J. Nonlinear Var. Anal. 11 (2027), No. 1, pp. 45-75
Available online at http://jnva.biemdas.com
https://doi.org/10.23952/jnva.11.2027.1.04

A UNIFIED STOCHASTIC APPROXIMATION FRAMEWORK FOR MEAN-SQUARE AND ALMOST SURE CONVERGENCE WITH RATES IN MULTIVALUED VARIATIONAL INEQUALITIES

SIXUAN BAI[1], AKHTAR A. KHAN[2,*], CHRISTIANE TAMMER[3], ADRIAN ZĂLINESCU[4]

[1]School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, China
[2]School of Mathematics and Statistics, Rochester Institute of Technology, Rochester, NY 14623, USA
[3]Institute of Mathematics, Martin-Luther-University of Halle-Wittenberg, Theodor-Lieser-Str. 5, D-06120 Halle-Saale, Germany
[4]Faculty of Computer Science, Alexandru Ioan Cuza University of Iași, Romania, and "O. Mayer" Institute of Mathematics, Romanian Academy, Iași, Romania

Abstract. This paper presents a unified stochastic approximation framework for solving multivalued variational inequalities in real separable Hilbert spaces. We propose an iterative projection scheme that operates over a sequence of expanding closed convex subsets, a formulation motivated by variable-mesh discretization strategies in infinite-dimensional optimization. The proposed algorithm utilizes noisy observations of the underlying operator, accommodating both intrinsic stochasticity (as in the minimization of expected-value functionals) and extrinsic computational errors. We establish the strong convergence of the generated sequence, in both the mean-square and almost sure senses, to the limit of the solutions of the discretized problems, which is a solution of the original variational inequality under standard approximation conditions. Furthermore, we derive explicit rates of convergence for both modes of convergence, and we treat the ill-posed case of vanishing regularization. The versatility of the framework is demonstrated through applications to non-smooth stochastic optimization, support vector machines, optimal control and inverse problems for partial differential equations with random data, and image deblurring.

Keywords. Almost sure convergence; Multivalued variational inequalities; Rate of convergence; Regularized Robbins–Monro approach; Stochastic approximation.

1. INTRODUCTION

Let $(\Omega, \mathscr{A}, \mathbb{P})$ be a probability space. Let H be a real separable Hilbert space equipped with the norm $\|\cdot\|$ and the scalar product $\langle\cdot,\cdot\rangle$. We denote by $L^2_H(\Omega,\mathscr{A},\mathbb{P})$ the Bochner space of equivalence classes of square-integrable random variables from Ω to H, where the norm is defined by $\|U\|_{L^2} := \left(\int_\Omega \|U(\omega)\|^2 d\mathbb{P}(\omega)\right)^{1/2}$. Let $F : H \rightrightarrows H$ be a multivalued mapping with domain $D(F)$, and let C be a non-empty, closed, and convex subset of H such that $C \subset D(F)$.

*Corresponding author.
E-mail address: bsx6621@163.com (S. Bai), aaksma@rit.edu (A.A. Khan) christiane.tammer@mathematik.uni-halle.de (C. Tammer) adrian.zalinescu@uaic.ro (A. Zălinescu).

Received 27 March 2026; Accepted 1 October 2026; Published online 10 November 2026.

©2027 Journal of Nonlinear and Variational Analysis

Our primary objective is to investigate the multivalued variational inequality which consists of finding a point $u \in C$ such that there exists $\eta \in F(u)$ satisfying

$$\langle \eta, v-u \rangle \geq 0, \quad \text{for every } v \in C. \tag{1.1}$$

To approximate a solution, we consider a sequence of discretized variational inequalities defined over expanding subspaces. Let $\{H_n\}_{n\in\mathbb{N}}$ be an increasing sequence of closed subspaces of H (i.e., $H_n \subset H_{n+1}$), and let P_n denote the orthogonal projection onto H_n. We consider a sequence of non-empty, closed, and convex subsets $\{C_n\}_{n\in\mathbb{N}}$ such that $C_n \subset H_n$ for all $n \geq 1$, and a sequence of multivalued mappings $\{F_n\}_{n\in\mathbb{N}}$ with $F_n : H_n \rightrightarrows H$ and $C_n \subset D(F_n)$. At each step $n \in \mathbb{N}$, the discretized problem is to find $w_n \in C_n$ such that for some $w_n^* \in F_n(w_n)$,

$$\langle w_n^*, v-w_n \rangle \geq 0, \quad \text{for every } v \in C_n. \tag{1.2}$$

We propose a unified stochastic approximation framework to solve (1.1). Let $\{t_n\}_{n\in\mathbb{N}}$ be a sequence of positive step sizes. We define the mapping $Q_n := P_{C_n}P_n$ as the composition of the metric projection P_{C_n} onto the set C_n and the orthogonal projection P_n onto the subspace H_n. Since $C_n \subset H_n$, for every $x \in H$ and $c \in C_n$ we have $\|x-c\|^2 = \|x-P_nx\|^2 + \|P_nx-c\|^2$, so that, in fact, $Q_n = P_{C_n}$. We nevertheless keep the factorized form, since it reflects the two-stage structure of the discretization: the update is first represented in the discrete space H_n and then projected onto the discrete constraint set C_n. Starting from an initial guess $u_1 \in L^2_H(\Omega, \mathscr{A}, \mathbb{P})$ with $u_1(\omega) \in C_1$ for almost every $\omega \in \Omega$, the sequence $\{u_n\}$ is generated by the update rule

$$u_{n+1}(\omega) := Q_n\left(u_n(\omega) - t_n\eta_n(\omega)\right). \tag{1.3}$$

Here, $\eta_n(\omega) := u_n^*(\omega) + \xi_n(\omega)$ represents a noisy observation of the operator at the current state. The term $u_n^*(\omega) \in F_n(u_n(\omega))$ is a measurable selection from the multivalued operator (its existence is shown in the proof of Theorem 2.1; note that $u_n(\omega) \in C_n \subset D(F_n)$ under the nestedness assumption (T_3) below), while $\xi_n(\omega)$ is a random noise variable. To ensure the stochastic process is rigorously defined despite the non-invertibility of the projection operator Q_n (see [1]), we let $\{\mathscr{F}_n\}_{n\geq 1}$ denote the filtration generated by the initial state and the history of the noise sequence: $\mathscr{F}_n := \sigma(u_1, \xi_1, \ldots, \xi_{n-1})$ for $n \geq 2$, with $\mathscr{F}_1 := \sigma(u_1)$. By construction, the current iterate u_n is $\mathscr{F}_n$-measurable (see the proof of Theorem 2.1). Furthermore, the noise ξ_n is $\mathscr{F}_{n+1}$-measurable, and we assume that the standard martingale difference condition $\mathbb{E}[\xi_n \mid \mathscr{F}_n] = 0$ holds. Consequently, the observation η_n is $\mathscr{F}_{n+1}$-measurable.

This formulation provides a versatile framework encompassing several important problem classes. If F and F_n are single-valued, (1.1) reduces to the classical variational inequality, and u_n^* is simply the evaluation $F_n(u_n)$. More critically, this setup addresses non-smooth stochastic optimization. Consider minimizing a convex functional defined as an expectation, $J(u) := \mathbb{E}[\psi(u,\zeta)]$, where ζ is a random variable. In this context, $F(u)$ is identified with the subdifferential $\partial J(u)$. Under standard regularity conditions, the subdifferential commutes with the expectation, that is, $\partial J(u) = \mathbb{E}[\partial_u \psi(u,\zeta)]$. This allows us to generate the noisy observation η_n by drawing a sample ζ_n, independent of $\mathscr{F}_n$, and selecting a stochastic subgradient $\eta_n := g_n \in \partial_u\psi(u_n,\zeta_n)$. The decomposition $\eta_n = u_n^* + \xi_n$, with $u_n^* := \mathbb{E}[g_n \mid \mathscr{F}_n] \in \partial J(u_n)$ and $\xi_n := g_n - u_n^*$, then naturally interprets u_n^* as the true expected subgradient and ξ_n as the zero-mean stochastic error.

Furthermore, our approach remains valid even when the operator F is deterministic but computationally expensive. If $F(u)$ represents the solution to a complex physical law or differential

equation, the "noise" ξ_n can model numerical approximation errors or sensor measurement noise. Provided these errors satisfy martingale-like bounds, our framework allows for the use of coarse approximations in early iterations that are progressively refined, guaranteeing convergence to the true solution despite extrinsic uncertainties.

We now illustrate the breadth of this framework with four specific applications.

Example 1.1 (Application to Support Vector Machines)**.** The proposed framework provides a rigorous foundation for training support vector machines (SVMs) using stochastic iterative algorithms. The standard primal formulation for a soft-margin linear SVM seeks a weight vector $u \in \mathbb{R}^d$ that minimizes the regularized hinge loss:

$$\min_{u\in\mathbb{R}^d} J(u) := \frac{\lambda}{2}\|u\|^2 + \mathbb{E}_{(x,y)\sim\mathscr{D}}\left[\max(0, 1 - y\langle x, u\rangle)\right],$$

where $\lambda > 0$ is the regularization parameter. Here, $x \in \mathbb{R}^d$ represents the feature vector and $y \in \{-1, +1\}$ denotes the binary class label. The notation $(x, y) \sim \mathscr{D}$ signifies that the data pairs are random variables distributed according to an unknown underlying probability distribution $\mathscr{D}$.

This optimization problem maps directly to our variational inequality framework (1.1) as follows:

- The operator $F(u) = \partial J(u)$ is the subdifferential of the objective function. Owing to the non-smoothness of the function $\max(0, \cdot)$ at the origin, the subdifferential of the sample loss is set-valued whenever $y\langle x, u\rangle = 1$; consequently, $F(u)$ may be set-valued as well (for instance, when $\mathscr{D}$ is an empirical distribution supported on finitely many data points).
- The stochastic update (1.3) corresponds to drawing a random data point (x_n, y_n) at iteration n and computing a stochastic subgradient:

$$\eta_n = \lambda u_n - \begin{cases} y_n x_n & \text{if } y_n\langle x_n, u_n\rangle < 1, \\ 0 & \text{otherwise.} \end{cases}$$

- The noise term $\xi_n = \eta_n - u_n^*$, where $u_n^* := \mathbb{E}[\eta_n \mid \mathscr{F}_n] \in \partial J(u_n)$, represents the sampling error introduced by using a single data point to estimate a subgradient of the entire population risk.

Thus, under the assumptions of our main results, such stochastic subgradient algorithms converge to the optimal SVM classifier in the mean-square and almost sure senses, even though the loss function is non-smooth and the data stream may be infinite or noisy.

Example 1.2 (Application to Sparse Control of PDEs with Random Coefficients)**.** Our framework is particularly well suited for optimal control problems constrained by partial differential equations (PDEs) with uncertain data, especially when sparse controls are desired. Consider the problem of minimizing an expected cost functional:

$$\min_{u\in U_{\text{ad}}} J(u) := \mathbb{E}\left[\frac{1}{2}\|y(u,\omega) - y_d\|_{L^2}^2\right] + \lambda\|u\|_{L^1},$$

subject to the stochastic elliptic constraint:

$$-\nabla\cdot(a(x,\omega)\nabla y(x,\omega)) = u(x) \quad \text{in } D, \quad y = 0 \text{ on } \partial D.$$

Here, $a(x,\omega)$ represents a random diffusion coefficient (e.g., uncertain permeability in a porous medium), $y(u,\omega)$ is the random state, and u is the deterministic control. The set U_{ad}, a non-empty, closed, and convex subset of $L^2(D)$, represents the set of admissible controls. The term $\lambda\|u\|_{L^1}$ is a non-smooth regularization promoting sparsity (e.g., minimizing the number of active actuators).

Writing $j(u,\omega) := \frac{1}{2}\|y(u,\omega) - y_d\|_{L^2}^2$, the optimality condition for this problem is the variational inequality: find $u \in U_{\text{ad}}$ and $\zeta \in \partial(\lambda\|\cdot\|_{L^1})(u)$ such that

$$\langle \mathbb{E}[\nabla_u j(u,\omega)] + \zeta, v - u\rangle \geq 0, \quad \text{for all } v \in U_{\text{ad}}.$$

This fits our general framework (1.1) where:

- The operator F is the sum of the expected gradient of the smooth part and the multivalued subdifferential of the L^1 norm.
- The stochastic iteration (1.3) corresponds to a projected stochastic subgradient method. At each step n, we sample a realization ω_n of the random coefficient, solve the corresponding deterministic PDE to obtain $y(u_n,\omega_n)$, compute the gradient $\nabla_u j(u_n,\omega_n)$ (e.g., via an adjoint equation), and add a subgradient of the L^1 term.
- The "noise" arises from approximating the full expectation $\mathbb{E}[\cdot]$ with a single sample solution.

Our convergence results justify the use of this computationally efficient "sample-then-optimize" strategy for finding sparse optimal controls in the presence of uncertainty.

Example 1.3 (Application to Inverse Problems with Random Sources)**.** A further significant application of our framework lies in solving inverse problems for partial differential equations where the source term is stochastic.

Consider the problem of identifying a deterministic diffusion coefficient $u(x)$ (the unknown parameter) in an elliptic equation, given noisy measurements of the state. The system is governed by

$$-\nabla\cdot(u(x)\nabla y) = f(x,\omega) \quad \text{in } D, \quad y = 0 \text{ on } \partial D,$$

where $f(x,\omega)$ represents a random source term (e.g., stochastic excitation or uncertain load). We seek to minimize the expected mismatch between the state $y(u,\omega)$ and some observed data z_d, typically with regularization:

$$\min_{u\in C_{\text{ad}}} J(u) := \mathbb{E}\left[\frac{1}{2}\|y(u,\omega) - z_d\|_{L^2}^2\right] + \mathscr{R}(u),$$

where $\mathscr{R}$ is a (possibly non-smooth) regularization functional, such as total variation (TV), used to preserve edges in the coefficient $u(x)$.

The set C_{ad} defines the admissible search space for the coefficient. To ensure that the PDE is well posed (uniformly elliptic), C_{ad} typically enforces uniform boundedness away from zero:

$$C_{\text{ad}} = \{u \in L^2(D) : 0 < \underline{\alpha} \leq u(x) \leq \overline{\alpha} \text{ a.e. in } D\},$$

which is a non-empty, closed, and convex subset of $L^2(D)$ contained in $L^\infty(D)$.

Since the coefficient-to-state map $u \mapsto y(u,\omega)$ is nonlinear, the output least-squares functional above is, in general, nonconvex, so that the monotonicity assumptions of our main results need not hold for it. A convex alternative is provided by the modified output least-squares (MOLS)

functional, which leads to a monotone operator F and is therefore covered by our analysis; see, e.g., [2, 3].

This problem fits the variational inequality framework (1.1) as follows:

- The randomness in the stochastic estimates of the operator $F(u) = \partial J(u)$ comes entirely from the stochastic source $f(\cdot, \omega)$, which propagates through the PDE solution map $u \mapsto y(u, \omega)$.
- The iterative scheme (1.3) allows us to reconstruct the coefficient $u(x)$ by repeatedly solving the forward PDE with sampled realizations of the source $f_n = f(\cdot, \omega_n)$ and updating the estimate via stochastic subgradient descent.
- The non-smooth regularization $\mathscr{R}(u)$ (e.g., for reconstructing blocky structures) is handled by the multivalued nature of the operator.

Example 1.4 (Application to Stochastic Image Deblurring)**.** The proposed framework is ideally suited for regularized image deblurring and restoration, particularly when dealing with large-scale images or hyperspectral data where computing full gradients is expensive.

