JOURNAL OF DECISION MAKING AND HEALTHCARE
Volume 3 (2026), No. 2, 78–107
DOI:https://doi.org/10.69829/jdmh-026-0302-ta02
Tulipa Opera Scholarum

C-FISTA-INSPIRED SINGLE-PROJECTION ALGORITHM FOR STOCHASTIC VARIATIONAL INEQUALITIES WITH APPLICATIONS IN NETWORK BANDWIDTH ALLOCATION PROBLEM

JEREMIAH NKWEGU EZEORA[1,*], FRANCIS ONYINYECHI NWAWURU[2], AND JEN-CHIH YAO[3]

[1] *Department of Mathematics and Statistics, University of Port Harcourt, Rivers State, Nigeria*
[2] *Analysis, Control System and Optimization Research Group, Department of Mathematics, Chukwuemeka Odumegwu Ojukwu University, Uli 432107, Anambra State, Nigeria*
[3] *Center for General Education, China Medical University, Taichung 40402, Taiwan*

Abstract. This paper investigates stochastic variational inequality problems arising in large-scale optimization, machine learning, and network systems, where exact operator evaluations are unavailable. Motivated by the need for computationally efficient and scalable algorithms, we propose a novel C-FISTA-inspired single-projection method that integrates inertial acceleration, adaptive stepsize selection, and variance-controlled stochastic approximation. Unlike classical extragradient-type methods that require multiple projections or prior knowledge of Lipschitz constants, the proposed scheme performs only one projection per iteration while maintaining strong convergence guarantees. The algorithm incorporates a momentum-driven correction mechanism and a data-dependent stepsize rule, ensuring robustness under stochastic noise. Under standard assumptions, we establish almost sure convergence of the generated sequences to a solution of the stochastic variational inequality. Furthermore, under strong pseudomonotonicity, we derive linear convergence rates in both almost sure and mean-square senses. Numerical experiments on stochastic complementarity problems, fractional programming models, and network bandwidth allocation demonstrate the superior performance of the proposed method in terms of accuracy, stability, and computational efficiency compared with existing algorithms.

Keywords. C-FISTA; Stochastic variational inequality; Stochastic approximation; Variance reduction.

1. Introduction

Stochastic variational inequality (SVI) problems constitute a fundamental class of models arising in optimization, machine learning, economics, and engineering. These problems extend deterministic variational inequalities by incorporating randomness through stochastic oracles, thereby enabling the modeling of large-scale systems where exact evaluations of the underlying operator are either unavailable or computationally prohibitive. In particular, given a nonempty closed convex set $X \subset \mathbb{R}^n$, ξ be a random vector with probability support on $\Xi \subset \mathbb{R}^n$ and $f : \mathbb{R}^n \times \Xi \to \mathbb{R}^n$ be a mapping that is integrable with respect to Ξ. The stochastic variational inequality finds a point $x^\star \in X$ such that

$$\langle F(x^\star), x - x^\star \rangle \geq 0, \quad \forall x \in X, \tag{1.1}$$

where $F(x^\star) = \mathbb{E}[f(x, \xi)] \in \mathbb{R}^n$. The model (1.1) has been extensively studied in literature by different authors, using different methods and has been applied to network bandwidth allocations, data science, economics and many more as can be seen in [1–4]. We note that if $\mathbb{E}[f(x, \xi)]$ can be easily computed, then (1.1) becomes a deterministic variational inequality problem (DVIP for short). In this case, numerous iterative algorithms such as [5, 6] developed to solve variational inequality could be deployed

*Corresponding author.
E-mail address: jeremiah.ezeora@uniport.edu.ng (J. N. Ezeora), fo.nwawuru@coou.edu.ng (F. O. Nwawuru), and yaojc@mail.cmu.edu.tw (J.-C. Yao)
Received: May 06, 2026, Revised: May 13, 2026, and Accepted: May 20, 2026.

to solve (1.1). However, because this is not always the case due to unknown probability distribution, computing $\mathbb{E}[f(x,\xi)]$ can be reasonably difficult, the methods for VIPs might fail to work for SVIP. Hence, there is need to develop a different method for solving (1.1).

Basically, there are two classical methods for solving SVIP: the stochastic average approximation (SAA) and stochastic approximation (SA). The SAA has been used to investigate solutions to stochastic generalized equations [7]. Furthermore, in the work of Liang [8], they transformed (1.1) into unconstrained optimization problem and applied SAA-based hybrid method Newton method. Other interesting results on SAA can be found in [9, 10] and cited references contained therein. On the other hand, SA has recently gained scientific attention because its superior scalability, online adaptability, and low per-iteration complexity, making it particularly suitable for large-scale and oracle-based stochastic optimization and variational inequality problems. In the early developemnt of this method, Jiang and Xu [11] proposed a single projection based SVIP with the update:

$$x_0 \in X \qquad x_{k+1} = \Pi_X(x_k - \lambda_k f(x_k, \xi_k)), \qquad k \in \mathbb{N}_0, \tag{1.2}$$

where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, Π_X is a projection from $\mathbb{R}^n$ onto X and λ_k is a nonnegative sequence satisfying $\sum_{k=0}^{\infty} \lambda_k = \infty$ and $\sum_{k=0}^{\infty} \lambda_k^2 < \infty$. Under strong monotone condition on the cost function F and Lipschitz continuity of F on X, the authors established that the recursive sequence generated by (1.2) converge almost surely (a.s) to the unique solution of the SVIP defined in (1.1). Since after this development, (1.2) has been modified incorporating extragradient method [12] (see more results on [13, 14]), subgradient extragradient method [15] (see also [5, 16]). These Classical methods for solving SVIs, often require multiple projections per iteration or rely on conservative stepsizes dictated by Lipschitz constants. While these approaches are theoretically sound, their computational burden becomes significant in high-dimensional settings or when projections onto the feasible set are expensive. Recent advances have therefore focused on designing single-projection methods and adaptive stepsize strategies that reduce per-iteration complexity while maintaining strong convergence guarantees [2, 3, 17].

Recently, fast iterative optimization algorithms have gained a considerable attention because of few iterations required for convergence. The celebrated heavy-ball method of Polyak [18], popularly called inertia method has continuously been employed to study many optimization problems in different settings [1, 18–21]. Furthermore, fast iterative shrinking-thresholding algorithm (FISTA) [22] has significantly improved the convergence speed of many algorithms. This technique has been modified and extended to enhance performance of iterative schemes. A significant variant of [22] is the C-FISTA [23], adopted for studying composite optimization models. The C-FISTA has the following structure:

$$\begin{cases} y^k = \frac{1}{1+\theta} x^k + \frac{\theta}{1+\theta} z^k \\ x^{k+1} = \arg\min_{x \in X} \left(H(\bar{y}^k) + \nabla H(\bar{y}^k)^\top (x - y^k) + \frac{rL}{2} \|x - y^k\|^2 + R(x) \right\} \\ x^{k+1} = \mathrm{Prox}_{R/(rL)} \left(y^k - \frac{1}{rL} \nabla H(\bar{y}^k) \right) \\ z^{k+1} = (1-\theta) z^k + \theta y^k + \alpha (x^{k+1} - y^k). \end{cases} \tag{1.3}$$

Under standard assumptions, the authors [23] obtained a global linear convergence.

Motivation and constributions to literature

Motivated by these developments, we propose a novel C-FISTA-inspired single-projection algorithm for stochastic variational inequalities. Our method has the following advantages:

(a) The single-projection structure reduces computational overhead, making the propsed method attractive for large-scale applications. At each iteration, the algorithm computes only one projection onto the feasible set, thereby significantly reducing computational cost compared to classical extragradient-type schemes [14], subgradient extragradient [1, 4] and cited references contained therein.

(b) The design of the correction direction combines both forward and backward information, enabling the method to effectively control stochastic errors while preserving stability.
(c) The adaptive stepsize rule ensures robustness to stochastic noise without requiring Lipschitz constants. This self-adaptive stepsize rule eliminates the need for prior knowledge of Lipschitz constants or linesearch methods that terminates at a certain predetermined criteria. Hence, making the proposed scheme computationally efficient in handling real-world problems and increases variance-controlled stochastic oracle based on minibatch sampling.
(d) Incorporation of momentum provides acceleration while maintaining theoretical guarantees, making the proposed algorithm more robust in dealing with large-scale problems.

These features distinguish the method from existing stochastic extragradient and proximal schemes. Overall, this work contributes to the growing literature on efficient stochastic methods for variational inequalities by providing a theoretically grounded and computationally efficient algorithm that balances accuracy, stability, and scalability.

Organization. The remaining part of this paper is designed as follows: In section 2, fundamental definitions and already established results in the form of lemmas are presented. The proposed algorithm and the required assumptions are carefully discussed in section 3. In the section 4, almost sure convergence and rate of convergence are rigorously derived. Applications of the proposed scheme and numerically illustration are also discussed in section 5 and conclusion of the paper is provided in section 6.

2. Preliminaries

Throughout of this paper, we shall take X to be a nonempty closed and convex subset of $\mathbb{R}^n$, $\langle , \rangle$ represents the inner product, while $\|.\|$ is the induced norm. Additionally, $i.d.d.$ denotes independent and identically distributed and $a.s$ means almost surely. We begin with essential definitions and terminologies. Given a σ-algebra $\mathcal{F}$ and a random variable ξ, we use $\mathbb{E}[\xi]$ to denote the expectation of ξ, $\mathbb{E}[\xi \mid \mathcal{F}]$ to denote the conditional expectation, $\xi \in \mathcal{F}$ to denote that ξ is $\mathcal{F}$-measurable. For $p \geq 1$, the L^p-norm of ξ is denoted by $|\xi|_p$. Let $(\Omega, \mathcal{F}, P)$ be a probability space with support Ξ, and let $\{\xi_j\}_{j=1}^N$ be an i.i.d. sample from Ξ. For $p \in [2, \infty)$, define $\sigma_p(x) = \left(\mathbb{E}\left[\|\epsilon(x,\xi)\|^p\right]\right)^{1/p}, \qquad \epsilon(x,\xi) = G(x,\xi) - F(x)$, where $G(x,\xi) = \frac{1}{N}\sum_{j=1}^N f(x,\xi_j)$.

Let X be a nonempty closed convex subset of $\mathbb{R}^n$. For any $x \in \mathbb{R}^n$, there exists an element $z \in X$ such that

$$\|z - x\| = \inf_{y \in X} \|y - x\|.$$

The element z is denoted by $\Pi_X(x)$. The mapping $\Pi_X : \mathbb{R}^n \to X$ is called a projection from $\mathbb{R}^n$ onto X.

Going forward from here, we consider the following definitions.

Definition 2.1. An operator $T : X \to X$ is said to be:

(1) monotone if

$$\langle T(x) - T(y),\, x - y\rangle \geq 0, \qquad \forall x, y \in X.$$

(2) pseudomonotone if,

$$\langle T(x),\, y - x\rangle \geq 0 \;\Rightarrow\; \langle T(y),\, y - x\rangle \geq 0, \qquad \forall x, y \in X.$$

Remark 2.2. We remark that, every monotone operator is pseudomonotone, but the converse is not true in general.

The following lemmas characterize the properties of Π_X.

Lemma 2.3. *[25] Let Π_X be the projection from $\mathbb{R}^n$ onto X. Let $x \in \mathbb{R}^n$. Then*

$$z = \Pi_X(x) \quad \Longleftrightarrow \quad \langle x - z, z - y\rangle \geq 0, \quad \forall y \in X.$$

Lemma 2.4. *[27] Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a mapping. For any $x \in \mathbb{R}^n$ and $\beta > 0$, define*

$$r_\beta(x) = x - \Pi_X(x - \beta F(x)).$$

Then

$$\min\{1, \beta\}\, \|r_1(x)\| \leq \|r_\beta(x)\| \leq \max\{1, \beta\}\, \|r_1(x)\|.$$

Lemma 2.5. *[26] Suppose that Assumption 2.1 holds. Then for any $q \in [p, 2p]$ with $p \geq 2$, there exists a constant $C_q > 0$ (depending only on q and $C_2 = 1$ if $p = q = 2$) such that for any $z, x \in \mathbb{R}^n$,*

$$\|\epsilon(z, \xi)\|^q \leq C_q \left(\frac{\sigma_q(x) + L_q \|z - x\|}{\sqrt{N}} \right),$$

where

$$L_q = \sqrt[q]{\mathbb{E}[\mathcal{L}(\xi)^q]} + L, \quad \text{with } L = \mathbb{E}[L(\xi)].$$

Lemma 2.6. *([28]) Let $\{v_k\}$, $\{\alpha_k\}$, $\{\beta_k\}$ and $\{\gamma_k\}$ be sequences of nonnegative random variables adapted to a filtration $\{\mathcal{F}_k\}$. Suppose that*

$$\mathbb{E}[v_{k+1} \mid \mathcal{F}_k] \leq (1 + \alpha_k) v_k - \beta_k + \gamma_k, \qquad \forall k \in \mathbb{N}.$$

Assume that

$$\sum_{k=0}^{\infty} \alpha_k < \infty, \qquad \sum_{k=0}^{\infty} \gamma_k < \infty \quad \text{a.s.}$$

Then $\{v_k\}$ converges almost surely to a nonnegative random variable, and

$$\sum_{k=0}^{\infty} \beta_k < \infty \quad \text{a.s.}$$

3. Assumptions and The Algorithm

Assumption 3.1. *Suppose that:*

(i) $X \subset \mathbb{R}^n$ *is nonempty, closed and convex.*

(ii) *The solution set* $X^* \neq \emptyset$.

(iii) $\langle F(y), y - x\rangle \geq 0, \qquad x \in X^*$ *and* $y \in X$.

(iv) *The mapping F is called pseudomonotone on X if for all $x, y \in X$,*

$$\langle F(x), y - x\rangle \geq 0 \implies \langle F(y), y - x\rangle \geq 0.$$

(v) *The stepsizes satisfy* $\lambda_k > 0$, $\sum_{k=0}^{\infty} \gamma_k < \infty$.

