J. Nonlinear Var. Anal. 7 (2023), No. 6, pp. 995-1016
Available online at http://jnva.biemdas.com
https://doi.org/10.23952/jnva.7.2023.6.08

A STOCHASTIC PROJECTION AND CONTRACTION ALGORITHM WITH INERTIAL EFFECTS FOR STOCHASTIC VARIATIONAL INEQUALITIES

LIYA LIU[1], XIAOLONG QIN[2,3,*]

[1]*School of Mathematics and Statistics, Southwest University, 400715, Chongqing, China*
[2]*Center for Advanced Information Technology, Kyung Hee University, Seoul 02447, Korea*
[3]*Department of Mathematics, Zhejiang Normal University, Jinhua 321004, China*

Abstract. In this paper, we investigate a stochastic approximation based algorithm for solving nonmonotone stochastic variational inequalities. Our algorithm combines the projection and contraction method with the inertial extrapolation technique. The self-adaptive step size sequence is generated by employing the Armijo's line search rule. We also investigate the almost sure convergence property without using the prior knowledge of the Lipschitz constant of the involved operator in our algorithm. Theoretical results related to the convergence rate and the oracle complexity are provided under mild assumptions. Primary numerical experiments are presented to demonstrate the efficiency of the algorithm.

Keywords. Convergence result; Inertial effects; Projection and contraction method; Stochastic variational inequality.

1. INTRODUCTION

Let $\mathbb{R}^d$ be a d-dimensional Euclidean space with inner product $\langle\cdot,\cdot\rangle$ and induced norm $\|\cdot\|$. Let X be a nonempty, closed, and convex set of $\mathbb{R}^d$. Given an operator $G:\mathbb{R}^d\to\mathbb{R}^d$, we consider the variational inequality problem consists of finding an $x^*\in X$ such that $\langle G(x^*),x-x^*\rangle\geq 0$ for all $x\in X$. It is known that the variational inequality is fundamental in applied mathematics and operations research. In view of the situations that many practical problems may involve random factors, Gurkan et al. [1] proposed a stochastic variational inequality.

In the stochastic case, we start with a probability space $(\Omega,\mathscr{F},\mathbb{P})$, where $\mathscr{F}$ denotes a σ-algebra on Ω, and $\mathbb{P}$ is the associated probability distribution. Let $(\Xi,\mathfrak{G})$ be a measurable space. Let $\xi:\Omega\to\Xi$ be a random variable defined on the probability space $(\Omega,\mathscr{F},\mathbb{P})$, which induces an expectation $\mathbb{E}$ and a distribution $\mathbb{P}_\xi$. One assumes that $F(\cdot,\cdot):\mathbb{R}^d\times\Xi\to\mathbb{R}^d$ is integrable on Ξ. Let $G:\mathbb{R}^d\to\mathbb{R}^d$ be defined by $G(x)=\mathbb{E}[F(x,\xi)]$ for all $x\in\mathbb{R}^d$. We include the formulation of the stochastic variational inequality (SVI, shortly) as follows

$$\text{Find an } x^*\in X,\ \langle\mathbb{E}[F(x^*,\xi)],x-x^*\rangle\geq 0,\ \forall x\in X. \tag{1.1}$$

In what follows, the solution set of (1.1) is denoted by $S(X,G)$. We denote by $M(X,G)$ the solution set of the involved Minty variational inequality: find a $z\in X$ such that $\langle G(x),x-z\rangle\geq 0$,

*Corresponding author.
E-mail address: qxlxajh@163.com (X. Qin).
Received March 4, 2023; Accepted October 22, 2023.

where z can be referred as a weak solution to (1.1). The SVI represents a unified form of optimality conditions of the stochastic optimization, which theoretically and algorithmically applied in various fields, such as engineering, management, and economics; see, e.g., [2–4].

The well-developed deterministic algorithms cannot directly resolve SVIs unless the involved expectation can be evaluated explicitly. It is of interest to devise stochastic algorithms that produce asymptotically solutions of SVIs. Based on how sampling is incorporated with the algorithm, there are two basic methodologies existed for solving SVIs: Sample average approximation (SAA, shortly) method and and stochastic approximation (SA, shortly) method.

The SAA method was studied in the stochastic programming; see, e.g., [5–7]. It replaces the expected-value operator $G(\cdot)$ with an empirical estimator of the form $\frac{1}{N}\sum_{j=1}^{N} F(\cdot,\xi_j)$ by using samples of $\xi_1,\xi_2,\cdots,\xi_N$. In this case, deterministic algorithms can be employed to solve SVIs. From this point of view, the SAA method is called an 'exterior' approach since it does not depend on the structure of the algorithm. In order to precisely approximate the exact solution of SVIs, it requires the sample size N to be relatively large, which results in a large computational complexity. The SA approach was initiated by Robbins and Monro [8] for solving the stochastic root-finding problem. Its strategy is to replace the expected-value operator $G(\cdot)$ along the iterations by the random operator $F(\cdot,\xi)$. Along the trajectory of the iterative procedure, the interior and online samples of ξ can bring out the oracle error, $\varepsilon(x,\xi) = F(x,\xi) - G(x)$, $\forall x \in X$. It is noted that the SA method can be viewed as an "interior" approach since it relies on the structure of the algorithm. In contrast with the SAA method, one sees that the SA method is easier to implement with the less memory. The efficiency of SA methods prompts the study of SA-based algorithms for dealing with SVIs [9–11]. Jiang and Xu [12] first extended the SA method to solve SVIs. The almost sure convergence of the resulting algorithm was proven under the assumption of the strongly monotonicity and the Lipschitz continuity on the expected-value operator $G(\cdot)$. Yousefian et al. [13] proposed an SA-based algorithm with adaptive step sizes for solving Cartesian SVIs under the strongly monotone assumption on the expected-value operator $G(\cdot)$. In order to relax the strongly monotonicity to the plain monotonicity, Koshal et al. [14] incorporated the Tikhonov regularization technique into an SA-based proximal point algorithm and guaranteed the almost sure convergence result.

