability in Hilbert spaces.

6.2.2.2 Preliminary results

A fundamental question regarding Algorithm 6.1 is whether x_{n+1} is well-defined and measurable for every $n \in \mathbb{N}$. To provide an affirmative answer, it is necessary to first investigate the properties of Q.

Lemma 6.2 ([4, Definition 29.24]) *Set* $\chi = \langle \mathsf{x}-\mathsf{y} \mid \mathsf{y}-\mathsf{z}\rangle_{\mathsf{H}}$, $\mu = \|\mathsf{x}-\mathsf{y}\|_{\mathsf{H}}^2$, $\nu = \|\mathsf{y}-\mathsf{z}\|_{\mathsf{H}}^2$, *and* $\rho = \mu\nu - \chi^2$. *Then*

$$\mathsf{Q}\,(\mathsf{x},\mathsf{y},\mathsf{z}) = \begin{cases} \mathsf{y}, & \text{if } \rho = 0 \text{ and } \chi < 0;\\ \mathsf{z}, & \text{if } \rho = 0 \text{ and } \chi \geqslant 0;\\ \mathsf{x} + (1+\chi/\nu)(\mathsf{z}-\mathsf{y}), & \text{if } \rho > 0 \text{ and } \chi\nu \geqslant \rho;\\ \mathsf{y} + (\nu/\rho)\big(\chi(\mathsf{x}-\mathsf{y}) + \mu(\mathsf{z}-\mathsf{y})\big), & \text{if } \rho > 0 \text{ and } \chi\nu < \rho. \end{cases} \tag{6.8}$$

Lemma 6.3 *Let* $\{x,y,z\} \subset L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$. *Suppose that there exists some* $u \in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ *such that* $u \in \mathsf{O}(x,y,z)$ P-a.s. *Then* $\mathsf{Q}(x,y,z) \in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$. *In particular,* $\mathsf{Q}(x,y,z) \in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$ *if there exists* $\mathsf{u} \in \mathsf{H}$ *such that* $\mathsf{u} \in \mathsf{O}(x,y,z)$ P-a.s.

Proof. Set $\chi = \langle x-y \mid y-z\rangle_{\mathsf{H}}$, $\mu = \|x-y\|_{\mathsf{H}}^2$, $\nu = \|y-z\|_{\mathsf{H}}^2$, and $\rho = \mu\nu - \chi^2$. The continuity of the addition, the scalar multiplication, the scalar product on H, and the norm on H, along with the measurability of x, y, and z, assure us that χ, μ, ν, and ρ are measurable. Define the disjoint measurable sets

$$\begin{cases} \mathsf{S}_1 = \big[\rho = 0 \text{ and } \chi < 0\big]; & \mathsf{S}_2 = \big[\rho = 0 \text{ and } \chi \geqslant 0\big];\\ \mathsf{S}_3 = \big[\rho > 0 \text{ and } \chi\nu \geqslant \rho\big]; & \mathsf{S}_4 = \big[\rho > 0 \text{ and } \chi\nu < \rho\big]. \end{cases} \tag{6.9}$$

Then it follows from Lemma 6.2 and (6.9) that we can write $\mathsf{Q}(x,y,z)$ as

$$\begin{aligned} \mathsf{Q}(x,y,z) = 1_{\mathsf{S}_1} y + 1_{\mathsf{S}_2} z + 1_{\mathsf{S}_3}\bigg(x + \Big(1 + \frac{\chi}{\nu + 1_{[\nu=0]}}\Big)(z-y)\bigg)&\\ + 1_{\mathsf{S}_4}\bigg(y + \Big(\frac{\nu}{\rho + 1_{[\rho=0]}}\big(\chi(x-y) + \mu(z-y)\big)\Big)\bigg)&\ \text{ P-a.s.,} \end{aligned} \tag{6.10}$$

which shows that $\mathsf{Q}(x,y,z)$ is measurable, as $(\Omega,\mathcal{F},\mathsf{P})$ is complete. On the other hand, we deduce from (6.1) and the assumptions that

$$\frac{1}{2}\mathsf{E}\big\|\mathsf{Q}(x,y,z) - x\big\|_{\mathsf{H}}^2 = \frac{1}{2}\mathsf{E}\big\|\mathrm{proj}_{\mathsf{O}(x,y,z)}\, x - x\big\|_{\mathsf{H}}^2 \leqslant \frac{1}{2}\mathsf{E}\|u - x\|_{\mathsf{H}}^2 \leqslant \mathsf{E}\|u\|_{\mathsf{H}}^2 + \mathsf{E}\|x\|_{\mathsf{H}}^2. \tag{6.11}$$

Hence, since $\{x,u\} \subset L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$, we conclude that $\mathsf{Q}(x,y,z) \in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$. ☐

Lemma 6.4 ([12, Lemma 2.8]) *Let* $\boldsymbol{x} = (x_1,\ldots,x_{\mathsf{N}})$ *be an* H^{N}*-valued random variable, let* $(\mathsf{K},\mathcal{K})$ *be a measurable space, and suppose that the random variable* $k\colon(\Omega,\mathcal{F}) \to (\mathsf{K},\mathcal{K})$ *is independent of* $\sigma(\boldsymbol{x})$. *Let* $\mathsf{f}\colon(\mathsf{K}\times\mathsf{H},\mathcal{K}\otimes\mathcal{B}_{\mathsf{H}}) \to \mathbb{R}$ *be measurable and such that* $\mathsf{E}|\mathsf{f}(k,x_1)| < +\infty$, *and define* $\mathsf{g}\colon\mathsf{H}\to\mathbb{R}\colon\mathsf{x}\mapsto\mathsf{E}\mathsf{f}(k,\mathsf{x})$. *Then, for* P*-almost every* $\omega' \in \Omega$,

$$\mathsf{E}\big(\mathsf{f}(k,x_1)\,\big|\,\sigma(\boldsymbol{x})\big)(\omega') = \int_\Omega \mathsf{f}\big(k(\omega), x_1(\omega')\big)\mathsf{P}(d\omega) = \mathsf{g}\big(x_1(\omega')\big). \tag{6.12}$$

6.2.3 Convergence analysis

In this section, we establish the strong convergence of the sequence $(x_n)_{n\in\mathbb{N}}$ generated by Algorithm 6.1 to the solution of the best approximation problem, in both the almost sure and $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ modes.

Theorem 6.5 *Let $(x_n)_{n\in\mathbb{N}}$ be the sequence generated by Algorithm 6.1. Then the following hold:*

(i) $(x_n)_{n\in\mathbb{N}}$ *is a well-defined sequence in* $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$.

(ii) $(\forall n \in \mathbb{N})\ \mathsf{Z} \subset \mathsf{H}(x_0, x_n) \cap \mathsf{H}(x_n, r_n)$ P-a.s.

(iii) $(\exists\, \ell \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathbb{R}))\ \|x_n - x_0\|_{\mathsf{H}} \uparrow \ell \leqslant \|\mathrm{proj}_{\mathsf{Z}}\, x_0 - x_0\|_{\mathsf{H}}$ P-a.s.

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

(v) $\sum_{n\in\mathbb{N}} \mathsf{E}\|x_{n+1} - x_n\|_{\mathsf{H}}^2 < +\infty$ *and* $\sum_{n\in\mathbb{N}} \mathsf{E}(\|x_{n+1} - x_n\|_{\mathsf{H}}^2 \,|\, \mathcal{X}_n) < +\infty$ P-a.s.

(vi) $\sum_{n\in\mathbb{N}} \mathsf{E}(\lambda_n^2\|a_n - x_n\|_{\mathsf{H}}^2) < +\infty$ *and* $\sum_{n\in\mathbb{N}} \mathsf{E}(\lambda_n^2\|a_n - x_n\|_{\mathsf{H}}^2 \,|\, \mathcal{X}_n) < +\infty$ P-a.s.

(vii) *Suppose that* $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \mathsf{Z}$ P-a.s. *Then* $(x_n)_{n\in\mathbb{N}}$ *converges strongly* P-a.s. *and strongly in* $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ *to* $\mathrm{proj}_{\mathsf{Z}}\, x_0$.