The image restoration problem seeks to recover an original image u from a noisy, blurred observation $y = Ku + \varepsilon$, where K is a linear blurring operator (convolution) and ε is the measurement noise. The standard variational formulation with non-smooth regularization (e.g., total variation or wavelet sparsity) is

$$\min_u J(u) := \frac{1}{2}\|Ku - y\|^2 + \lambda \mathscr{R}(u).$$

This fits our variational inequality framework (1.1) precisely:

- Since the operator K is linear, the data fidelity term $J_{\text{data}}(u) := \frac{1}{2}\|Ku - y\|^2$ is a convex quadratic functional, and its gradient $\nabla J_{\text{data}}(u) = K^*(Ku - y)$ is a globally monotone operator.
- The regularization $\mathscr{R}(u)$ (e.g., $\mathscr{R}(u) = \|u\|_{TV}$ or $\|u\|_1$) is convex but non-differentiable, requiring the set-valued subdifferential $F(u) = K^*(Ku - y) + \lambda \partial \mathscr{R}(u)$.
- For very large images, evaluating the full residual $Ku - y$ and the back-projection K^* is costly. We can interpret the data term as a sum over the discrete set of pixel indices $\mathscr{I}$:

$$\|Ku - y\|^2 = \sum_{i \in \mathscr{I}} ((Ku)_i - y_i)^2.$$

- Our iterative scheme (1.3) allows us to perform stochastic (sub)gradient descent: at each step, we sample a random subset of pixels (a batch $S_n \subset \mathscr{I}$) to compute the partial error, rescaled so that the resulting gradient estimate is unbiased, and update the image. Under the assumptions of our main results, this guarantees convergence to a minimizer of J without processing the full dataset at every iteration.

Remark 1.1. It is important to note that while non-smooth regularization functionals (such as total variation) are often defined on Banach spaces in the continuous setting, our computational framework operates on their finite-dimensional discretizations. In practice, continuous fields are approximated by vectors in $\mathbb{R}^N$, which is naturally a Hilbert space. In this discrete setting, the subdifferential operators are well-defined set-valued mappings, ensuring that our Hilbert space convergence analysis remains fully rigorous for practical numerical implementations.

A distinguishing feature of our work is the integration of these stochastic methods with a variable-mesh strategy. Standard stochastic approximation methods typically operate on a fixed discretization, which leads to a dilemma: coarse meshes introduce a large discretization bias, while fine meshes make every iteration computationally expensive. Assuming that the constraint sets are nested ($C_n \subset C_{n+1}$) and approximate C as $n \to \infty$ allows us to model a hierarchical refinement strategy, similar to adaptive mesh refinement in finite element methods. By coordinating the step-size decay with the expansion of the discretization space, our algorithm generates a sequence that converges strongly in H to the limit of the discrete solutions, and hence to a solution of the continuous problem whenever the discretization is consistent. This ensures asymptotic consistency and allows for significant computational savings by performing early iterations on coarse, inexpensive grids.

The iterative schemes proposed herein fall within the broader framework of stochastic approximation. This dynamic field, initiated by the seminal work of Robbins and Monro [4], provides a robust methodology for finding roots or extrema of functions when only noisy evaluations are available. The field possesses a rich history with applications across diverse domains; see [5, 6, 7, 8, 9, 10] and the references therein. In recent years, stochastic approximation has garnered significant renewed interest, driven largely by advancements in large-scale machine learning, stochastic gradient descent (SGD) variants, and stochastic variational inequalities [11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24]. Some recent developments can be found in [25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35].

Although the stochastic approximation framework is naturally suited for optimization problems involving noisy experimental data, suggesting strong utility for inverse and control problems under uncertainty, its application to stochastic partial differential equations (SPDEs), understood here as PDEs with random coefficients or data, is a surprisingly recent development. Unlike traditional sample average approximation (SAA) methods, which can be computationally prohibitive in high dimensions, stochastic approximation offers an efficient, iterative approach to handling the infinite-dimensional noise inherent in SPDEs. Specifically, Geiersbach and Pflug [36] and Martin, Krumscheid, and Nobile [37] independently pioneered the stochastic approximation approach to optimal control in SPDEs (see also [38, 39]). Furthermore, the application of these methods to inverse problems involving parameter estimation in SPDEs is detailed in [2, 3, 40, 41, 42, 43, 44, 45].

Main Contributions. In this work, we develop a unified framework for analyzing the convergence of stochastic approximation schemes applied to multivalued variational inequalities. Our analysis bridges the gap between standard mean-square error estimates and the stricter requirements of almost sure convergence, with a particular focus on the challenging regime of vanishing regularization. Notably, we extend the seminal results of Kluge and Telschow [46]. The main results are organized as follows:

- **Mean-Square Convergence Analysis:** We begin by establishing the fundamental L^2 properties of the algorithm. In Theorem 2.1, we prove the mean-square convergence of the sequence under general conditions on the operator monotonicity and noise variance. We then refine this analysis in Theorem 2.2, deriving explicit algebraic convergence rates $\mathcal{O}(n^{-q})$ for the well-posed case, where p and r denote the decay exponents of the step size and the spatial discretization error: the rate $q = \min(p, r)$ holds in general, and it improves to $q = \min(p, 2r)$ when the discrete solutions vary slowly between

consecutive iterations. Finally, in Theorem 2.3, we extend the mean-square analysis to vanishing regularization ($c_n \to 0$), where the consistency condition on the discretization error is replaced by a tracking condition on the increments of the evolving regularized solutions.

- **Almost Sure (Pathwise) Convergence:** Moving beyond expected values, we address the stronger requirement of trajectory-wise convergence. Theorem 3.1 establishes almost sure convergence under summability hypotheses on the step sizes and mesh errors. Complementing this, Theorem 3.2 provides pathwise convergence rates, ensuring that the error decays polynomially along almost every realization of the stochastic process, again with improved rates when the discrete solutions vary slowly.
- **Vanishing Regularization and Target Stability:** A central contribution of this work is the rigorous treatment of the shifting target problem inherent to vanishing regularization. Theorem 3.3 proves almost sure convergence to the minimum-norm solution even when the regularization parameter decays to zero. To make the abstract conditions of this theorem actionable, we provide a stability analysis of the underlying deterministic trajectory. Lemma 3.1 quantifies the discrete velocity of the regularized solutions, showing that it decays as $\mathcal{O}(1/n)$ for polynomial regularization schedules.
- **Optimal Parameter Selection:** Finally, we translate our theoretical findings into concrete guidelines for parameter selection. In Theorem 3.4, we prove that for polynomial step sizes $t_n \propto n^{-p}$ and regularization parameters $c_n \propto n^{-k}$, the condition $p+k<1$ (with $\frac{1}{2} < p < 1$) is sufficient to guarantee almost sure convergence. This result delineates a "feasibility region" where the algorithm successfully balances noise suppression against the drift of the moving target.

The remainder of this paper is organized as follows. Section 2 focuses on the analysis in the mean-square sense; we establish the strong convergence of the proposed iterative scheme, derive explicit a priori estimates for the rate of convergence, and treat the case of vanishing regularization. In Section 3, we extend this framework to the almost sure setting, utilizing martingale convergence arguments to establish analogous convergence results and rate estimates. This section also treats vanishing regularization in the almost sure setting, analyzes the stability of the regularized solutions with respect to the regularization parameter, and derives explicit polynomial parameter schedules. Section 4 presents some concluding remarks. Appendix A collects the auxiliary results used throughout, including recursive inequalities for real sequences and the Robbins–Siegmund theorem.

2. Analysis of Mean-Square Convergence

In this section, we establish the convergence of the proposed stochastic approximation scheme in the mean-square sense. Our analysis relies on controlling the error between the stochastic iterate u_n and the theoretical limit $\bar{u}$, which is decomposed into an approximation error (arising from the discretization w_n) and a stochastic error (arising from noisy observations). To derive these results, we utilize a suite of recursive inequality lemmas provided in Appendix A. Specifically, we employ Lemma A.1 to establish the asymptotic convergence of the recursive sequence and Lemma A.2 to ensure uniform boundedness in the presence of nonlinear perturbations. Furthermore, we characterize the polynomial decay rate of the error using Lemma A.3. The

Robbins–Siegmund theorem (Lemma A.4) will be used in Section 3 to guarantee the almost sure convergence of the iterative stochastic dynamics.

We now state the main convergence theorem.

Theorem 2.1 (Mean-Square Convergence)**.** *Assume that the following conditions hold:*

(T_1) *For every $n \in \mathbb{N}$, the multivalued operator F_n has non-empty, closed, and convex values on C_n, is measurable, and is strongly monotone on C_n. Specifically, there exists a sequence of constants $c_n > 0$ such that for all $u, v \in C_n$ and all $u^* \in F_n(u)$, $v^* \in F_n(v)$,*

$$\langle u^* - v^*, u - v \rangle \geq c_n \|u - v\|^2.$$

(T_2) *There exist constants $\kappa, \ell \geq 0$ such that, for all $n \in \mathbb{N}$, $\|v^*\| \leq \kappa \|v\| + \ell$, for all $v \in C_n$ and $v^* \in F_n(v)$.*

(T_3) *The sequence of feasible sets is nested, satisfying $C_n \subseteq C_{n+1}$ for all $n \geq 1$.*

(T_4) *For every n, the discretized problem (1.2) admits a solution $w_n \in C_n$. Furthermore, $\{w_n\}$ converges strongly to a limit $\bar{u} \in H$: $\lim_{n\to\infty} \|w_n - \bar{u}\| = 0$.*

(T_5) *The sequence of monotonicity constants $\{c_n\}_{n\in\mathbb{N}}$ is bounded. Furthermore, the step sizes $t_n > 0$ and the constants c_n satisfy the following conditions:*

$$\lim_{n\to\infty} \frac{t_n}{c_n} = 0, \quad \sum_{n=1}^{+\infty} t_n c_n = +\infty,$$

and the discretization error decays sufficiently fast relative to the conditioning:

$$\lim_{n\to\infty} \frac{\|w_n - \bar{u}\|}{c_n} = 0.$$

(T_6) *The stochastic errors $\{\xi_n\}_{n\in\mathbb{N}}$ form a martingale difference sequence in $L^2_H(\Omega, \mathscr{A}, \mathbb{P})$ with respect to the filtration $\{\mathscr{F}_n\}_{n\geq 1}$ defined in Section 1. More precisely, there exists a finite constant $\sigma > 0$ such that*

$$\mathbb{E}[\xi_n \mid \mathscr{F}_n] = 0 \quad \text{and} \quad \mathbb{E}[\|\xi_n\|^2 \mid \mathscr{F}_n] \leq \sigma^2 \quad \text{a.s.}$$

Then the sequence $\{u_n\}$ generated by the stochastic iteration (1.3) is well defined in $L^2_H(\Omega, \mathscr{A}, \mathbb{P})$ and converges in mean square to $\bar{u}$:

$$\lim_{n\to\infty} \|u_n - \bar{u}\|_{L^2}^2 = \lim_{n\to\infty} \mathbb{E}\left[\|u_n - \bar{u}\|^2\right] = 0.$$

Proof. We first show by induction that the iterates are well defined and satisfy $u_n(\omega) \in C_n$ for almost every ω. By assumption, the initial guess $u_1 \in L^2_H(\Omega, \mathscr{A}, \mathbb{P})$ is $\mathscr{F}_1$-measurable and satisfies $u_1(\omega) \in C_1$ a.s. Assume that $u_n \in L^2_H(\Omega, \mathscr{A}, \mathbb{P})$ is $\mathscr{F}_n$-measurable with $u_n(\omega) \in C_n \subset D(F_n)$ a.s. By hypothesis (T1), the operator F_n has non-empty, closed, and convex values on C_n. Because F_n is a measurable set-valued mapping and u_n is $\mathscr{F}_n$-measurable with $u_n(\omega) \in C_n$ a.s., the composition $F_n \circ u_n$ is a measurable set-valued map with non-empty closed images. Therefore, by the Measurable Selection Theorem (see, e.g., Aubin and Frankowska [47, Theorem 8.1.3]), there exists an $\mathscr{F}_n$-measurable selection $u_n^* \in F_n(u_n)$.

Furthermore, because the noise ξ_n is $\mathscr{F}_{n+1}$-measurable, the vector $u_n - t_n(u_n^* + \xi_n)$ is $\mathscr{F}_{n+1}$-measurable. As $Q_n = P_{C_n} P_n$ is a continuous mapping, it preserves measurability, and hence u_{n+1} is $\mathscr{F}_{n+1}$-measurable. By the linear growth condition (T_2) and the square-integrability of both u_n and ξ_n, the next iterate u_{n+1} has a finite second moment, so that $u_{n+1} \in L^2_H(\Omega, \mathscr{A}, \mathbb{P})$. Finally, $u_{n+1}(\omega) \in C_n \subseteq C_{n+1}$ by the definition of Q_n and the nestedness assumption (T_3).

We divide the convergence proof into four steps. Throughout, $\eta_n = u_n^* + \xi_n$ denotes the noisy observation used in (1.3), and $w_n^* \in F_n(w_n)$ is the element associated with the solution w_n of (1.2).

Step 1. We establish three elementary estimates. Since $u_n \in C_n \subset H_n$, we have $P_n u_n = u_n$ and $Q_n(u_n) = P_{C_n}(P_n u_n) = P_{C_n} u_n = u_n$.

(a) Using the non-expansivity of Q_n, we obtain

$$\|u_{n+1} - u_n\| = \|Q_n(u_n - t_n\eta_n) - Q_n(u_n)\| \le \|(u_n - t_n\eta_n) - u_n\| = t_n\|\eta_n\|. \tag{2.1}$$

Note that this estimate relies on the nested property (T_3). Without it, an additional interpolation error term involving $\|u_n - P_{C_n}(u_n)\|$ would be required.

(b) By the characteristic property of the metric projection P_{C_n} (which acts as the outer component of $Q_n = P_{C_n}P_n$), for any vector $z \in H$ and any point $v \in C_n$, the variational inequality $\langle P_{C_n}z - z, P_{C_n}z - v\rangle \le 0$ holds. We apply this with $z = P_n(u_n - t_n\eta_n)$ and $v = w_n \in C_n$. Utilizing the algebraic identity $\langle a-b, c\rangle = \langle a-z, c\rangle + \langle z-b, c\rangle$ with $a = u_{n+1}$ and $b = u_n$, we derive

$$\begin{aligned}\langle u_{n+1} - u_n, u_{n+1} - w_n\rangle &= \langle u_{n+1} - z, u_{n+1} - w_n\rangle + \langle z - u_n, u_{n+1} - w_n\rangle \\ &\le \langle P_n(u_n - t_n\eta_n) - u_n, u_{n+1} - w_n\rangle.\end{aligned}$$

Since $P_n u_n = u_n$, the linearity and self-adjointness of P_n (noting that $u_{n+1} - w_n \in H_n$) give

$$\begin{aligned}\langle u_{n+1} - u_n, u_{n+1} - w_n\rangle &\le \langle P_n(-t_n\eta_n), u_{n+1} - w_n\rangle = -t_n\langle \eta_n, u_{n+1} - w_n\rangle \\ &= -t_n\langle \eta_n, u_{n+1} - u_n\rangle - t_n\langle \eta_n, u_n - w_n\rangle.\end{aligned}$$

By the Cauchy–Schwarz inequality and (2.1), $-t_n\langle \eta_n, u_{n+1} - u_n\rangle \le t_n\|\eta_n\|\|u_{n+1} - u_n\| \le t_n^2\|\eta_n\|^2$, and therefore

$$\langle u_{n+1} - u_n, u_{n+1} - w_n\rangle \le t_n^2\|\eta_n\|^2 - t_n\langle u_n^* + \xi_n, u_n - w_n\rangle. \tag{2.2}$$

(c) By the definition of w_n in (1.2), we have $\langle w_n^*, v - w_n\rangle \ge 0$ for all $v \in C_n$, and in particular $\langle w_n^*, u_n - w_n\rangle \ge 0$, since $u_n \in C_n$. Together with the strong monotonicity condition (T_1), this yields

$$-\langle u_n^*, u_n - w_n\rangle = -\langle u_n^* - w_n^*, u_n - w_n\rangle - \langle w_n^*, u_n - w_n\rangle \le -c_n\|u_n - w_n\|^2.$$

Using $\|u_n - w_n\| \ge |\,\|u_n - \bar{u}\| - \|w_n - \bar{u}\|\,|$, we have

$$-\|u_n - w_n\|^2 \le -\|u_n - \bar{u}\|^2 + 2\|u_n - \bar{u}\|\|w_n - \bar{u}\| - \|w_n - \bar{u}\|^2,$$

and discarding the last (negative) term, we obtain

$$-\langle u_n^*, u_n - w_n\rangle \le -c_n\|u_n - \bar{u}\|^2 + 2c_n\|u_n - \bar{u}\|\|w_n - \bar{u}\|. \tag{2.3}$$

Step 2. We now estimate the conditional expectation $\mathbb{E}[\|u_{n+1} - \bar{u}\|^2 \mid \mathscr{F}_n]$, where $\mathscr{F}_n$ is the σ-algebra generated by the history of the process up to iteration n.