In much of the prior work, SA-based algorithms for solving SVIs were limited to monotone operators. However, the monotonicity is somewhat strong, which restricts the applications of the algorithms. Iusem et al. [15] devised an SA-based extragradient algorithm for solving pseudomonotone SVIs. The iterative steps can be expressed as: $\forall x_0 \in X$,

$$\begin{cases} y_n = \Pi_X\left(x_n - \frac{\gamma_n}{S_n}\sum_{j=1}^{S_n} F(x_n,\xi_n^j)\right), \\ x_{n+1} = \Pi_X\left(x_n - \frac{\gamma_n}{S_n}\sum_{j=1}^{S_n} F(y_n,\eta_n^j)\right), \end{cases} \tag{1.2}$$

where $\xi_n = \{\xi_n^j\}_j^{S_n}$ and $\eta_n = \{\eta_n^j\}_j^{S_n}$ are independent identically distributed (i.i.d., shortly) samples, L is the Lipschitz constant of the involved expected-value operator $\mathbb{E}[F(\cdot,\xi)]$, and $\gamma_n = \mathscr{O}\left(\frac{1}{L}\right)$ is the step size. The almost sure convergence requires $\mathbb{E}[F(\cdot,\xi)]$ to be pseudomonotone and Lipschitz continuous. It is noted that the projection onto the feasible set X is either unavailable or computationally expensive when feasible set X has a complex expression. In such situation, the efficiency of the extragradient method may be affected, due to the fact that the projection onto feasible set X has to be calculated twice per iteration. How to reduce the

number of projections onto feasible set X performed at each iteration has been one of the important tasks. Along this way, it has led to some projection-based methods with one projection onto the feasible set X per iteration; see, e.g., [16–18].

It is noted that the projection onto a specific half-space can be calculated by an explicit formula. In this spirit, Yang et al. [17] developed an SA-based subgradient extragradient algorithm, which reads as follows

$$\begin{cases} y_n = \Pi_X \left(x_n - \frac{\gamma_n}{S_n} \sum_{j=1}^{S_n} F(x_n, \xi_n^j) \right), \\ x_{n+1} = \Pi_{C_n} \left(x_n - \frac{\gamma_n}{S_n} \sum_{j=1}^{S_n} F(y_n, \eta_n^j) \right), \end{cases} \tag{1.3}$$

where

$$C_n := \left\{ z \in \mathbb{R}^d : \left\langle x_n - \frac{\gamma_n}{S_n} \sum_{j=1}^{S_n} F(x_n, \xi_n^j) - y_n, y_n - z \right\rangle \geq 0 \right\}.$$

One salient advantage of Algorithm (1.3) is the evaluation of only one projection onto the feasible set X, which improves the computational efficiency. As a modification of Algorithm (1.2), the SA-based projection and contraction algorithm [19] replaces the second projection onto feasible set X by computing the search direction and the step size. The resulting algorithm involves only one projection onto feasible set X and no further projections onto a priori half-space. The saving of one projection step allows the algorithm to be more potentially efficient than Algorithm (1.2). Through continuous research in this aspect, we focus on an extension of the SA-based projection and contraction algorithm for solving nonmonotone SVIs.

The update of the step size requires the Lipschitz constant of the involved operator. However, the Lipschitz constant is necessarily unknown or difficult to estimate. To overcome this drawback, the construction of self-adaptive step sizes has aroused numerous interest among researchers; see, e.g., [16–18]. They adopted an SA-version of Armijo's line search criterion to determine the step size. More precisely, choose γ_n as the maximum $\gamma \in \{\mu \nu^\iota : \iota \in \mathbb{N}_0\}$ such that

$$\gamma \left\| \frac{1}{S_n} \sum_{j=1}^{S_n} f(x_n, \xi_n^j) - \frac{1}{S_n} \sum_{j=1}^{S_n} f(z_n^\lambda, \xi_n^j) \right\| \leq \lambda \|x_n - z_n^\lambda\|, \tag{1.4}$$

where $z_n^\gamma = \Pi_X \left(x_n - \frac{\gamma}{S_n} \sum_{j=1}^{S_n} f(x_n, \xi_n^j) \right)$, $\nu \in (0,1)$, $\mu \in (0,1)$, and $\lambda \in (0,1)$.

It is noted that the inertial extrapolation as one of accelerated techniques has attracted the attention of researchers in building fast algorithms. Such inertial extrapolation was initially proposed by Polyak [20], acting as an acceleration process. It is based on a discrete version of a second-order dissipative dynamical system. The main feature of the inertial procedure is that it involves two iterative steps and the next iterative step is obtained by using the combination of previous two iterates. Recently, various types of inertial algorithms have been devised and applied to various optimization problems, such as split feasibility problems, equilibrium problems, variational inequalities, and fixed point problems; see, e.g., [21–25]. An interesting problem is how do we establish an inertial type algorithm for solving stochastic variational inequalities.

Motivated and inspired by the above results, our interest lies in developing an SA-based projection and contraction algorithm incorporated with inertial effects and step size rules. It is proved that the iterative sequence converges to a solution of SVIs in an almost sure sense. Some numerical experiments are presented to demonstrate the efficiency and advantages of the proposed algorithm. Our main contributions in this paper can be summarized as follows

(i) Our algorithm incorporates inertial terms, which accelerates the convergence speed. The test results demonstrate the benefits of inertial terms on the algorithm.
(ii) Our algorithm does not require a projection step onto a priori half-space, which reduces the computational cost, in contrast to algorithms in [16–18].
(iii) The almost sure convergence analysis of the algorithm is performed without using the prior knowledge of the Lipschitz constant of the operator involved.