Proof. (i) and (ii): Suppose that, for some $n \in \mathbb{N}$, $x_n \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ and $\mathsf{Z} \subset \mathsf{H}(x_0, x_n)$ P-a.s. It is clear from (6.5) that $r_n \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. On the other hand, for P-almost every $\omega \in \Omega$,

$$
\begin{aligned}
&\mathsf{H}\big((x_n(\omega), a_n(\omega)\big)\\
&\quad= \big\{\mathsf{z} \in \mathsf{H} \;\big|\; \langle \mathsf{z} - a_n(\omega) \,|\, x_n(\omega) - a_n(\omega)\rangle_{\mathsf{H}} \leqslant 0\big\}\\
&\quad= \big\{\mathsf{z} \in \mathsf{H} \;\big|\; \langle \mathsf{z} - a_n(\omega) \,|\, x_n(\omega) - x_n(\omega) - \lambda_n(a_n(\omega) - x_n(\omega))\rangle_{\mathsf{H}} \leqslant 0\big\}\\
&\quad= \big\{\mathsf{z} \in \mathsf{H} \;\big|\; \langle \mathsf{z} - r_n(\omega) \,|\, x_n(\omega) - r_n(\omega)\rangle_{\mathsf{H}} \leqslant \langle a_n(\omega) - r_n(\omega) \,|\, x_n(\omega) - r_n(\omega)\rangle_{\mathsf{H}}\big\}\\
&\quad= \big\{\mathsf{z} \in \mathsf{H} \;\big|\; \langle \mathsf{z} - r_n(\omega) \,|\, x_n(\omega) - r_n(\omega)\rangle_{\mathsf{H}} \leqslant -\lambda_n(1 - \lambda_n)\|x_n(\omega) - a_n(\omega)\|_{\mathsf{H}}^2\big\}\\
&\quad\subset \mathsf{H}\big(x_n(\omega), r_n(\omega)\big).
\end{aligned}
\tag{6.13}
$$

We deduce from (6.5) and (6.13) that $\mathsf{Z} \subset \mathsf{H}(x_n, r_n)$ P-a.s. Hence there exists a subset $\Omega_n' \in \mathcal{F}$ such that $\mathsf{P}(\Omega_n') = 1$ and $(\forall \omega \in \Omega_n')\ \ \mathsf{Z} \subset \mathsf{H}(x_n(\omega), r_n(\omega))$. Likewise, there exists $\Omega_n'' \in \mathcal{F}$ such that $\mathsf{P}(\Omega_n') = 1$ and $(\forall \omega \in \Omega_n'')\ \mathsf{Z} \subset \mathsf{H}(x_0(\omega), x_n(\omega))$. Let $\omega \in \Omega_n' \cap \Omega_n''$. It follows from [4, Theorem 3.16] and (6.5) that

$$
\begin{aligned}
\mathsf{Z} \subset \mathsf{H}\big(x_0(\omega), x_n(\omega)\big) \cap \mathsf{H}\big(x_n(\omega), r_n(\omega)\big) \;&\Rightarrow \mathsf{Z} \subset \mathsf{H}\big(x_0(\omega), \mathsf{Q}(x_0(\omega), x_n(\omega), r_n(\omega))\big)\\
&\Leftrightarrow \mathsf{Z} \subset \mathsf{H}(x_0(\omega), x_{n+1}(\omega)).
\end{aligned}
\tag{6.14}
$$

Since $\mathsf{P}(\Omega_n' \cap \Omega_n'') = 1$, we get $\varnothing \neq \mathsf{Z} \subset \mathsf{H}(x_0, x_{n+1})$ P-a.s. Hence, Lemma 6.3 yields $x_{n+1} \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. Since $x_0 \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ and $\mathsf{Z} \subset \mathsf{H}(x_0, x_0) = \mathsf{H}$ P-a.s., we conclude by an inductive argument that $(x_n)_{n\in\mathbb{N}}$ and $(r_n)_{n\in\mathbb{N}}$ are sequences in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$, and, for every $n \in \mathbb{N}$, $\mathsf{Z} \subset \mathsf{H}(x_0, x_n) \cap \mathsf{H}(x_n, r_n)$ P-a.s.

(iii): Set, for every $n \in \mathbb{N}$, $\mathcal{X}_n = \sigma(x_0, \ldots, x_n)$, and consider the random process $(\|x_n - x_0\|_H, \mathcal{X}_n)_{n\in\mathbb{N}}$. Let us show that this process is a submartingale. Let $n \in \mathbb{N}$. It follows from (ii) that

$$O(x_0, x_n, r_n) = H(x_0, x_n) \cap H(x_n, r_n) \text{ P-a.s. and } x_{n+1} \in H(x_0, x_n) \cap H(x_n, r_n) \text{ P-a.s.} \tag{6.15}$$

We note from (6.1) that $x_n = \operatorname{proj}_{H(x_0,x_n)} x_0$ P-a.s., and from (6.15) that $x_{n+1} \in H(x_0, x_n)$ P-a.s. Then

$$(\forall n \in \mathbb{N}) \quad \|x_n - x_0\|_H \leqslant \|x_{n+1} - x_0\|_H \quad \text{P-a.s.} \tag{6.16}$$

Hence

$$(\forall n \in \mathbb{N}) \quad \|x_n - x_0\|_H = E(\|x_n - x_0\|_H \,|\, \mathcal{X}_n) \leqslant E(\|x_{n+1} - x_0\|_H \,|\, \mathcal{X}_n) \quad \text{P-a.s.} \tag{6.17}$$

Therefore $(\|x_n - x_0\|_H, \mathcal{X}_n)_{n\in\mathbb{N}}$ is a positive submartingale. On the other hand, for every $n \in \mathbb{N}$, $\operatorname{proj}_Z x_0 \in Z \subset H(x_0, x_n)$ P-a.s. Then, for every $n \in \mathbb{N}$, $\|x_n - x_0\|_H \leqslant \|\operatorname{proj}_Z x_0 - x_0\|_H$ P-a.s., which shows that

$$(\forall n \in \mathbb{N}) \quad E\|x_n - x_0\|_H^2 \leqslant E\|\operatorname{proj}_Z x_0 - x_0\|_H^2 < +\infty. \tag{6.18}$$

Consequently, $\sup_{n\in\mathbb{N}} E\|x_n - x_0\|_H^2 < +\infty$ and we deduce from [16, §IV Theorems 4.1s(i) and 4.1s(iii)] that $(\|x_n - x_0\|_H)_{n\in\mathbb{N}}$ converges P-a.s. and converges in $L^2(\Omega, \mathcal{F}, P; \mathbb{R})$ to a random variable $\ell \in L^2(\Omega, \mathcal{F}, P; \mathbb{R})$ which satisfies $\ell \leqslant \|\operatorname{proj}_Z x_0 - x_0\|_H$ P-a.s.

(iv): Note that, for every $n \in \mathbb{N}$,

$$\|x_n\|_H \leqslant \|x_n - x_0\|_H + \|x_0\|_H \ \text{ P-a.s.} \tag{6.19}$$

We deduce from (6.19) and (iii) that $(\|x_n\|_H)_{n\in\mathbb{N}}$ is bounded P-a.s.

(v): Let $n \in \mathbb{N}$. Since $x_{n+1} \in H(x_0, x_n)$ P-a.s., we have

$$\begin{aligned} \|x_{n+1} - x_n\|_H^2 &\leqslant \|x_{n+1} - x_n\|_H^2 + 2\langle x_{n+1} - x_n \,|\, x_n - x_0\rangle_H \\ &= \|x_{n+1} - x_0\|_H^2 - \|x_n - x_0\|_H^2 \ \text{ P-a.s.} \end{aligned} \tag{6.20}$$

Hence, $\sum_{j=0}^{n} \|x_{j+1} - x_j\|_H^2 \leqslant \|x_{n+1} - x_0\|_H^2 \leqslant \|\operatorname{proj}_Z x_0 - x_0\|_H^2$ P-a.s. Therefore, by taking expected value and limit $n \to +\infty$, we have

$$\sum_{j\in\mathbb{N}} E\Big(E(\|x_{j+1} - x_j\|_H^2 \,|\, \mathcal{X}_j)\Big) = \sum_{j\in\mathbb{N}} E\|x_{j+1} - x_j\|_H^2 \leqslant E\|\operatorname{proj}_Z x_0 - x_0\|_H^2 < +\infty. \tag{6.21}$$

We conclude that $\sum_{j\in\mathbb{N}} E\|x_{j+1} - x_j\|_H^2 < +\infty$ and that $\sum_{j\in\mathbb{N}} E(\|x_{j+1} - x_j\|_H^2 \,|\, \mathcal{X}_j) < +\infty$ P-a.s.