(vi) *The relaxation parameters satisfy*

$$\theta_k \in (\underline{\theta}, 1), \quad \underline{\theta} > 0.$$

(vii) *The momentum parameter satisfies*

$$0 \leq \eta \leq \frac{1}{\phi}, \quad \phi = \frac{1 + \sqrt{5}}{2}.$$

(viii) *Let $(\Omega, \mathcal{F}, P)$ be a probability space with support Ξ, and let $f(\cdot, \cdot) : \mathbb{R}^n \times \Xi \to \mathbb{R}^n$ be a measurable function. For almost every $\xi \in \Xi$,*

$$\|f(x, \xi) - f(y, \xi)\| \leq \mathcal{L}(\xi)\|x - y\|, \quad \forall x, y \in \mathbb{R}^n,$$

where $\mathcal{L} : \Xi \to [0, \infty)$ is a measurable function such that $\mathcal{L}(\xi) \geq 1$ for almost every $\xi \in \Xi$ and $\mathbb{E}\big[\mathcal{L}(\xi)^p\big] < \infty, \quad \forall p \geq 2$.

(ix) There exists $z \in \mathbb{R}^n$ such that

$$\mathbb{E}\big[\|f(z,\xi)\|^p\big] < \infty.$$

Algorithm 1 C-FISTA-Inspired Single-Projection Algorithm for Stochastic Variational Inequalities

1: Choose $x_0 = z_0 \in \mathbb{R}^n$, $\lambda_0 > 0$, $\mu \in (0,1)$, $\eta \in [0, 1/\phi]$, and $\{\theta_k\} \subset (\underline{\theta}, 1)$ with $\underline{\theta} > 0$. Choose $\{\gamma_k\} \subset (0,\infty)$ and $\{N_k\} \subset \mathbb{N}$ such that

$$\sum_{k=0}^{\infty} \gamma_k < \infty, \qquad \sum_{k=0}^{\infty} \frac{1}{N_k} < \infty.$$

2: **for** $k = 0, 1, 2, \ldots$ **do**,

3: compute the momentum point

$$s_k = (1-\eta)x_k + \eta z_k.$$

4: Draw an i.i.d. sample $\xi_k = \{\xi_{1,k}, \ldots, \xi_{N_k,k}\}$ and compute

$$G(s_k, \xi_k) = \frac{1}{N_k} \sum_{j=1}^{N_k} f(s_k, \xi_{j,k}).$$

5: Compute the single projection

$$y_k = P_X\big(s_k - \lambda_k G(s_k, \xi_k)\big).$$

6: Evaluate

$$G(y_k, \xi_k) = \frac{1}{N_k} \sum_{j=1}^{N_k} f(y_k, \xi_{j,k}),$$

and define

$$d_k = s_k - y_k - \lambda_k\big(G(s_k, \xi_k) - G(y_k, \xi_k)\big).$$

7: Compute

$$x_{k+1} = s_k - \theta_k d_k.$$

8: Compute the auxiliary sequence

$$z_{k+1} = \frac{\eta}{1+\eta} z_k + \frac{1}{1+\eta} s_k + \eta(x_{k+1} - s_k).$$

9: Update the stepsize

$$\lambda_{k+1} = \begin{cases} \min\left\{\lambda_k + \gamma_k, \ \dfrac{\mu\|s_k - y_k\|}{\|G(s_k,\xi_k) - G(y_k,\xi_k)\|}\right\}, & \text{if } G(s_k,\xi_k) \neq G(y_k,\xi_k), \\ \lambda_k + \gamma_k, & \text{otherwise.} \end{cases}$$

10: **end for**

Remark 3.2. To the best of our knowledge, Algorithm 1 is first algorithm constructed by integrating the acceleration mechanism of the C-FISTA framework for stochastic optimization. Its distinctive features can be summarized as follows:

- **Single-projection structure:** In contrast to classical stochastic extragradient-type methods (e.g., [27]), which require multiple projections per iteration, the proposed algorithm performs only one projection step, thereby significantly reducing computational cost. This technique is totally different from the recent single projection discussed in [24].

- **Inertial stabilization via momentum:** The introduction of the momentum point $s_k = (1-\eta)x_k + \eta z_k$ serves a dual purpose: it mitigates stochastic noise through averaging effects and accelerates convergence via inertial extrapolation.
- **Golden-ratio-type parameterization:** The algorithm employs a parameter regime inspired by the C-FISTA framework, where the inertial weight η and the update coefficients are chosen in connection with a golden-ratio-type constant [29]. This structure ensures a delicate balance between stability and acceleration, and plays a crucial role in guaranteeing convergence while maintaining a single-projection architecture.
- **Extragradient-type correction without extra projection:** The direction d_k is carefully designed to capture the discrepancy between stochastic operator evaluations at s_k and y_k, effectively avoiding additional projection steps.
- **Adaptive stepsize under stochastic noise:** The stepsize update rule incorporates sample-based operator differences, allowing the method to adapt to stochastic variability without requiring prior knowledge of Lipschitz constants, thus enhancing robustness and practical implementability.

Consequently, Algorithm 1 can be interpreted as a stochastic extension of the C-FISTA framework with a single-projection architecture, offering a computationally efficient and theoretically sound alternative to existing stochastic variational inequality methods.

4. Convergence Analysis

In this section, we derive almost sure convergence and convergence rate. Two theorems are stated and proved under certain assumptions. We present preliminary results as lemmas in this section which are used to establish the theorems.

4.1. Almost sure convergence.

We begin with the following preliminary results. The a.s. convergence of the iterative sequences $\{x_k, z_k\}$ generated by Algorithm 1 is established with respect to the following filtration. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We define the filtration

$$\mathcal{F}_0 := \sigma(x_0, z_0),$$

and for each $k \geq 1$,

$$\mathcal{F}_k := \sigma(x_0, z_0, \xi_0, \xi_1, \ldots, \xi_{k-1}).$$

Then, for every $k \in \mathbb{N}$,

$$x_k,\ z_k,\ s_k,\ \lambda_k \in \mathcal{F}_k.$$

Moreover, after drawing ξ_k, the quantities y_k, d_k, x_{k+1}, and z_{k+1} are $\mathcal{F}_{k+1}$-measurable.
Oracle errors. Define the oracle errors at iteration k by

$$\epsilon_{1,k} := G(s_k, \xi_k) - F(s_k), \qquad \epsilon_{2,k} := G(y_k, \xi_k) - F(y_k), \quad \forall k \in \mathbb{N}. \tag{4.1}$$

We now begin with the following lemma regarding the choice for the chosen stepsize.

Lemma 4.1. *Suppose that Assumption 3.1 holds. Let $\{\lambda_k\}$ be generated by Algorithm 1. Define*

$$\widehat{L}_k := \frac{1}{N_k} \sum_{j=1}^{N_k} L(\xi_{j,k}), \qquad k \in \mathbb{N}. \tag{4.2}$$

Then, for every $k \in \mathbb{N}$, the following hold:

$$\lambda_{k+1} \leq \lambda_k + \gamma_k, \tag{4.3}$$

$$\lambda_{k+1} \geq \min\left\{\lambda_k, \frac{\mu}{\widehat{L}_k}\right\}. \tag{4.4}$$

Consequently, there exists a nonnegative random variable λ_* such that

$$\lambda_k \to \lambda_* \qquad a.s. \tag{4.5}$$

Proof. From the stepsize update in Algorithm 1, we immediately obtain

$$\lambda_{k+1} \leq \lambda_k + \gamma_k,$$

which proves (4.3).

Next, by Assumption 3.1(viii), for each $j = 1, \ldots, N_k$,

$$\|f(s_k, \xi_{j,k}) - f(y_k, \xi_{j,k})\| \leq L(\xi_{j,k})\|s_k - y_k\|.$$

Averaging over $j = 1, \ldots, N_k$, we get

$$\|G(s_k, \xi_k) - G(y_k, \xi_k)\| \leq \frac{1}{N_k}\sum_{j=1}^{N_k} L(\xi_{j,k})\|s_k - y_k\| = \widehat{L}_k\|s_k - y_k\|.$$

If $G(s_k, \xi_k) \neq G(y_k, \xi_k)$, then the stepsize rule gives

$$\lambda_{k+1} = \min\left\{\lambda_k + \gamma_k,\ \mu\frac{\|s_k - y_k\|}{\|G(s_k, \xi_k) - G(y_k, \xi_k)\|}\right\}.$$

Since $\gamma_k > 0$, it follows that

$$\lambda_{k+1} \geq \min\left\{\lambda_k,\ \mu\frac{\|s_k - y_k\|}{\|G(s_k, \xi_k) - G(y_k, \xi_k)\|}\right\}.$$

Using the previous estimate for $\|G(s_k, \xi_k) - G(y_k, \xi_k)\|$, we obtain

$$\mu\frac{\|s_k - y_k\|}{\|G(s_k, \xi_k) - G(y_k, \xi_k)\|} \geq \frac{\mu}{\widehat{L}_k},$$

and therefore

$$\lambda_{k+1} \geq \min\left\{\lambda_k, \frac{\mu}{\widehat{L}_k}\right\}.$$

If $G(s_k, \xi_k) = G(y_k, \xi_k)$, then

$$\lambda_{k+1} = \lambda_k + \gamma_k \geq \lambda_k \geq \min\left\{\lambda_k, \frac{\mu}{\widehat{L}_k}\right\}.$$

Hence (4.4) holds in all cases.

Finally, since λ_k is $\mathcal{F}_k$-measurable, nonnegative, one obtains

$$\mathbb{E}[\lambda_{k+1} \mid \mathcal{F}_k] \leq \lambda_k + \gamma_k,$$

which is the desired result. Lemma 2.6 applies with $v_k = \lambda_k$, $\alpha_k = 0$, $\beta_k = 0$, and the same sequence γ_k. Because $\sum_{k=0}^{\infty}\gamma_k < \infty$ a.s., we conclude that $\{\lambda_k\}$ converges almost surely to a nonnegative random variable λ_*. Thus (4.5) follows. □

The next lemma forms a fundamental part in the Theorem 4.6 hypothesis and its proof.

Lemma 4.2. *Suppose that Assumption 3.1 holds. Let $\{x_k\}, \{y_k\}$ and $\{s_k\}$ be generated by Algorithm 1. Then, for each $x \in X^*$ and each $k \in \mathbb{N}$,*

$$\|x_{k+1} - x\|^2 \leq \|s_k - x\|^2 - \theta_k\left(1 - \frac{\mu^2\lambda_k^2}{\lambda_{k+1}^2}\right)\|s_k - y_k\|^2 + 2\lambda_k\|y_k - x\|\,\|\epsilon_{2,k}\|. \tag{4.6}$$

Proof. Define

$$\omega_k := y_k - \lambda_k\big(G(y_k,\xi_k) - G(s_k,\xi_k)\big). \tag{4.7}$$

Then

$$d_k = s_k - \omega_k,$$

and hence

$$x_{k+1} = s_k - \theta_k d_k = (1-\theta_k)s_k + \theta_k\omega_k.$$

Therefore,

$$\begin{aligned}\|x_{k+1}-x\|^2 &= \|(1-\theta_k)(s_k-x) + \theta_k(\omega_k - x)\|^2\\ &= (1-\theta_k)\|s_k-x\|^2 + \theta_k\|\omega_k - x\|^2 - \theta_k(1-\theta_k)\|\omega_k - s_k\|^2.\end{aligned} \tag{4.8}$$

Next, by (4.7),

$$\begin{aligned}\|\omega_k - x\|^2 &= \|y_k - x\|^2 + \lambda_k^2\|G(y_k,\xi_k) - G(s_k,\xi_k)\|^2\\ &\quad -2\lambda_k\langle y_k - x,\ G(y_k,\xi_k) - G(s_k,\xi_k)\rangle.\end{aligned} \tag{4.9}$$

Also,

$$\|y_k - x\|^2 = \|s_k - x\|^2 - \|s_k - y_k\|^2 + 2\langle y_k - s_k,\ y_k - x\rangle. \tag{4.10}$$

Substituting (4.10) into (4.9), we obtain

$$\begin{aligned}\|\omega_k - x\|^2 &= \|s_k - x\|^2 - \|s_k - y_k\|^2 + 2\langle y_k - s_k,\ y_k - x\rangle\\ &\quad + \lambda_k^2\|G(y_k,\xi_k) - G(s_k,\xi_k)\|^2 - 2\lambda_k\langle y_k - x,\ G(y_k,\xi_k) - G(s_k,\xi_k)\rangle.\end{aligned}$$

Since

$$y_k = P_X\big(s_k - \lambda_k G(s_k,\xi_k)\big),$$

Lemma 2.3 yields

$$\langle s_k - \lambda_k G(s_k,\xi_k) - y_k,\ y_k - x\rangle \geq 0,$$

that is,

$$\langle y_k - s_k,\ y_k - x\rangle \leq -\lambda_k\langle G(s_k,\xi_k),\ y_k - x\rangle.$$

Hence

$$\begin{aligned}\|\omega_k - x\|^2 &\leq \|s_k - x\|^2 - \|s_k - y_k\|^2 + \lambda_k^2\|G(y_k,\xi_k) - G(s_k,\xi_k)\|^2\\ &\quad -2\lambda_k\langle G(s_k,\xi_k),\ y_k - x\rangle - 2\lambda_k\langle y_k - x,\ G(y_k,\xi_k) - G(s_k,\xi_k)\rangle\\ &= \|s_k - x\|^2 - \|s_k - y_k\|^2 + \lambda_k^2\|G(y_k,\xi_k) - G(s_k,\xi_k)\|^2\\ &\quad -2\lambda_k\langle y_k - x,\ G(y_k,\xi_k)\rangle.\end{aligned} \tag{4.11}$$

Now, by Assumption 3.1(iii), since $x \in X^*$ and $y_k \in X$,

$$\langle F(y_k),\ y_k - x\rangle \geq 0.$$

Since

$$G(y_k,\xi_k) = F(y_k) + \epsilon_{2,k},$$

we have

$$\langle y_k - x,\ G(y_k,\xi_k)\rangle = \langle y_k - x,\ F(y_k)\rangle + \langle y_k - x,\ \epsilon_{2,k}\rangle \geq \langle y_k - x,\ \epsilon_{2,k}\rangle.$$

Therefore,

$$-2\lambda_k\langle y_k - x,\ G(y_k,\xi_k)\rangle \leq -2\lambda_k\langle y_k - x,\ \epsilon_{2,k}\rangle.$$

On the other hand, if $G(s_k,\xi_k) \neq G(y_k,\xi_k)$, then the stepsize update gives

$$\lambda_{k+1} \leq \mu\frac{\|s_k - y_k\|}{\|G(s_k,\xi_k) - G(y_k,\xi_k)\|},$$

and so

$$\lambda_k^2\|G(y_k,\xi_k)-G(s_k,\xi_k)\|^2 \le \frac{\mu^2\lambda_k^2}{\lambda_{k+1}^2}\|s_k-y_k\|^2. \tag{4.12}$$

If $G(s_k,\xi_k)=G(y_k,\xi_k)$, then (4.12) is trivial.