The paper is organized as follows. Some basic definitions and lemmas are presented in Section 2. The algorithm and its almost sure convergence analysis are provided in Section 3. We focus on analyzing the convergence rate and the oracle complexity of the proposed algorithm in Section 4. We provide some numerical experiments based on the randomly generated data to support the theoretical results in Section 5. Concluding remarks are reported in Section 6, the last section.

2. Preliminaries

Let us start this section by introducing the notations to facilitate our analysis. Let $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, where $\mathbb{N}$ is the set of all positive integers. Given a sequence $\{x_k\}$, the notation $x_k = \mathscr{O}(\mathscr{B})$ denotes that there exists a constant $C > 0$ such that $\|x_k\| \leq C\mathscr{B}$ for all $n \in \mathbb{N}_0$. Given a σ-algebra $\mathscr{F}$ and a random variable ξ, we denote by $\mathbb{E}[\xi]$, $\mathbb{E}[\xi \mid \mathscr{F}]$, and $\mathbb{V}[\xi]$, the expectation, the conditional expectation, and the variance, respectively. The notations $\xi \in \mathscr{F}$ and $\xi \perp\!\!\!\perp \mathscr{F}$ denote that ξ is $\mathscr{F}$-measurable and ξ is independent of $\mathscr{F}$. We denote by $\sigma(\xi_1, \cdots, \xi_n)$ the σ-algebra generated by random variables $\xi_1, \cdots, \xi_n$. Given the random variable ξ and $p \geq 1$, $|\xi|_p$ is the L_p-norm of ξ and $|\xi \mid \mathscr{F}|_p := \sqrt[p]{\mathbb{E}[|\xi|^p \mid \mathscr{F}]}$ is the L_p-norm of ξ conditional to the σ-algebra $\mathscr{F}$. The symbol "a.s." is the abbreviation for "almost surely". We use $\Pi_X(x)$ to denote the projection of x onto a closed and convex set X, i.e., $\Pi_X(x) := \arg\min_{y \in X} \|y - x\|^2$.

We recall some properties of projections; see [26] for more details.

Lemma 2.1. *Let $X \subseteq \mathbb{R}^d$ be a nonempty, convex and closed set. Then*

(i) Given $x \in \mathbb{R}^d$, $\langle x - \Pi_X(x), \Pi_X(x) - y\rangle \geq 0$ for all $y \in X$.
(ii) For all $x, y \in \mathbb{R}^d$, $\|\Pi_X(x) - \Pi_X(y)\| \leq \|x - y\|$.
(iii) For all $x \in X$ and $y \in \mathbb{R}^d$, $\|x - \Pi_X(y)\| \leq \langle x - \Pi_X(y), x - y\rangle$.
(iv) Given $\gamma > 0$ and $G : \mathbb{R}^d \to \mathbb{R}^d$, $S(X, G) = \{x \in \mathbb{R}^d : x = \Pi_X(x - \gamma G(x))\}$.

For a given operator $G : \mathbb{R}^d \to \mathbb{R}^d$ and a constant $\gamma > 0$, we define the natural residual function associated to SVI (1.1) as $\mathfrak{R}_\gamma(x) := \|x - \Pi_X(x - \gamma G(x))\|$ for all $x \in \mathbb{R}^d$. For simplicity, we denote by $\mathfrak{R}(x) = \mathfrak{R}_1(x)$.

Lemma 2.2. *[26] Given $x \in \mathbb{R}^d$, the function $\gamma \mapsto \frac{\mathfrak{R}_\gamma(x)}{\gamma}$ is non-increasing over $(0, \infty)$.*

Definition 2.1. [27] An operator $G : X \to \mathbb{R}^d$ is said to be

(i) μ-strongly monotone on X if there exists $\mu > 0$ such that $\langle G(x) - G(y), x - y\rangle \geq \mu\|x - y\|^2$ for all $x, y \in X$;
(ii) monotone on X if $\langle G(x) - G(y), x - y\rangle \geq 0$ for all $x, y \in X$;
(iii) pseudomonotone on X if $\langle G(y), x - y\rangle \geq 0 \Rightarrow \langle G(x), x - y\rangle \geq 0$ for all $x, y \in X$;
(iv) quasimonotone on X if $\langle G(y), x - y\rangle > 0 \Rightarrow \langle G(x), x - y\rangle \geq 0$ for all $x, y \in X$;
(v) nonmonotone on X if there exists $y \in X$ such that $\langle G(x), x - y\rangle \geq 0$ for all $x \in X$.

Remark 2.1. The relation between different kinds of monotonicity can be described as: strong monotonicity $\Rightarrow$ monotonicity $\Rightarrow$ pseudomonotonicity $\Rightarrow$ quasimonotonicity. In addition, if X is weakly compact, then quasimonotonicity $\Rightarrow$ nonmonotonicity; see [28], while the reverse is not always true (see Example 5.1 below.)

Remark 2.2. If G is pseudomonotone, then $S(X,G) \subseteq M(X,G)$, while the reverse case may be not true; see [29,30]. This explains why the solution of the Minty variational inequality can be referred as a weak solution of the corresponding SVI. While, in the nonmonotone setting, if G is continuous, then $M(X,G) \subseteq S(X,G)$; see [30,31].