We begin with the algebraic expansion of the squared norm $\|u_{n+1} - \bar{u}\|^2$. To facilitate the use of the projection property from Step 1, we decompose the cross term relative to the auxiliary point w_n:

$$\begin{aligned}\|u_{n+1} - \bar{u}\|^2 &= \|(u_{n+1} - u_n) + (u_n - \bar{u})\|^2 \\ &= \|u_n - \bar{u}\|^2 + \|u_{n+1} - u_n\|^2 + 2\langle u_{n+1} - u_n, u_n - \bar{u}\rangle.\end{aligned}$$

We decompose the inner product by adding and subtracting w_n and u_{n+1} within the second slot:

$$\begin{aligned}2\langle u_{n+1}-u_n, u_n-\bar{u}\rangle &= 2\langle u_{n+1}-u_n, u_{n+1}-w_n\rangle + 2\langle u_{n+1}-u_n, w_n-u_{n+1}+u_n-\bar{u}\rangle \\ &= 2\langle u_{n+1}-u_n, u_{n+1}-w_n\rangle + 2\langle u_{n+1}-u_n, w_n-\bar{u}\rangle - 2\|u_{n+1}-u_n\|^2.\end{aligned}$$

Substituting this back into the norm expansion yields

$$\begin{aligned}\|u_{n+1}-\bar{u}\|^2 \le \|u_n-\bar{u}\|^2 + \|u_{n+1}-u_n\|^2 + 2\langle u_{n+1}-u_n, u_{n+1}-w_n\rangle \\ + 2\langle u_{n+1}-u_n, w_n-\bar{u}\rangle.\end{aligned}$$

We now apply (2.1) to the second term, (2.2) to the third term, and the Cauchy–Schwarz inequality together with (2.1) to the last term. This yields

$$\begin{aligned}\|u_{n+1}-\bar{u}\|^2 \le \|u_n-\bar{u}\|^2 + t_n^2\|u_n^*+\xi_n\|^2 - 2t_n\langle u_n^*, u_n-w_n\rangle - 2t_n\langle \xi_n, u_n-w_n\rangle \\ + 2t_n^2\|u_n^*+\xi_n\|^2 + 2t_n\|u_n^*+\xi_n\|\|w_n-\bar{u}\|.\end{aligned}$$

Grouping the squared norm terms, we obtain the pathwise estimate

$$\begin{aligned}\|u_{n+1}-\bar{u}\|^2 \le \|u_n-\bar{u}\|^2 + 3t_n^2\|u_n^*+\xi_n\|^2 - 2t_n\langle u_n^*, u_n-w_n\rangle - 2t_n\langle \xi_n, u_n-w_n\rangle \\ + 2t_n\|u_n^*+\xi_n\|\|w_n-\bar{u}\|. \qquad (2.4)\end{aligned}$$

We now take the conditional expectation $\mathbb{E}[\cdot \mid \mathscr{F}_n]$, recalling that u_n and u_n^* are $\mathscr{F}_n$-measurable and w_n is deterministic. By (T_6), the noise ξ_n has the following properties:

(1) Zero mean: $\mathbb{E}[\xi_n \mid \mathscr{F}_n] = 0$. Consequently, the linear noise interaction term vanishes:

$$\mathbb{E}[-2t_n\langle \xi_n, u_n-w_n\rangle \mid \mathscr{F}_n] = 0.$$

(2) Bounded variance: $\mathbb{E}[\|\xi_n\|^2 \mid \mathscr{F}_n] \le \sigma^2$. Using the inequality $\|a+b\|^2 \le 2\|a\|^2+2\|b\|^2$, we bound the squared norm term:

$$\mathbb{E}[\|u_n^*+\xi_n\|^2 \mid \mathscr{F}_n] \le 2\|u_n^*\|^2 + 2\sigma^2.$$

(3) First moment bound: By Jensen's inequality, $\mathbb{E}[\|\xi_n\| \mid \mathscr{F}_n] \le (\mathbb{E}[\|\xi_n\|^2 \mid \mathscr{F}_n])^{1/2} \le \sigma$. Thus,

$$\mathbb{E}[\|u_n^*+\xi_n\| \mid \mathscr{F}_n] \le \|u_n^*\| + \sigma.$$

Applying these properties, we have

$$\begin{aligned}\mathbb{E}[\|u_{n+1}-\bar{u}\|^2 \mid \mathscr{F}_n] \le \|u_n-\bar{u}\|^2 + 3t_n^2(2\|u_n^*\|^2+2\sigma^2) \\ - 2t_n\langle u_n^*, u_n-w_n\rangle + 2t_n(\|u_n^*\|+\sigma)\|w_n-\bar{u}\|.\end{aligned}$$

Next, we apply the structural conditions on the operator F_n. By (2.3),

$$-2t_n\langle u_n^*, u_n-w_n\rangle \le -2t_nc_n\|u_n-\bar{u}\|^2 + 4t_nc_n\|u_n-\bar{u}\|\|w_n-\bar{u}\|.$$

By the linear growth condition (T_2) and the triangle inequality $\|u_n\| \le \|u_n-\bar{u}\| + \|\bar{u}\|$, we have $\|u_n^*\| \le \kappa\|u_n-\bar{u}\| + C_0$, where $C_0 := \kappa\|\bar{u}\| + \ell$. This implies $\|u_n^*\|^2 \le 2\kappa^2\|u_n-\bar{u}\|^2 + 2C_0^2$. Substituting these estimates into the expectation inequality, we obtain

$$\begin{aligned}\mathbb{E}[\|u_{n+1}-\bar{u}\|^2 \mid \mathscr{F}_n] \le \|u_n-\bar{u}\|^2 + 3t_n^2\left(2(2\kappa^2\|u_n-\bar{u}\|^2+2C_0^2)+2\sigma^2\right) \\ + \left(-2t_nc_n\|u_n-\bar{u}\|^2 + 4t_nc_n\|u_n-\bar{u}\|\|w_n-\bar{u}\|\right) \\ + 2t_n(\kappa\|u_n-\bar{u}\| + C_0 + \sigma)\|w_n-\bar{u}\|.\end{aligned}$$

Finally, we group the terms by powers of the error norm $\|u_n - \bar{u}\|$. The coefficient of $\|u_n - \bar{u}\|^2$ collects terms from the identity, the strong monotonicity (negative), and the squared operator bound (positive, of order t_n^2):

$$\begin{aligned}\mathbb{E}[\|u_{n+1} - \bar{u}\|^2 \mid \mathscr{F}_n] \leq \left[1 - 2t_n c_n + 12t_n^2\kappa^2\right] \|u_n - \bar{u}\|^2 \\ + \left[4t_n c_n \|w_n - \bar{u}\| + 2t_n\kappa\|w_n - \bar{u}\|\right] \|u_n - \bar{u}\| \\ + \left[3t_n^2(2\sigma^2 + 4C_0^2) + 2t_n(C_0 + \sigma)\|w_n - \bar{u}\|\right]. \end{aligned} \tag{2.5}$$

This inequality is now in the form required to apply Lemma A.2.

Step 3. To apply Lemma A.2, we derive a deterministic recursive inequality by taking the total expectation of the estimate obtained in Step 2. Let $s_n := \mathbb{E}[\|u_n - \bar{u}\|^2]$ denote the mean squared error at iteration n.

First, we apply the law of total expectation $\mathbb{E}[\cdot] = \mathbb{E}[\mathbb{E}[\cdot \mid \mathscr{F}_n]]$ to inequality (2.5). The term involving the first moment of the error, $\mathbb{E}[\|u_n - \bar{u}\|]$, requires attention. Since the square root function is concave, Jensen's inequality implies

$$\mathbb{E}[\|u_n - \bar{u}\|] \leq \sqrt{\mathbb{E}[\|u_n - \bar{u}\|^2]} = \sqrt{s_n}.$$

Applying this to the result from Step 2 yields

$$\begin{aligned} s_{n+1} \leq \left(1 - 2t_n c_n + 12\kappa^2 t_n^2\right) s_n + \left(4t_n c_n\|w_n - \bar{u}\| + 2t_n\kappa\|w_n - \bar{u}\|\right)\sqrt{s_n} \\ + \left(3t_n^2(2\sigma^2 + 4C_0^2) + 2t_n(C_0 + \sigma)\|w_n - \bar{u}\|\right). \end{aligned} \tag{2.6}$$

To bring this inequality into a standard form suitable for convergence analysis, we simplify the coefficient of s_n. We analyze the contraction term $1 - 2t_n c_n + 12\kappa^2 t_n^2$. We require the first-order negative term to dominate the second-order positive term. Specifically, we seek an integer N_0 such that, for all $n \geq N_0$, $1 - 2t_n c_n + 12\kappa^2 t_n^2 \leq 1 - t_n c_n$. This inequality holds if and only if $12\kappa^2 t_n^2 \leq t_n c_n$, or equivalently $12\kappa^2 t_n \leq c_n$. Given the assumption (T_5) that the step size decays faster than the monotonicity modulus, i.e., $\lim_{n\to\infty}(t_n/c_n) = 0$, this condition is satisfied for sufficiently large n. Moreover, since $\{c_n\}$ is bounded, $t_n c_n = (t_n/c_n)\, c_n^2 \to 0$, so that, after enlarging N_0 if necessary, $t_n c_n \in (0,1)$ for all $n \geq N_0$.

Consequently, for $n \geq N_0$, we define the deterministic sequences corresponding to the notation in Lemma A.2:

- $\nu_n = t_n c_n$,
- $\delta_n = 2t_n(2c_n + \kappa)\|w_n - \bar{u}\|$,
- $\gamma_n = 6t_n^2(\sigma^2 + 2C_0^2) + 2t_n(C_0 + \sigma)\|w_n - \bar{u}\|$.

Substituting these definitions results in the following recursive relation for the mean squared error:

$$s_{n+1} \leq (1 - \nu_n)s_n + \delta_n\sqrt{s_n} + \gamma_n, \quad n \geq N_0. \tag{2.7}$$

We are now in a position to apply Lemma A.2. We verify the requisite conditions:

(a) We examine the ratio δ_n/ν_n:

$$\frac{\delta_n}{t_n c_n} = \frac{2t_n(2c_n + \kappa)\|w_n - \bar{u}\|}{t_n c_n} = 2\left(2 + \frac{\kappa}{c_n}\right)\|w_n - \bar{u}\|.$$

Using assumption (T_5), specifically that the discretization error decays faster than the conditioning parameter ($\lim_{n\to\infty} \|w_n - \bar{u}\|/c_n = 0$), together with $\|w_n - \bar{u}\| \to 0$, this ratio tends to 0. Thus, for any $\theta \in (0,1)$, we have $\delta_n \leq \theta \nu_n$ for large n.

(b) We examine the ratio γ_n/ν_n. The term γ_n consists of two components.
- The first component is of order $\mathcal{O}(t_n^2)$. Its ratio to ν_n is proportional to $t_n^2/(t_n c_n) = t_n/c_n$, which tends to 0 by (T_5).
- The second component is of order $\mathcal{O}(t_n\|w_n - \bar{u}\|)$. Its ratio to ν_n is proportional to $\|w_n - \bar{u}\|/c_n$, which also tends to 0 by (T_5).

Consequently, $\lim_{n\to\infty} \gamma_n/\nu_n = 0$, which implies $\sup_n \gamma_n/\nu_n < +\infty$.

Since all conditions of Lemma A.2 are satisfied, the sequence $\{s_n\}_{n\geq N_0}$ is bounded. As the finitely many terms $s_1, \ldots, s_{N_0}$ are finite (because $u_n \in L_H^2(\Omega, \mathscr{A}, \mathbb{P})$ for every n), we conclude that

$$M_x := \sup_{n\geq 1} s_n < +\infty. \tag{2.8}$$

Step 4. We now combine the recursive inequality (2.6) with the bound (2.8) to prove that the mean squared error converges to zero.

Recall inequality (2.6), written with the first moment $\mathbb{E}[\|u_n - \bar{u}\|]$ before the application of Jensen's inequality:

$$\begin{aligned} s_{n+1} \leq \left(1 - 2t_n c_n + 12\kappa^2 t_n^2\right) s_n + \left(4t_n c_n \|w_n - \bar{u}\| + 2t_n \kappa \|w_n - \bar{u}\|\right) \mathbb{E}[\|u_n - \bar{u}\|] \\ + 3t_n^2(2\sigma^2 + 4C_0^2) + 2t_n(C_0 + \sigma)\|w_n - \bar{u}\|. \end{aligned} \tag{2.9}$$

To apply Lemma A.1, we isolate the first-order contraction term $-2t_n c_n$. The term $12\kappa^2 t_n^2 s_n$ is of higher order (since $t_n \to 0$). Using the bound $s_n \leq M_x$ from Step 3, we move this term into the additive perturbation sequence to simplify the multiplicative coefficient. We also use the bound $\mathbb{E}[\|u_n - \bar{u}\|] \leq \sqrt{s_n} \leq \sqrt{M_x}$. Substituting these bounds, the inequality simplifies to the standard form

$$s_{n+1} \leq (1 - 2t_n c_n) s_n + a_n, \tag{2.10}$$

where the total perturbation term a_n collects all non-contractive terms:

$$\begin{aligned} a_n = \underbrace{12\kappa^2 t_n^2 M_x}_{\text{from coefficient}} + \underbrace{3t_n^2(2\sigma^2 + 4C_0^2)}_{\text{noise variance}} \\ + \|w_n - \bar{u}\| \left(\underbrace{4t_n c_n \sqrt{M_x} + 2t_n \kappa \sqrt{M_x}}_{\text{cross terms}} + \underbrace{2t_n(C_0 + \sigma)}_{\text{linear growth/noise}} \right). \end{aligned} \tag{2.11}$$

Grouping terms by powers of t_n for clarity, we write

$$a_n = t_n^2 \underbrace{\left(12\kappa^2 M_x + 6\sigma^2 + 12C_0^2\right)}_{:=K_2} + t_n \|w_n - \bar{u}\| \underbrace{\left(4c_n\sqrt{M_x} + 2\kappa\sqrt{M_x} + 2C_0 + 2\sigma\right)}_{:=K_{3,n}}. \tag{2.12}$$

We now invoke Lemma A.1: if a sequence of non-negative real numbers $\{s_n\}$ satisfies $s_{n+1} \leq (1 - \alpha_n)s_n + a_n$ with $\alpha_n \in (0,1)$ for all sufficiently large n, $\sum \alpha_n = +\infty$, and $\lim_{n\to\infty}(a_n/\alpha_n) = 0$, then $\lim_{n\to\infty} s_n = 0$.

We verify these conditions with $\alpha_n = 2t_n c_n$. As shown in Step 3, $t_n c_n \to 0$, so $\alpha_n \in (0,1)$ for all sufficiently large n.