Combining the above estimates, we get

$$\|\omega_k-x\|^2 \le \|s_k-x\|^2-\left(1-\frac{\mu^2\lambda_k^2}{\lambda_{k+1}^2}\right)\|s_k-y_k\|^2-2\lambda_k\langle y_k-x,\ \epsilon_{2,k}\rangle.$$

Substituting this into (4.8), we obtain

$$\begin{aligned}\|x_{k+1}-x\|^2 \le \|s_k-x\|^2-\theta_k\left(1-\frac{\mu^2\lambda_k^2}{\lambda_{k+1}^2}\right)\|s_k-y_k\|^2\\ -2\theta_k\lambda_k\langle y_k-x,\ \epsilon_{2,k}\rangle-\theta_k(1-\theta_k)\|\omega_k-s_k\|^2.\end{aligned}$$

Since $0<\theta_k<1$, it follows that

$$-\theta_k(1-\theta_k)\|\omega_k-s_k\|^2\le 0.$$

Moreover, by Cauchy–Schwarz,

$$-2\theta_k\lambda_k\langle y_k-x,\ \epsilon_{2,k}\rangle \le 2\theta_k\lambda_k\|y_k-x\|\,\|\epsilon_{2,k}\| \le 2\lambda_k\|y_k-x\|\,\|\epsilon_{2,k}\|.$$

Hence

$$\|x_{k+1}-x\|^2 \le \|s_k-x\|^2-\theta_k\left(1-\frac{\mu^2\lambda_k^2}{\lambda_{k+1}^2}\right)\|s_k-y_k\|^2+2\lambda_k\|y_k-x\|\,\|\epsilon_{2,k}\|,$$

which proves Lemma 4.2. □

The next lemma deals with the stochastic error arising from Lemma 4.2.

Lemma 4.3. *Suppose that Assumption 3.1 holds. Let $\{x_k\},\{y_k\}$ and $\{s_k\}$ be generated by Algorithm 1. Then, for each $x\in X^*$ and each $k\in\mathbb{N}$, there exist a constant $m_{1,x}>0$ depending on x and a constant $m_2>0$ independent of k such that*

$$\mathbb{E}[\langle y_k-x,\epsilon_{2,k}\rangle \mid \mathcal{F}_k] \le \frac{m_{1,x}}{N_k}+\frac{m_2}{N_k}\|s_k-x\|^2, \tag{4.13}$$

where

$$\epsilon_{2,k} := G(y_k,\xi_k)-F(y_k).$$

Proof. The proof is similar to the work of [1, Lemma 9] and [4, Lemma 3.4] so we skip it. □

Lemma 4.4. *Suppose that Assumption 3.1 holds. Let $\{x_k\},\{y_k\}$ and $\{s_k\}$ be generated by Algorithm 1. Define*

$$r_1(x) := \|x-P_X(x-F(x))\|, \qquad \epsilon_{1,k} := G(s_k,\xi_k)-F(s_k).$$

Then, for each $k\in\mathbb{N}$,

$$\|s_k-y_k\|^2 \ge \frac{1}{2}\big(\min\{1,\lambda_k\}\big)^2 r_1(s_k)^2-\lambda_k^2\|\epsilon_{1,k}\|^2. \tag{4.14}$$

Proof. Since

$$y_k = P_X\big(s_k-\lambda_k G(s_k,\xi_k)\big) = P_X\big(s_k-\lambda_k(F(s_k)+\epsilon_{1,k})\big),$$

we also have

$$r_{\lambda_k}(s_k) = s_k-P_X\big(s_k-\lambda_k F(s_k)\big).$$

By Lemma 2.4,

$$\big(\min\{1,\lambda_k\}\big)^2 r_1(s_k)^2 \le \|r_{\lambda_k}(s_k)\|^2. \tag{4.15}$$

Now,

$$\begin{aligned}\|r_{\lambda_k}(s_k)\|^2 &= \|s_k - P_X(s_k - \lambda_k F(s_k))\|^2 \\ &= \|s_k - y_k + y_k - P_X(s_k - \lambda_k F(s_k))\|^2 \\ &\leq 2\|s_k - y_k\|^2 + 2\|y_k - P_X(s_k - \lambda_k F(s_k))\|^2.\end{aligned}$$

Since P_X is nonexpansive,

$$\begin{aligned}\|y_k - P_X(s_k - \lambda_k F(s_k))\| &= \|P_X(s_k - \lambda_k G(s_k, \xi_k)) - P_X(s_k - \lambda_k F(s_k))\| \\ &\leq \lambda_k \|G(s_k, \xi_k) - F(s_k)\| \\ &= \lambda_k \|\epsilon_{1,k}\|.\end{aligned}$$

Therefore,

$$\|r_{\lambda_k}(s_k)\|^2 \leq 2\|s_k - y_k\|^2 + 2\lambda_k^2 \|\epsilon_{1,k}\|^2. \tag{4.16}$$

Combining (4.15) and (4.16), we obtain

$$\big(\min\{1, \lambda_k\}\big)^2 r_1(s_k)^2 \leq 2\|s_k - y_k\|^2 + 2\lambda_k^2 \|\epsilon_{1,k}\|^2,$$

which is the desired result of Lemma 4.4. □

The next technical lemma provides an essential link between the auxiliary sequence and the main sequence introduced in the Algorithm 1 and it is a tool mostly needed for proving the first theorem of this paper.

Lemma 4.5. *Suppose that Assumption 3.1 holds. Then, for each $x \in X^*$ and each $k \in \mathbb{N}$,*

$$\|s_k - x\|^2 \leq (1 - \eta)\|x_k - x\|^2 + \eta\|z_k - x\|^2, \tag{4.17}$$

and

$$\|z_{k+1} - x\|^2 \leq \eta\|z_k - x\|^2 + \|s_k - x\|^2 + \eta(1 + \eta)\|x_{k+1} - s_k\|^2. \tag{4.18}$$

Proof. Fix $x \in X^*$.

Since

$$s_k = (1 - \eta)x_k + \eta z_k,$$

and $0 \leq \eta \leq 1/\phi < 1$, s_k is a convex combination of x_k and z_k. Hence

$$s_k - x = (1 - \eta)(x_k - x) + \eta(z_k - x),$$

and by convexity of the squared norm,

$$\|s_k - x\|^2 \leq (1 - \eta)\|x_k - x\|^2 + \eta\|z_k - x\|^2.$$

Thus (4.17) holds.

Next, from Step 8 of Algorithm 1,

$$z_{k+1} = \frac{\eta}{1 + \eta} z_k + \frac{1}{1 + \eta} s_k + \eta(x_{k+1} - s_k).$$

Subtracting x from both sides, we get

$$z_{k+1} - x = \frac{\eta}{1 + \eta}(z_k - x) + \frac{1}{1 + \eta}(s_k - x) + \eta(x_{k+1} - s_k).$$

Set

$$a_k := \frac{\eta}{1 + \eta}(z_k - x) + \frac{1}{1 + \eta}(s_k - x), \qquad b_k := \eta(x_{k+1} - s_k). \tag{4.19}$$

Then

$$z_{k+1} - x = a_k + b_k. \tag{4.20}$$

Hence

$$\|z_{k+1} - x\|^2 = \|a_k + b_k\|^2.$$

Applying Young's inequality

$$2\langle u, v\rangle \leq \eta\|u\|^2 + \eta^{-1}\|v\|^2$$

to $u = a_k$ and $v = b_k$, we obtain

$$\|z_{k+1} - x\|^2 \leq (1+\eta)\|a_k\|^2 + \left(1 + \frac{1}{\eta}\right)\|b_k\|^2. \tag{4.21}$$

Now a_k is a convex combination of $z_k - x$ and $s_k - x$, since

$$a_k = \frac{\eta}{1+\eta}(z_k - x) + \frac{1}{1+\eta}(s_k - x)$$

and

$$\frac{\eta}{1+\eta} + \frac{1}{1+\eta} = 1.$$

Therefore, by convexity of the squared norm,

$$\|a_k\|^2 \leq \frac{\eta}{1+\eta}\|z_k - x\|^2 + \frac{1}{1+\eta}\|s_k - x\|^2,$$

which implies

$$(1+\eta)\|a_k\|^2 \leq \eta\|z_k - x\|^2 + \|s_k - x\|^2. \tag{4.22}$$

Also,

$$\left(1 + \frac{1}{\eta}\right)\|b_k\|^2 = \eta(1+\eta)\|x_{k+1} - s_k\|^2. \tag{4.23}$$

Substituting (4.22) and (4.23) into (4.21), we obtain

$$\|z_{k+1} - x\|^2 \leq \eta\|z_k - x\|^2 + \|s_k - x\|^2 + \eta(1+\eta)\|x_{k+1} - s_k\|^2,$$

that is, (4.18). This completes the proof of Lemma 4.5. □

Having obtained sufficient preliminary results in this section (see, Lemmas 4.1–4.5), we now state and prove the first theorem of the paper.

Theorem 4.6. *Let $\{x_k\}, \{y_k\}, \{s_k\}$ and $\{z_k\}$ be generated by Algorithm 1, and suppose that Assumption 3.1 holds. Suppose that Lemmas 4.1–4.5 hold. Then the sequences $\{x_k\}$ and $\{z_k\}$ converge almost surely to a common point $x^\star \in X^*$.*

Proof. Fix $x \in X^*$. By Lemma 4.1, there exists $\bar{\lambda} > 0$ such that

$$\lambda_k \leq \bar{\lambda}, \qquad \forall k \in \mathbb{N}, \quad \text{a.s.} \tag{4.24}$$

Moreover, since $\lambda_k \to \lambda_*$ a.s., and $\lambda_* > 0$ a.s., for any $\bar{\mu} \in (\mu^2, 1)$ there exist $\hat{\lambda} > 0$ and a random integer K_0 such that, for all $k \geq K_0$,

$$\lambda_k \geq \hat{\lambda}, \qquad 1 - \frac{\mu^2\lambda_k^2}{\lambda_{k+1}^2} \geq 1 - \bar{\mu}, \qquad \text{a.s.} \tag{4.25}$$

Next, rewrite the z-update as

$$z_{k+1} = \frac{\eta}{1+\eta}z_k + \frac{1-\eta-\eta^2}{1+\eta}s_k + \eta x_{k+1}.$$

Since $0 \leq \eta \leq 1/\phi$, we have $1 - \eta - \eta^2 \geq 0$, and

$$\frac{\eta}{1+\eta} + \frac{1-\eta-\eta^2}{1+\eta} + \eta = 1.$$

Hence, by convexity of the squared norm,

$$\|z_{k+1} - x\|^2 \leq \frac{\eta}{1+\eta}\|z_k - x\|^2 + \frac{1-\eta-\eta^2}{1+\eta}\|s_k - x\|^2 + \eta\|x_{k+1} - x\|^2. \tag{4.26}$$

Let

$$a := \frac{\eta(1+\eta)}{1-\eta}, \qquad V_k := \|x_k - x\|^2 + a\|z_k - x\|^2. \tag{4.27}$$

Then

$$a\frac{\eta}{1+\eta} = \frac{\eta^2}{1-\eta}, \qquad 1 + a\eta + a\frac{1-\eta-\eta^2}{1+\eta} = \frac{1}{1-\eta}. \tag{4.28}$$

Now, by Lemma 4.2, Lemma 4.3, Lemma 4.4, and (4.25), for all $k \geq K_0$,

$$\begin{aligned}\mathbb{E}\big[\|x_{k+1} - x\|^2 \mid \mathcal{F}_k\big] &\leq \|s_k - x\|^2 - \theta_k\left(1 - \frac{\mu^2\lambda_k^2}{\lambda_{k+1}^2}\right)\|s_k - y_k\|^2 + \frac{m_{1,x}}{N_k} + \frac{m_2}{N_k}\|s_k - x\|^2 \\ &\leq \left(1 + \frac{m_2}{N_k}\right)\|s_k - x\|^2 - \theta(1-\bar{\mu})\|s_k - y_k\|^2 + \frac{m_{1,x}}{N_k}.\end{aligned} \tag{4.29}$$

On the other hand, by Lemma 4.4 and (4.24)–(4.25),

$$\|s_k - y_k\|^2 \geq \frac{1}{2}\big(\min\{1, \hat{\lambda}\}\big)^2 r_1(s_k)^2 - \bar{\lambda}^2\|\epsilon_{1,k}\|^2.$$

Substituting the estimate from Lemma 4.4 into (4.29), we obtain

$$\begin{aligned}\mathbb{E}\big[\|x_{k+1} - x\|^2 \mid \mathcal{F}_k\big] \leq \left(1 + \frac{m_2}{N_k}\right)\|s_k - x\|^2 - \frac{\theta(1-\bar{\mu})}{2}\big(\min\{1, \hat{\lambda}\}\big)^2 r_1(s_k)^2 \\ + \theta(1-\bar{\mu})\bar{\lambda}^2\mathbb{E}\big[\|\epsilon_{1,k}\|^2 \mid \mathcal{F}_k\big] + \frac{m_{1,x}}{N_k}.\end{aligned}$$

By Lemma 2.5 with $q = 2$,

$$\mathbb{E}\big[\|\epsilon_{1,k}\|^2 \mid \mathcal{F}_k\big] \leq \frac{2\sigma_2(x)^2 + 2L_2^2\|s_k - x\|^2}{N_k}.$$

Hence

$$\mathbb{E}\big[\|x_{k+1} - x\|^2 \mid \mathcal{F}_k\big] \leq \left(1 + \frac{c_1}{N_k}\right)\|s_k - x\|^2 - c_2 r_1(s_k)^2 + \frac{c_{3,x}}{N_k}, \tag{4.30}$$

where

- $c_1 := m_2 + 2\theta(1-\bar{\mu})\bar{\lambda}^2 L_2^2$,
- $c_2 := \frac{\theta(1-\bar{\mu})}{2}\big(\min\{1, \hat{\lambda}\}\big)^2$, and
- $c_{3,x} := m_{1,x} + 2\theta(1-\bar{\mu})\bar{\lambda}^2\sigma_2(x)^2$.