(vi): Let $n \in \mathbb{N}$. Since $x_{n+1} \in H(x_n, r_n)$ P-a.s., we have

$$\|r_n - x_n\|_H^2 \leqslant \|x_{n+1} - r_n\|_H^2 + \|x_n - r_n\|_H^2$$

$$\begin{aligned} &\leqslant \|x_{n+1} - r_n\|_H^2 + 2\langle x_{n+1} - r_n \mid r_n - x_n\rangle_H + \|x_n - r_n\|_H^2 \\ &= \|x_{n+1} - x_n\|_H^2 \quad \text{P-a.s.} \end{aligned} \tag{6.22}$$

It follows then from (v) and (6.5) that

$$\sum_{j\in\mathbb{N}} \mathsf{E}\Big(\mathsf{E}(\lambda_j^2\|a_j - x_j\|_H^2 \mid \mathcal{X}_j)\Big) = \sum_{j\in\mathbb{N}} \mathsf{E}(\lambda_j^2\|a_j - x_j\|_H^2) = \sum_{j\in\mathbb{N}} \mathsf{E}\|r_j - x_j\|_H^2 < +\infty. \tag{6.23}$$

Hence, $\sum_{j\in\mathbb{N}} \mathsf{E}(\lambda_j^2\|a_j - x_j\|_H^2) < +\infty$ and $\sum_{j\in\mathbb{N}} \mathsf{E}(\lambda_j^2\|a_j - x_j\|_H^2 \mid \mathcal{X}_j) < +\infty$ P-a.s.

(vii): It follows from (iv) that $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \neq \varnothing$ P-a.s. Now suppose that $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset Z$ P-a.s. Then there exists $\Omega' \in \mathcal{F}$ such that $\mathsf{P}(\Omega') = 1$ and $(\forall \omega \in \Omega')\ \varnothing \neq \mathfrak{W}(x_n(\omega))_{n\in\mathbb{N}} \subset Z$. In addition, consider from (iii) the event $\Omega'' \in \mathcal{F}$ such that $\mathsf{P}(\Omega'') = 1$ and

$$(\forall \omega \in \Omega'') \quad \begin{cases} \|x_n(\omega) - x_0(\omega)\|_H \to \ell(\omega); \\ \ell(\omega) \leqslant \|\mathrm{proj}_Z\, x_0(\omega) - x_0(\omega)\|_H. \end{cases} \tag{6.24}$$

Let $\omega \in \Omega' \cap \Omega''$ and let $x \in Z$ be a weak sequential cluster point of $(x_n(\omega))_{n\in\mathbb{N}}$, say $x_{k_n}(\omega) \rightharpoonup x$. Then, the weak lower semicontinuity of $\|\cdot\|_H$ and (6.24) imply that

$$\begin{aligned} \|x - x_0(\omega)\|_H &\leqslant \varliminf_{n\in\mathbb{N}} \|x_{k_n}(\omega) - x_0(\omega)\|_H \\ &= \ell(\omega) \\ &\leqslant \|\mathrm{proj}_Z\, x_0(\omega) - x_0(\omega)\|_H \\ &= \inf_{y\in Z} \|y - x_0(\omega)\|_H \\ &\leqslant \|x - x_0(\omega)\|_H. \end{aligned} \tag{6.25}$$

Therefore $x = \mathrm{proj}_Z\, x_0(\omega)$ is the only weak sequential cluster point of the sequence $(x_n(\omega))_{n\in\mathbb{N}}$ and $x_n(\omega) \rightharpoonup \mathrm{proj}_Z\, x_0(\omega)$. Hence, $x_n(\omega) - x_0(\omega) \rightharpoonup \mathrm{proj}_Z\, x_0(\omega) - x_0(\omega)$ and (6.25) yields $\|x_n(\omega) - x_0(\omega)\|_H \to \|\mathrm{proj}_Z\, x_0(\omega) - x_0(\omega)\|_H$. Then, by [4, Lemma 2.41(i)], we get that $x_n(\omega) - x_0(\omega) \to \mathrm{proj}_Z\, x_0(\omega) - x_0(\omega)$ and, therefore, $x_n(\omega) \to \mathrm{proj}_Z\, x_0(\omega)$. Since $\mathsf{P}(\Omega' \cap \Omega'') = 1$, we deduce that $x_n \to \mathrm{proj}_Z\, x_0$ P-a.s. Furthermore, note from (iii) that, for every $n \in \mathbb{N}$,

$$\begin{aligned} \|\mathrm{proj}_Z\, x_0 - x_n\|_H &\leqslant \|\mathrm{proj}_Z\, x_0 - x_0\|_H + \|x_n - x_0\|_H \\ &\leqslant 2\|\mathrm{proj}_Z\, x_0 - x_0\|_H \quad \text{P-a.s.} \end{aligned} \tag{6.26}$$

Since $\|\mathrm{proj}_Z\, x_0 - x_0\|_H \in L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathbb{R})$, the dominated convergence theorem guarantees that $(x_n)_{n\in\mathbb{N}}$ converges strongly in $L^2(\Omega, \mathcal{F}, \mathsf{P}; H)$ to $\mathrm{proj}_Z\, x_0$. ☐

6.2.4 Application to common fixed point best approximation

In this section we specialize the best approximation problem to the following setting.

Problem 6.6 Let $(\mathsf{K},\mathcal{K})$ be a measurable space and let $(\mathsf{T}_k)_{k\in\mathsf{K}}$ be a family of firmly quasinonexpansive operators such that $\mathsf{T}\colon(\mathsf{K}\times\mathsf{H},\mathcal{K}\otimes\mathcal{B}_{\mathsf{H}})\to(\mathsf{H},\mathcal{B}_{\mathsf{H}})\colon(k,x)\mapsto\mathsf{T}_k x$ is measurable and, for every $k\in\mathsf{K}$, $\mathrm{Id}-\mathsf{T}_k$ is demiclosed at 0. Let k be a K-valued random variable, suppose that

$$\mathsf{Z}=\big\{z\in\mathsf{H}\;\big|\;z\in\operatorname{Fix}\mathsf{T}_{k}\ \mathsf{P}\text{-a.s.}\big\}\neq\varnothing, \tag{6.27}$$

and let $x_0\in L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$. The goal is to find $\operatorname{proj}_{\mathsf{Z}}x_0$.

To emphasize the versatility of Problem 6.6, we present some examples of firmly quasinonexpansive operators commonly used in nonlinear analysis and optimization methods.

Example 6.7 ([3, Proposition 2.3]) Let $\mathsf{T}\colon\mathsf{H}\to\mathsf{H}$. Then T is firmly quasinonexpansive if one of the following holds:

(i) Z is a nonempty closed convex subset of H and $\mathsf{T}=\operatorname{proj}_{\mathsf{Z}}$. Here, $\operatorname{Fix}\mathsf{T}=\mathsf{Z}$.

(ii) $f\colon\mathsf{H}\to\left]-\infty,+\infty\right]$ is a proper lower semicontinuous convex function and

$$\mathsf{T}=\operatorname{prox}_f\colon\mathsf{H}\to\mathsf{H}\colon x\mapsto\underset{y\in\mathsf{H}}{\operatorname{argmin}}\Big(f(y)+\frac{1}{2}\|x-y\|_{\mathsf{H}}^2\Big). \tag{6.28}$$

Here, $\operatorname{Fix}\mathsf{T}=\operatorname{Argmin}f$.

(iii) $\mathsf{A}\colon\mathsf{H}\to2^{\mathsf{H}}$ is maximally monotone and $\mathsf{T}=\mathsf{J}_{\mathsf{A}}=(\mathrm{Id}+\mathsf{A})^{-1}$. Here, $\operatorname{Fix}\mathsf{T}=\big\{z\in\mathsf{H}\mid 0\in\mathsf{A}z\big\}$.

(iv) $f\colon\mathsf{H}\to\mathbb{R}$ is a continuous convex function, $s\colon\mathsf{H}\to\mathsf{H}\colon x\mapsto s(x)\in\partial f(x)$ is a selection of ∂f, and

$$\mathsf{T}=\mathsf{G}_f\colon\mathsf{H}\to\mathsf{H}\colon x\mapsto\begin{cases}x-\dfrac{f(x)}{\|s(x)\|_{\mathsf{H}}^2}s(x), & \text{if } f(x)>0;\\[2mm] x, & \text{if } f(x)\leqslant 0,\end{cases}$$

is the subgradient projector onto $\operatorname{Fix}\mathsf{T}=\big\{x\in\mathsf{H}\mid f(x)\leqslant0\big\}$.