We define the oracle error map $\varepsilon : X \times \Xi \to \mathbb{R}^d$ by

$$\varepsilon(x,\xi) := F(x,\xi) - G(x),\ \forall x \in X, \forall \xi \in \Xi.$$

For a given constant $p \geq 2$, the oracle error's p-moment function is defined by

$$\sigma_p(x) := \sqrt[p]{\mathbb{E}[\|\varepsilon(x,\xi)\|^p]},\ \forall x \in X. \tag{2.1}$$

Given an i.i.d. sample $\xi_S = (\xi_j)_{j=1}^S$ drawn from Ξ, the empirical operator and the oracle's empirical mean error associated to ξ_S are respectively defined by

$$\hat{F}(x,\xi_S) := \frac{1}{S}\sum_{j=1}^S f(x,\xi_j) \text{ and } \hat{\varepsilon}(x,\xi_S) := \frac{1}{S}\sum_{j=1}^S \varepsilon(x,\xi_j),\ \forall x \in X. \tag{2.2}$$

We give the following standard assumption that are assumed to hold through the rest of this paper.

Assumption 2.1. (i) For almost every $\xi \in \Xi$, $F : \mathbb{R}^d \times \Xi \to \mathbb{R}^d$ satisfies

$$\|F(x,\xi) - F(y,\xi)\| \leq \mathscr{L}(\xi)\|x-y\|, \text{ for all } x,y \in \mathbb{R}^d,$$

where $\mathscr{L} : \Xi \to \mathbb{R}_+$ is a measurable function such that $\mathscr{L}(\xi) \geq 1$ for almost every $\xi \in \Xi$. Then the expected-value operator $G(x) = \mathbb{E}[F(x,\xi)]$ is Lipschitz continuous with the constant $L = \mathbb{E}[\mathscr{L}(\xi)]$. For $p \geq 2$, the oracle error's p-moment function $\sigma_p(\cdot)$ is Lipschitz continuous with the constant $L_p = |\mathscr{L}(\cdot)|_p + L$.
(ii) There exist $z \in \mathbb{R}^d$ and $p \geq 2$ such that $\mathbb{E}[\|F(z,\xi)\|^p] < \infty$ and $\mathbb{E}[\mathscr{L}(\xi)^p] < \infty$.

The following lemmas are crucial in our subsequent analysis.

Lemma 2.3. *[32] For any $q \geq 2$, there exists $C_q > 0$ such that, for any vector-valued martingale $\{w_j\}_{j=1}^S \subseteq \mathbb{R}^d$ adapted to the filtration $\{\mathscr{F}_j\}_{j=0}^S$ with $w_0 = 0$,*

$$\left|\sup_{j\leq S}\|w_j\|\right|_q \leq C_q \left|\sqrt{\sum_{j=1}^S \|w_j - w_{j-1}\|^2}\right|_q \leq C_q \sqrt{\sum_{j=1}^S \left|\|w_j - w_{j-1}\|^2\right|_q}.$$

Lemma 2.4. *[18] Let $\xi_S := (\xi_j)_{j=1}^S$ be an i.i.d. sample drawn from Ξ. Given $p \geq 2$, suppose that Assumption 2.1 holds and q is in $[p,2p]$ such that the integrability condition in Assumption 2.1 is satisfied. Let $\hat{\varepsilon}(\cdot,\cdot)$ be defined in Definition (2.2) and C_q be defined in Lemma 2.3. Set $C_2 = 1$ if $q = p = 2$. It holds that, for all $x \in X$,*

$$\left|\|\hat{\varepsilon}(x,\xi_S)\|\right|_q \leq C_q \frac{L_q\|x-z\| + \sigma_q(z)}{\sqrt{S}},\ \forall z \in X.$$

Lemma 2.5. *[18] Assume that the solution set, $S(X,G)$, of SVI (1.1) is $\emptyset$. Suppose that Assumption 2.1 holds. Let $\xi := \{\xi_j\}_{j=1}^{S}$ be an i.i.d sample drawn from Ξ and $\gamma : \Xi \to [0,\hat{\gamma}]$ be a random variable for some $0 < \hat{\gamma} \leq 1$. Let $(\gamma, x) \in [0,\hat{\gamma}] \times X$ and $y(x,\gamma,\xi) := \Pi_X(x - \gamma\hat{F}(x,\xi))$. There exist positive constants $\{c_j\}_{j=1}^{3}$ such that, for any $x \in X$,*

$$|||\varepsilon(y(x,\gamma,\xi),\xi)|||_2 \leq \frac{c_1\sigma_4(x^*) + (c_2L_2 + c_3L_4)\|x - x^*\|}{\sqrt{S}}, \forall x^* \in S(X,G).$$

Lemma 2.6. *[33] Let the filtration $\{\mathscr{F}_n : n \in \mathbb{N}_0\}$ be a collection of σ-fields satisfying $\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \cdots \subseteq \mathscr{F}_n \subseteq \cdots \subseteq \mathscr{F}$. Let $\{\mu_n\}$, $\{\nu_n\}$, $\{\alpha_n\}$, and $\{\beta_n\}$ be sequences of nonnegative random variables adapted to the filtration $\{\mathscr{F}_n\}$. Suppose that a.s. $\sum_{n=0}^{\infty}\alpha_n < \infty$, a.s. $\sum_{n=0}^{\infty}\beta_n < \infty$. For each $n \in \mathbb{N}_0$, it holds that $\mathbb{E}[\nu_{k+1} \mid \mathscr{F}_n] \leq (1+\alpha_n)\nu_n - \mu_n + \beta_n$. Then, $\{\nu_n\}$ a.s. converges as $n \to \infty$ and a.s. $\sum_{n=0}^{\infty}\mu_n < \infty$.*

3. The Algorithm and Its Convergence Analysis

In this section, we propose an SA-based projection and contraction algorithm with inertial effects for solving nonmonotone SVI (1.1).