Multiplying (4.26) by a and adding to (4.30), we obtain

$$\begin{aligned}\mathbb{E}[V_{k+1} \mid \mathcal{F}_k] &\leq (1 + a\eta)\mathbb{E}\big[\|x_{k+1} - x\|^2 \mid \mathcal{F}_k\big] + a\frac{1-\eta-\eta^2}{1+\eta}\|s_k - x\|^2 + a\frac{\eta}{1+\eta}\|z_k - x\|^2 \\ &\leq \left(1 + a\eta + a\frac{1-\eta-\eta^2}{1+\eta}\right)\|s_k - x\|^2 + a\frac{\eta}{1+\eta}\|z_k - x\|^2 \\ &\quad + (1+a\eta)\frac{c_1}{N_k}\|s_k - x\|^2 - (1+a\eta)c_2 r_1(s_k)^2 + (1+a\eta)\frac{c_{3,x}}{N_k}.\end{aligned}$$

Substituting (4.30) into the preceding inequality, we obtain

$$\begin{aligned}\mathbb{E}[V_{k+1} \mid \mathcal{F}_k] \leq (1+a\eta)\left(1 + \frac{c_1}{N_k}\right)\|s_k - x\|^2 - (1+a\eta)c_2 r_1(s_k)^2 \\ + (1+a\eta)\frac{c_{3,x}}{N_k} + a\frac{1-\eta-\eta^2}{1+\eta}\|s_k - x\|^2 + a\frac{\eta}{1+\eta}\|z_k - x\|^2.\end{aligned}$$

Using

$$1 + a\eta + a\frac{1-\eta-\eta^2}{1+\eta} = \frac{1}{1-\eta}, \qquad a\frac{\eta}{1+\eta} = \frac{\eta^2}{1-\eta},$$

it follows that

$$\begin{aligned}\mathbb{E}[V_{k+1} \mid \mathcal{F}_k] \leq{}& \frac{1}{1-\eta}\|s_k - x\|^2 + \frac{\eta^2}{1-\eta}\|z_k - x\|^2 \\ &+ \frac{(1+a\eta)\big(m_2 + 2\theta(1-\bar{\mu})\bar{\lambda}^2 L_2^2\big)}{N_k}\|s_k - x\|^2 \\ &- (1+a\eta)\frac{\theta(1-\bar{\mu})}{2}\big(\min\{1,\hat{\lambda}\}\big)^2 r_1(s_k)^2 \\ &+ \frac{(1+a\eta)\big(m_{1,x} + 2\theta(1-\bar{\mu})\bar{\lambda}^2\sigma_2(x)^2\big)}{N_k}.\end{aligned} \tag{4.31}$$

Now Lemma 4.5 gives

$$\|s_k - x\|^2 \leq (1-\eta)\|x_k - x\|^2 + \eta\|z_k - x\|^2.$$

Therefore,

$$\frac{1}{1-\eta}\|s_k - x\|^2 + \frac{\eta^2}{1-\eta}\|z_k - x\|^2 \leq \|x_k - x\|^2 + \frac{\eta(1+\eta)}{1-\eta}\|z_k - x\|^2 = V_k.$$

Since also $\|s_k - x\|^2 \leq V_k$, we deduce from (4.31) that

$$\mathbb{E}[V_{k+1} \mid \mathcal{F}_k] \leq \left(1 + \frac{c_4}{N_k}\right) V_k - c_5 r_1(s_k)^2 + \frac{c_{6,x}}{N_k}, \qquad \forall k \geq K_0, \tag{4.32}$$

where

- $c_4 := (1+a\eta)\big(m_2 + 2\theta(1-\bar{\mu})\bar{\lambda}^2 L_2^2\big)$.
- $c_5 := (1+a\eta)\frac{\theta(1-\bar{\mu})}{2}\big(\min\{1,\hat{\lambda}\}\big)^2$.
- $c_{6,x} := (1+a\eta)\big(m_{1,x} + 2\theta(1-\bar{\mu})\bar{\lambda}^2\sigma_2(x)^2\big)$.

Because $\sum_{k=0}^{\infty} N_k^{-1} < \infty$, Lemma 2.6 applied to (4.32) yields that $\{V_k\}$ converges almost surely and

$$\sum_{k=0}^{\infty} r_1(s_k)^2 < \infty \qquad \text{a.s.} \tag{4.33}$$

Hence

$$r_1(s_k) \to 0 \qquad \text{a.s.} \tag{4.34}$$

Since $\{V_k\}$ converges a.s., the sequences $\{x_k\}$ and $\{z_k\}$ are almost surely bounded, and so is $\{s_k\}$. Let ω belong to a full-measure set on which (4.34) holds and $\{s_k(\omega)\}$ is bounded. Then there exists a subsequence $\{s_{k_j}(\omega)\}$ converging to some $\bar{x}$. By continuity of r_1, (4.34) implies

$$r_1(\bar{x}) = 0,$$

hence

$$\bar{x} = P_X(\bar{x} - F(\bar{x})),$$

which is equivalent to $\bar{x} \in X^*$ by Lemma 2.3.

It remains to show that $\{x_k\}$ and $\{z_k\}$ have the same cluster points as $\{s_k\}$.

Since

$$y_k = P_X\big(s_k - \lambda_k G(s_k, \xi_k)\big),$$

we have

$$\|s_k - y_k\| \leq \big\|s_k - P_X(s_k - \lambda_k F(s_k))\big\| + \lambda_k\|\epsilon_{1,k}\|.$$

By Lemma 2.4,

$$\big\|s_k - P_X(s_k - \lambda_k F(s_k))\big\| \leq \max\{1, \lambda_k\}\, r_1(s_k).$$

Hence

$$\|s_k - y_k\| \leq \max\{1, \lambda_k\}\, r_1(s_k) + \lambda_k\|\epsilon_{1,k}\|.$$

Therefore, using (4.34), the boundedness of $\{\lambda_k\}$, and the fact from Lemma 4.2 and 4.3, we obtain

$$\|s_k - y_k\| \to 0 \qquad \text{a.s.}$$

Since

$$d_k = s_k - y_k - \lambda_k\big(G(s_k,\xi_k) - G(y_k,\xi_k)\big),$$

we have

$$\|d_k\| \le \|s_k - y_k\| + \lambda_k\|G(s_k,\xi_k) - G(y_k,\xi_k)\|.$$

By Assumption 3.1(viii),

$$\|G(s_k,\xi_k) - G(y_k,\xi_k)\| \le \widehat{L}_k\|s_k - y_k\|, \qquad \widehat{L}_k := \frac{1}{N_k}\sum_{j=1}^{N_k} L(\xi_{j,k}), \qquad k \in \mathbb{N}.$$

Hence

$$\|d_k\| \le \big(1 + \lambda_k\widehat{L}_k\big)\|s_k - y_k\|.$$

Using (4.24)–(4.25), we have $\lambda_k \le \bar{\lambda}$ and $\widehat{L}_k \le \mu/\hat{\lambda}$ for all $k \ge K_0$, and therefore

$$\|d_k\| \le \left(1 + \frac{\mu\bar{\lambda}}{\hat{\lambda}}\right)\|s_k - y_k\|, \qquad \forall k \ge K_0, \quad \text{a.s.}$$

Thus one may take

$$c_7 := 1 + \frac{\mu\bar{\lambda}}{\hat{\lambda}},$$

so that

$$\|d_k\| \le c_7\|s_k - y_k\|, \qquad \forall k \ge K_0, \quad \text{a.s.}$$

Therefore, we get from the Algorithm 1 that

$$\|x_{k+1} - s_k\| = \theta_k\|d_k\| \to 0 \qquad \text{a.s.} \tag{4.35}$$

Also, from the z-update,

$$z_k - s_{k-1} = \frac{\eta}{1+\eta}(z_{k-1} - s_{k-1}) + \eta(x_k - s_{k-1}),$$

and (4.35) gives $x_k - s_{k-1} \to 0$ a.s. Since $\eta/(1+\eta) \in [0,1)$, it follows that

$$z_k - s_{k-1} \to 0 \qquad \text{a.s.} \tag{4.36}$$

Consequently,

$$x_k - z_k = (x_k - s_{k-1}) - (z_k - s_{k-1}) \to 0 \qquad \text{a.s.}$$

Hence

$$x_k - s_k = \eta(x_k - z_k) \to 0, \qquad z_k - s_k = (1-\eta)(z_k - x_k) \to 0 \qquad \text{a.s.} \tag{4.37}$$

Therefore, every cluster point of $\{x_k\}$ and $\{z_k\}$ is also a cluster point of $\{s_k\}$, hence belongs to X^*. Let $x^\star$ be one such cluster point. Since $V_k(x^\star) = \|x_k - x^\star\|^2 + a\|z_k - x^\star\|^2$ converges a.s., and along a subsequence converging to $x^\star$ it tends to 0, we conclude that

$$V_k(x^\star) \to 0 \qquad \text{a.s.}$$

Thus

$$x_k \to x^\star, \qquad z_k \to x^\star \qquad \text{a.s.}$$

Since

$$s_k = (1-\eta)x_k + \eta z_k,$$

it follows that

$$s_k \to x^\star \qquad \text{a.s.}$$

Moreover, from $\|s_k - y_k\| \to 0$ a.s., we have

$$\|y_k - x^\star\| \leq \|y_k - s_k\| + \|s_k - x^\star\| \to 0 \qquad \text{a.s.}$$

Hence

$$y_k \to x^\star \qquad \text{a.s.}$$

This completes the proof of Theorem 4.6. □

4.2. **Rate of convergence.** In this section, we discuss the convergence rate of the proposed method. To achieve this, we introduce a new assumption.

Assumption 4.7. *The cost function $F : X \to \mathcal{H}$ is said to be strongly pseudomonotone on X if there exists a constant $\delta > 0$ such that*

$$\langle F(y),\, y - x\rangle \geq \delta\|y - x\|^2, \qquad \forall x \in X^*, y \in X. \tag{4.38}$$

The constant δ is called the modulus of strong pseudomonotonicity.

Now, let $\{x_k\}$ be a sequence in $\mathbb{R}^n$. Then

(i) The sequence $\{x_k\}$ is said to converge $R-$linearly to x^* with rate $\rho \in [, 0, 1)$ if there exists a constant $c > 0$ such that $\|x_k - x^*\| \leq c\rho^k, \qquad \forall n \in \mathbb{N}$.

(ii) $\{x_k\}$ is said to converge $Q-$linearly to x^* with rate $\rho \in [0, 1)$ if $\|x_{k+1} - x^*\| \leq \rho\|x_k - x^*\|, \qquad \forall n \in \mathbb{N}$.

Theorem 4.8. *Suppose that Assumption 3.1 and Assumption 4.7 hold. Let $\{x_k\}, \{y_k\}, \{s_k\}$ and $\{z_k\}$ be generated by Algorithm 1. Let $\bar{\mu} \in (\mu^2, 1)$ and choose $\hat{\lambda} > 0$ such that, almost surely, $\lambda_k \geq \hat{\lambda}$ for all sufficiently large k. Define*

$$\Delta := \min\left\{\frac{1 - \bar{\mu}}{2},\, \hat{\lambda}\delta\right\}. \tag{4.39}$$

Assume that there exist a constant $\rho \in (0, 1)$ and an integer $K \in \mathbb{N}$ such that

$$\tau_k := \frac{1 - \theta_k\Delta}{1 - \eta} + \frac{\eta^2}{1 - \eta} < \rho, \qquad \forall k \geq K. \tag{4.40}$$

If, in addition,

$$\sum_{i=0}^{\infty} \frac{1}{\rho^i N_{K+i}} < \infty, \tag{4.41}$$

then the sequences $\{x_k\}$ and $\{z_k\}$ converge almost surely and in mean square to the same point $x^\star \in X^*$, and there exists a constant $M_x > 0$ such that

$$\mathbb{E}\big[V_{k+1}(x^\star)\big] \leq \rho^{k-K} M_x, \qquad \forall k \geq K, \tag{4.42}$$

where

$$V_k(x) := \|x_k - x\|^2 + \frac{\eta(1 + \eta)}{1 - \eta}\|z_k - x\|^2. \tag{4.43}$$

Proof. Let $x^\star \in X^*$ be the almost sure limit given by Theorem 4.6. Define

$$a := \frac{\eta(1 + \eta)}{1 - \eta}, \qquad V_k := V_k(x^\star) := \|x_k - x^\star\|^2 + a\|z_k - x^\star\|^2. \tag{4.44}$$

By Lemma 4.1, there exists $\bar{\lambda} > 0$ such that

$$\lambda_k \leq \bar{\lambda}, \qquad \forall k \in \mathbb{N}, \quad \text{a.s.} \tag{4.45}$$

Moreover, since $\lambda_k \to \lambda_*$ a.s. and $\lambda_* > 0$ a.s., for each $\bar{\mu} \in (\mu^2, 1)$ there exists an integer $K_0 \in \mathbb{N}$ such that

$$\lambda_k \geq \hat{\lambda}, \qquad 1 - \frac{\mu^2 \lambda_k^2}{\lambda_{k+1}^2} \geq 1 - \bar{\mu}, \qquad \forall k \geq K_0, \quad \text{a.s.} \tag{4.46}$$

We now derive the key estimate. From the proof of Lemma 4.2, for each $k \geq K_0$,

$$\|\omega_k - x^\star\|^2 \leq \|s_k - x^\star\|^2 - \left(1 - \frac{\mu^2 \lambda_k^2}{\lambda_{k+1}^2}\right) \|s_k - y_k\|^2 - 2\lambda_k \langle y_k - x^\star, G(y_k, \xi_k)\rangle, \tag{4.47}$$

where

$$\omega_k := y_k - \lambda_k \big(G(y_k, \xi_k) - G(s_k, \xi_k)\big).$$

Since

$$G(y_k, \xi_k) = F(y_k) + \epsilon_{2,k},$$

we have

$$\langle y_k - x^\star, G(y_k, \xi_k)\rangle = \langle y_k - x^\star, F(y_k)\rangle + \langle y_k - x^\star, \epsilon_{2,k}\rangle.$$