6.2.4.1 Algorithm and convergence

Since the set Z of Problem 6.6 is a nonempty closed convex subset of H [12, Remark 5.3], we invoke Theorem 6.5 to introduce the first randomly selected block-iterative fixed point algorithm to solve Problem 6.6, which guarantees both strong almost sure convergence and strong convergence in $L^2(\Omega,\mathcal{F},\mathsf{P};\mathsf{H})$. We emphasize that this method is novel even for K finite or countably infinite. Furthermore, unlike the deterministic algorithms, this stochastic method is able to solve Problem 6.6 in its generality by random activation of block of operators.

Theorem 6.8 *In the setting of Problem 6.6, let* $0 < \mathsf{M} \in \mathbb{N}$, $\delta \in \left]0, 1/\mathsf{M}\right[$, *and* $\varepsilon \in \left]0, 1\right[$. *Iterate*

$$
\begin{array}{l}
\text{for } \mathsf{n} = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\mathcal{X}_\mathsf{n} = \sigma(x_0, \ldots, x_\mathsf{n}) \\
\textit{for } \mathsf{i} = 1, \ldots, \mathsf{M} \\
\left\lfloor
\begin{array}{l}
k_{\mathsf{i},\mathsf{n}} \text{ is a copy of } k \text{ and is independent of } \mathcal{X}_\mathsf{n} \\
p_{\mathsf{i},\mathsf{n}} = \mathsf{T}_{k_{\mathsf{i},\mathsf{n}}} x_\mathsf{n}
\end{array}
\right. \\
(\beta_{\mathsf{i},\mathsf{n}})_{1 \leqslant \mathsf{i} \leqslant \mathsf{M}} \text{ are } [0,1]\text{-valued random variables such that} \\
\qquad \sum_{\mathsf{i}=1}^{\mathsf{M}} \beta_{\mathsf{i},\mathsf{n}} = 1 \text{ P-a.s. and } (\forall \mathsf{i} \in \{1, \ldots, \mathsf{M}\})\ \beta_{\mathsf{i},\mathsf{n}} \geqslant \delta 1_{[\|p_{\mathsf{i},\mathsf{n}} - x_\mathsf{n}\|_\mathsf{H} = \max_{1 \leqslant \mathsf{j} \leqslant \mathsf{M}} \|p_{\mathsf{j},\mathsf{n}} - x_\mathsf{n}\|_\mathsf{H}]} \\
p_\mathsf{n} = \sum_{\mathsf{i}=1}^{\mathsf{M}} \beta_{\mathsf{i},\mathsf{n}} p_{\mathsf{i},\mathsf{n}} \\
L_\mathsf{n} = \dfrac{\sum_{\mathsf{i}=1}^{\mathsf{M}} \beta_{\mathsf{i},\mathsf{n}} \|p_{\mathsf{i},\mathsf{n}} - x_\mathsf{n}\|_\mathsf{H}^2 + 1_{[p_\mathsf{n} = x_\mathsf{n}]}}{\|p_\mathsf{n} - x_\mathsf{n}\|_\mathsf{H}^2 + 1_{[p_\mathsf{n} = x_\mathsf{n}]}} \\
a_\mathsf{n} = x_\mathsf{n} + L_\mathsf{n}(p_\mathsf{n} - x_\mathsf{n}) \\
\text{take } \lambda_\mathsf{n} \in L^\infty(\Omega, \mathcal{F}, \mathsf{P}; [\varepsilon, 1]) \\
r_\mathsf{n} = x_\mathsf{n} + \lambda_\mathsf{n}(a_\mathsf{n} - x_\mathsf{n}) \\
x_{\mathsf{n}+1} = \mathsf{Q}(x_0, x_\mathsf{n}, r_\mathsf{n}).
\end{array}
\right.
\end{array}
\tag{6.29}
$$

Then $(x_\mathsf{n})_{\mathsf{n} \in \mathbb{N}}$ *converges strongly* P-a.s. *and strongly in* $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ *to* $\mathrm{proj}_\mathsf{Z}\, x_0$.

Proof. We will show that the sequence constructed by (6.29) corresponds to a sequence generated by Algorithm 6.1. First, let us show that

$$(\forall \mathsf{n} \in \mathbb{N}) \quad L_\mathsf{n} \geqslant 1 \ \text{ P-a.s.} \tag{6.30}$$

Fix $\mathsf{z} \in \mathsf{Z}$ and $\mathsf{n} \in \mathbb{N}$. For every $\mathsf{i} \in \{1, \ldots, \mathsf{M}\}$, let $\Omega_{\mathsf{i},\mathsf{n}} \in \mathcal{F}$ be such that

$$\mathsf{P}(\Omega_{\mathsf{i},\mathsf{n}}) = 1 \quad \text{and} \quad (\forall \omega \in \Omega_{\mathsf{i},\mathsf{n}}) \quad \mathsf{z} \in \mathrm{Fix}\, \mathsf{T}_{k_{\mathsf{i},\mathsf{n}}(\omega)}. \tag{6.31}$$

Thanks to (6.31), we then choose $\Omega_\mathsf{n} \in \mathcal{F}$ such that

$$\mathsf{P}(\Omega_\mathsf{n}) = 1 \quad \text{and} \quad (\forall \omega \in \Omega_\mathsf{n}) \quad \bigcap_{1 \leqslant \mathsf{i} \leqslant \mathsf{M}} \mathrm{Fix}\, \mathsf{T}_{k_{\mathsf{i},\mathsf{n}}(\omega)} \neq \varnothing \quad \text{and} \quad \sum_{\mathsf{i}=1}^{\mathsf{M}} \beta_{\mathsf{i},\mathsf{n}}(\omega) = 1. \tag{6.32}$$

Given $\omega \in \Omega_\mathsf{n}$, we study the following two cases:

- Suppose that $p_\mathsf{n}(\omega) = x_\mathsf{n}(\omega)$. Then [7, Proposition 2.4] shows that

$$x_\mathsf{n}(\omega) \in \mathrm{Fix}\left(\sum_{\mathsf{i}=1}^{\mathsf{M}} \beta_{\mathsf{i},\mathsf{n}}(\omega) \mathsf{T}_{k_{\mathsf{i},\mathsf{n}}(\omega)}\right) = \bigcap_{1 \leqslant \mathsf{i} \leqslant \mathsf{M}} \mathrm{Fix}\, \mathsf{T}_{k_{\mathsf{i},\mathsf{n}}(\omega)}, \tag{6.33}$$

hence, for every $i \in \{1,\dots,M\}$, $x_n(\omega) = p_{i,n}(\omega)$. Thus,

$$L_n(\omega) = \frac{\sum_{i=1}^{M} \beta_{i,n}(\omega)\|p_{i,n}(\omega) - x_n(\omega)\|_H^2 + 1_{[p_n = x_n]}(\omega)}{\|p_n(\omega) - x_n(\omega)\|_H^2 + 1_{[p_n = x_n]}(\omega)} = \frac{1_{[p_n = x_n]}(\omega)}{1_{[p_n = x_n]}(\omega)} = 1. \tag{6.34}$$

- Suppose that $p_n(\omega) \neq x_n(\omega)$. Then the convexity of $\|\cdot\|_H^2$ yields

$$0 < \|p_n(\omega) - x_n(\omega)\|_H^2 = \left\|\sum_{i=1}^{M} \beta_{i,n}(\omega)\big(p_{i,n}(\omega) - x_n(\omega)\big)\right\|_H^2 \leqslant \sum_{i=1}^{M} \beta_{i,n}(\omega)\|p_{i,n}(\omega) - x_n(\omega)\|_H^2, \tag{6.35}$$

  which implies that $L_n(\omega) \geqslant 1$.