Algorithm 1

Step 1: Choose $x_0, x_1 \in \mathbb{R}^d$. Take $\zeta_n \in (0,\infty), \forall n \in \mathbb{N}$ and $\theta \in (0,\infty)$. Take μ, ν, $\lambda \in (0,1)$ and $\tau \in \left(0, \frac{2(1-\lambda)}{1+\lambda}\right)$. Take the sample rate $\{S_n\} \subset \mathbb{N}$, and set $n = 1$.

Step 2: Compute $w_n = x_n + \eta_n(x_n - x_{n-1})$, where

$$\eta_n = \begin{cases} \min\left\{\theta, \frac{\zeta_n}{\|x_n - x_{n-1}\|}\right\}, & \text{if } x_n \neq x_{n-1}, \\ \theta, & \text{otherwise.} \end{cases}$$

Step 3: Generate an i.i.d. sample $\xi_n := (\xi_n^j)_{j=1}^{S_n}$ from Ξ and compute

$$y_n = \Pi_X(w_n - \gamma_n(\hat{F}(w_n,\xi_n))),$$

where $\hat{F}(w_n,\xi_n) = \frac{1}{S_n}\sum_{j=1}^{S_n}F(w_n,\xi_n^j)$ and γ_n is the largest $\gamma \in \{\mu\nu^k : k \in \mathbb{N}_0\}$ such that

$$\gamma\|\hat{F}(w_n,\xi_n) - \hat{F}(y_n,\xi_n)\| \leq \lambda\|w_n - y_n\|. \tag{3.1}$$

If $y_n = w_n$, then stop. Otherwise, go to ***Step 4***.

Step 4: Compute $\hat{F}(y_n,\xi_n) = \frac{1}{S_n}\sum_{j=1}^{S_n}F(y_n,\xi_n^j)$ and

$$d_n(w_n,y_n) = (w_n - y_n) - \gamma_n(\hat{F}(w_n,\xi_n) - \hat{F}(y_n,\xi_n)). \tag{3.2}$$

If $d_n(w_n,y_n) = 0$, then stop. Otherwise, go to ***Step 5***.

Step 5: Compute

$$x_{n+1} = w_n - \tau\beta_n d_n(w_n,y_n), \tag{3.3}$$

where $\beta_n = \sqrt{1-\lambda}\frac{\|w_n - y_n\|}{\|d_n(w_n,y_n)\|}$. Set $n := n+1$ and return back to ***Step 2***.

Remark 3.1. (i) The inertial terminology is believed to improve the computational performance of the algorithm.

(ii) The adaptive step size usage avoids the requirement of a priori Lipschitz constant of the associated operator.
(iii) The implementation of Algorithm 1 needs to compute the projection onto the feasible set X only once in each iteration.

In the convergence analysis, the following assumption is considered.

Assumption 3.1. (i) $F : \mathbb{R}^d \times \Xi \to \mathbb{R}^d$ is a Garathéodory operator;
(ii) $M(X,F) \neq \emptyset$;
(iii) $S_n \in (0,\infty)$, $\forall n \in \mathbb{N}$ and $\sum_{n=1}^{\infty} \frac{1}{\sqrt{S_n}} < \infty$;
(iv) $\zeta_n \in (0,\infty)$, $\forall n \in \mathbb{N}$ and $\sum_{k=1}^{\infty} \zeta_n < \infty$.

We respectively define two oracle errors $\{\varepsilon_n^{'}\}$ and $\{\varepsilon_n^{''}\}$ by

$$\varepsilon_n^{'} := \hat{F}(w_n,\xi_n) - G(w_n) \text{ and } \varepsilon_n^{''} := \hat{F}(y_n,\xi_n) - G(y_n), \forall n \in \mathbb{N}.$$

Define the σ-algebras by $\mathscr{F}_0 := \sigma(x_0)$, $\mathscr{F}_1 := \sigma(x_0,x_1)$, and $\mathscr{F}_n := \sigma(x_0,x_1,\xi_1,\cdots,\xi_{n-1})$ for all $n > 1$. We observe that by induction, $x_n \in \mathscr{F}_n$, $w_n \in \mathscr{F}_n$, $y_n \in \mathscr{F}_{n+1}$ and $\gamma_n \in \mathscr{F}_{n+1}$.

The following lemmas are essential in establishing the almost sure convergence of Algorithm 1.

Lemma 3.1. *[18] Suppose that Assumption 2.1 holds. Then the line search rule (3.1) in Algorithm 1 terminates after a finite number of iterations. Define $L_n := \frac{1}{S_n}\sum_{j=1}^{S_n} \mathscr{L}(\xi_n^j)$. If Algorithm 1 does not stop at iteration $n+1$, then a.s. $\gamma_n \geq \min\{\mu, \frac{\lambda\nu}{L_n}\}$. Moreover, $|\gamma_n \mid \mathscr{F}_n|_2 \cdot |\mathscr{L}(\xi)|_2 \geq \min\{\mu,\lambda\nu\}$.*

Lemma 3.2. *Let $\{d_n\}$ and $\{\beta_n\}$ be two sequences generated by Algorithm 1. Then*
(i) $d_n(w_n,y_n) = 0$ if and only if $w_n = y_n$;
(ii) $\beta_n \in \left(\frac{1-\lambda}{1+\lambda}, 1\right)$, $\forall n \in \mathbb{N}$.