By Assumption 4.7,

$$\langle F(y_k), y_k - x^\star\rangle \geq \delta \|y_k - x^\star\|^2,$$

and hence

$$-2\lambda_k \langle y_k - x^\star, G(y_k, \xi_k)\rangle \leq -2\lambda_k \delta \|y_k - x^\star\|^2 - 2\lambda_k \langle y_k - x^\star, \epsilon_{2,k}\rangle. \tag{4.48}$$

Substituting (4.48) into (4.47) and using (4.46), we obtain

$$\|\omega_k - x^\star\|^2 \leq \|s_k - x^\star\|^2 - (1 - \bar{\mu})\|s_k - y_k\|^2 - 2\hat{\lambda}\delta \|y_k - x^\star\|^2 - 2\lambda_k \langle y_k - x^\star, \epsilon_{2,k}\rangle. \tag{4.49}$$

Now define

$$\Delta := \min\left\{\frac{1 - \bar{\mu}}{2},\ \hat{\lambda}\delta\right\}. \tag{4.50}$$

Then

$$1 - \bar{\mu} \geq 2\Delta, \qquad 2\hat{\lambda}\delta \geq 2\Delta,$$

so (4.49) gives

$$\|\omega_k - x^\star\|^2 \leq \|s_k - x^\star\|^2 - 2\Delta \|s_k - y_k\|^2 - 2\Delta \|y_k - x^\star\|^2 - 2\lambda_k \langle y_k - x^\star, \epsilon_{2,k}\rangle. \tag{4.51}$$

Since

$$s_k - x^\star = (s_k - y_k) + (y_k - x^\star),$$

we have

$$\|s_k - x^\star\|^2 \leq 2\|s_k - y_k\|^2 + 2\|y_k - x^\star\|^2,$$

that is,

$$\|s_k - y_k\|^2 + \|y_k - x^\star\|^2 \geq \frac{1}{2}\|s_k - x^\star\|^2. \tag{4.52}$$

Combining (4.51) and (4.52), we get

$$\|\omega_k - x^\star\|^2 \leq (1 - \Delta)\|s_k - x^\star\|^2 - 2\lambda_k \langle y_k - x^\star, \epsilon_{2,k}\rangle. \tag{4.53}$$

Next, since

$$x_{k+1} = (1 - \theta_k) s_k + \theta_k \omega_k,$$

we have

$$\|x_{k+1} - x^\star\|^2 = (1 - \theta_k)\|s_k - x^\star\|^2 + \theta_k \|\omega_k - x^\star\|^2 - \theta_k(1 - \theta_k)\|\omega_k - s_k\|^2.$$

Dropping the last nonpositive term and using (4.53), we obtain

$$\|x_{k+1} - x^\star\|^2 \leq (1 - \theta_k \Delta)\|s_k - x^\star\|^2 - 2\theta_k \lambda_k \langle y_k - x^\star, \epsilon_{2,k}\rangle, \qquad \forall k \geq K_0. \tag{4.54}$$

Taking conditional expectation in (4.54) and using Lemma 4.3, together with

$$0 < \theta_k < 1, \qquad \lambda_k \leq \bar{\lambda},$$

we get, for all $k \geq K_0$,

$$\begin{aligned}\mathbb{E}\left[\|x_{k+1} - x^\star\|^2 \,\middle|\, \mathcal{F}_k\right] &\leq (1 - \theta_k\Delta)\|s_k - x^\star\|^2 + 2\bar{\lambda}\,\mathbb{E}[\langle y_k - x^\star, \epsilon_{2,k}\rangle \mid \mathcal{F}_k] \\ &\leq \left(1 - \theta_k\Delta + \frac{2\bar{\lambda}m_2}{N_k}\right)\|s_k - x^\star\|^2 + \frac{2\bar{\lambda}m_{1,x^\star}}{N_k}.\end{aligned} \tag{4.55}$$

On the other hand, from the z-update,

$$z_{k+1} = \frac{\eta}{1+\eta}z_k + \frac{1-\eta-\eta^2}{1+\eta}s_k + \eta x_{k+1}.$$

Since $0 \leq \eta \leq 1/\phi$, we have $1 - \eta - \eta^2 \geq 0$, and

$$\frac{\eta}{1+\eta} + \frac{1-\eta-\eta^2}{1+\eta} + \eta = 1.$$

Hence, by convexity of the squared norm,

$$\|z_{k+1} - x^\star\|^2 \leq \frac{\eta}{1+\eta}\|z_k - x^\star\|^2 + \frac{1-\eta-\eta^2}{1+\eta}\|s_k - x^\star\|^2 + \eta\|x_{k+1} - x^\star\|^2. \tag{4.56}$$

Multiplying (4.56) by a and adding to (4.55), we obtain

$$\mathbb{E}[V_{k+1} \mid \mathcal{F}_k] \leq (1 + a\eta)\mathbb{E}\left[\|x_{k+1} - x^\star\|^2 \,\middle|\, \mathcal{F}_k\right] + a\frac{1-\eta-\eta^2}{1+\eta}\|s_k - x^\star\|^2 + a\frac{\eta}{1+\eta}\|z_k - x^\star\|^2. \tag{4.57}$$

Using

$$a\frac{\eta}{1+\eta} = \frac{\eta^2}{1-\eta}, \qquad 1 + a\eta + a\frac{1-\eta-\eta^2}{1+\eta} = \frac{1}{1-\eta}, \tag{4.58}$$

Substituting (4.55) into (4.57), we obtain

$$\begin{aligned}\mathbb{E}[V_{k+1} \mid \mathcal{F}_k] \leq{}& (1 + a\eta)\left(1 - \theta_k\Delta + \frac{2\bar{\lambda}m_2}{N_k}\right)\|s_k - x^\star\|^2 \\ &+ (1 + a\eta)\frac{2\bar{\lambda}m_{1,x^\star}}{N_k} + a\frac{1-\eta-\eta^2}{1+\eta}\|s_k - x^\star\|^2 + a\frac{\eta}{1+\eta}\|z_k - x^\star\|^2.\end{aligned}$$

Using

$$1 + a\eta + a\frac{1-\eta-\eta^2}{1+\eta} = \frac{1}{1-\eta}, \qquad a\frac{\eta}{1+\eta} = \frac{\eta^2}{1-\eta},$$

it follows that

$$\mathbb{E}[V_{k+1} \mid \mathcal{F}_k] \leq \left(\frac{1-\theta_k\Delta}{1-\eta} + \frac{\eta^2}{1-\eta}\right)\|s_k - x^\star\|^2 + \frac{\eta^2}{1-\eta}\|z_k - x^\star\|^2 + \frac{k_1}{N_k}\|s_k - x^\star\|^2 + \frac{k_2}{N_k}, \tag{4.59}$$

where

$$k_1 := (1 + a\eta)\,2\bar{\lambda}m_2, \qquad k_2 := (1 + a\eta)\,2\bar{\lambda}m_{1,x^\star}.$$

Now Lemma 4.5 gives

$$\|s_k - x^\star\|^2 \leq (1-\eta)\|x_k - x^\star\|^2 + \eta\|z_k - x^\star\|^2. \tag{4.60}$$

Using (4.60) and the definition of τ_k, we obtain, for all $k \geq K$,

$$\mathbb{E}[V_{k+1} \mid \mathcal{F}_k] \leq \tau_k V_k + \frac{k_1}{N_k}V_k + \frac{k_2}{N_k} \leq \rho V_k + \frac{k_1}{N_k}V_k + \frac{k_2}{N_k}. \tag{4.61}$$

Since Theorem 4.6 yields $V_k \to 0$ a.s., the sequence $\{V_k\}$ is almost surely bounded. Thus there exists a finite random variable Γ such that

$$V_k \le \Gamma, \qquad \forall k \in \mathbb{N}, \quad \text{a.s.}$$

Taking expectation in (4.61), we obtain

$$\mathbb{E}[V_{k+1}] \le \rho\,\mathbb{E}[V_k] + \frac{k_1}{N_k}\mathbb{E}[V_k] + \frac{k_2}{N_k}, \qquad \forall k \ge K.$$

Now set

$$M_V := \sup_{k \ge K} \mathbb{E}[V_k].$$

Since $\{V_k\}$ is nonnegative and convergent almost surely, it is bounded almost surely; moreover, under the standing moment assumptions we may assume that $M_V < \infty$. Therefore,

$$\mathbb{E}[V_{k+1}] \le \rho\,\mathbb{E}[V_k] + \frac{k_1 M_V + k_2}{N_k}, \qquad \forall k \ge K.$$

Thus, by defining

$$k_3 := k_1 M_V + k_2,$$

we arrive at

$$\mathbb{E}[V_{k+1}] \le \rho, \mathbb{E}[V_k] + \frac{k_3}{N_k}, \qquad \forall k \ge K \tag{4.62}$$

Applying (4.62) to $\mathbb{E}[V_k]$, we obtain

$$\mathbb{E}[V_k] \le \rho\mathbb{E}[V_{k-1}] + \frac{k_3}{N_{k-1}}. \tag{4.63}$$

Substituting (4.63) into (4.62) yields

$$\mathbb{E}[V_{k+1}] \le \rho\left(\rho\mathbb{E}[V_{k-1}] + \frac{k_3}{N_{k-1}}\right) + \frac{k_3}{N_k} \quad = \rho^2\mathbb{E}[V_{k-1}] + \rho\frac{k_3}{N_{k-1}} + \frac{k_3}{N_k}. \tag{4.64}$$

Applying the same argument once more, we get

$$\mathbb{E}[V_{k+1}] \le \rho^3\mathbb{E}[V_{k-2}] + \rho^2\frac{k_3}{N_{k-2}} + \rho\frac{k_3}{N_{k-1}} + \frac{k_3}{N_k}. \tag{4.65}$$

Continuing this recursion inductively up to the index K, we obtain

$$\mathbb{E}[V_{k+1}] \le \rho^{k-K}\mathbb{E}[V_K] + k_3\sum_{i=0}^{k-K}\rho^i\frac{1}{N_{k-i}}. \tag{4.66}$$

Re-indexing the sum,

$$\sum_{i=0}^{k-K}\rho^i\frac{1}{N_{k-i}} = \rho^{k-K}\sum_{i=0}^{k-K}\frac{1}{\rho^i N_{K+i}}.$$

It follows that

$$\mathbb{E}[V_{k+1}] \le \rho^{k-K}M_{x^\star}, \qquad \forall k \ge K, \qquad M_{x^\star} > 0. \tag{4.67}$$

Finally, since $V_k \to 0$ a.s. by Theorem 4.6, it follows that

$$x_k \to x^\star, \qquad z_k \to x^\star \qquad \text{a.s.}$$

Hence the sequences $\{x_k\}$ and $\{z_k\}$ converge almost surely and in mean square to the same point $x^\star \in X^*$. This completes the proof. □

TABLE 1. Numerical parameter settings for the stochastic complementarity problem

Algorithm	Assigned Parameter Values
Proposed	$\lambda_0 = 0.005, \quad \mu = 0.5, \quad \eta = \frac{1}{2\varphi}, \quad \varphi = \frac{1+\sqrt{5}}{2},$ $\theta_k = 0.9 + \frac{1}{50k}, \quad \gamma_k = \frac{1}{k^2}, \quad N_k = \lfloor 0.99^{1-k} \rfloor.$
SDIM	$\lambda_1 = 0.005, \quad \mu = 0.5, \quad \alpha = 0.01, \quad \beta = 1,$ $\theta_k = 0.9 + \frac{1}{50k}, \quad \zeta_k = \chi_k = \frac{50}{(k+100)^2},$ $\gamma_k = \frac{1}{k^2}, \quad N_k = \lfloor 0.99^{1-k} \rfloor.$
NFPM	$\lambda_1 = 0.005, \quad \mu = 0.5, \quad \tau = \frac{\sqrt{2}}{4}, \quad \gamma = 1, \quad \phi = \frac{1+\sqrt{5}}{2},$ $\tau_k = 0.5 + \frac{1}{k}, \quad p_k = \frac{100}{k}, \quad q_k = \frac{1}{k^2}, \quad \gamma_k = 0.1 + \frac{1}{k},$ $N_k = \lfloor 0.99^{1-k} \rfloor.$
SASSEM	$\lambda_0 = 0.1, \quad \tau = 0.1, \quad N_k = \lfloor 0.99^{1-k} \rfloor.$

5. Applications and Numerical Illustratons

In this section, we apply the proposed algorithm to solve some real-word problems and carry out numerical experiments to justify the superiority of our algorithm. To be precise, the performance Algorithm 1 will be compared with [3, Algorithm 3.1] which will be denoted with the code SDIM, [2, Algorithm 3.1] with the code NFPM, [4, Algorithm 3.1] with the code SASSEM. Experiments are conducted on a 64-bit Windows platform equipped with an Intel(R) Core(TM) i7-6600U CPU @ 2.60GHz (2 cores, 4 threads) and 8 GB of RAM. All numerical computations, data analysis, and visualizations are implemented in a Python 3.9 environment using standard scientific libraries, including NumPy, SciPy, Pandas, Scikit-learn, and Matplotlib.