Next, we will show by induction that $(x_n)_{n\in\mathbb{N}}$ and $(a_n)_{n\in\mathbb{N}}$ are in $L^2(\Omega,\mathcal{F},P;H)$. Let $i \in \{1,\dots,M\}$ and suppose that $x_n \in L^2(\Omega,\mathcal{F},P;H)$. Then $T_{k_{i,n}}x_n = T \circ (k_{i,n}, x_n)$ is measurable. Furthermore, for every $\omega \in \Omega_{i,n}$, $2T_{k_{i,n}(\omega)} - \mathrm{Id}$ is quasinonexpansive with $\mathrm{Fix}(2T_{k_{i,n}(\omega)} - \mathrm{Id}) = \mathrm{Fix}\, T_{k_{i,n}(\omega)}$ [7, Proposition 2.2(v)]. Hence

$$\begin{aligned} 2\|p_{i,n}(\omega)\|_H^2 &= \frac{1}{2}\|2T_{k_{i,n}(\omega)}x_n(\omega)\|_H^2 \\ &\leqslant \|(2T_{k_{i,n}(\omega)} - \mathrm{Id})x_n(\omega) - z\|_H^2 + \|x_n(\omega) + z\|_H^2 \\ &\leqslant \|x_n(\omega) - z\|_H^2 + \|x_n(\omega) + z\|_H^2. \end{aligned} \tag{6.36}$$

Thus, since $x_n \in L^2(\Omega,\mathcal{F},P;H)$ and $z \in H$, we have $p_{i,n} \in L^2(\Omega,\mathcal{F},P;H)$. Hence, (6.29) yields that $p_n \in L^2(\Omega,\mathcal{F},P;H)$ and that a_n is measurable. To show that $a_n \in L^2(\Omega,\mathcal{F},P;H)$, we first note that

$$\begin{aligned} x_n - a_n &= L_n(x_n - p_n) \\ &= \frac{\sum_{i=1}^{M} \beta_{i,n}\|p_{i,n} - x_n\|_H^2 + 1_{[p_n = x_n]}}{\|p_n - x_n\|_H^2 + 1_{[p_n = x_n]}}(x_n - p_n) \\ &= \frac{\sum_{i=1}^{M} \beta_{i,n}\big(\langle x_n \mid x_n - p_{i,n}\rangle_H - \langle p_{i,n} \mid x_n - p_{i,n}\rangle_H\big)}{\|x_n - p_n\|_H^2 + 1_{[p_n = x_n]}}(x_n - p_n) \\ &= \frac{\langle x_n \mid x_n - p_n\rangle_H - \sum_{i=1}^{M} \beta_{i,n}\langle p_{i,n} \mid x_n - p_{i,n}\rangle_H}{\|x_n - p_n\|_H^2 + 1_{[p_n = x_n]}}(x_n - p_n) \qquad (6.37) \\ &= \frac{\langle x_n \mid x_n - p_n\rangle_H - \sum_{i=1}^{M} \beta_{i,n}\langle p_{i,n} - z \mid x_n - p_{i,n}\rangle_H - \sum_{i=1}^{M} \beta_{i,n}\langle z \mid x_n - p_{i,n}\rangle_H}{\|x_n - p_n\|_H^2 + 1_{[p_n = x_n]}}(x_n - p_n) \\ &= \frac{\langle x_n - z \mid x_n - p_n\rangle_H - \sum_{i=1}^{M} \beta_{i,n}\langle p_{i,n} - z \mid x_n - p_{i,n}\rangle_H}{\|x_n - p_n\|_H^2 + 1_{[p_n = x_n]}}(x_n - p_n) \quad P\text{-a.s.} \qquad (6.38) \end{aligned}$$

On the other hand, it follows from [4, Proposition 4.2(iv)] that

$$(\forall i \in \{1,\ldots,M\}) \quad \big\langle p_{i,n} - z \,\big|\, x_n - p_{i,n}\big\rangle_H = \big\langle T_{k_{i,n}} x_n - z \,\big|\, x_n - T_{k_{i,n}} x_n\big\rangle_H \geqslant 0 \;\; \text{P-a.s.} \tag{6.39}$$

In turn, the concavity of $y \mapsto \langle y - z \mid x_n(\omega) - y\rangle_H$ yields

$$0 \leqslant \sum_{i=1}^{M} \beta_{i,n} \big\langle p_{i,n} - z \,\big|\, x_n - p_{i,n}\big\rangle_H \leqslant \big\langle p_n - z \,\big|\, x_n - p_n\big\rangle_H \;\; \text{P-a.s.} \tag{6.40}$$

and therefore it follows from the convexity of the norm square and (6.38) that

$$\begin{aligned}
\frac{1}{2}\mathsf{E}\|x_n - a_n\|_H^2 &= \frac{1}{2}\mathsf{E}\left\|\frac{\langle x_n - z \mid x_n - p_n\rangle_H - \sum_{i=1}^{M} \beta_{i,n}\langle p_{i,n} - z \mid x_n - p_{i,n}\rangle_H}{\|x_n - p_n\|_H^2 + 1_{[p_n = x_n]}}(x_n - p_n)\right\|_H^2 \\
&= \frac{1}{2}\mathsf{E}\left|\frac{\langle x_n - z \mid x_n - p_n\rangle_H - \sum_{i=1}^{M} \beta_{i,n}\langle p_{i,n} - z \mid x_n - p_{i,n}\rangle_H}{\|x_n - p_n\|_H + 1_{[p_n = x_n]}}\right|^2 \\
&\leqslant \mathsf{E}\left|\frac{\langle x_n - z \mid x_n - p_n\rangle_H}{\|x_n - p_n\|_H + 1_{[p_n = x_n]}}\right|^2 + \mathsf{E}\left|\frac{\sum_{i=1}^{M} \beta_{i,n}\langle p_{i,n} - z \mid x_n - p_{i,n}\rangle_H}{\|x_n - p_n\|_H + 1_{[p_n = x_n]}}\right|^2 \\
&\leqslant \mathsf{E}\left|\frac{\|x_n - z\|_H \|x_n - p_n\|_H}{\|x_n - p_n\|_H + 1_{[p_n = x_n]}}\right|^2 + \mathsf{E}\left|\frac{\langle p_n - z \mid x_n - p_n\rangle_H}{\|x_n - p_n\|_H + 1_{[p_n = x_n]}}\right|^2 \\
&\leqslant \mathsf{E}\|x_n - z\|_H^2 + \mathsf{E}\left|\frac{\|p_n - z\|_H \|x_n - p_n\|_H}{\|x_n - p_n\|_H + 1_{[p_n = x_n]}}\right|^2 \\
&\leqslant \mathsf{E}\|x_n - z\|_H^2 + \mathsf{E}\|p_n - z\|_H^2 \\
&< +\infty.
\end{aligned} \tag{6.41}$$

Since $\{x_n, p_n\} \subset L^2(\Omega, \mathcal{F}, \mathsf{P}; H)$ and $z \in H$, we thus obtain $x_n - a_n \in L^2(\Omega, \mathcal{F}, \mathsf{P}; H)$ and hence $a_n \in L^2(\Omega, \mathcal{F}, \mathsf{P}; H)$. Moreover, it follows from (6.29) that $r_n \in L^2(\Omega, \mathcal{F}, \mathsf{P}; H)$. On the other hand, we deduce from (6.29) and (6.37) that

$$\begin{aligned}
\langle a_n \mid x_n - a_n\rangle_H &= L_n\big\langle x_n + L_n(p_n - x_n) \,\big|\, x_n - p_n\big\rangle_H \\
&= L_n\Big(\langle x_n \mid x_n - p_n\rangle_H - L_n\|p_n - x_n\|_H^2\Big) \\
&= L_n\Big(\langle x_n \mid x_n - p_n\rangle_H - \langle x_n \mid x_n - p_n\rangle_H + \sum_{i=1}^{M} \beta_{i,n}\langle p_{i,n} \mid x_n - p_{i,n}\rangle_H\Big) \\
&= L_n \sum_{i=1}^{M} \beta_{i,n}\langle p_{i,n} \mid x_n - p_{i,n}\rangle_H \;\; \text{P-a.s.}
\end{aligned} \tag{6.42}$$