Proof. (i) By the definition of d_n, we have that

$$\begin{aligned} \|w_n - y_n\| - \gamma_n\|\hat{F}(w_n,\xi_n) - \hat{F}(y_n,\xi_n)\| &\leq \|d_n(w_n,y_n)\|, \\ \|w_n - y_n\| + \gamma_n\|\hat{F}(w_n,\xi_n) - \hat{F}(y_n,\xi_n)\| &\geq \|d_n(w_n,y_n)\|. \end{aligned} \tag{3.4}$$

As a consequence of (3.4), we have that

$$(1-\lambda)\|w_n - y_n\| \leq \|d_n(w_n,y_n)\| \leq (1+\lambda)\|w_n - y_n\|. \tag{3.5}$$

In view of (3.5) and $\lambda \in (0,1)$, we obtain that $d_n(w_n,y_n) = 0$ if and only if $w_n = y_n$.

(ii) If $d_n(w_n,y_n) \neq 0$, by using (3.5) and the definition of β_n, one sees that

$$\frac{1-\lambda}{1+\lambda} \leq \beta_n \leq 1.$$

□

Proposition 3.1. *Suppose that Assumptions 2.1 and 3.1 hold. Let $\{x_n\}$ be a sequence generated by Algorithm 1. Then, for any $z \in M(X,G)$, it holds that*

$$\begin{aligned} \|x_{n+1} - z\|^2 \leq &(1+\zeta_n)\|x_n - z\|^2 + \zeta_n(1+\zeta_n) + \tau(\tau - 2\beta_n)(1-\lambda)\|w_n - y_n\|^2 \\ &- 2\tau\beta_n\gamma_n\langle\varepsilon_n^{''}, y_n - z\rangle. \end{aligned} \tag{3.6}$$

Proof. By using (3.3), we have that

$$\begin{aligned}\|x_{n+1}-z\|^2 =&\|w_n-\tau\beta_n d_n(w_n,y_n)-z\|^2\\ =&\underbrace{\|w_n-z\|^2}_{\text{Term 1}}-2\tau\beta_n\underbrace{\langle d_n(w_n,y_n),w_n-z\rangle}_{\text{Term 2}}+\tau^2\beta_n^2\|d_n(w_n,y_n)\|^2.\end{aligned}\tag{3.7}$$

Now, we estimate the Term 1 in (3.7). From the definition of x_n, we have that

$$\begin{aligned}\text{Term 1} =&\|x_n+\eta_n(x_n-x_{n-1})-z\|^2\\ \le&\|x_n-z\|^2+\|\eta_n(x_n-x_{n-1})\|^2+2\eta_n\|x_n-x_{n-1}\|\|x_n-z\|\\ \le&\|x_n-z\|^2+\zeta_n^2+2\zeta_n\|x_n-z\|\\ \le&(1+\zeta_n)\|x_n-z\|^2+\zeta_n(1+\zeta_n).\end{aligned}\tag{3.8}$$

Next, we estimate the Term 2 in (3.7). By combining (3.1) with (3.2), we obtain that

$$\begin{aligned}\langle d_n(w_n,y_n),w_n-y_n\rangle =&\langle w_n-y_n-\gamma_n(\hat{F}(w_n,\xi_n)-\hat{F}(y_n,\xi_n)),w_n-y_n\rangle\\ =&\|w_n-y_n\|^2-\gamma_n\langle\hat{F}(w_n,\xi_n)-\hat{F}(y_n,\xi_n),w_n-y_n\rangle\\ \ge&(1-\lambda)\|w_n-y_n\|^2.\end{aligned}\tag{3.9}$$

By recalling the definition of y_n and using Lemma 2.1 (i), we have that

$$\langle w_n-\gamma_n(\hat{F}(w_n,\xi_n))-y_n,y_n-z\rangle\ge 0.\tag{3.10}$$

Since $z\in M(X,G)$ and $y_n\in X$, one gets that $\langle G(y_n),y_n-z\rangle\ge 0$. By using the fact that $G(y_n)=\hat{F}(y_n,\xi_n)-\varepsilon_n''$, one further derives that

$$\langle\hat{F}(y_n,\xi_n),y_n-z\rangle\ge\langle\varepsilon_n'',y_n-z\rangle.\tag{3.11}$$

By using (3.10), (3.11), and the definition of $d_n(w_n,y_n)$, one sees that

$$\langle d_n(w_n,y_n),y_n-z\rangle\ge\gamma_n\langle\varepsilon_n'',y_n-z\rangle.\tag{3.12}$$

By summing up both sides of (3.9) and (3.12), we obtain that

$$\begin{aligned}\text{Term 2} =&\langle d_n(w_n,y_n),w_n-y_n\rangle+\langle d_n(w_n,y_n),y_n-z\rangle\\ \ge&(1-\lambda)\|w_n-y_n\|^2+\gamma_n\langle\varepsilon_n'',y_n-z\rangle.\end{aligned}\tag{3.13}$$

By using (3.7), (3.8), (3.13), and the definition of β_n, we have that

$$\begin{aligned}&\|x_{n+1}-z\|^2\\ \le&\|w_n-z\|^2-2\tau\beta_n\left[(1-\lambda)\|w_n-y_n\|^2+\gamma_n\langle\varepsilon_n'',y_n-z\rangle\right]+\tau^2(1-\lambda)\|w_n-y_n\|^2\\ \le&(1+\zeta_n)\|x_n-z\|^2+\zeta_n(1+\zeta_n)+\tau(\tau-2\beta_n)(1-\lambda)\|w_n-y_n\|^2-2\tau\beta_n\gamma_n\langle\varepsilon_n'',y_n-z\rangle.\end{aligned}$$

We hence have the desired result. □

Proposition 3.2. *Under Assumptions 2.1 and 3.1, let $\{x_n\}$ be an infinite sequence generated by Algorithm 1. Let $z\in M(X,G)$ be arbitrarily chosen. Let c_1,c_2, and c_3 be the same as in*