Example 5.1. *([2, 3]) Consider the stochastic complementarity problem defined by*

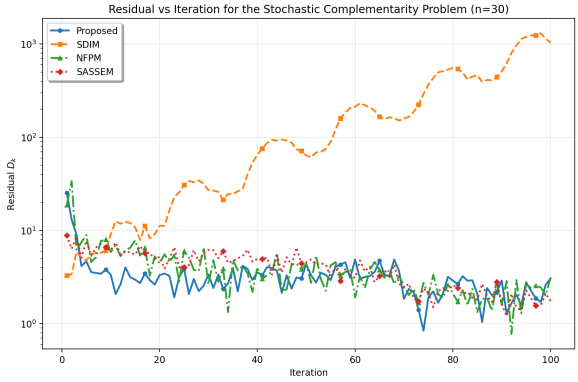
$$x \in \mathbb{R}^n_+, \qquad \mathbb{E}[f(x,\xi)] \geq 0, \qquad x^\top \mathbb{E}[f(x,\xi)] = 0,$$

where the mapping $f : \mathbb{R}^n \times \Omega \to \mathbb{R}^n$ is given by

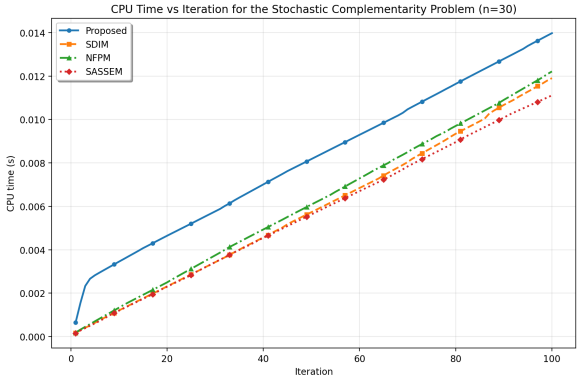
$$f(x,\xi) = D(x,\xi) + M(\xi)x + q(\xi), \qquad X = \mathbb{R}^n.$$

The components are defined as follows: The nonlinear term $D(x,\xi)$ is defined by $D(x,\xi) = \left(d_i(\xi)\arctan(a_i(\xi)x_i)\right)_{i=1}^n$, where $d(\xi)$ and $a(\xi)$ are random vectors whose components are uniformly distributed on $(0,1)$. The matrix $M(\xi)$ is defined by $M(\xi) = B + Y(\xi)$, where $B \in \mathbb{R}^{n\times n}$ is a skew-symmetric matrix whose entries are generated from the uniform distribution on $(0,3)$, and $Y(\xi)$ is a diagonal matrix whose entries are generated from the uniform distribution on $(0,2)$. The vector $q(\xi) \in \mathbb{R}^n$ has components generated from the uniform distribution on $(-2,2)$. For this example, we choose the parameters $\mu = 0.5$, $\gamma_k = \frac{1}{k^2}$. The initial points are taken as $x_0 = e$, $x_1 = 0.1e$, where $e = (1,1,1,...,1)^T \in \mathbb{R}^n$, for Algorithm 1 (or Proposed), while for SDIM, NFPM, and SASSEM, we take $x_0 = e$. The stopping criterion for all algorithms is the maximum number of iterations $k = 100$. The residual used for performance evaluation is defined by $D_k = \|x_k - \Pi_X(x_k - G(x_k,\xi_k))\|$.

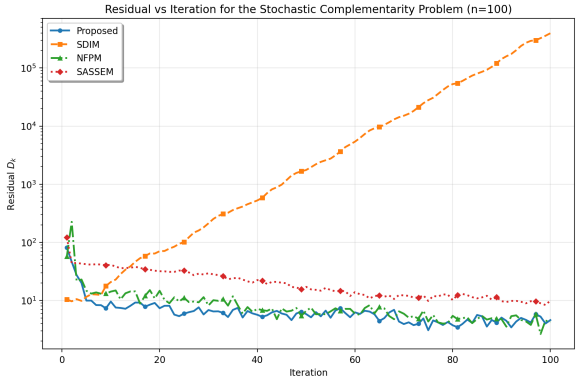
The results in Figure 1 show consistent decay of the residual across all dimensions, confirming convergence of the four methods. The **Proposed** algorithm exhibits faster reduction in D_k and smoother



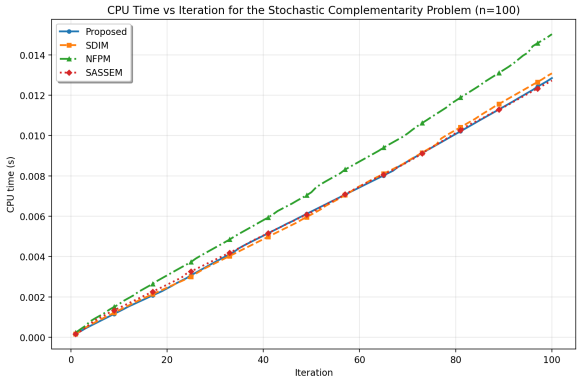
(A) D_k vs iteration ($n = 30$)



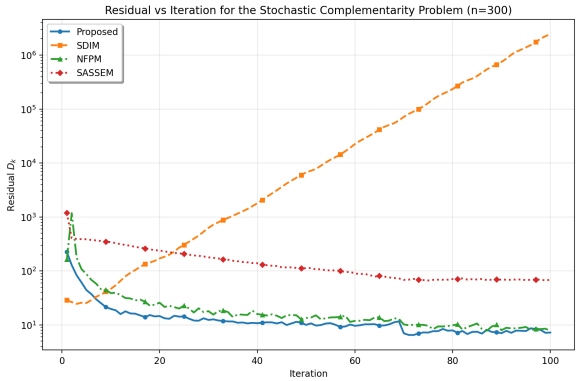
(B) CPU time vs iteration ($n = 30$)



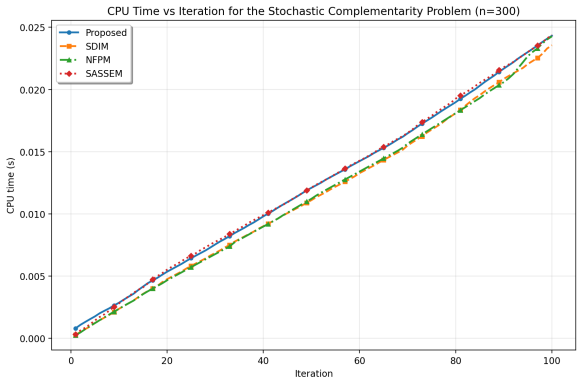
(C) D_k vs iteration ($n = 100$)



(D) CPU time vs iteration ($n = 100$)



(E) D_k vs iteration ($n = 300$)

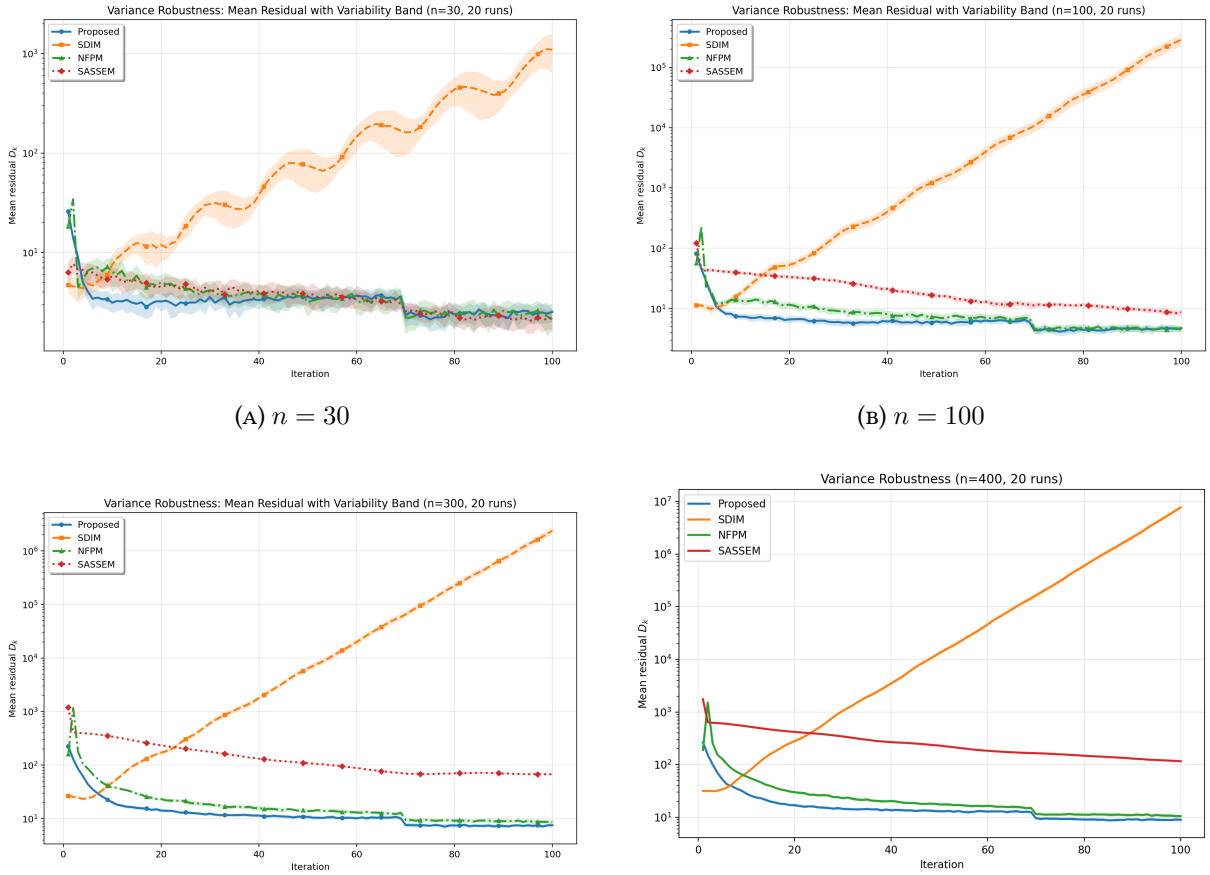


(F) CPU time vs iteration ($n = 300$)

FIGURE 1. Performance comparison of the algorithms for different problem dimensions

trajectories, indicating improved stability under stochastic noise. As the dimension increases, all methods incur higher computational cost; however, the single-projection structure ensures that the Proposed method remains competitive in CPU time. SDIM and NFPM perform comparably but display mild oscillations, while SASSEM converges more conservatively due to its restrictive stepsize strategy.

In the Figure 3 shows different choices of η illustrate the impact of inertial acceleration on convergence behavior and computational efficiency. The results show that appropriate momentum selection significantly improves convergence speed while maintaining computational stability.



(A) $n = 30$ (B) $n = 100$

(C) $n = 300$ (D) $n = 400$

FIGURE 2. Variance robustness analysis of the Proposed, SDIM, NFPM, and SASSEM algorithms for the stochastic complementarity problem. Each plot shows the mean residual D_k over 20 independent runs, with shaded regions representing one standard deviation. The results demonstrate the stability and robustness of the algorithms under stochastic perturbations as the problem dimension increases

The Table 2 indicate that the Proposed method achieves consistently low residuals across all dimensions, demonstrating stable convergence behavior. In contrast, SDIM exhibits extremely large residual values, suggesting instability under the considered stochastic setting. NFPM performs competitively with the Proposed method, while SASSEM shows acceptable performance at lower dimensions but deteriorates as n increases. In terms of CPU time, all methods remain comparable, confirming that the improved accuracy of the Proposed method does not incur additional computational cost.

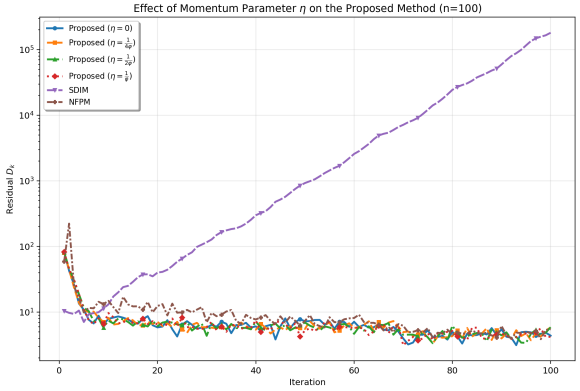
Example 5.2. *Consider the following stochastic fractional convex quadratic problem:*

$$\min_{x\in X} \mathbb{E}\left[\frac{\phi(x,\xi)}{\psi(x)}\right],$$

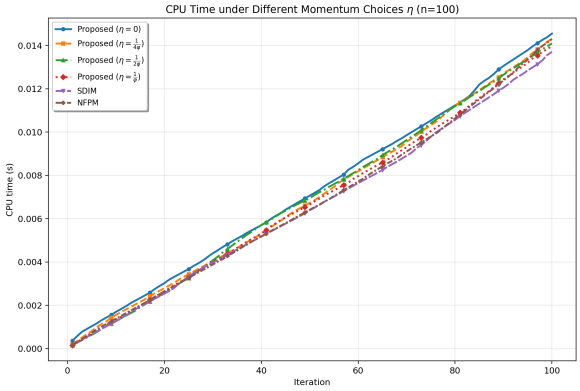
where

$$\phi(x,\xi) = \frac{1}{2}x^\top\left(0.025UU^\top + 0.025\frac{\|UU^\top\|_F}{\|V(\xi)\|_F}V(\xi)\right)x + \frac{1}{2}\left((c+\bar{c}(\xi))^\top x + 4n\right)^2,$$

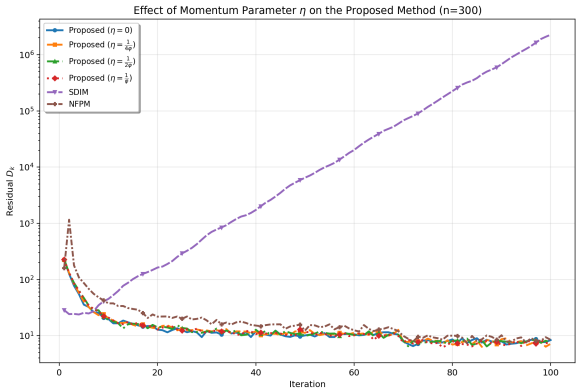
$$\psi(x) = r^\top x + t + 4n,$$



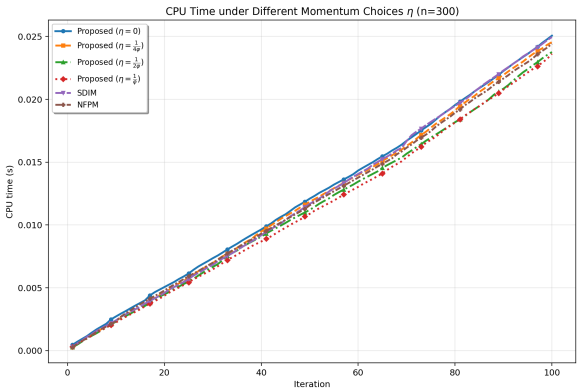
(A) Residual D_k vs iteration ($n = 100$)



(B) CPU time vs iteration ($n = 100$)



(C) Residual D_k vs iteration ($n = 300$)



(D) CPU time vs iteration ($n = 300$)