Furthermore, (6.30), (6.40), and (6.42) yield

$$\langle z \mid x_n - a_n\rangle_H = L_n\langle z \mid x_n - p_n\rangle_H$$

$$
\begin{aligned}
&= L_n \sum_{i=1}^{M} \beta_{i,n} \langle z \mid x_n - p_{i,n} \rangle_H \\
&\leqslant L_n \sum_{i=1}^{M} \beta_{i,n} \langle p_{i,n} \mid x_n - p_{i,n} \rangle_H \\
&= \langle a_n \mid x_n - a_n \rangle_H \ \ \text{P-a.s.,}
\end{aligned} \tag{6.43}
$$

which shows that $Z \subset H(x_n, a_n)$ P-a.s. We can repeat the arguments in (6.13)–(6.14) to show that $x_{n+1} \in L^2(\Omega, \mathcal{F}, P; H)$. Since $x_0 \in L^2(\Omega, \mathcal{F}, P; H)$, we conclude inductively that

$$
(\forall n \in \mathbb{N}) \quad \{x_n, a_n\} \subset L^2(\Omega, \mathcal{F}, P; H) \ \text{ and } \ Z \subset H(x_n, a_n) \text{ P-a.s.} \tag{6.44}
$$

Therefore, the sequence $(x_n)_{n \in \mathbb{N}}$ constructed by (6.29) corresponds to one generated by Algorithm 6.1. It is therefore enough to show that $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset Z$ P-a.s. to prove the claim. For this purpose, we infer first from (6.29) that

$$
\begin{aligned}
E(\|a_n - x_n\|_H^2 \mid \mathcal{X}_n) &= E\Big(\|L_n(p_n - x_n)\|_H^2 \,\Big|\, \mathcal{X}_n\Big) \\
&= E\Big(|L_n|^2 \|p_n - x_n\|_H^2 \,\Big|\, \mathcal{X}_n\Big) \\
&= E\Big(L_n \sum_{i=1}^{M} \beta_{i,n} \|p_{i,n} - x_n\|_H^2 \,\Big|\, \mathcal{X}_n\Big) \\
&\geqslant E\Big(\sum_{i=1}^{M} \beta_{i,n} \|p_{i,n} - x_n\|_H^2 \,\Big|\, \mathcal{X}_n\Big) \\
&\geqslant E\Big(\delta \max_{1\leqslant j\leqslant M} \|p_{j,n} - x_n\|_H^2 \,\Big|\, \mathcal{X}_n\Big) \\
&\geqslant \delta E\Big(\|p_{1,n} - x_n\|_H^2 \,\Big|\, \mathcal{X}_n\Big) \\
&= \delta E\Big(\|T_{k_{1,n}} x_n - x_n\|_H^2 \,\Big|\, \mathcal{X}_n\Big).
\end{aligned} \tag{6.45}
$$

However, $k_{1,n}$ is independent of $\mathcal{X}_n$ and is a copy of k. Thus, Lemma 6.4 guarantees that, for P-almost every $\omega' \in \Omega$,

$$
\begin{aligned}
E\Big(\|T_{k_{1,n}} x_n - x_n\|_H^2 \,\Big|\, \mathcal{X}_n\Big)(\omega') &= \int_\Omega \|T_{k_{1,n}(\omega)} x_n(\omega') - x_n(\omega')\|_H^2 P(d\omega) \\
&= \int_\Omega \|T_{k(\omega)} x_n(\omega') - x_n(\omega')\|_H^2 P(d\omega).
\end{aligned} \tag{6.46}
$$

Therefore, for P-almost every $\omega' \in \Omega$, (6.45) yields

$$
E(\lambda_n^2 \|a_n - x_n\|_H^2 \mid \mathcal{X}_n)(\omega') \geqslant \varepsilon^2 \delta \int_\Omega \|T_{k(\omega)} x_n(\omega') - x_n(\omega')\|_H^2 P(d\omega) \ \text{ P-a.s.} \tag{6.47}
$$

Taking the expected value in (6.47), summing over $n \in \mathbb{N}$, and applying Theorem 6.5(vi), we find that

$$\mathsf{E}\left(\sum_{n\in\mathbb{N}} \int_{\Omega} \|\mathsf{T}_{k(\omega)} x_n - x_n\|_{\mathsf{H}}^2 \mathsf{P}(d\omega)\right) = \sum_{n\in\mathbb{N}} \mathsf{E}\left(\int_{\Omega} \|\mathsf{T}_{k(\omega)} x_n - x_n\|_{\mathsf{H}}^2 \mathsf{P}(d\omega)\right) < +\infty, \tag{6.48}$$

which implies

$$\sum_{n\in\mathbb{N}} \int_{\Omega} \|\mathsf{T}_{k(\omega)} x_n - x_n\|_{\mathsf{H}}^2 \mathsf{P}(d\omega) < +\infty \quad \mathsf{P}\text{-a.s.} \tag{6.49}$$

This leads to the existence of a set $\Omega' \in \mathcal{F}$ such that $\mathsf{P}(\Omega') = 1$ and, for every $\omega' \in \Omega'$, the series $\sum_{n\in\mathbb{N}} \int_{\Omega} \|\mathsf{T}_{k(\omega)} x_n(\omega') - x_n(\omega')\|_{\mathsf{H}}^2 \mathsf{P}(d\omega)$ converges. Furthermore, in view of Theorem 6.5(iv), we can assume that $\mathfrak{W}(x_n(\omega'))_{n\in\mathbb{N}} \neq \varnothing$ on Ω'. Fix $\omega' \in \Omega'$ and let $x(\omega') \in \mathfrak{W}(x_n(\omega'))_{n\in\mathbb{N}}$, say $x_{j_n}(\omega') \rightharpoonup x(\omega')$. It follows from the monotone convergence theorem that

$$\int_{\Omega} \sum_{n\in\mathbb{N}} \|\mathsf{T}_{k(\omega)} x_n(\omega') - x_n(\omega')\|_{\mathsf{H}}^2 \mathsf{P}(d\omega) = \sum_{n\in\mathbb{N}} \int_{\Omega} \|\mathsf{T}_{k(\omega)} x_n(\omega') - x_n(\omega')\|_{\mathsf{H}}^2 \mathsf{P}(d\omega) < +\infty. \tag{6.50}$$

Hence, for P-almost every $\omega \in \Omega$, $\sum_{n\in\mathbb{N}} \|\mathsf{T}_{k(\omega)} x_n(\omega') - x_n(\omega')\|_{\mathsf{H}}^2 < +\infty$. Therefore, there exists $\Omega'' \in \mathcal{F}$ such that $\mathsf{P}(\Omega'') = 1$ and

$$(\forall \omega \in \Omega'') \quad \mathsf{T}_{k(\omega)} x_n(\omega') - x_n(\omega') \to 0. \tag{6.51}$$

It then follows from the demiclosedness of the operators $(\mathrm{Id} - \mathsf{T}_k)_{k\in K}$ at 0 that

$$(\forall \omega \in \Omega'') \quad \mathsf{T}_{k(\omega)} x(\omega') = x(\omega'). \tag{6.52}$$

Therefore $x(\omega') \in \{\mathsf{z} \in \mathsf{H} \mid \mathsf{z} \in \operatorname{Fix} \mathsf{T}_k \ \mathsf{P}\text{-a.s.}\} = \mathsf{Z}$. Since ω' is arbitrarily taken in Ω', we conclude that $\mathfrak{W}(x_n)_{n\in\mathbb{N}} \subset \mathsf{Z}$ P-a.s., and the claim thus follows from Theorem 6.5(vii). ☐

Remark 6.9 The algorithm of Theorem 6.8 is inspired by the deterministic extrapolated algorithm proposed in [6, Section 6.5], where the collection of sets is at most countably infinite and the control rule $k_{i,n}$, the weights $\beta_{i,n}$, and the relaxation parameters λ_n are deterministic.

Remark 6.10 The random activation of indices in Theorem 6.8 replaces the control rule strategies of deterministic methods. Thus, the same approach to randomize the control rule can be used, for instance, to design randomized block-iterative strongly convergent algorithms to find the best approximation from the Kuhn-Tucker set associated with a primal-dual monotone inclusion problem [5, 9, 10].

6.2.4.2 Best approximation from a finite collection of closed convex sets

To connect Theorem 6.8 with Haugazeau's original work, we specialize it to solve the following simple best approximation problem.

Problem 6.11 Let $(\mathsf{Z}_k)_{1\leqslant k\leqslant p}$ be a finite collection of closed convex subsets of H such that $\mathsf{Z} = \bigcap_{1\leqslant k\leqslant p} \mathsf{Z}_k \neq \varnothing$ and let $\mathsf{x}_0 \in \mathsf{H}$. The goal is to find $\operatorname{proj}_{\mathsf{Z}} \mathsf{x}_0$.