Lemma 2.5. We make use of the following definitions for the sake of clarity:

$$\begin{aligned}
&\chi := \tau\left(\frac{2(1-\lambda)}{1+\lambda} - \tau\right)(1-\lambda),\\
&p := \max\{2\tau\mu(1+\mu L), c_1L_2 + c_2L_4, 2\tau\mu^2L_2\},\\
&\varpi := 4\chi\mu^2(L_2)^2 + 8p^2,\\
&q(z) := \max\{1, c_3\sigma_4(z), 2\tau\mu^2\sigma_2(z), c_3\sigma_4(z)\},\\
&\varphi(z) := 2\chi\mu^2\sigma_2(z)^2 + 4q(z)^2.
\end{aligned}$$

Then, for any $n \in \mathbb{N}$,

$$\begin{aligned}
&\mathbb{E}[\|x_{n+1} - z\|^2 \mid \mathscr{F}_n]\\
&\leq \left(1 + \zeta_n + \frac{\varpi}{\sqrt{S_n}}\right)\|x_n - z\|^2 + \zeta_n(1+\zeta_n) + \frac{\varpi\zeta_n^2 + \varphi(z)}{\sqrt{S_n}} - \frac{\chi}{2}\mathbb{E}[\gamma_n^2\mathscr{R}(w_n)^2 \mid \mathscr{F}_n].
\end{aligned} \tag{3.14}$$

Proof. From the definition of y_n and Lemma 2.1 (ii) and (iv), we deduce that

$$\begin{aligned}
\|y_n - z\| &= \left\|\Pi_X(w_n - \gamma_n(\hat{F}(w_n, \xi_n))) - \Pi_X(z - \gamma_n G(z))\right\|\\
&\leq \left\|w_n - \gamma_n(\hat{F}(w_n, \xi_n)) - (z - \gamma_n G(z))\right\|\\
&\leq \|w_n - z\| + \gamma_n\|G(w_n) - G(z)\| + \gamma_n\|\varepsilon_n^{'}\|\\
&\leq (1 + \gamma_n L)\|w_n - z\| + \gamma_n\|\varepsilon_n^{'}\|.
\end{aligned} \tag{3.15}$$

By using (3.15), Lemma 3.2 (ii), and $\gamma_n \in (0, \mu)$, we concludes that

$$\begin{aligned}
-2\tau\beta_n\gamma_n\langle\varepsilon_n^{''}, y_n - z\rangle &\leq 2\tau\mu\|\varepsilon_n^{''}\|\|y_n - z\|\\
&\leq 2\tau\mu\|\varepsilon_n^{''}\|\left((1+\mu L)\|w_n - z\| + \mu\|\varepsilon_n^{'}\|\right).
\end{aligned} \tag{3.16}$$

Taking $\mathbb{E}[\cdot \mid \mathscr{F}_n]$ in (3.16) and using Hölder's inequality, we see that

$$\begin{aligned}
&\mathbb{E}[-2\tau\beta_n\gamma_n\langle\varepsilon_n^{''}, y_n - z\rangle]\\
&\leq 2\tau\mu\mathbb{E}[(1+\mu L)\|\varepsilon_n^{''}\|\|w_n - z\| + \mu\|\varepsilon_n^{'}\|\|\varepsilon_n^{''}\| \mid \mathscr{F}_n]\\
&\leq 2\tau\mu\left[(1+\mu L)\|w_n - z\|\mathbb{E}[\|\varepsilon_n^{''}\| \mid \mathscr{F}_n] + \mu\mathbb{E}[\|\varepsilon_n^{'}\|\|\varepsilon_n^{''}\| \mid \mathscr{F}_n]\right]\\
&\leq 2\tau\mu(1+\mu L)\|w_n - z\|\||\varepsilon_n^{''}\| \mid \mathscr{F}_n|_2 + 2\tau\mu^2\||\varepsilon_n^{'}\| \mid \mathscr{F}_n|_2\||\varepsilon_n^{''}\| \mid \mathscr{F}_n|_2.
\end{aligned} \tag{3.17}$$

Since $w_n \in \mathscr{F}_n$ and $\xi_n \perp\!\!\!\perp \mathscr{F}_n$, we find from Lemma 2.4 that

$$\||\varepsilon_n^{'}\| \mid \mathscr{F}_n|_2 \leq \frac{L_2\|w_n - z\| + \sigma_2(z)}{\sqrt{S_n}}. \tag{3.18}$$

Since $w_n \in \mathscr{F}_n$, $\xi_n \perp\!\!\!\perp \mathscr{F}_n$, $\gamma_n \in (0, \mu] \subset (0,1)$, and $y_n = y(w_n, \gamma_n, \xi_n)$, we see that Lemma 2.5 yields that

$$\left\|\|\varepsilon_n^{''}\| \mid \mathscr{F}_n\right|_2 \leq \frac{(c_1L_2 + c_2L_4)\|w_n - z\| + c_3\sigma_4(z)}{\sqrt{S_n}}, \tag{3.19}$$

where c_1, c_2, and c_3 are given in Lemma 2.5. The definition of w_n asserts that