1. By assumption (T_5), $\sum_{n=1}^{\infty} t_n c_n = +\infty$. Thus, $\sum \alpha_n = +\infty$.

2. We analyze the limit of the ratio a_n/α_n:

$$\begin{aligned}\frac{a_n}{2t_nc_n} &= \frac{t_n^2K_2}{2t_nc_n} + \frac{t_n\|w_n-\bar{u}\|K_{3,n}}{2t_nc_n} \\ &= \frac{K_2}{2}\underbrace{\left(\frac{t_n}{c_n}\right)}_{\to 0} + \frac{\|w_n-\bar{u}\|}{2c_n}\left(4c_n\sqrt{M_x}+2\kappa\sqrt{M_x}+2C_0+2\sigma\right) \\ &= \frac{K_2}{2}\left(\frac{t_n}{c_n}\right) + 2\sqrt{M_x}\|w_n-\bar{u}\| + \underbrace{\left(\frac{\|w_n-\bar{u}\|}{2c_n}\right)}_{\to 0}(2\kappa\sqrt{M_x}+2C_0+2\sigma). \end{aligned} \tag{2.13}$$

- The first term vanishes because of assumption (T_5): $\lim(t_n/c_n)=0$.
- The second term vanishes because $w_n \to \bar{u}$.
- The third term vanishes because of the consistency assumption $\lim(\|w_n-\bar{u}\|/c_n)=0$.

Since both conditions are satisfied, Lemma A.1 implies

$$\lim_{n\to\infty} s_n = \lim_{n\to\infty} \mathbb{E}[\|u_n-\bar{u}\|^2] = 0. \tag{2.14}$$

This confirms that the sequence $\{u_n\}$ converges strongly to $\bar{u}$ in the mean-square sense. The proof is complete. □

Remark 2.1 (On the Hypotheses of Theorem 2.1)**.** (1) Theorem 2.1 identifies the limit of the iterates as the limit $\bar{u}$ of the discrete solutions. Under standard consistency conditions on the discretization, $\bar{u}$ is a solution of the original problem (1.1). Suppose, for instance, that

- (C_1) every $v \in C$ is the strong limit of some sequence $v_n \in C_n$, and whenever $\{n_k\}_{k\in\mathbb{N}}$ is a strictly increasing sequence of indices and $x_{n_k} \in C_{n_k}$ converge weakly to x, then $x \in C$;
- (C_2) whenever $\{n_k\}_{k\in\mathbb{N}}$ is a strictly increasing sequence of indices, $x_{n_k} \in C_{n_k}$ converge strongly to x, and $x^*_{n_k} \in F_{n_k}(x_{n_k})$ converge weakly to x^*, then $x^* \in F(x)$.

Since $w_n \in C_n$ and $w_n \to \bar{u}$ strongly, we have $\bar{u} \in C$ by (C_1). Moreover, $\{w_n\}$ is bounded, being convergent, and hence, by (T_2), so is $\{w_n^*\}$. Therefore, there exists a subsequence $\{w^*_{n_k}\}$ converging weakly to some $\eta \in H$, and $\eta \in F(\bar{u})$ by (C_2). For $v \in C$, choose $v_n \in C_n$ with $v_n \to v$ by (C_1). Since $w^*_{n_k} \rightharpoonup \eta$ and $v_{n_k} - w_{n_k} \to v-\bar{u}$ strongly, passing to the limit as $k\to\infty$ in $\langle w^*_{n_k}, v_{n_k}-w_{n_k}\rangle \geq 0$ yields $\langle \eta, v-\bar{u}\rangle \geq 0$. Hence, $\bar{u}$ solves (1.1).

(2) Square-summability of the step sizes, $\sum t_n^2 < +\infty$, is not among the hypotheses of Theorem 2.1, since it is not required for mean-square convergence. It will be needed for the almost sure convergence results in Section 3.

Remark 2.2 (Measurability and Well-definedness)**.** The measurability of the iterates $\{u_n\}$ with respect to the filtration $\{\mathscr{F}_n\}$, as established in the proof of Theorem 2.1, serves as a foundational property for the entire study. Consequently, for the subsequent convergence results presented in this work, including those establishing almost sure convergence via the Robbins–Siegmund lemma, we will omit the explicit verification of measurability. It is understood that the stochastic process $\{u_n\}$ remains well defined in the Bochner space $L^2_H(\Omega,\mathscr{A},\mathbb{P})$ throughout the iterative scheme.

Remark 2.3 (Extensions to General Noise Conditions)**.** The proof presented above utilizes a uniform bounded variance assumption (Condition (T_6)) to streamline the derivation of the recursive error bounds. However, Theorem 2.1 can be extended to broader stochastic settings with minor modifications to the algebra in Steps 2 and 3:

(1) The uniform bound $\mathbb{E}[\|\xi_n\|^2 \mid \mathscr{F}_n] \le \sigma^2$ can be relaxed to allow the noise intensity to grow with the iterate: $\mathbb{E}[\|\xi_n\|^2 \mid \mathscr{F}_n] \le \sigma_0^2 + \sigma_1^2\|u_n\|^2$. Since $\|u_n\|^2 \le 2\|u_n - \bar{u}\|^2 + 2\|\bar{u}\|^2$, this produces in Step 2 an additional term of order $t_n^2\sigma_1^2\|u_n - \bar{u}\|^2$, which is absorbed into the contraction coefficient in the same way as the term $12\kappa^2 t_n^2$ in Step 3, together with an additional cross term of order $t_n\sigma_1\|u_n - \bar{u}\|\|w_n - \bar{u}\|$, which is handled like the corresponding term involving κ. Since $t_n/c_n \to 0$ by (T_5), the dissipativity of the operator dominates the noise growth for all sufficiently large n (i.e., $t_n c_n > C t_n^2\sigma_1^2$ for some constant C independent of n), and the conclusion of Theorem 2.1 remains valid.

(2) The uniform bound can also be replaced by a deterministic, possibly unbounded, sequence: if $\mathbb{E}[\|\xi_n\|^2 \mid \mathscr{F}_n] \le \sigma_n^2$ a.s., then the proof carries over verbatim with σ replaced by σ_n, provided that

$$\lim_{n\to\infty} \frac{t_n\sigma_n^2}{c_n} = 0 \quad \text{and} \quad \lim_{n\to\infty} \frac{\sigma_n\|w_n - \bar{u}\|}{c_n} = 0.$$

These conditions guarantee that $\gamma_n/\nu_n \to 0$ in Step 3 and $a_n/\alpha_n \to 0$ in Step 4.

In Theorem 2.1, we established the strong convergence of the proposed scheme, confirming that the sequence of iterates $\{u_n\}$ indeed approaches the limit $\bar{u}$. However, this result did not provide any information regarding the speed of this convergence. The previous result relied essentially on qualitative conditions (the divergence of $\sum t_n c_n$ and the vanishing of the ratios t_n/c_n and $\|w_n - \bar{u}\|/c_n$) to ensure asymptotic stability.

In the following, we advance the analysis from qualitative validity to quantitative efficiency. We impose specific polynomial decay rates on the step size and on the mesh approximation error, and, for a sharper estimate, on the increments of the discrete solutions, to derive explicit asymptotic rates for the mean squared error. This analysis is crucial for understanding the interplay between the two primary sources of error: the variance of the stochastic noise and the bias introduced by the spatial discretization. We first isolate a one-step estimate of the iterates relative to the discrete solutions w_n, which will be used both for the rate estimates and for the analysis of regularized schemes.

Lemma 2.1 (One-Step Estimate Relative to the Discrete Solutions)**.** *Assume that conditions* (T_1)*–*(T_4) *and* (T_6) *of Theorem 2.1 hold, and set* $K := \sup_{n\in\mathbb{N}} \left(2(\kappa\|w_n\| + \ell)^2 + \sigma^2\right)$*, which is finite since the sequence* $\{w_n\}$ *converges by* (T_4)*. Then, for every* $n \in \mathbb{N}$*,*

$$\mathbb{E}[\|u_{n+1} - w_n\|^2 \mid \mathscr{F}_n] \le (1 - 2t_n c_n + 2\kappa^2 t_n^2)\|u_n - w_n\|^2 + K t_n^2 \quad a.s. \tag{2.15}$$

Proof. Recall the update $u_{n+1} = Q_n(u_n - t_n\eta_n)$, where $\eta_n = u_n^* + \xi_n$. Since $w_n \in C_n \subset H_n$, it is a fixed point of Q_n, and the non-expansivity of Q_n yields

$$\begin{aligned}
\|u_{n+1} - w_n\|^2 &= \|Q_n(u_n - t_n\eta_n) - Q_n(w_n)\|^2 \\
&\le \|u_n - w_n - t_n(u_n^* + \xi_n)\|^2 \\
&= \|u_n - w_n\|^2 - 2t_n\langle u_n^* + \xi_n, u_n - w_n\rangle + t_n^2\|u_n^* + \xi_n\|^2.
\end{aligned}$$

We take the conditional expectation $\mathbb{E}[\cdot \mid \mathscr{F}_n]$ and use (T_6). Since u_n and u_n^* are $\mathscr{F}_n$-measurable and the noise has zero conditional mean, the terms that are linear in ξ_n vanish, and in particular $\mathbb{E}[\|u_n^* + \xi_n\|^2 \mid \mathscr{F}_n] = \|u_n^*\|^2 + \mathbb{E}[\|\xi_n\|^2 \mid \mathscr{F}_n] \le \|u_n^*\|^2 + \sigma^2$. Hence

$$\mathbb{E}[\|u_{n+1} - w_n\|^2 \mid \mathscr{F}_n] \le \|u_n - w_n\|^2 - 2t_n\langle u_n^*, u_n - w_n\rangle + t_n^2(\|u_n^*\|^2 + \sigma^2).$$

Since w_n solves (1.2), we have $\langle w_n^*, u - w_n\rangle \ge 0$ for all $u \in C_n$, and in particular for $u = u_n$, which belongs to C_n (see the proof of Theorem 2.1). Together with the strong monotonicity condition (T_1), this gives

$$-\langle u_n^*, u_n - w_n\rangle = -\langle u_n^* - w_n^*, u_n - w_n\rangle - \langle w_n^*, u_n - w_n\rangle \le -c_n\|u_n - w_n\|^2.$$

For the squared norm term, we use the linear growth condition (T_2) and the triangle inequality $\|u_n\| \le \|u_n - w_n\| + \|w_n\|$ to obtain

$$\|u_n^*\|^2 \le (\kappa\|u_n - w_n\| + \kappa\|w_n\| + \ell)^2 \le 2\kappa^2\|u_n - w_n\|^2 + 2(\kappa\|w_n\| + \ell)^2.$$

Substituting these bounds into the previous inequality and using the definition of K yields (2.15). □

Theorem 2.2 (Asymptotic Convergence Rates)**.** *Assume that conditions (T_1)–(T_4) and (T_6) of Theorem 2.1 hold and that the monotonicity moduli are bounded from above and away from zero, i.e., there exist constants $0 < c \le \bar{c}$ such that $c \le c_n \le \bar{c}$ for all $n \in \mathbb{N}$. Suppose that the step size and the mesh approximation error satisfy the following decay rates:*

- (R_1) *The step size sequence is given by $t_n = t_0 n^{-p}$ with $0 < p \le 1$ and $t_0 > 0$.*
- (R_2) *The discrete solutions satisfy the approximation bound $\|w_n - \bar{u}\| \le \widetilde{C} n^{-r}$ with $r > 0$.*

Then the following assertions hold.

(a) Let $q_a := \min(p, r)$, and, in the case $p = 1$, assume that $2ct_0 > q_a$. Then

$$\limsup_{n\to\infty} n^{q_a}\,\mathbb{E}\left[\|u_n - \bar{u}\|^2\right] < +\infty, \quad \textit{i.e.,} \quad \mathbb{E}\left[\|u_n - \bar{u}\|^2\right] = \mathscr{O}\left(n^{-\min(p,r)}\right).$$

(b) Assume, in addition, that the discrete solutions vary slowly in the sense that

- (R_3) *$\|w_{n+1} - w_n\| \le \widehat{C} n^{-(r+1)}$ for all $n \in \mathbb{N}$.*

Let $q_b := \min(p, 2r)$, and, in the case $p = 1$, assume that $2ct_0 > q_b$. Then

$$\limsup_{n\to\infty} n^{q_b}\,\mathbb{E}\left[\|u_n - \bar{u}\|^2\right] < +\infty, \quad \textit{i.e.,} \quad \mathbb{E}\left[\|u_n - \bar{u}\|^2\right] = \mathscr{O}\left(n^{-\min(p,2r)}\right).$$

Proof. Throughout the proof, $s_n := \mathbb{E}[\|u_n - \bar{u}\|^2]$. We first note that, under (R_1) and (R_2), all conditions of (T_5) are satisfied: $\{c_n\}$ is bounded, $t_n/c_n \le t_n/c \to 0$, $\sum t_n c_n \ge c\,t_0 \sum n^{-p} = +\infty$ since $p \le 1$, and $\|w_n - \bar{u}\|/c_n \le \widetilde{C} n^{-r}/c \to 0$. Hence all hypotheses of Theorem 2.1 are satisfied.

Proof of (a).

Step 1: Since the hypotheses of Theorem 2.1 hold, Step 3 of its proof yields the uniform bound (2.8), i.e., $M_x = \sup_{n\ge1} s_n < +\infty$.

Step 2: We start from inequality (2.6) derived in the proof of Theorem 2.1. Since $s_n \ge 0$ and $c_n \ge c$, we have $(1 - 2t_n c_n)s_n \le (1 - 2t_n c)s_n$. Moreover, $12\kappa^2 t_n^2 s_n \le 12\kappa^2 M_x t_n^2$, $\sqrt{s_n} \le \sqrt{M_x}$, and $4c_n + 2\kappa \le 4\bar{c} + 2\kappa$. Therefore,

$$s_{n+1} \le (1 - 2c\,t_n)s_n + K_a t_n^2 + L_a t_n\|w_n - \bar{u}\|, \tag{2.16}$$

where

$$K_a := 12\kappa^2 M_x + 6\sigma^2 + 12C_0^2, \qquad L_a := (4\bar{c} + 2\kappa)\sqrt{M_x} + 2(C_0 + \sigma).$$

We emphasize that the discretization error enters (2.16) linearly, through the term $L_a t_n \|w_n - \bar{u}\|$, which originates from the cross term $2\langle u_{n+1} - u_n, w_n - \bar{u}\rangle$ in Step 2 of the proof of Theorem 2.1.

Step 3: Substituting (R_1) and (R_2) into (2.16), the driving terms satisfy

$$K_a t_n^2 = K_a t_0^2 n^{-2p}, \qquad L_a t_n \|w_n - \bar{u}\| \le L_a t_0 \widetilde{C} n^{-(p+r)}.$$

Hence, with $\lambda_a := p + \min(p, r)$ and a suitable constant $K' > 0$,

$$s_{n+1} \le \left(1 - \frac{2ct_0}{n^p}\right) s_n + \frac{K'}{n^{\lambda_a}},$$

where $1 - 2ct_0 n^{-p} \in [0, 1)$ for all sufficiently large n. By Lemma A.3, the rate is determined by the difference between the source decay power λ_a and the contraction power p, namely $q_a = \lambda_a - p = \min(p, r)$, provided that, in the case $p = 1$, the contraction constant exceeds this difference, i.e., $2ct_0 > q_a$. This proves (a).

Proof of (b).