To implement Theorem 6.8, we need first to embed Problem 6.11 into the setting of Problem 6.6. Let us define then

$$\begin{cases} \mathsf{K} = \{1, \ldots, p\}; \\ \mathcal{K} = 2^{\mathsf{K}}; \\ k \text{ is any } \mathsf{K}\text{-valued random variable such that } (\forall \mathsf{k} \in \{1, \ldots, p\})\ \mathsf{P}\big([k = \mathsf{k}]\big) > 0. \end{cases} \tag{6.53}$$

Then $\mathsf{Z} = \big\{\mathsf{z} \in \mathsf{H} \mid \mathsf{z} \in \mathsf{Z}_k\ \mathsf{P}\text{-a.s.}\big\} = \big\{\mathsf{z} \in \mathsf{H} \mid \mathsf{z} \in \operatorname{Fix} \operatorname{proj}_{\mathsf{Z}_k}\ \mathsf{P}\text{-a.s.}\big\}$ where, following Example 6.7(i), the firmly quasinonexpansive operators correspond to the projectors. We remark that any random variable k satisfying (6.53) suffices to define the randomized method. In particular, k may be uniformly distributed over K, i.e., for every $\mathsf{k} \in \{1, \ldots, p\}$, $\mathsf{P}([k = \mathsf{k}]) = 1/p$. However, (6.53) also allows for the implementation of nonuniform weights.

We implement Theorem 6.8 with $\lambda_n \equiv 1$ and $\mathsf{M} = 1$, which implies that $L_n \equiv 1$ and $r_n \equiv a_n \equiv p_n \equiv p_{1,n}$.

Corollary 6.12 *In the setting of Problem 6.11 and (6.53), iterate*

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} k_n \text{ is a copy of } k \text{ and is independent of } \sigma(\mathsf{x}_0, \ldots, x_n) \\ x_{n+1} = \mathsf{Q}\big(\mathsf{x}_0, x_n, \operatorname{proj}_{\mathsf{Z}_{k_n}} x_n\big). \end{array} \right. \end{array} \tag{6.54}$$

Then $(x_n)_{n\in\mathbb{N}}$ *converges strongly* P-a.s. *and strongly in* $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$ *to* $\operatorname{proj}_{\mathsf{Z}} \mathsf{x}_0$.

Remark 6.13

(i) The proposed randomized strategy differs fundamentally from Haugazeau's original cyclic approach. The cyclic method follows a deterministic order, ensuring that every set is activated exactly once every p iterations. This places a strict bound on how long any specific set is ignored. On the other hand, the random strategy selects the active index independently at each iteration. While this approach does not enforce a fixed time between visits to a specific set, it still ensures that all sets are visited infinitely often with probability one. Hence, it provides a simpler alternative that avoids rigid patterns. Moreover, the expected waiting time to activate a specific index k remains finite since it is the inverse of the probability of activation, e.g., it is equal to p in the case of uniformly distributed k.

(ii) Another way to analyze the asymptotic behavior of the algorithm is to fix $\omega \in \Omega$ and study the behavior of $(x_n(\omega))_{n\in\mathbb{N}}$ via the control rule $(k_n(\omega))_{n\in\mathbb{N}}$. In this case, the convergence of $(x_n(\omega))_{n\in\mathbb{N}}$ cannot be established by appealing to deterministic results since $(k_n(\omega))_{n\in\mathbb{N}}$ fails to satisfy the necessary assumptions; see [6, Definition 3.1 and Theorem 3.1].

(iii) To compare Haugazeau's original cyclic approach with the randomized strategy, we use $\lambda_n \equiv 1$ and $M = 1$ in Corollary 6.12. However, randomly selected block-iterative versions can also be implemented, where the extrapolation $L_n \geqslant 1$ will significantly accelerate the convergence, as shown for deterministic algorithms [8, 14, 28].

6.2.5 Application to the computation of a Chebyshev center

We specialize the setting to $H = \mathbb{R}^N$ and let S be a nonempty bounded subset of H. A Chebyshev center of S and the Chebyshev radius of S are a solution and the optimal value, respectively, of the problem

$$\underset{x \in H}{\text{minimize}} \ \sup_{y \in S} \|x - y\|_H. \tag{6.55}$$

We refer to [1, Chapter 15] for further background on Chebyshev centers. The goal is to find a Chebyshev center of S and the Chebyshev radius of S. However, this task can be difficult, particularly if the set S is nonconvex, which makes (6.55) nonconvex as well. Another way to rewrite this problem is as the following constrained minimization in a higher space:

$$\underset{(x,\rho) \in Z}{\text{minimize}} \ \rho, \quad \text{where} \ Z = \bigcap_{y \in S} \big\{(z, \xi) \in H \times \mathbb{R} \mid \|z - y\|_H \leqslant \xi\big\}. \tag{6.56}$$

For every $y \in S$, the set $\big\{(z, \xi) \in H \times \mathbb{R} \mid \|z - y\|_H \leqslant \xi\big\}$ is closed and convex, and so is the intersection. We propose the following α-approximation of (6.56).

Problem 6.14 Set $K = S$, let $\mathcal{K}$ be the Borel σ-algebra of H restricted to S, and let k be an S-valued random variable with uniform distribution over S. Let $\alpha \in \,]0, +\infty[$ and set $(x_0, \rho_0) = (0, -\alpha)$. The task is to

$$\underset{(x,\rho) \in Z}{\text{minimize}} \ \big\|(x, \rho) - (x_0, \rho_0)\big\|^2_{H \times \mathbb{R}}, \quad \text{where} \ Z = \big\{(z, \xi) \in H \times \mathbb{R} \mid \|z - k\|_H \leqslant \xi \ \ \text{P-a.s.}\big\}. \tag{6.57}$$

This α-approximation of (6.56) is a best approximation problem in the sense of Problem 6.6 under the quasinonexpansive operators defined by the projectors, as in Example 6.7(i). The objective function

$$(x, \rho) \mapsto \big\|(x, \rho) - (x_0, \rho_0)\big\|^2_{H \times \mathbb{R}} = \|x\|^2_H + |\rho|^2 + 2\alpha\rho + |\alpha|^2 \tag{6.58}$$

corresponds to a regularization of the objective function of (6.56) and the distance between their minimizers vanishes as $\alpha \uparrow +\infty$. Furthermore, we replace the strict intersection over every $y \in S$ with the almost sure intersection defined by k.

In our experiment, we solve three instances of Problem 6.14 for $H = \mathbb{R}^2$, $\alpha = 200$, and where S is generated randomly. For every $y \in S$, the set $\big\{(z, \xi) \in H \times \mathbb{R} \mid \|z - y\|_H \leqslant \xi\big\}$ corresponds to the translated second-order cone and its projector operator is known explicitly [17, Proposition 3.3]. We employ Theorem 6.8 with $M = 1$ and $\lambda_n \equiv 1$.

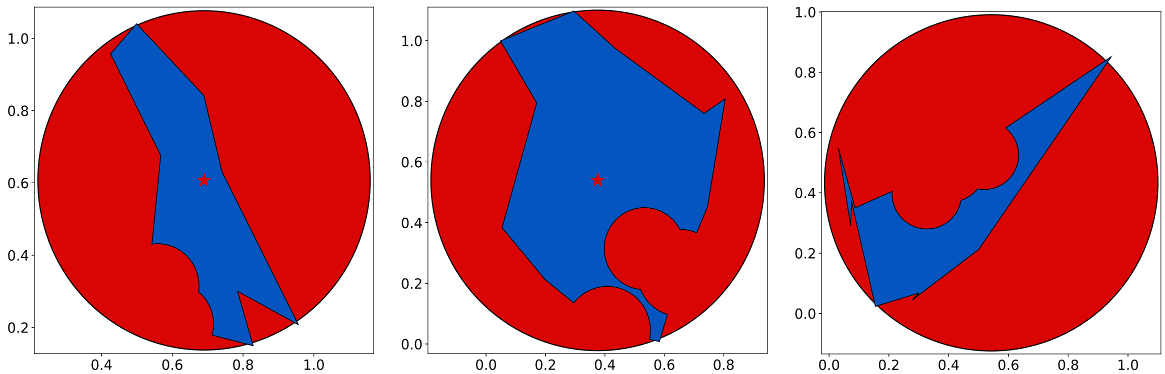


Figure 6.1 Three instances of the experiment of Section 6.2.5. Blue: The subset S. Red: The ball with center and radius given by the α-approximation.