FIGURE 3. Effect of the momentum parameter η in the Proposed method compared with SDIM and NFPM for the stochastic complementarity problem

TABLE 2. Performance comparison of Proposed, SDIM, NFPM, and SASSEM on the stochastic complementarity problem at the final iteration $k = 100$

Dimension n	**Algorithm**	**Final residual** D_{100}	**CPU time (s)**
30	Proposed	3.0607	0.0152
	SDIM	1031.7339	0.0153
	NFPM	3.1152	0.0126
	SASSEM	1.7530	0.0117
100	Proposed	4.5998	0.0131
	SDIM	394959.2308	0.0131
	NFPM	4.2575	0.0135
	SASSEM	9.9384	0.0130
300	Proposed	7.2232	0.0212
	SDIM	2431817.9283	0.0178
	NFPM	7.7529	0.0186
	SASSEM	68.8552	0.0173

and

$$X = \{x \in \mathbb{R}^n : Ax \leq b,\ \|x\| \leq 1\}.$$

The parameters are generated as follows:

- U *and* c *are deterministic vectors generated from the standard normal distribution;*
- $V(\xi)$ *is a random square matrix with entries drawn from a standard normal distribution;*
- $\bar{c}(\xi)$ *is a random vector with entries uniformly distributed on* $(0, 1)$*;*
- r *is a deterministic vector and* t *is a scalar, both drawn from the uniform distribution on* $(0, 5)$*;*
- $A \in \mathbb{R}^{\lfloor n/10 \rfloor \times n}$ *and* $b \in \mathbb{R}^{\lfloor n/10 \rfloor}$ *are generated from the standard normal distribution.*

For numerical experiments, we use the following common settings for all algorithms (Proposed, SDIM, NFPM, and SASSEM):

$$\mu = 0.5, \qquad \gamma_k = \frac{1}{(50+k)^2}, \qquad N_k = \lfloor 0.99^{1-k} \rfloor.$$

The initial points are chosen as $x_0 = e, \quad x_1 = 0.5e$ *(for Proposed and SDIM), and* $x_0 = e$ *(for NFPM and SASSEM). The maximum iteration* $k = 200$ *is used as the stopping criterion.*

We note that the stochastic fractional programming problem in Example 5.2 can be reformulated as a stochastic variational inequality problem. Indeed, by applying a Dinkelbach-type transformation [30], the problem is equivalent to minimizing

$$\mathbb{E}[\phi(x, \xi) - \theta^\star \psi(x)],$$

whose first-order optimality condition leads to the stochastic variational inequality

$$\langle F(x^\star), x - x^\star \rangle \geq 0, \quad \forall x \in X,$$

where

$$F(x) = \mathbb{E}[\nabla_x \phi(x, \xi) - \theta^\star \nabla \psi(x)].$$

Thus, the proposed method can be directly applied to solve this class of stochastic fractional problems.

The plots in Figure 4 show that all methods decrease the objective value $F(x_k)$ over iterations, confirming convergence. The Proposed algorithm consistently achieves a faster reduction, indicating the effectiveness of its inertial mechanism. SDIM and NFPM exhibit comparable behavior but converge more gradually, with occasional mild fluctuations. In terms of CPU time, the Proposed method attains lower objective values within similar computational cost, demonstrating improved efficiency. This highlights the advantage of combining single-projection updates with momentum in stochastic fractional settings.

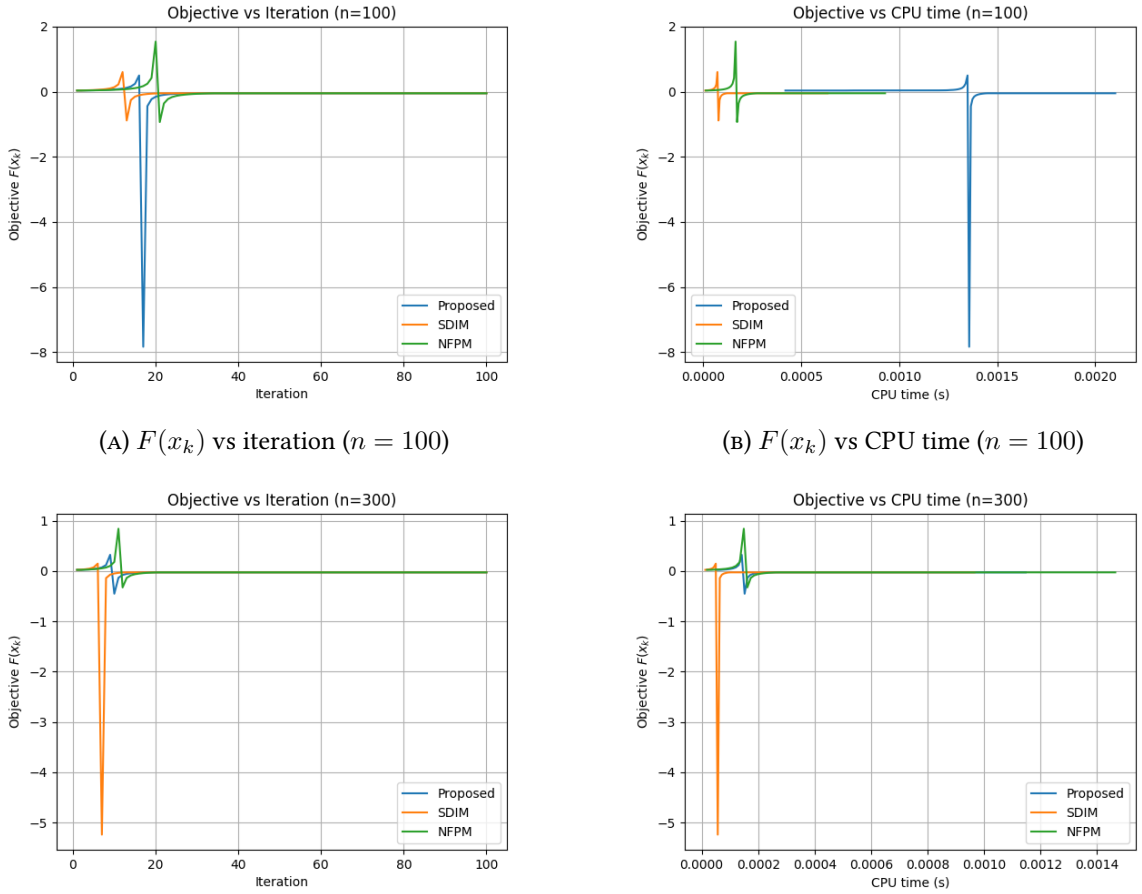
The Figure 5 shows that all algorithms reduce the residual D_k over iterations, confirming convergence. The **Proposed** method achieves faster decay and smoother behavior, indicating improved stability. CPU time remains comparable across methods, demonstrating efficiency. As dimension increases, the Proposed algorithm maintains strong performance, highlighting its scalability and robustness.

The table 3, shows consistent residual reduction across all methods. The **Proposed** algorithm demonstrates stable convergence with competitive CPU time. SDIM attains faster reduction, while NFPM and SASSEM exhibit slower decay, especially as the dimension increases.

Example 5.3. *In this section, the Proposed method is applied to a bandwidth allocation problem, and its performance is compared with SDIM, NFPM, and SASSEM.*

In a communication network, selfish users compete for network bandwidth. The following notations are used:

- $S = \{1, 2, \ldots, m\}$*: the set of all users in the network;*
- $\mathcal{R}(s)$*: the set of routes governed by user* $s \in S$*;*
- $|\mathcal{R}(s)|$*: the number of routes available to user* s*;*



(A) $F(x_k)$ vs iteration ($n = 100$)

(B) $F(x_k)$ vs CPU time ($n = 100$)

(C) $F(x_k)$ vs iteration ($n = 300$)

(D) $F(x_k)$ vs CPU time ($n = 300$)

FIGURE 4. Objective value convergence of the Proposed, SDIM, and NFPM algorithms for the stochastic fractional programming problem

- $x_s^{(r)}$: *the flow rate of user* s *through route* $r \in \mathcal{R}(s)$;
- $\xi_s^{(r)}$: *the random weighted parameter of route* $r \in \mathcal{R}(s)$;
- $L = \{1, 2, \ldots, k\}$: *the set of all links;*
- $R = \bigcup_{s \in S} \mathcal{R}(s)$: *the set of all routes;*
- $L_s^{(l)}$: *the set of links through which route* $l \in \mathcal{R}(s)$ *goes.*

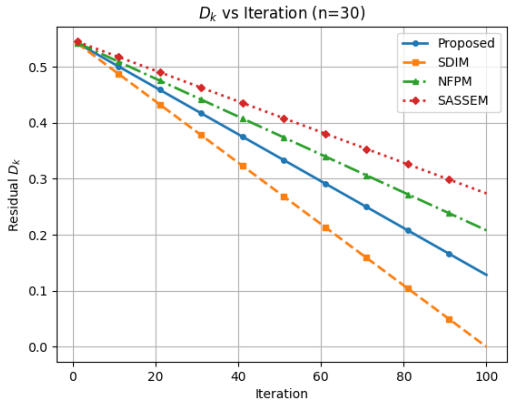
The utility of user s *is modeled by the concave function*

$$f_s(x_s, \xi_s) = \sum_{r \in \mathcal{R}(s)} \xi_s^{(r)} \ln(1 + x_s^{(r)}), \qquad \forall s \in S,$$

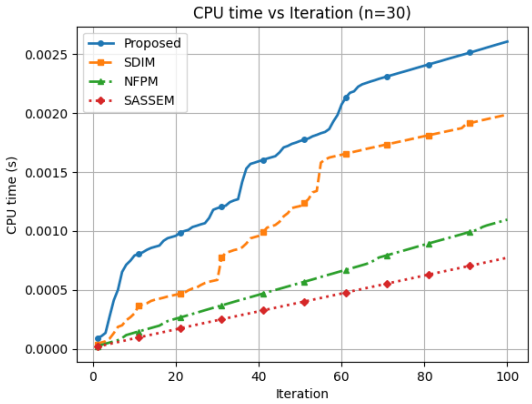
where $x_s = (x_s^{(r)})_{r \in \mathcal{R}(s)}$ *is the flow-rate decision vector of user* s.

The network bandwidth allocation model is formulated as the stochastic optimization problem

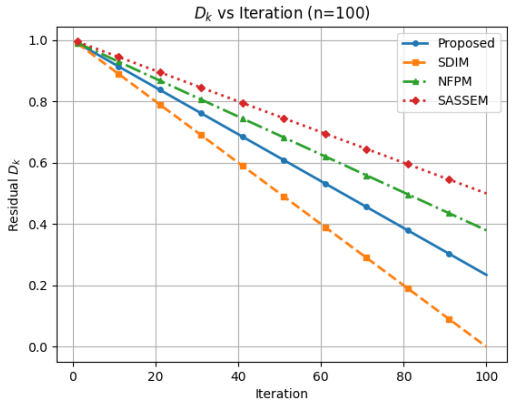
$$\max \sum_{s \in S} \mathbb{E}[f_s(x_s, \xi_s)] - g(x) \quad \textit{s.t.} \quad x \in \bigcap_{l \in L} X_l, \tag{5.1}$$



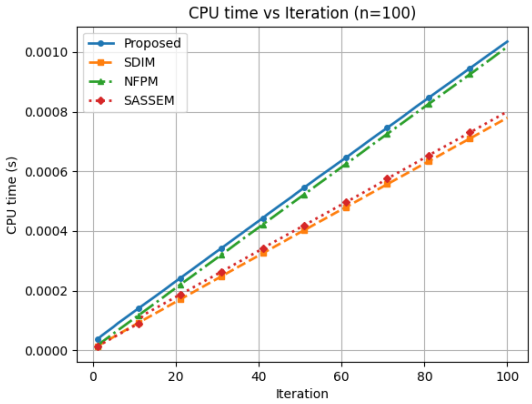
(A) D_k vs iteration ($n = 30$)



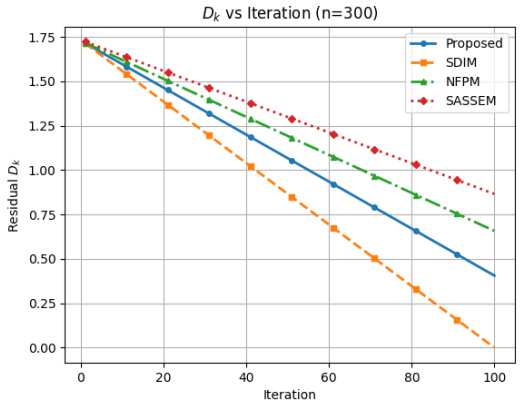
(B) CPU time vs iteration ($n = 30$)



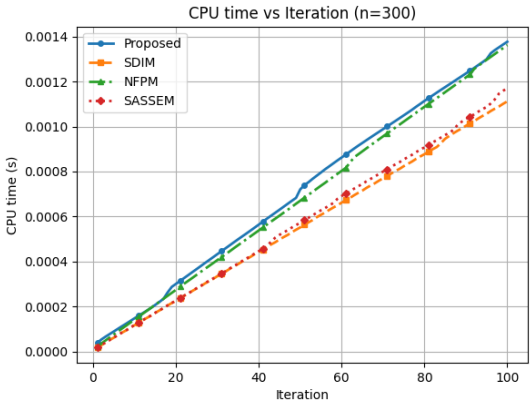
(C) D_k vs iteration ($n = 100$)



(D) CPU time vs iteration ($n = 100$)



(E) D_k vs iteration ($n = 300$)



(F) CPU time vs iteration ($n = 300$)

FIGURE 5. Performance comparison of Proposed, SDIM, NFPM, and SASSEM for the stochastic complementarity problem