Step 4: If $p < 1$, set $\varepsilon := \frac{1}{2}$; if $p = 1$, choose $\varepsilon \in (0, 1)$ so small that $2(1-\varepsilon)ct_0 > q_b$, which is possible since $2ct_0 > q_b$. Since $t_n \to 0$, there exists $N_0 \ge 2$ such that $2\kappa^2 t_n \le \varepsilon c$ and $2ct_n \le 1$ for all $n \ge N_0$. For such n, we have $-2t_n c_n + 2\kappa^2 t_n^2 \le -2ct_n + \varepsilon c t_n$, and Lemma 2.1 yields

$$\mathbb{E}[\|u_{n+1} - w_n\|^2 \mid \mathscr{F}_n] \le (1 - (2-\varepsilon)c t_n)\, \|u_n - w_n\|^2 + K t_n^2. \tag{2.17}$$

We define $e_n^2 := \mathbb{E}[\|u_n - w_{n-1}\|^2]$ for $n \ge 2$. Using the relaxed triangle inequality $\|a+b\|^2 \le (1+\delta)\|a\|^2 + (1+\frac{1}{\delta})\|b\|^2$ with $\delta = \varepsilon c t_n$, we obtain

$$\|u_n - w_n\|^2 \le (1 + \varepsilon c t_n)\|u_n - w_{n-1}\|^2 + \left(1 + \frac{1}{\varepsilon c t_n}\right) \|w_{n-1} - w_n\|^2.$$

We insert this bound into (2.17) and take full expectations. With $x := c t_n \in (0, \frac{1}{2}]$, we have

$$(1 - (2-\varepsilon)x)\,(1 + \varepsilon x) \le 1 - (2 - 2\varepsilon)x \quad \text{and} \quad (1 - (2-\varepsilon)x)\left(1 + \frac{1}{\varepsilon x}\right) \le 1 + \frac{1}{\varepsilon x} \le \frac{2}{\varepsilon x}.$$

Consequently, for all $n \ge N_0$,

$$e_{n+1}^2 \le (1 - 2(1-\varepsilon)c t_n)\, e_n^2 + \frac{2}{\varepsilon c t_n} \|w_{n-1} - w_n\|^2 + K t_n^2. \tag{2.18}$$

In contrast to (2.16), the discretization now enters only through the increments $\|w_{n-1} - w_n\|$ of the moving target.

Step 5: By (R_3), for $n \ge 2$ we have $\|w_{n-1} - w_n\| \le \widehat{C}(n-1)^{-(r+1)} \le 2^{r+1}\widehat{C} n^{-(r+1)}$. Substituting this and (R_1) into (2.18), the driving terms satisfy

$$\frac{2}{\varepsilon c t_n} \|w_{n-1} - w_n\|^2 \le \frac{2^{2r+3}\widehat{C}^2}{\varepsilon c t_0} n^{p-2r-2}, \qquad K t_n^2 = K t_0^2 n^{-2p}.$$

Since $p \le 1$, we have $p - 2r - 2 \le -p - 2r \le -(p + q_b)$ and $-2p \le -(p + q_b)$. Hence, with a suitable constant $K'' > 0$,

$$e_{n+1}^2 \le \left(1 - \frac{2(1-\varepsilon)c t_0}{n^p}\right) e_n^2 + \frac{K''}{n^{p+q_b}}, \qquad n \ge N_0.$$

If $p = 1$, the choice of ε guarantees $2(1-\varepsilon)ct_0 > q_b$. Therefore, Lemma A.3 yields $e_n^2 = \mathcal{O}(n^{-q_b})$.

Step 6. Finally, by the elementary inequality $\|a+b\|^2 \le 2\|a\|^2 + 2\|b\|^2$ and (R_2),

$$s_n \le 2e_n^2 + 2\|w_{n-1} - \bar{u}\|^2 \le 2e_n^2 + 2^{2r+1}\widetilde{C}^2 n^{-2r}, \qquad n \ge 2.$$

Since $e_n^2 = \mathcal{O}(n^{-q_b})$ and $2r \ge q_b$, we conclude that $s_n = \mathcal{O}(n^{-q_b})$. This proves (b) and completes the proof. □

Remark 2.4 (On the Hypotheses of Theorem 2.2)**.** (1) Condition (R_3) implies condition (R_2). Indeed, since $w_n \to \bar{u}$,

$$\|w_n - \bar{u}\| \le \sum_{k=n}^{\infty} \|w_{k+1} - w_k\| \le \widehat{C}\left(n^{-(r+1)} + \int_n^{\infty} x^{-(r+1)}\,dx\right) \le \widehat{C}\left(1 + \frac{1}{r}\right) n^{-r}.$$

Thus, part (b) requires only a bound on the increments of the discrete solutions. Moreover, the proof of part (b) does not use the uniform bound (2.8).

(2) Since (T_5) does not require $\sum t_n^2 < +\infty$, step sizes $t_n = t_0 n^{-p}$ with $0 < p \le \frac{1}{2}$ are admissible in Theorem 2.2, in contrast to the almost sure results of Section 3.

(3) More generally, if $\|w_{n+1} - w_n\| \le \widehat{C} n^{-\rho}$ with $\rho > p$, the argument of part (b) yields the rate $\mathcal{O}\left(n^{-\min(p, 2r, 2\rho - 2p)}\right)$, under the condition $2ct_0 > \min(1, 2r, 2\rho - 2)$ when $p = 1$.

Remark 2.5 (Competition between Error Sources)**.** Theorem 2.2 highlights a critical competition between two distinct error components inherent to the algorithm:

(1) The error arising from the noisy operator observations, which decays at the rate $\mathcal{O}(n^{-p})$.

(2) The structural error arising from projecting the infinite-dimensional problem onto finite-dimensional subspaces. Its effective contribution depends on the available information about the discrete solutions. Under (R_2) alone, the discretization error enters the recursion linearly, leading to the effective rate $\mathcal{O}(n^{-r})$. If, in addition, the discrete solutions vary slowly in the sense of (R_3), the iterates track the moving target w_n, and the discretization error enters only through $\|w_n - \bar{u}\|^2$, leading to the effective rate $\mathcal{O}(n^{-2r})$.

Unlike the previous strong convergence result, which simply required both errors to vanish asymptotically, this theorem demonstrates that the overall convergence rate is bottlenecked by the slower of the two components. This implies that refining the mesh faster than required (i.e., choosing $r > p$ in part (a), or $2r > p$ in part (b)) yields no improvement in the rate, as the convergence is then dominated solely by the stochastic noise. Conversely, an overly coarse mesh prevents the algorithm from achieving the optimal stochastic rate $\mathcal{O}(n^{-p})$. Condition (R_3) is natural when the discretization is refined gradually along the iteration. If the mesh is instead refined abruptly at isolated iterations, the increments $\|w_{n+1} - w_n\|$ may be of order n^{-r} at those iterations, and the general estimate of part (a) applies.

We now extend the analysis to ill-posed problems where the strong monotonicity condition does not hold uniformly. We employ a regularization strategy (e.g., Tikhonov regularization) characterized by a vanishing modulus $c_n \to 0$. In this setting, the sequence $\{w_n\}$ (the unique solutions to the regularized problems on C_n) acts as a "moving target" that converges to $\bar{u}$, which in the Tikhonov setting is the minimum-norm solution of (1.1). In contrast to Theorem 2.1, the condition $\|w_n - \bar{u}\|/c_n \to 0$ is replaced by a tracking condition on the increments $\|w_{n-1} - w_n\|$ of the moving target, which remains meaningful when w_n converges slowly to $\bar{u}$, as is typical for ill-posed problems.

Theorem 2.3 (Convergence for Regularized Schemes)**.** *Let the assumptions of Theorem 2.1 hold, with the modifications that the monotonicity modulus c_n is now a sequence of positive numbers decaying to zero ($\lim_{n\to\infty} c_n = 0$), and that the conditions on the step sizes and discretization error in (T_5) are replaced by the following consistency conditions.*

Let $\{w_n\}$ denote the sequence of solutions to the regularized variational inequalities (1.2). Suppose the step size t_n and the regularization parameter c_n satisfy

$$\sum_{n=1}^{\infty} t_n c_n = +\infty, \quad \lim_{n\to\infty} \frac{t_n}{c_n} = 0, \quad \lim_{n\to\infty} \frac{\|w_{n-1} - w_n\|}{t_n c_n} = 0. \tag{2.19}$$

Then the iteration converges to $\bar{u}$ in mean square: $\lim_{n\to\infty} \mathbb{E}[\|u_n - \bar{u}\|^2] = 0$.

Proof. We proceed in several steps to handle the moving target sequence $\{w_n\}$. Note first that $t_n = (t_n/c_n)\, c_n \to 0$, and hence $t_n c_n \to 0$.

Step 1. By Lemma 2.1, for every $n \in \mathbb{N}$,

$$\mathbb{E}[\|u_{n+1} - w_n\|^2 \mid \mathscr{F}_n] \le (1 - 2t_n c_n + 2\kappa^2 t_n^2)\|u_n - w_n\|^2 + K t_n^2.$$

By assumption, $\lim_{n\to\infty} t_n/c_n = 0$. Therefore, for sufficiently large n, we have $2\kappa^2 t_n^2 \le t_n c_n$, and the coefficient simplifies $1 - 2t_n c_n + 2\kappa^2 t_n^2 \le 1 - t_n c_n$. This yields the fundamental inequality relative to the target w_n, valid for all sufficiently large n:

$$\mathbb{E}[\|u_{n+1} - w_n\|^2 \mid \mathscr{F}_n] \le (1 - t_n c_n)\|u_n - w_n\|^2 + K t_n^2. \tag{2.20}$$

Step 2. We define the mean squared error relative to the previous target as $e_n^2 := \mathbb{E}[\|u_n - w_{n-1}\|^2]$. Consequently, after taking full expectations, the left-hand side of (2.20) becomes $e_{n+1}^2 = \mathbb{E}[\|u_{n+1} - w_n\|^2]$.

To relate the term $\|u_n - w_n\|^2$ on the right-hand side to the error metric e_n^2, we use the relaxed triangle inequality $\|a+b\|^2 \le (1+\delta)\|a\|^2 + (1+\frac{1}{\delta})\|b\|^2$. To retain a first-order contraction, we select $\delta = \frac{1}{2} t_n c_n$:

$$\begin{aligned} \|u_n - w_n\|^2 &= \|(u_n - w_{n-1}) + (w_{n-1} - w_n)\|^2 \\ &\le \left(1 + \frac{1}{2} t_n c_n\right) \|u_n - w_{n-1}\|^2 + \left(1 + \frac{2}{t_n c_n}\right) \|w_{n-1} - w_n\|^2. \end{aligned} \tag{2.21}$$

Substituting this bound into (2.20) and taking full expectations yields

$$e_{n+1}^2 \le (1 - t_n c_n) \left[\left(1 + \frac{1}{2} t_n c_n\right) e_n^2 + \left(1 + \frac{2}{t_n c_n}\right) \|w_{n-1} - w_n\|^2 \right] + K t_n^2. \tag{2.22}$$

We simplify the multiplicative coefficient. Observing that $(1-x)(1+x/2) = 1 - x/2 - x^2/2 \le 1 - x/2$, we have

$$(1 - t_n c_n)\left(1 + \frac{1}{2} t_n c_n\right) \le 1 - \frac{1}{2} t_n c_n.$$

This allows us to write the recursive inequality as

$$e_{n+1}^2 \le \left(1 - \frac{1}{2} t_n c_n\right) e_n^2 + \beta_n, \tag{2.23}$$

where the total perturbation term β_n collects the tracking error and the noise:

$$\beta_n = (1 - t_n c_n)\left(1 + \frac{2}{t_n c_n}\right)\|w_{n-1} - w_n\|^2 + K t_n^2 \le \frac{3}{t_n c_n}\|w_{n-1} - w_n\|^2 + K t_n^2,$$

where we used $t_n c_n \le 1$ for all sufficiently large n.

Step 3. The recurrence derived in Step 2 is of the form $e_{n+1}^2 \le (1-\alpha_n)e_n^2 + \beta_n$ with $\alpha_n = \frac{1}{2}t_n c_n \in (0,1)$ for all sufficiently large n. We now apply Lemma A.1 and verify its conditions:

(1) Since $\sum t_n c_n = +\infty$, it follows that $\sum \alpha_n = +\infty$.
(2) We must show that $\lim_{n\to\infty} \beta_n/\alpha_n = 0$. We have

$$\begin{aligned}\frac{\beta_n}{\alpha_n} &\le \frac{\frac{3}{t_n c_n}\|w_{n-1} - w_n\|^2 + K t_n^2}{\frac{1}{2}t_n c_n} \\ &= 6\left(\frac{\|w_{n-1} - w_n\|}{t_n c_n}\right)^2 + 2K\left(\frac{t_n}{c_n}\right). \end{aligned} \tag{2.24}$$

Both terms on the right-hand side converge to zero by the consistency conditions (2.19):

- the first term vanishes by the target tracking condition $\lim_{n\to\infty}\|w_{n-1} - w_n\|/(t_n c_n) = 0$;
- the second term vanishes by the noise control condition $\lim_{n\to\infty} t_n/c_n = 0$.

Thus, Lemma A.1 implies $\lim_{n\to\infty} e_n^2 = \lim_{n\to\infty}\mathbb{E}[\|u_n - w_{n-1}\|^2] = 0$.

Step 4. Finally, to establish convergence to $\bar{u}$, we apply the elementary inequality $\|a+b\|^2 \le 2\|a\|^2 + 2\|b\|^2$:

$$\mathbb{E}[\|u_n - \bar{u}\|^2] \le 2\mathbb{E}[\|u_n - w_{n-1}\|^2] + 2\|w_{n-1} - \bar{u}\|^2. \tag{2.25}$$

The first term on the right-hand side converges to 0 by Step 3, and the second term converges to 0 by (T_4). Therefore, $\lim_{n\to\infty}\mathbb{E}[\|u_n - \bar{u}\|^2] = 0$. The proof is complete. □

Remark 2.6 (Parameter Selection Strategy)**.** The conflicting requirements of noise suppression ($t_n/c_n \to 0$) and target tracking can be satisfied by choosing polynomial decay rates. Let $t_n = n^{-a}$ and $c_n = n^{-b}$ with $a, b > 0$. The conditions of the theorem impose the following constraints on the exponents:

(1) $\sum t_n c_n = +\infty$ requires $a + b \le 1$.
(2) $\lim t_n/c_n = 0$ requires $a > b$.
(3) For Tikhonov-type regularization, the regularized solutions $w(c)$ satisfy the standard estimate $\|w(c) - w(c')\| \le \|\bar{u}\|\,|c - c'|/c$ for $0 < c \le c'$. Hence, ignoring the contribution of the changing discretization,

$$\|w_{n-1} - w_n\| \lesssim \frac{|c_{n-1} - c_n|}{c_n} = \mathcal{O}(n^{-1}),$$

in agreement with the velocity estimate of Lemma 3.1 in Section 3. The tracking condition then requires

$$\frac{n^{-1}}{n^{-a}n^{-b}} = n^{a+b-1} \to 0, \quad \text{i.e.,} \quad a + b < 1.$$

Thus, a valid selection requires $b < a$ and $a + b < 1$. For example, the choice $t_n = n^{-0.6}$ and $c_n = n^{-0.2}$ ensures mean-square convergence.

Remark 2.7 (Absence of Convergence Rate)**.** Unlike the strongly monotone case, where we derived rates of the form $\mathcal{O}(n^{-q})$ in Theorem 2.2, this theorem only establishes asymptotic convergence ($\lim_{n\to\infty}\mathbb{E}[\|u_n-\bar{u}\|^2]=0$). To derive a specific rate of convergence for the ill-posed case, one would require additional source conditions (smoothness assumptions on the solution $\bar{u}$) to quantify the behavior of the bias term $\|w_n-\bar{u}\|$, which is beyond the scope of this general consistency result.

3. Almost Sure Convergence via Robbins–Siegmund

In the preceding section, we established the convergence of the algorithm in the mean-square sense. While this result guarantees that the average error over many realizations decays to zero, it does not strictly characterize the behavior of individual trajectories. Mean-square convergence allows for the theoretical possibility of rare, large deviations. To ensure the reliability of the method for any single execution, we now strengthen our analysis to establish almost sure convergence (convergence with probability one).