Acknowledgement

This work is part of the author's Ph.D. dissertation. The author gratefully acknowledges the guidance of his Ph.D. advisor P. L. Combettes throughout this work. This work was supported by the National Science Foundation under grant DMS-2513409.

References

[1] A. R. Alimov and I. G. Tsar'kov, *Geometric Approximation Theory*. Springer, Cham, 2022.

[2] A. Alotaibi, P. L. Combettes, and N. Shahzad, Best approximation from the Kuhn–Tucker set of composite monotone inclusions, *Numer. Funct. Anal. Optim.*, vol. 36, pp. 1513–1532, 2015.

[3] H. H. Bauschke and P. L. Combettes, A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces, *Math. Oper. Res.*, vol. 26, pp. 248–264, 2001.

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

[5] M. N. Bùi and P. L. Combettes, Multivariate monotone inclusions in saddle form, *Math. Oper. Res.*, vol. 47, pp. 1082–1109, 2022.

[6] P. L. Combettes, Strong convergence of block-iterative outer approximation methods for convex optimization, *SIAM J. Control Optim.*, vol. 38, pp. 538–565, 2000.

[7] P. L. Combettes, Quasi-Fejérian analysis of some optimization algorithms, in: *Inherently Parallel Algorithms for Feasibility and Optimization*, (D. Butnariu, Y. Censor, and S. Reich, eds.), pp. 115–152. Elsevier, New York, 2001.

[8] P. L. Combettes, A block-iterative surrogate constraint splitting method for quadratic signal recovery, *IEEE Trans. Signal Process.*, vol. 51, pp. 1771–1782, 2003.

[9] P. L. Combettes, The geometry of monotone operator splitting methods, *Acta Numer.*, vol. 33, pp. 487–632, 2024.

[10] P. L. Combettes and J. Eckstein, Asynchronous block-iterative primal-dual decomposition methods for monotone inclusions, *Math. Program.*, vol. B168, pp. 645–672, 2018.

[11] P. L. Combettes and J. I. Madariaga, Almost-surely convergent randomly activated monotone operator splitting methods, *SIAM J. Imaging Sci.*, vol. 18, pp. 2177–2205, 2025.

[12] P. L. Combettes and J. I. Madariaga, A geometric framework for stochastic iterations, *Math. Comp.*, to appear.

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

[14] P. L. Combettes and Z. C. Woodstock, Reconstruction of functions from prescribed proximal points, *J. Approx. Theory*, vol. 268, art. 105606, 26 pp., 2021.

[15] A. Dieuleveut, G. Fort, E. Moulines, and H.-T. Wai, Stochastic approximation beyond gradient for signal processing and machine learning, *IEEE Trans. Signal Process.*, vol. 71, pp. 3117–3148, 2023.

[16] J. L. Doob, *Stochastic Processes*. Wiley, New York, 1953.

[17] M. Fukushima, Z.-Q. Luo, and P. Tseng, Smoothing functions for second-order-cone complementarity problems, *SIAM J. Optim.*, vol. 12, pp. 436–460, 2002.

[18] Y. Haugazeau, *Sur les Inéquations Variationnelles et la Minimisation de Fonctionnelles Convexes.* Thèse, Université de Paris, 1968.

[19] N. Hermer, D. R. Luke, and A. Sturm, Random function iterations for consistent stochastic feasibility, *Numer. Funct. Anal. Optim.*, vol. 40, pp. 386–420, 2019.

[20] T. Hytönen, J. van Neerven, M. Veraar, and L. Weis, *Analysis in Banach Spaces. Volume I: Martingales and Littlewood–Paley Theory.* Springer, New York, 2016.

[21] P. R. Johnstone, J. Eckstein, T. Flynn, and S. Yoo, Stochastic projective splitting, *Comput. Optim. Appl.*, vol. 87, pp. 397–437, 2024.

[22] V. R. Kostić and S. Salzo, The method of randomized Bregman projections for stochastic feasibility problems, *Numer. Algorithms*, vol. 93, pp. 1269–1307, 2023.

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

[24] D. R. Luke, S. Schultze, and H. Grubmüller, Stochastic algorithms for large-scale composite optimization: the case of likelihood maximization for X-FEL imaging, *Math. Program.*, to appear.

[25] M. Marques Alves, J. E. Navarro Caballero, and R. T. Marcavillaca, An inexact inertial projective splitting algorithm with strong convergence, *J. Optim. Theory Appl.*, vol. 207, art. 65, 2025.

[26] M. Marques Alves, J. E. Navarro Caballero, M. Geremia, and R. T. Marcavillaca, A strongly convergent inertial inexact proximal-point algorithm for monotone inclusions with applications to variational inequalities, *Optimization*, vol. 75, pp. 543–570, 2026.

[27] I. Necoara and A. Nedić, Minibatch stochastic subgradient-based projection algorithms for feasibility problems with convex inequalities, *Comput. Optim. Appl.*, vol. 80, pp. 121–152, 2021.

[28] G. Pierra, Éclatement de contraintes en parallèle pour la minimisation d'une forme quadratique, *Lecture Notes in Comput. Sci.*, vol. 41. Springer, New York, 1976, pp. 200–218.

[29] M. V. Solodov and B. F. Svaiter, Forcing strong convergence of proximal point iterations in a Hilbert space, *Math. Program.*, vol. A87, pp. 189-202, 2000.

Chapter 7

CONCLUSION

7.1 Summary

In this dissertation, we have proposed frameworks for the analysis and the design of stochastic iterative methods and, in doing so, we have addressed the open questions **(Q1)–(Q5)** of Chapter 1. First, we have presented a geometric algorithmic framework to construct stochastic quasi-Fejér monotone sequences. We have established the convergence principles for the constructed sequences. As applications of these frameworks, we have developed new stochastic extrapolated parallel algorithms for feasibility problems and randomized block-iterative projective splitting methods for solving systems of coupled inclusions. In addition, we applied the framework to provide novel approaches to classic algorithms such as the stochastic proximal point algorithm and the stochastic gradient method. Finally, we have proposed an abstract stochastic version of the Haugazeau method and investigated its asymptotic behavior.

7.2 Future work

Direction 7.1 We have proposed stochastic methods for a wide range of problems in Hilbertian nonlinear analysis. A natural extension of this work would be the study of stochastic iterative algorithms in Banach spaces.

Direction 7.2 In Section 3.2.5, we studied stochastic splitting algorithms for finding a common point of a family of sets. On the other hand, Section 4.2.5 is dedicated to investigating stochastic methods for finding a zero of a highly structured multivariate monotone inclusion problem. A new approach could emerge from combining these two problems into a new feasibility-composite inclusion problem, i.e., each set is the set of zeros of a composite monotone inclusion problem.

Direction 7.3 In Chapters 2 and 4, we studied different approaches for designing stochastic splitting algorithms for the sum of maximally monotone operators. Recently, paper [1] studied composite inclusion problems defined over the integral of maximally monotone operators; see [1, Example 5.8]. Another interesting question is to extend the proposed stochastic splitting algorithms beyond the (finite) sum of operators onto the integral of operators.

Direction 7.4 Our asymptotic analysis of the proposed methods consists of guaranteeing weak/strong convergence P-a.s. as well as weak/strong/linear convergence in $L^2(\Omega, \mathcal{F}, \mathsf{P}; \mathsf{H})$. Recent papers have studied stochastic methods in which the generated sequence $(x_\mathsf{n})_{\mathsf{n}\in\mathbb{N}}$ fails to converge in any of these modes, but converges in distribution to some invariant probability measure; see, e.g., [2, 3]. This topic could extend the framework of Chapter 3 to incorporate convergence in distribution under weaker conditions.

Raleigh, April 17, 2026

References

[1] M. N. Bùi and P. L. Combettes, Hilbert direct integrals of monotone operators, *Canad. J. Math.*, vol. 77, pp. 1425–1456, 2025.

[2] N. Hermer, D. R. Luke, and A. Sturm, Nonexpansive Markov operators and random function iterations for stochastic fixed point problems, *J. Convex Anal.*, vol. 30, pp. 1073–1114, 2023.

[3] D. R. Luke, Convergence in distribution of randomized algorithms: The case of partially separable optimization, *Math. Program.*, vol. A212, pp. 763–798, 2025.