TABLE 3. Performance summary of Proposed, SDIM, NFPM, and SASSEM for the stochastic complementarity problem. The quantities D_1, D_{50}, and D_{100} denote the residual values at iterations 1, 50, and 100, respectively, while CPU_{50} and CPU_{100} denote the corresponding CPU times in seconds

n	**Algorithm**	D_1	D_{50}	D_{100}	CPU_{50} **(s)**	CPU_{100} **(s)**
30	Proposed	0.542245	0.337523	0.128312	0.002433	0.003077
	SDIM	0.542245	0.273861	0.000000	0.000564	0.001015
	NFPM	0.542245	0.377174	0.207918	0.000521	0.001071
	SASSEM	0.544984	0.410792	0.273861	0.000401	0.000814
100	Proposed	0.990000	0.616231	0.234265	0.000598	0.001172
	SDIM	0.990000	0.500000	0.000000	0.000410	0.000832
	NFPM	0.990000	0.688622	0.379605	0.000530	0.001066
	SASSEM	0.995000	0.750000	0.500000	0.000446	0.000845
300	Proposed	1.714730	1.067343	0.405758	0.000777	0.001485
	SDIM	1.714730	0.866025	0.000000	0.000578	0.001115
	NFPM	1.714730	1.192729	0.657496	0.000704	0.001444
	SASSEM	1.723391	1.299038	0.866025	0.000529	0.001056

where the congestion cost is

$$g(x) = \Upsilon \sum_{\hat{s}\in S}\sum_{d\in L} x_{\hat{s}}^{(r)}\varphi_{\hat{s},d}^{(l)}\left(\sum_{s\in S} x_s^{(r)}\varphi_{s,d}^{(l)}\right),$$

and

$$X_l = \left\{ x \in \mathbb{R}_+^{\sum_{s\in S}|\mathcal{R}(s)|} : \sum_{s\in S} x_s^{(r)}\varphi_d^{(r)} \le c_l \right\}$$

where Υ *is the scaling parameter,* $c_l(\ge 0)$ *is the capacity of the link* $l \in \mathcal{L}$ *and* $\varphi_{s,d}^{(l)} = 1$ *if* $d \in \mathcal{L}_s^{(s)}$ *and* $\varphi_{s,d}^{(l)} = 0$ *otherwise. Let* $X = \bigcap_{l\in L} X_l$*. Then problem (5.1) is equivalent to the stochastic variational inequality*

$$\text{find } x^* \in X \text{ such that } \langle F(x^*), x - x^*\rangle \ge 0, \qquad \forall x \in X, \tag{5.2}$$

where

$$F(x) = \mathbb{E}[f(x,\xi)], \quad f(x,\xi) = -\left(\frac{\xi_s^{(r)}}{1+x_s^{(r)}}\right)_{s\in S} + 2\Upsilon M^\top M x.$$

Now consider the bandwidth allocation problem on the network, which consists of 8 *routes,* 14 *links and* 5 *users. In this case,*

$$S = \{1,2,\ldots,5\}, \qquad L = \{1,2,\ldots,14\}, \qquad R = \{1,2,\ldots,8\},$$

with

$$\mathcal{R}(1) = \mathcal{R}(2) = \mathcal{R}(5) = \{1,2\}, \qquad \mathcal{R}(3) = \mathcal{R}(4) = \{1\},$$

TABLE 4. Uniform distributions of the random weighted parameters

Random parameter	$\xi_1^{(1)}$	$\xi_1^{(2)}$	$\xi_2^{(1)}$	$\xi_2^{(2)}$	$\xi_3^{(1)}$	$\xi_4^{(1)}$	$\xi_5^{(1)}$	$\xi_5^{(2)}$
Uniform distribution	$U(18,20)$	$U(15,17)$	$U(16,18)$	$U(17,19)$	$U(11,13)$	$U(10,12)$	$U(12,14)$	$U(10,12)$

TABLE 5. Link capacities

Link no.	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}	c_{14}
Capacity	56	55	53	55	54	53	56	57	53	42	59	58	49	51

and adjacency matrix

$$M = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \end{bmatrix}.$$

It is known that F is pseudomonotone and Lipschitz continuous.

Numerical setting. *We consider a network with 8 routes, 14 links, and 5 users. The scaling parameter is set to $\Upsilon = 0.15$, while the random parameters $\xi_s^{(r)}$ follow the uniform distributions specified in Table 4. The link capacities are given in Table 5.*

Algorithm parameters. *For all methods (Proposed, SDIM, NFPM, and SASSEM), we use the parameter settings described in Table 1. In particular, we set*

$$\mu = 0.5, \qquad \gamma_k = \frac{1}{k^2}, \qquad N_k = 2\lfloor (k+3)(\ln(k+3))^2 \rfloor.$$

The initial points are randomly generated in $(0,10)^8$, and the stopping criterion is given by

$$D_k < 5 \times 10^{-3} \quad \text{or} \quad k \geq 150.$$

The uniform distributions of the random weighted parameters in the above Table 4 reflect varying user priorities across routes. Higher intervals indicate stronger preference or demand, introducing stochastic variability that influences route selection and overall network equilibrium behavior.

The link capacity Table 5 shows moderate variability across links, indicating heterogeneous resource availability. Links with lower capacities, such as c_{10}, are more prone to congestion, while higher-capacity links provide more stable and reliable transmission pathways.

From the network structure in Figure 6, Users 1 and 5 appear to be the most strategic, as they have access to multiple routes and connections to central links, allowing flexible bandwidth allocation. Routes 3 and 5 seem more efficient due to their shorter and less congested paths. In contrast, Routes 7 and 8 pass through several shared links, making them more prone to congestion. Among the links, Links 1,

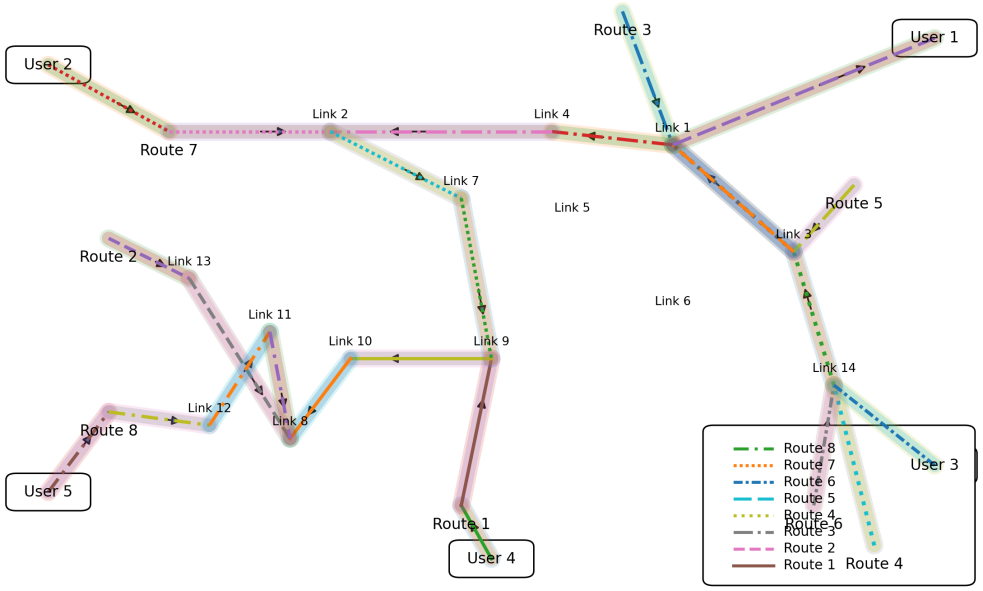


FIGURE 6. Bandwidth network structure

3, and 14 are heavily connected and thus critical but risky, while peripheral links such as Links 11 and 12 are comparatively safer due to lower traffic interaction.

6. CONCLUSION

In this paper, we proposed a novel C-FISTA-inspired single-projection algorithm for solving stochastic variational inequality problems. The method integrates inertial acceleration, adaptive stepsize control, and a carefully designed correction mechanism, resulting in a computationally efficient framework that avoids multiple projections and does not require prior knowledge of Lipschitz constants. The single-projection structure significantly reduces per-iteration cost, making the algorithm suitable for large-scale stochastic environments. We established almost sure convergence of the generated sequences under standard assumptions using a supermartingale framework. In addition, under strong pseudomonotonicity, we derived linear convergence rates, demonstrating the theoretical strength of the proposed method. The analysis highlights the role of momentum and adaptive stepsizes in balancing convergence speed and stability. Extensive numerical experiments on stochastic complementarity problems, fractional programming, and network bandwidth allocation confirm the effectiveness of the proposed algorithm. In particular, the method consistently achieves faster convergence, improved robustness to stochastic noise, and competitive computational cost compared with SDIM, NFPM, and SASSEM.

STATEMENTS AND DECLARATIONS

The authors declare that they have no conflict of interest, and the manuscript has no associated data.

ACKNOWLEDGMENTS

There is no funding for this project.

References

[1] M. Dilshad, I. Al-Dayel, F. O. Nwawuru, and J. N. Ezeora. On quasi-monotone stochastic variational inequalities with applications. *Axioms*, 14:Article ID 912, 2025.

[2] S. Wang, Y. Zhang, and Y. J. Cho. A new infeasible projection method for stochastic variational inequality problem. *Journal of Optimization Theory and Applications*, 208:Article ID 6, 2026.

[3] S. Wang, Y. Zhang, and Y. J. Cho. A stochastic double inertial method for solving stochastic variational inequality problem. *Bulletin of the Malaysian Mathematical Sciences Society*, 49:Article ID 26, 2026.

[4] S. Wang, H. Tao, R. Lin, and Y. J. Cho. A self-adaptive stochastic subgradient extragradient algorithm for the stochastic pseudomonotone variational inequality problem with application. *Zeitschrift für Angewandte Mathematik und Physik*, 73:Article ID 164, 2022.

[5] F. O. Nwawuru, G. N. Echezona, and C. C. Okeke. Finding a common solution of variational inequality and fixed point problems using subgradient extragradient techniques. *Rendiconti del Circolo Matematico di Palermo Series 2*, 73:1255-1275, 2024.

[6] L. Liu, S. Y. Cho, and J.-C. Yao. Convergence analysis of an inertial Tseng's extragradient algorithm for solving pseudomonotone variational inequalities and applications. *Journal of Nonlinear and Variational Analysis*, 5:627-644, 2021.

[7] A. Shapiro. Monte Carlo sampling methods. *Handbook of Operations Research and Management Science*, 10:353-425, 2003.

[8] Y. C. Liang, Q. N. Fan, and P. P. Shen. A hybrid Newton method for stochastic variational inequality problems and application to traffic equilibrium. *Asia-Pacific Journal of Operational Research*, 38:Article ID 2050036, 2021.

[9] H. Xu. Sample average approximation methods for a class of stochastic variational inequality problems. *Asia-Pacific Journal of Operational Research*, 27:103-119, 2010.

[10] S. He, P. Zhang, X. Hu, and R. Hu. A sample average approximation method based on a D-gap function for stochastic variational inequality problems. *Journal of Industrial and Management Optimization*, 10(3):977-987, 2014.

[11] H. Jiang and H. Xu. Stochastic approximation approaches to the stochastic variational inequality problem. *IEEE Transactions on Automatic Control*, 53:1462-1475, 2008.

[12] G. M. Korpelevich. The extragradient method for finding saddle points and other problems. *Ekonomika i Matematicheskie Metody*, 12:747-756, 1976.

[13] F. O. Nwawuru and J. N. Ezeora. Inertial-based extragradient algorithm for approximating a common solution of split-equilibrium problems and fixed-point problems of nonexpansive semigroups. *Journal of Inequalities and Applications*, 2023:Article ID 22, 2023.

[14] A. Kannan and U. V. Shanbhag. Optimal stochastic extragradient schemes for pseudomonotone stochastic variational inequality problems and their variants. *Computational Optimization and Applications*, 74:779-820, 2019.

[15] Y. Censor, A. Gibali, and S. Reich. The subgradient extragradient method for solving variational inequalities in Hilbert space. *Journal of Optimization Theory and Applications*, 148:318-335, 2011.

[16] F. O. Nwawuru, J. N. Ezeora, H. ur Rehman, and J.-C. Yao. Self-adaptive subgradient extragradient algorithm for solving equilibrium and fixed point problems in Hilbert spaces. *Numerical Algorithms*, 101:1567-1605, 2025.

[17] P. Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM Journal on Control and Optimization*, 38:431-446, 2000.

[18] B. T. Polyak. Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics*, 4:1-17, 1964.

[19] F. O. Nwawuru, O. K. Narain, M. Dilshad, and J. N. Ezeora. Splitting method involving two-step inertial for solving inclusion and fixed point problems with applications. *Fixed Point Theory and Algorithms for Sciences and Engineering*, 2025:Article ID 8, 2025.

[20] Y. Yao, N. Shahzad, M. Postolache, and J.-C. Yao. Convergence of self-adaptive Tseng-type algorithms for split variational inequalities and fixed point problems. *Carpathian Journal of Mathematics*, 40:753-768, 2024.

[21] H. ur Rehman, D. Ghosh, J.-C. Yao, and X. Zhao. Solving equilibrium and fixed-point problems in Hilbert spaces: a class of strongly convergent Mann-type dual-inertial subgradient extragradient methods. *Journal of Computational and Applied Mathematics*, 464:Article ID 116509, 2025.

[22] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2:183-202, 2009.

[23] C. Garner and S. Zhang. Linearly-convergent FISTA variant for composite optimization with duality. *Journal of Scientific Computing*, 94:Article ID 65, 2023.

[24] M. Dilshad, I. Al-Dayel, F. O. Nwawuru, and P. Agarwal. A single-projection proximal algorithm for stochastic mixed variational inequalities with applications to breast cancer screening. *AIMS Mathematics*, 11:10533-10565, 2026.

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

[26] C. Fang and S. Chen. Some extragradient algorithms for variational inequalities. In W. Han, S. Migórski, and M. Sofonea, editors, *Advances in Variational and Hemivariational Inequalities. Advances in Mechanics and Mathematics*, vol 33, pages 145–171, 2015. Springer, Cham.

[27] A. N. Iusem, A. Jofré, R. I. Oliveira, and P. Thompson. Variance-based extragradient methods with line search for stochastic variational inequalities. *SIAM Journal on Optimization*, 29:175-206, 2019.

[28] L. C. Ceng, P. Cubiotti, and J.-C. Yao. An implicit iterative scheme for monotone variational inequalities and fixed point problems. *Nonlinear Analysis: Theory, Methods & Applications*, 11:2445-2457, 2008.

[29] Y. Malitsky. Golden ratio algorithms for variational inequalities. *Mathematical Programming*, 184:383-410, 2020.

[30] W. Dinkelbach. On nonlinear fractional programming. *Management Science*, 13:492-498, 1967.