Theorem 3.1 (Almost Sure Convergence)**.** *Let the assumptions (T_1), (T_2), (T_3), (T_4), and (T_6) be satisfied. Assume that the following variant of (T_5) holds:*

$(\tilde{T}_5)$ *The sequence of monotonicity constants $\{c_n\}_{n\in\mathbb{N}}$ is bounded above by a constant $\bar{c}>0$. Furthermore, the step sizes $t_n>0$ and the constants $c_n>0$ satisfy*

$$\sum_{n=1}^{+\infty} t_n c_n = +\infty, \quad \sum_{n=1}^{+\infty} t_n^2 < +\infty,$$

and the mesh approximation satisfies the summability condition

$$\sum_{n=1}^{\infty} t_n \|w_n-\bar{u}\| < +\infty.$$

Then the sequence $\{u_n\}$ generated by the stochastic iteration (1.3) converges almost surely to $\bar{u}$:

$$\mathbb{P}\left(\lim_{n\to\infty}\|u_n-\bar{u}\|=0\right)=1.$$

Proof. We divide the proof into four steps. Throughout, $V_n := \|u_n-\bar{u}\|^2$ and $C_0 := \kappa\|\bar{u}\|+\ell$.
Step 1. We start from the pathwise estimate (2.4) obtained in Step 2 of the proof of Theorem 2.1:

$$\begin{aligned}\|u_{n+1}-\bar{u}\|^2 \le \|u_n-\bar{u}\|^2 + 3t_n^2\|\eta_n\|^2 - 2t_n\langle u_n^*, u_n-w_n\rangle \\ - 2t_n\langle \xi_n, u_n-w_n\rangle + 2t_n\|\eta_n\|\|w_n-\bar{u}\|,\end{aligned}$$

where $\eta_n = u_n^*+\xi_n$ and $u_n^*\in F_n(u_n)$. We take the conditional expectation $\mathbb{E}[\cdot\mid\mathscr{F}_n]$. By the martingale difference property (T_6), the term linear in ξ_n vanishes, and

$$\mathbb{E}[\|\eta_n\|^2\mid\mathscr{F}_n] = \|u_n^*\|^2+\mathbb{E}[\|\xi_n\|^2\mid\mathscr{F}_n] \le \|u_n^*\|^2+\sigma^2, \qquad \mathbb{E}[\|\eta_n\|\mid\mathscr{F}_n]\le\|u_n^*\|+\sigma.$$

By (2.3), which follows from the strong monotonicity condition (T_1) and the variational characterization of w_n,

$$-2t_n\langle u_n^*, u_n-w_n\rangle \le -2t_nc_nV_n + 4t_nc_n\sqrt{V_n}\,\|w_n-\bar{u}\|.$$

Finally, by the linear growth condition (T_2) and the triangle inequality $\|u_n\| \le \|u_n - \bar{u}\| + \|\bar{u}\|$, we have $\|u_n^*\| \le \kappa\sqrt{V_n} + C_0$, and hence $\|u_n^*\|^2 \le 2\kappa^2 V_n + 2C_0^2$. Combining these estimates, we obtain

$$\begin{aligned}\mathbb{E}[V_{n+1} \mid \mathscr{F}_n] \le \big(1 - 2t_n c_n + 6\kappa^2 t_n^2\big) V_n + (4c_n + 2\kappa)\, t_n \|w_n - \bar{u}\| \sqrt{V_n} \\ + 3(2C_0^2 + \sigma^2)\, t_n^2 + 2(C_0 + \sigma)\, t_n \|w_n - \bar{u}\|. \end{aligned} \tag{3.1}$$

To decouple the product $\sqrt{V_n}\,\|w_n - \bar{u}\|$, we use the elementary inequality $v \le 1 + v^2$ for $v \ge 0$, which gives $\sqrt{V_n} \le 1 + V_n$. Substituting this bound into (3.1) and grouping the terms, we arrive at

$$\begin{aligned}\mathbb{E}[V_{n+1} \mid \mathscr{F}_n] \le \big(1 + 6\kappa^2 t_n^2 + (4c_n + 2\kappa) t_n \|w_n - \bar{u}\|\big) V_n - 2t_n c_n V_n \\ + 3(2C_0^2 + \sigma^2)\, t_n^2 + (4c_n + 2\kappa + 2C_0 + 2\sigma)\, t_n \|w_n - \bar{u}\|. \end{aligned}$$

Step 2. The inequality obtained in Step 1 is in the canonical form of the Robbins–Siegmund Lemma A.4:

$$\mathbb{E}[V_{n+1} \mid \mathscr{F}_n] \le (1 + \beta_n) V_n + \chi_n - \zeta_n,$$

with the non-negative sequences

- (multiplicative perturbation) $\beta_n = 6\kappa^2 t_n^2 + (4c_n + 2\kappa)\, t_n \|w_n - \bar{u}\|$,
- (additive perturbation) $\chi_n = 3(2C_0^2 + \sigma^2)\, t_n^2 + (4c_n + 2\kappa + 2C_0 + 2\sigma)\, t_n \|w_n - \bar{u}\|$,
- (drift) $\zeta_n = 2t_n c_n V_n$.

The sequences β_n and χ_n are deterministic, and ζ_n is $\mathscr{F}_n$-measurable.

Step 3. We verify the summability of β_n and χ_n. Since $c_n \le \bar{c}$, the terms involving $\|w_n - \bar{u}\|$ are bounded by constant multiples of $t_n \|w_n - \bar{u}\|$; for example,

$$\sum_{n=1}^{\infty} 4t_n c_n \|w_n - \bar{u}\| \le 4\bar{c} \sum_{n=1}^{\infty} t_n \|w_n - \bar{u}\| < +\infty.$$

Hence both β_n and χ_n are bounded by linear combinations of the terms of the convergent series $\sum t_n^2$ and $\sum t_n \|w_n - \bar{u}\|$, and therefore

$$\sum_{n=1}^{\infty} \beta_n < +\infty \quad \text{and} \quad \sum_{n=1}^{\infty} \chi_n < +\infty.$$

The Robbins–Siegmund Lemma A.4 thus implies:

(1) the sequence $\{V_n\}$ converges almost surely to a finite random variable $V_\infty \ge 0$;
(2) the drift terms are summable almost surely: $\sum_{n=1}^{\infty} t_n c_n V_n < +\infty$ a.s.

Step 4. It remains to show that $V_\infty = 0$ almost surely. Let Ω_0 be the event of probability one on which both assertions of Step 3 hold, and fix $\omega \in \Omega_0$. If $V_\infty(\omega) > 0$, then $V_n(\omega) \ge \frac{1}{2} V_\infty(\omega)$ for all sufficiently large n, and hence $\sum_n t_n c_n V_n(\omega) = +\infty$, because $\sum t_n c_n = +\infty$. This contradicts the second assertion of Step 3. Therefore, $V_\infty = 0$ almost surely, which completes the proof. □

Remark 3.1. In contrast to (T_5), condition $(\tilde{T}_5)$ imposes no requirement on the ratios t_n/c_n and $\|w_n - \bar{u}\|/c_n$; the summability conditions compensate for them. If, in addition, these ratios tend to zero, then Theorem 2.1 also yields mean-square convergence.

After establishing that the algorithm converges to $\bar{u}$ with probability one, a natural question arises regarding the speed of this convergence along individual trajectories. Although the mean-square analysis provides a rate for the average error, individual sample paths can exhibit

significant fluctuations. The following theorem quantifies the almost sure polynomial decay rate of the squared error. In analogy with Theorem 2.2, it explicitly links the convergence speed to the step-size decay parameter p and the mesh approximation quality r, showing how the "slower" of these two factors bottlenecks the pathwise performance.

Theorem 3.2 (Almost Sure Convergence Rates)**.** *Assume that conditions* (T_1)*–*(T_4) *and* (T_6) *hold and that there exist constants* $0 < c \le \bar{c}$ *such that* $c \le c_n \le \bar{c}$ *for all* $n \in \mathbb{N}$*. Let the step size be* $t_n = t_0 n^{-p}$ *with* $t_0 > 0$ *and* $\frac{1}{2} < p \le 1$*.*

(a) Suppose that $\|w_n - \bar{u}\| \le \widetilde{C} n^{-r}$ *with* $r > 1 - p$*, and let* q *satisfy*

$$0 < q < \min(2p-1,\, p+r-1).$$

In the case $p = 1$*, assume in addition that* $2ct_0 > q$*. Then*

$$\|u_n - \bar{u}\|^2 = \mathscr{O}(n^{-q}) \quad \textit{almost surely.}$$

(b) Suppose that $\|w_{n+1} - w_n\| \le \widehat{C} n^{-(r+1)}$ *for all* $n \in \mathbb{N}$*, with* $r > 0$*, and let* q *satisfy*

$$0 < q < \min(2p-1,\, 2r+1-p).$$

In the case $p = 1$*, assume in addition that* $2ct_0 > q$*. Then*

$$\|u_n - \bar{u}\|^2 = \mathscr{O}\left(n^{-\min(q,2r)}\right) \quad \textit{almost surely.}$$

Proof. If $p < 1$, set $\varepsilon := 1$; if $p = 1$, choose $\varepsilon \in (0,1)$ so small that $(2-\varepsilon)ct_0 > q$, which is possible since $2ct_0 > q$. Since $t_n \to 0$, there exists $N_0 \ge 2$ such that

$$6\kappa^2 t_n \le \frac{\varepsilon c}{2} \quad \text{and} \quad 2\bar{c}\, t_n \le 1 \qquad \text{for all } n \ge N_0. \tag{3.2}$$

Step 1. For a non-negative adapted sequence $\{X_n\}$, we consider the scaled sequence $Z_n := n^q X_n$. Since $0 < q < 2p - 1 \le 1$, Bernoulli's inequality gives

$$\frac{(n+1)^q}{n^q} = \left(1 + \frac{1}{n}\right)^q \le 1 + \frac{q}{n}.$$

Suppose that, for all $n \ge N_0$,

$$\mathbb{E}[X_{n+1} \mid \mathscr{F}_n] \le (1 - (2-\varepsilon)c\, t_n) X_n + D_n, \tag{3.3}$$

with deterministic $D_n \ge 0$. By (3.2), we have $(2-\varepsilon)c\, t_n \le 2\bar{c}\, t_n \le 1$, so the coefficient of X_n is non-negative, and therefore

$$\mathbb{E}[Z_{n+1} \mid \mathscr{F}_n] \le \rho_n Z_n + (n+1)^q D_n, \qquad \rho_n := \left(1 + \frac{q}{n}\right)\left(1 - \frac{(2-\varepsilon)ct_0}{n^p}\right).$$

We have $\rho_n \le 1 + \frac{q}{n} - \frac{(2-\varepsilon)ct_0}{n^p}$. If $p < 1$, the negative term decays more slowly than $\frac{q}{n}$, so $\rho_n \le 1$ for all sufficiently large n. If $p = 1$, then $\rho_n \le 1 + \frac{1}{n}(q - (2-\varepsilon)ct_0) \le 1$ by the choice of ε. Hence, enlarging N_0 if necessary, $\mathbb{E}[Z_{n+1} \mid \mathscr{F}_n] \le Z_n + (n+1)^q D_n$ for all $n \ge N_0$. If $\sum_n (n+1)^q D_n < +\infty$, the Robbins–Siegmund Lemma A.4 (with $\beta_n = \zeta_n = 0$) implies that $\{Z_n\}$ converges almost surely to a finite random variable. In particular, $\sup_n Z_n < +\infty$ a.s., that is, $X_n = \mathscr{O}(n^{-q})$ almost surely.

Step 2 (proof of (a)). Let $X_n := V_n = \|u_n - \bar{u}\|^2$, and set $A := 4\bar{c} + 2\kappa$, $K_1 := 3(2C_0^2 + \sigma^2)$, and $L := 2(C_0 + \sigma)$. We start from (3.1). By (3.2), $6\kappa^2 t_n^2 \le \frac{\varepsilon c}{2} t_n$, and by Young's inequality,

$$A\, t_n \|w_n - \bar{u}\| \sqrt{V_n} \le \frac{\varepsilon c}{2} t_n V_n + \frac{A^2}{2\varepsilon c} t_n \|w_n - \bar{u}\|^2.$$

Since $c_n \ge c$, we obtain (3.3) with

$$D_n = K_1 t_n^2 + \frac{A^2}{2\varepsilon c} t_n \|w_n - \bar{u}\|^2 + L t_n \|w_n - \bar{u}\|.$$

The discretization error enters linearly through the last term. Using $t_n = t_0 n^{-p}$ and $\|w_n - \bar{u}\|^2 \le \widetilde{C}^2 n^{-2r} \le \widetilde{C}^2 n^{-r}$, we find a constant $K' > 0$ such that $D_n \le K' \left(n^{-2p} + n^{-(p+r)} \right)$. Consequently,

$$(n+1)^q D_n \le 2^q K' \left(n^{q-2p} + n^{q-p-r} \right),$$

which is summable if and only if $q - 2p < -1$ and $q - p - r < -1$, i.e., $q < 2p - 1$ and $q < p + r - 1$. Both conditions hold by assumption, and Step 1 yields (a).

Step 3 (proof of (b)). Let $X_n := W_n := \|u_n - w_{n-1}\|^2$ for $n \ge 2$, and $\Delta_n := \|w_{n-1} - w_n\|$. By Lemma 2.1, $c_n \ge c$, and $2\kappa^2 t_n \le \frac{\varepsilon c}{2}$ (a consequence of (3.2)), we have for $n \ge N_0$

$$\mathbb{E}[\|u_{n+1} - w_n\|^2 \mid \mathscr{F}_n] \le \left(1 - \left(2 - \tfrac{\varepsilon}{2}\right) c\, t_n\right) \|u_n - w_n\|^2 + K t_n^2.$$

By the relaxed triangle inequality with $\delta = \frac{\varepsilon}{2} c\, t_n$,

$$\|u_n - w_n\|^2 \le \left(1 + \tfrac{\varepsilon}{2} c\, t_n\right) W_n + \left(1 + \frac{2}{\varepsilon c\, t_n}\right) \Delta_n^2.$$

With $x := c\, t_n \le \frac{1}{2}$, we have $\left(1 - (2 - \frac{\varepsilon}{2})x\right)(1 + \frac{\varepsilon}{2}x) \le 1 - (2 - \varepsilon)x$ and $\left(1 - (2 - \frac{\varepsilon}{2})x\right)(1 + \frac{2}{\varepsilon x}) \le \frac{3}{\varepsilon x}$. Hence (3.3) holds with

$$D_n = \frac{3}{\varepsilon c\, t_n} \Delta_n^2 + K t_n^2.$$

Since $\Delta_n \le \widehat{C}(n-1)^{-(r+1)} \le 2^{r+1} \widehat{C} n^{-(r+1)}$, there is a constant $K'' > 0$ such that

$$(n+1)^q D_n \le K'' \left(n^{q+p-2r-2} + n^{q-2p} \right),$$

which is summable if and only if $q < 2r + 1 - p$ and $q < 2p - 1$. By Step 1, $W_n = \mathscr{O}(n^{-q})$ almost surely. Finally, the bound on the increments implies $\|w_n - \bar{u}\| \le \widehat{C}(1 + \frac{1}{r}) n^{-r}$ by Remark 2.4, and therefore

$$\|u_n - \bar{u}\|^2 \le 2W_n + 2\|w_{n-1} - \bar{u}\|^2 = \mathscr{O}(n^{-q}) + \mathscr{O}(n^{-2r}) = \mathscr{O}\left(n^{-\min(q, 2r)}\right) \quad \text{a.s.}$$

The proof is complete. □

Remark 3.2. Under the hypotheses of part (a), we have $p + r > 1$, so that $\sum t_n \|w_n - \bar{u}\| < +\infty$ and all conditions of Theorem 3.1 are satisfied. The almost sure rates are slightly slower than the corresponding mean-square rates of Theorem 2.2 (note that $2p - 1 < p$ for $p < 1$), which is the typical price for controlling individual trajectories. As in the mean-square case, the additional information on the increments in part (b) allows the discretization error to enter only quadratically.

The previous theorem quantified the convergence rate under the assumption of persistent strong monotonicity ($c_n \geq c > 0$). Solving ill-posed problems, however, necessitates that the regularization parameter decays to zero ($c_n \to 0$). In this regime, the stabilizing effect of the regularizer diminishes over time, rendering the algorithm increasingly sensitive to stochastic noise. Consequently, establishing almost sure convergence requires stricter control over the interplay between the step size, the regularization decay, and the trajectory of the approximating sequence. The following theorem presents such sufficient conditions.

Theorem 3.3 (Almost Sure Convergence with Vanishing Regularization)**.** *Assume that conditions* (T_1)*–*(T_4) *and* (T_6) *hold, where the monotonicity moduli* $c_n > 0$ *satisfy* $\lim_{n\to\infty} c_n = 0$*. Suppose that*

$$\sum_{n=1}^{\infty} t_n c_n = +\infty, \quad \lim_{n\to\infty} \frac{t_n}{c_n} = 0, \quad \sum_{n=1}^{\infty} t_n^2 < +\infty, \tag{3.4}$$

and that the increments of the regularized solutions satisfy the summability condition

$$\sum_{n=2}^{\infty} \frac{\|w_{n-1} - w_n\|^2}{t_n c_n} < +\infty. \tag{3.5}$$

Then the sequence $\{u_n\}$ *converges almost surely to* $\bar{u}$ *(in the Tikhonov setting, the minimum-norm solution):*

$$\mathbb{P}\left(\lim_{n\to\infty} \|u_n - \bar{u}\| = 0\right) = 1.$$

Proof. We divide the proof into four steps.

Step 1. By Lemma 2.1 and the condition $t_n/c_n \to 0$, the argument of Step 1 of the proof of Theorem 2.3 shows that, for all sufficiently large n,

$$\mathbb{E}[\|u_{n+1} - w_n\|^2 \mid \mathscr{F}_n] \leq (1 - t_n c_n)\|u_n - w_n\|^2 + K t_n^2, \tag{3.6}$$

where $K > 0$ is the constant from Lemma 2.1. Note also that $t_n = (t_n/c_n)\, c_n \to 0$, so that $t_n c_n \to 0$.

Step 2. Let $V_n := \|u_n - w_{n-1}\|^2$ for $n \geq 2$. By definition, $V_{n+1} = \|u_{n+1} - w_n\|^2$, so the left-hand side of (3.6) is precisely $\mathbb{E}[V_{n+1} \mid \mathscr{F}_n]$.

To relate the term $\|u_n - w_n\|^2$ on the right-hand side to V_n, we expand the shift in the target sequence. With $\Delta_n := \|w_{n-1} - w_n\|$,

$$\begin{aligned} \|u_n - w_n\|^2 &= \|(u_n - w_{n-1}) + (w_{n-1} - w_n)\|^2 \\ &\leq \|u_n - w_{n-1}\|^2 + 2\|u_n - w_{n-1}\|\|w_{n-1} - w_n\| + \|w_{n-1} - w_n\|^2 \\ &= V_n + 2\sqrt{V_n}\,\Delta_n + \Delta_n^2. \end{aligned} \tag{3.7}$$

Substituting this expansion into (3.6), we obtain

$$\mathbb{E}[V_{n+1} \mid \mathscr{F}_n] \leq (1 - t_n c_n)\left(V_n + 2\sqrt{V_n}\,\Delta_n + \Delta_n^2\right) + K t_n^2. \tag{3.8}$$

To handle the cross term $2\sqrt{V_n}\,\Delta_n$, we apply the weighted Young inequality $2ab \leq \varepsilon a^2 + \frac{1}{\varepsilon} b^2$ with $\varepsilon = \frac{1}{2} t_n c_n$, which ensures that the resulting V_n term can be absorbed by the contraction factor:

$$2\sqrt{V_n}\,\Delta_n \leq \left(\frac{1}{2} t_n c_n\right) V_n + \left(\frac{2}{t_n c_n}\right) \Delta_n^2.$$

Substituting this bound into (3.8) and distributing the factor $(1-t_nc_n)$, the right-hand side of (3.8) is bounded by

$$\begin{aligned}&(1-t_nc_n)\left(1+\frac{1}{2}t_nc_n\right)V_n+(1-t_nc_n)\left(1+\frac{2}{t_nc_n}\right)\Delta_n^2+Kt_n^2\\&\leq\left(1-\frac{1}{2}t_nc_n-\frac{1}{2}(t_nc_n)^2\right)V_n+\left(1+\frac{2}{t_nc_n}\right)\Delta_n^2+Kt_n^2.\end{aligned}$$

We drop the negative quadratic term $-\frac{1}{2}(t_nc_n)^2$. For the coefficient of Δ_n^2, since $t_nc_n\to 0$, we have $t_nc_n\leq 1$ for all sufficiently large n, and hence $1+\frac{2}{t_nc_n}\leq\frac{3}{t_nc_n}$. This yields the recursive inequality

$$\mathbb{E}[V_{n+1}\mid\mathscr{F}_n]\leq\left(1-\frac{1}{2}t_nc_n\right)V_n+\underbrace{\frac{3\|w_{n-1}-w_n\|^2}{t_nc_n}+Kt_n^2}_{=:\chi_n}. \tag{3.9}$$

Step 3. Inequality (3.9) is of the Robbins–Siegmund form with $\beta_n=0$ and drift $\zeta_n=\frac{1}{2}t_nc_nV_n$:

$$\mathbb{E}[V_{n+1}\mid\mathscr{F}_n]\leq V_n+\chi_n-\tfrac{1}{2}t_nc_nV_n.$$

The additive term χ_n is summable:

$$\sum_n\chi_n=3\sum_n\frac{\|w_{n-1}-w_n\|^2}{t_nc_n}+K\sum_n t_n^2<+\infty,$$

where the first sum is finite by (3.5) and the second by (3.4). Therefore, the Robbins–Siegmund Lemma A.4 implies that V_n converges almost surely to a finite random variable $V_\infty\geq 0$, and that $\sum_n t_nc_nV_n<+\infty$ almost surely.

Step 4. Since $\sum t_nc_n=+\infty$, the argument of Step 4 of the proof of Theorem 3.1 shows that $V_\infty=0$ almost surely. Finally, by the triangle inequality,

$$\|u_n-\bar{u}\|\leq\|u_n-w_{n-1}\|+\|w_{n-1}-\bar{u}\|=\sqrt{V_n}+\|w_{n-1}-\bar{u}\|.$$

Since $V_n\to 0$ almost surely and $w_{n-1}\to\bar{u}$ by (T_4), we conclude that

$$\lim_{n\to\infty}\|u_n-\bar{u}\|=0\quad\text{almost surely.}$$

The proof is complete. □

Remark 3.3. Theorem 3.3 does not require the tracking condition $\|w_{n-1}-w_n\|/(t_nc_n)\to 0$ of Theorem 2.3; it is replaced by the summability condition (3.5). If both conditions hold, the iterates converge both almost surely and in mean square.

Although the preceding theorem provides abstract criteria for almost sure convergence, practical implementation requires specific schedules for the step size t_n and the regularization parameter c_n. The most common and robust choice is polynomial decay, which allows for a straightforward trade-off between noise suppression and convergence speed. To verify the summability condition (3.5) under such schedules, we must first understand the behavior of the underlying deterministic reference trajectory $\{w_n\}$. To avoid the difficulties associated with evaluating the discrete velocity of the target across strictly expanding constraint sets, we restrict our analysis to the case of a fixed constraint set C. This ensures that consecutive regularized solutions can be tested against one another in their respective variational inequalities. The following lemma

quantifies the drift of this trajectory, showing that the discrete velocity $\|w_{n-1}-w_n\|$ decays at a controlled rate.

Lemma 3.1 (Stability of Regularized VI Trajectory)**.** *Let H be a real Hilbert space and $C \subseteq H$ be a non-empty, closed, and convex constraint set. Consider the sequence of regularized variational inequality problems: find $w_n \in C$ such that there exists $g_n \in S(w_n)$ satisfying*

$$\langle g_n + c_n T(w_n), z - w_n \rangle \geq 0, \quad \text{for all } z \in C, \tag{3.10}$$

where $S : C \rightrightarrows H$ is a monotone set-valued operator, $T : C \to H$ is a single-valued, strongly monotone operator with modulus $\mu > 0$, and $\{c_n\}$ is a positive regularization sequence given by $c_n = c_0 n^{-k}$ with $c_0, k > 0$. Assume that these problems admit solutions w_n such that $M := \sup_n \|T(w_n)\| < +\infty$.

Then, for all $n \geq 2$, the increments of the regularized solutions satisfy

$$\|w_{n-1} - w_n\| \leq \frac{M}{\mu} \left| \frac{c_{n-1} - c_n}{c_n} \right| = \mathscr{O}\left(\frac{1}{n}\right). \tag{3.11}$$

Proof. Let w_{n-1} and w_n be the solutions to the regularized VI at steps $n-1$ and n. Because the constraint set C is fixed, both w_{n-1} and w_n belong to C, so each can be used as a test function in the other's variational inequality. By definition, there exist elements $g_{n-1} \in S(w_{n-1})$ and $g_n \in S(w_n)$ such that for all $z \in C$,

$$\langle g_{n-1} + c_{n-1} T(w_{n-1}), z - w_{n-1} \rangle \geq 0, \tag{3.12}$$

$$\langle g_n + c_n T(w_n), z - w_n \rangle \geq 0. \tag{3.13}$$

We test the inequalities against each other's solutions. Substituting $z = w_n \in C$ into (3.12) gives

$$\langle g_{n-1} + c_{n-1} T(w_{n-1}), w_n - w_{n-1} \rangle \geq 0, \tag{3.14}$$

and substituting $z = w_{n-1} \in C$ into (3.13) gives

$$\langle g_n + c_n T(w_n), w_{n-1} - w_n \rangle \geq 0. \tag{3.15}$$

Summing (3.14) and (3.15) and rearranging the terms (noting that $w_n - w_{n-1} = -(w_{n-1} - w_n)$) yields

$$-\langle g_{n-1} - g_n, w_{n-1} - w_n \rangle - \langle c_{n-1} T(w_{n-1}) - c_n T(w_n), w_{n-1} - w_n \rangle \geq 0,$$

that is,

$$\langle g_{n-1} - g_n, w_{n-1} - w_n \rangle + \langle c_{n-1} T(w_{n-1}) - c_n T(w_n), w_{n-1} - w_n \rangle \leq 0. \tag{3.16}$$

By the monotonicity of S, the first term is non-negative: $\langle g_{n-1} - g_n, w_{n-1} - w_n \rangle \geq 0$. Therefore, the second term must be non-positive:

$$\langle c_{n-1} T(w_{n-1}) - c_n T(w_n), w_{n-1} - w_n \rangle \leq 0.$$

We decompose the operator difference by adding and subtracting $c_n T(w_{n-1})$:

$$(c_{n-1} - c_n) \langle T(w_{n-1}), w_{n-1} - w_n \rangle + c_n \langle T(w_{n-1}) - T(w_n), w_{n-1} - w_n \rangle \leq 0.$$

Using the strong monotonicity of T (with modulus μ) in the second term,

$$\langle T(w_{n-1}) - T(w_n), w_{n-1} - w_n \rangle \geq \mu \|w_{n-1} - w_n\|^2,$$

and isolating the squared norm, we obtain

$$c_n \mu \|w_{n-1} - w_n\|^2 \leq -(c_{n-1} - c_n) \langle T(w_{n-1}), w_{n-1} - w_n \rangle.$$

Taking absolute values of the right-hand side and applying the Cauchy–Schwarz inequality gives

$$c_n\mu\|w_{n-1}-w_n\|^2 \le |c_{n-1}-c_n|\|T(w_{n-1})\|\|w_{n-1}-w_n\|.$$

If $w_{n-1}=w_n$, there is nothing to prove. Otherwise, dividing by $c_n\mu\|w_{n-1}-w_n\|$, we get

$$\|w_{n-1}-w_n\| \le \frac{\|T(w_{n-1})\|}{\mu}\frac{|c_{n-1}-c_n|}{c_n} \le \frac{M}{\mu}\left|\frac{c_{n-1}-c_n}{c_n}\right|.$$

Finally, for the polynomial decay $c_n = c_0 n^{-k}$ and $n \ge 2$, the mean value theorem gives

$$\left|\frac{c_{n-1}-c_n}{c_n}\right| = \left(1-\frac{1}{n}\right)^{-k} - 1 \le \frac{k}{n}\left(1-\frac{1}{n}\right)^{-k-1} \le \frac{k\,2^{k+1}}{n} = \mathcal{O}\left(\frac{1}{n}\right).$$

This completes the proof. □

We can now translate the abstract conditions of Theorem 3.3 into explicit conditions on polynomial schedules.

Theorem 3.4 (Sufficient Conditions for Almost Sure Convergence)**.** *Let conditions* (T_1)*–*(T_4) *and* (T_6) *hold with a fixed constraint set* $C_n = C$ *(so that* $C \subset H_n$ *for all* n*) and regularized operators* $F_n = S + c_n T$ *for all* n*, where* S*,* T*, and* μ *are as in Lemma 3.1 and* $\sup_n \|T(w_n)\| < +\infty$*. Let* $t_n = t_0 n^{-p}$ *and* $c_n = c_0 n^{-k}$ *with* $t_0, c_0 > 0$*. If the exponents satisfy*

$$\frac{1}{2} < p < 1, \quad 0 < k < 1-p, \tag{3.17}$$

(which implies $p+k<1$*), then the sequence* $\{u_n\}$ *converges almost surely to* $\bar{u}$ *(in the Tikhonov setting, the minimum-norm solution):*

$$\mathbb{P}\left(\lim_{n\to\infty}\|u_n-\bar{u}\|=0\right)=1. \tag{3.18}$$

Proof. Since T is strongly monotone with modulus μ and S is monotone, the operator $F_n = S + c_n T$ is strongly monotone with modulus μc_n. Thus (T_1) holds with μc_n in place of c_n, and we verify the hypotheses of Theorem 3.3 for this modulus. Since all conditions in (3.4) and (3.5) are invariant under multiplication of c_n by a positive constant, it suffices to verify them for $c_n = c_0 n^{-k}$.

Step 1. Since $p > \frac{1}{2}$, we have $1-p<p$, and hence $k<p$. Therefore, $t_n/c_n = (t_0/c_0)\,n^{k-p} \to 0$. Moreover, $\sum_n t_n c_n = t_0 c_0 \sum_n n^{-(p+k)} = +\infty$ since $p+k<1$, and $\sum_n t_n^2 = t_0^2 \sum_n n^{-2p} < +\infty$ since $p > \frac{1}{2}$. The latter condition controls the accumulated variance of the stochastic noise. Thus, all conditions in (3.4) hold, and $\lim_{n\to\infty} c_n = 0$.

Step 2. By Lemma 3.1, which applies since the constraint set is fixed, there exists a constant $C_3 > 0$ such that $\|w_{n-1}-w_n\| \le C_3 n^{-1}$ for all $n \ge 2$. Consequently,

$$\frac{\|w_{n-1}-w_n\|^2}{t_n c_n} \le \frac{C_3^2}{t_0 c_0} n^{p+k-2}.$$

The series $\sum_n n^{p+k-2}$ converges if and only if $p+k-2<-1$, i.e., $p+k<1$. Hence (3.5) holds; this is the origin of the upper bound on the regularization decay rate k.

Step 3. All hypotheses of Theorem 3.3 are satisfied, and the assertion follows. We note that, in addition, $\|w_{n-1}-w_n\|/(t_n c_n) \le \frac{C_3}{t_0 c_0} n^{p+k-1} \to 0$, so that the conditions of Theorem 2.3 also hold and the iterates converge in mean square as well. □

Remark 3.4. The case $p = 1$ is excluded in Theorem 3.4, since the conditions $k > 0$ and $k < 1 - p$ are then incompatible. This reflects the fact that, with vanishing regularization, the step sizes must decay more slowly than n^{-1} in order to maintain the divergence of $\sum t_n c_n$.

4. Concluding Remarks

In this paper, we presented a comprehensive study on the convergence of stochastic approximation schemes for solving multivalued variational inequalities. By analyzing the interplay between stochastic noise, spatial discretization errors, and regularization parameters, we established a unified framework that guarantees convergence in both the mean-square and almost sure senses. Our analysis progressed from the well-posed setting, where we derived explicit polynomial convergence rates, with improved rates when the discrete solutions vary slowly, to the more challenging ill-posed regime requiring vanishing regularization. A key contribution of this work was the derivation of precise consistency conditions—specifically the tracking requirement $p + k < 1$—that ensure the algorithm successfully pursues the shifting set of regularized solutions. We further supported these theoretical results with a stability analysis of the deterministic trajectory, proving that the discrete velocity of the target solutions decays sufficiently fast to permit tracking. Several avenues for future research naturally emerge from this framework. On the theoretical side, it would be of interest to extend the stability analysis of the regularized solutions to expanding constraint sets C_n, thereby combining vanishing regularization and mesh refinement in explicit parameter schedules. Regarding numerical applications, we plan to conduct extensive numerical experiments applying the proposed scheme to stochastic inverse problems and optimal control models. These simulations will serve to validate the theoretical convergence rates and demonstrate the practical robustness of the algorithm under varying noise levels. Furthermore, investigating iterate averaging (Polyak–Ruppert) and adaptive step-size rules could further enhance practical performance and robustness compared to fixed polynomial schedules.

Acknowledgements
The authors are grateful to the reviewers for their careful reading of the manuscript and for their valuable suggestions, which have significantly improved this work.

References

[1] J. Gwinner, B. Jadamba, A. A. Khan, F. Raciti, Uncertainty Quantification in Variational Inequalities, CRC Press, 2021.
[2] B. Jadamba, A. A. Khan, M. Sama, H.-J. Starkloff, C. Tammer, A convex optimization framework for the inverse problem of identifying a random parameter in a stochastic partial differential equation, SIAM/ASA J. Uncertainty Quantification 9 (2) (2021) 922–952.
[3] B. Jadamba, A. A. Khan, M. Richards, M. Sama, A convex inversion framework for identifying parameters in saddle point problems with applications to inverse incompressible elasticity, Inverse Problems 36 (7) (2020) 074003, 25.
[4] H. Robbins, S. Monro, A stochastic approximation method, Ann. Math. Statistics 22 (1951) 400–407.
[5] K. Barty, J.-S. Roy, C. Strugarek, Hilbert-valued perturbed subgradient algorithms, Math. Oper. Res. 32 (3) (2007) 551–562.
[6] D. P. Bertsekas, J. N. Tsitsiklis, Gradient convergence in gradient methods with errors, SIAM J. Optim. 10 (3) (2000) 627–642.

[7] J. Dippon, H. Walk, The averaged Robbins-Monro method for linear problems in a Banach space, J. Theoret. Probab. 19 (1) (2006) 166–189.
[8] J.-P. Bertran, Optimisation stochastique dans un espace de Hilbert. Méthode de gradient, C. R. Acad. Sci. Paris Sér. A-B 276 (1973) A613–A616.
[9] L. Goldstein, Minimizing noisy functionals in Hilbert space: an extension of the Kiefer-Wolfowitz procedure, J. Theoret. Probab. 1 (2) (1988) 189–204.
[10] J.-B. Hiriart-Urruty, Algorithmes stochastiques de résolution d'équations et d'inéquations variationnelles, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 33 (3) (1975/76) 167–186.
[11] A. Barbagallo, S. G. L. Bianco, Stochastic variational formulation for a general random time-dependent economic equilibrium problem, Optim. Lett. 14 (8) (2020) 2479–2493.
[12] R. I. Bot, P. Mertikopoulos, M. Staudigl, P. T. Vuong, Forward-backward-forward methods with variance reduction for stochastic variational inequalities, arXiv preprint arXiv:1902.03355.
[13] A. Dieuleveut, A. Durmus, F. Bach, Bridging the gap between constant step size stochastic gradient descent and Markov chains, Ann. Statist. 48 (3) (2020) 1348–1382.
[14] A. N. Iusem, A. Jofré, R. I. Oliveira, P. Thompson, Extragradient method with variance reduction for stochastic variational inequalities, SIAM J. Optim. 27 (2) (2017) 686–724.
[15] A. N. Iusem, A. Jofré, R. I. Oliveira, P. Thompson, Variance-based extragradient methods with line search for stochastic variational inequalities, SIAM J. Optim. 29 (1) (2019) 175–206.
[16] H. Jiang, H. Xu, Stochastic approximation approaches to the stochastic variational inequality problem, IEEE Trans. Automat. Control 53 (6) (2008) 1462–1475.
[17] J. Koshal, A. Nedić, U. V. Shanbhag, Regularized iterative stochastic approximation methods for stochastic variational inequality problems, IEEE Trans. Automat. Control 58 (3) (2013) 594–609.
[18] P. Mertikopoulos, M. Staudigl, Stochastic mirror descent dynamics and their convergence in monotone variational inequalities, J. Optim. Theory Appl. 179 (3) (2018) 838–867.
[19] M. Passacantando, F. Raciti, Optimal road maintenance investment in traffic networks with random demands, Optim. Lett. 15 (5) (2021) 1799–1819.
[20] L. Rosasco, S. Villa, B. C. Vũ, Convergence of stochastic proximal gradient algorithm, Appl. Math. Optim. 82 (3) (2020) 891–917.
[21] D. Scieur, A. d'Aspremont, F. Bach, Regularized nonlinear acceleration, Math. Program. 179 (1-2, Ser. A) (2020) 47–83.
[22] F. Q. Xia, Q. H. Ansari, J. C. Yao, A new incremental constraint projection method for solving monotone variational inequalities, Optim. Methods Softw. 32 (3) (2017) 470–502.
[23] Z.-P. Yang, J. Zhang, Y. Wang, G.-H. Lin, Variance-based subgradient extragradient method for stochastic variational inequality problems, J. Sci. Comput. 89 (1) (2021) Paper No. 4, 28.
[24] X.-J. Zhang, X.-W. Du, Z.-P. Yang, G.-H. Lin, An infeasible stochastic approximation and projection algorithm for stochastic variational inequalities, J. Optim. Theory Appl. 183 (3) (2019) 1053–1076.
[25] R. Bollapragada, T. Chen, R. Ward, On the fast convergence of minibatch heavy ball momentum, IMA J. Numer. Anal. 45 (3) (2025) 1397–1424.
[26] S. Duan, Z. Wu, Mean-field forward-backward stochastic variational inequalities with oblique subgradients and nonlocal parabolic variational inequalities, J. Math. Anal. Appl. 559 (1) (2026) Paper No. 130505, 34.
[27] Y. Fu, J. R. Angus, H. Qin, V. I. Geyko, Energy-momentum-conserving stochastic differential equations and algorithms for the nonlinear Landau-Fokker-Planck equation, Phys. Rev. E 111 (2) (2025) Paper No. 025211, 15.
[28] Z.-J. Gong, A. A. Khan, M. Sama, H.-J. Starkloff, Estimating mean and variance of random coefficients in stochastic variational problems using second-order methods, J. Optim. Theory Appl. 208 (1) (2026) Paper No. 12, 48.
[29] Z.-J. Gong, A. A. Khan, M. Sama, A. Zălinescu, Estimating random coefficients in mixed variational problems with an application to the stochastic elasticity imaging inverse problem for tumor identification, J. Convex Anal. 32 (2) (2025) 321–357.
[30] T. Li, X. Cai, D. Han, A Resampling-Free Stochastic Projection Contraction Algorithm for Solving Stochastic Variational Inequalities, Numer. Math. Theory Methods Appl. 19 (2) (2026) 275–297.

[31] J. Li, Y.-b. Xiao, A modified Popov algorithm for non-monotone and non-Lipschitzian stochastic variational inequalities, J. Optim. Theory Appl. 205 (3) (2025) Paper No. 54, 26.
[32] C. Liu, Y. Huang, Stochastic Polyak stepsize for SGD with variance reduction and momentum, Optim. Lett. 19 (8) (2025) 1601–1624.
[33] H. Song, H. Wang, J. Wu, J. Yang, An efficient ADAM-type algorithm with finite elements discretization technique for random elliptic optimal control problems, J. Comput. Appl. Math. 454 (2025) Paper No. 116199, 16.
[34] Y. Zeng, D. Han, Y. Su, J. Xie, On adaptive stochastic heavy ball momentum for solving linear systems, SIAM J. Matrix Anal. Appl. 45 (3) (2024) 1259–1286.
[35] B. Zhou, J. Jiang, Y. Song, H. Sun, Variance-based stochastic projection gradient method for two-stage cocoercive stochastic variational inequalities, Numer. Algorithms 98 (1) (2025) 1–33.
[36] C. Geiersbach, G. C. Pflug, Projected stochastic gradients for convex constrained problems in Hilbert spaces, SIAM J. Optim. 29 (3) (2019) 2079–2099.
[37] M. Martin, S. Krumscheid, F. Nobile, Complexity analysis of stochastic gradient methods for PDE-constrained optimal control problems with uncertain parameters, ESAIM Math. Model. Numer. Anal. 55 (4) (2021) 1599–1633.
[38] M. Morzfeld, X. Tu, J. Wilkening, A. J. Chorin, Parameter estimation by implicit sampling, Commun. Appl. Math. Comput. Sci. 10 (2) (2015) 205–225.
[39] B. Pedretscher, B. Kaltenbacher, O. Pfeiler, Parameter identification and uncertainty quantification in stochastic state space models and its application to texture analysis, Appl. Numer. Math. 146 (2019) 38–54.
[40] M. Dambrine, A. A. Khan, M. Sama, A stochastic regularized second-order iterative scheme for optimal control and inverse problems in stochastic partial differential equations, Philos. Trans. Roy. Soc. A 380 (2236) (2022) Paper No. 20210352, 13.
[41] J. Dippon, J. Gwinner, A. A. Khan, M. Sama, A new regularized stochastic approximation framework for stochastic inverse problems, Nonlinear Anal. Real World Appl. 73 (2023) Paper No. 103869, 29.
[42] B. Jadamba, A. A. Khan, A. Oberai, M. Sama, First-order and second-order adjoint methods for parameter identification problems with an application to the elasticity imaging inverse problem, Inverse Problems in Science and Engineering 25 (2017) 1768–1787.
[43] B. Jadamba, A. A. Khan, M. Sama, Y. Yang, An iteratively regularized stochastic gradient method for estimating a random parameter in a stochastic pde. A variational inequality approach, J. Nonlinear Var. Anal. 5 (2021) 865–880.
[44] B. Jadamba, A. A. Khan, F. Raciti, M. Sama, A variational inequality based stochastic approximation for estimating the flexural rigidity in random fourth-order models, Communications in Nonlinear Science and Numerical Simulation 111 (2022) 106406.
[45] R. Hawks, B. Jadamba, A. A. Khan, M. Sama, Y. Yang, A variational inequality based stochastic approximation for inverse problems in stochastic partial differential equations, in: Nonlinear Analysis and Global Optimization, Springer Optimization and Its Applications, Springer, Heidelberg, 2021, pp. 207–226.
[46] R. Kluge, G. Telschow, On the convergence and speed of convergence of some iteration methods for variational inequalities. I, in: Theory of nonlinear operators (Proc. Fourth Internat. Summer School, Acad. Sci., Berlin, 1975), Abhandlungen der Akademie der Wissenschaften der DDR, Abteilung Mathematik, Naturwissenschaften, Technik, No. 1, Akademie-Verlag, Berlin, 1977, pp. 135–163.
[47] J.-P. Aubin, H. Frankowska, Set-valued analysis, Vol. 2 of Systems & Control: Foundations & Applications, Birkhäuser Boston, Inc., Boston, MA, 1990.
[48] H. Robbins, D. Siegmund, A convergence theorem for non negative almost supermartingales and some applications, in: J. S. Rustagi (Ed.), Optimizing Methods in Statistics (Proc. Sympos., Ohio State Univ., Columbus, Ohio, 1971), Academic Press, New York, 1971, pp. 233–257.

APPENDIX A. AUXILIARY LEMMAS

To rigorously establish the convergence of the proposed stochastic approximation scheme, we rely on the following fundamental lemmas regarding recursive inequalities. These results provide the essential analytical tools to characterize the asymptotic behavior of the error sequence, allowing us to disentangle the deterministic contraction properties from the accumulation of stochastic noise. Since the conclusions of Lemmas A.1–A.4 are asymptotic, it suffices that the respective inequalities and conditions hold for all $n \geq N_0$, for some fixed $N_0 \in \mathbb{N}$; this is how they are applied in the main text.

Lemma A.1 (Convergence of Recursive Inequalities [46])**.** *Let $\{s_n\}_{n\geq 1}$ be a sequence of non-negative real numbers satisfying the inequality $s_{n+1} \leq (1-\alpha_n)s_n + \varepsilon_n$, $n \in \mathbb{N}$, where $\{\alpha_n\}$ and $\{\varepsilon_n\}$ are non-negative sequences. Assume that $\alpha_n \in [0,1)$ and $\sum_{n=1}^{\infty} \alpha_n = +\infty$. If either $\sum_{n=1}^{\infty} \varepsilon_n < +\infty$, or $\alpha_n > 0$ for all n and $\lim_{n\to\infty} \varepsilon_n/\alpha_n = 0$, then $\lim_{n\to\infty} s_n = 0$.*

Lemma A.2 (Boundedness of Nonlinear Recursive Inequalities [46])**.** *Let $\{s_n\}_{n\geq 1}$ be a sequence of non-negative real numbers satisfying $s_{n+1} \leq (1-\nu_n)s_n + \delta_n\sqrt{s_n} + \gamma_n$, $n \in \mathbb{N}$, where $\nu_n \in (0,1)$ and $\delta_n, \gamma_n \geq 0$. Assume that there exists a constant $\theta \in (0,1)$ such that $\delta_n \leq \theta\nu_n$ for all sufficiently large n, and that $\sup_n \gamma_n/\nu_n < +\infty$. Then the sequence $\{s_n\}$ is bounded.*

Lemma A.3 (Rate of Convergence [46])**.** *Let $\{s_n\}_{n\geq 1}$ be a sequence of non-negative real numbers satisfying $s_{n+1} \leq (1-an^{-p})s_n + bn^{-(p+q)}$, $n \in \mathbb{N}$, where $a > 0$, $b > 0$, $p \in (0,1]$, and $q > 0$. Then $s_n = \mathcal{O}(n^{-q})$, provided that $a > q$ whenever $p = 1$.*

Lemma A.4 (Robbins–Siegmund [48])**.** *Let $\{\mathcal{F}_n\}$ be a filtration on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let V_n, β_n, χ_n, and ζ_n be non-negative $\mathcal{F}_n$-measurable random variables such that*

$$\mathbb{E}[V_{n+1} \mid \mathcal{F}_n] \leq (1+\beta_n)V_n + \chi_n - \zeta_n, \quad n \in \mathbb{N}.$$

If $\sum_{n=1}^{\infty} \beta_n < +\infty$ and $\sum_{n=1}^{\infty} \chi_n < +\infty$ almost surely, then:

(i) V_n converges almost surely to a random variable V_∞ with $V_\infty < +\infty$ a.s.;

(ii) $\sum_{n=1}^{\infty} \zeta_n < +\infty$ almost surely.