$$\|w_n - z\| \leq \|x_n - z\| + \zeta_n. \tag{3.20}$$

Putting together (3.17), (3.18), (3.19), and (3.20), we obtain that

$$\begin{aligned}
&\mathbb{E}[-2\tau\beta_n\gamma_n\langle\varepsilon_n^{''}, y_n - z\rangle] \\
\leq& 2\tau\mu(1+\mu L)\|w_n - z\|\frac{(c_1L_2 + c_2L_4)\|w_n - z\| + c_3\sigma_4(z)}{\sqrt{S_n}} \\
&+ 2\tau\mu^2\frac{L_2\|w_n - z\| + \sigma_2(z)}{\sqrt{S_n}} \times \frac{(c_1L_2 + c_2L_4)\|w_n - z\| + c_3\sigma_4(z)}{\sqrt{S_n}} \\
\leq& \frac{4p^2\|w_n - z\|^2 + 4q(z)^2}{\sqrt{S_n}} \\
\leq& \frac{8p^2\|x_n - z\|^2 + 4q(z)^2 + 8p^2\zeta_n^2}{\sqrt{S_n}},
\end{aligned} \tag{3.21}$$

where

$$p := \max\{2\tau\mu(1+\mu L), c_1L_2 + c_2L_4, 2\tau\mu^2L_2\}$$

and

$$q(z) := \max\{1, c_3\sigma_4(z), 2\tau\mu^2\sigma_2(z), c_3\sigma_4(z)\}.$$

By the definition of the natural residual function and the definition of y_n, we have that

$$\begin{aligned}
&\gamma_n^2\mathscr{R}(w_n)^2 \leq \mathscr{R}_{\gamma_n}(w_n)^2 = \|w_n - \Pi_X(w_n - \gamma_nG(w_n))\|^2 \\
\leq& 2\|w_n - y_n\|^2 + 2\|\Pi_X(w_n - \gamma_nG(w_n)) - \Pi_X(w_n - \gamma_n(G(w_n) + \varepsilon_n^{'}))\|^2 \\
\leq& 2\|w_n - y_n\|^2 + 2\gamma_n^2\|\varepsilon_n^{'}\|^2.
\end{aligned} \tag{3.22}$$

Noting that $\gamma_n \in (0, \mu]$, it follows from (3.22) that

$$-\|w_n - y_n\|^2 \leq -\frac{1}{2}\gamma_n^2\mathscr{R}(w_n)^2 + \mu^2\|\varepsilon_n^{'}\|^2. \tag{3.23}$$

Recalling (3.8) and (3.18), one sees that

$$\begin{aligned}
\mathbb{E}[\|\varepsilon_n^{'}\|^2 \mid \mathscr{F}_n] \leq& \frac{2(L_2)^2\|w_n - z\|^2 + 2\sigma_2(z)^2}{S_n} \\
\leq& \frac{4(L_2)^2\|x_n - z\|^2 + 4(L_2)^2\zeta_n^2 + 2\sigma_2(z)^2}{S_n}.
\end{aligned} \tag{3.24}$$

Taking $\mathbb{E}[\cdot \mid \mathscr{F}_n]$ in (3.6) and using (3.21), (3.23), (3.24) and Lemma 3.2 (ii), one finds that

$$\begin{aligned}
&\mathbb{E}[\|x_{n+1} - z\|^2 \mid \mathscr{F}_n] \\
\leq& (1+\zeta_n)\|x_n - z\|^2 + \zeta_n(1+\zeta_n) - \chi\mathbb{E}[\|w_n - y_n\|^2 \mid \mathscr{F}_n] + \mathbb{E}[-2\tau\beta_n\gamma_n\langle\varepsilon_n^{''}, y_n - z\rangle \mid \mathscr{F}_n] \\
\leq& (1+\zeta_n)\|x_n - z\|^2 + \zeta_n(1+\zeta_n) - \frac{\chi}{2}\mathbb{E}[\gamma_n^2\mathscr{R}(w_n)^2 \mid \mathscr{F}_n] \\
&+ 2\chi\mu^2\frac{2(L_2)^2\|x_n - z\|^2 + 2(L_2)^2\zeta_n^2 + \sigma_2(z)^2}{S_n} \\
&+ \frac{8p^2\|x_n - z\|^2 + 4q(z)^2 + 8p^2\zeta_n^2}{\sqrt{S_n}} \\
=& \left(1 + \zeta_n + \frac{\varpi}{\sqrt{S_n}}\right)\|x_n - z\|^2 + \zeta_n(1+\zeta_n) + \frac{\varpi\zeta_n^2 + \varphi(z)}{\sqrt{S_n}} - \frac{\chi}{2}\mathbb{E}[\gamma_n^2\mathscr{R}(w_n)^2 \mid \mathscr{F}_n],
\end{aligned}$$

where $\chi := \tau\left(\frac{2(1-\lambda)}{1+\lambda} - \tau\right)(1-\lambda)$, $\varpi := 4\chi\mu^2(L_2)^2 + 8p^2$, and $\varphi(z) := 2\chi\mu^2\sigma_2(z)^2 + 4q(z)^2$. The proof is completed. □

Theorem 3.1. *Let Assumptions 2.1 and 3.1 hold. Then, either Algorithm 1 stops at iteration $n+1$, that is, w_n is a solution in $S(X,G)$, or Algorithm 1 generates an infinite sequence $\{x_n\}$, that is, every cluster point of $\{x_n\}$ a.s. belongs to $S(X,G)$.*

Proof. If Algorithm 1 stops at iteration $n+1$, one sees that $w_n = \Pi_X(w_n - \mu(\hat{F}(w_n,\xi_n)))$. This together with Lemma 2.1 (i) implies that, for all $x \in X$,

$$\langle \hat{F}(w_n,\xi_n), x - w_n\rangle \geq 0. \tag{3.25}$$

Noting $\mathbb{E}[\varepsilon_n'|\mathscr{F}_n] = 0$, $\hat{F}(w_n,\xi_n) = \varepsilon_n' + G(w_n)$, and $w_n \in \mathscr{F}_n$ and taking $\mathbb{E}[\cdot|\mathscr{F}_n]$ in (3.25), one obtains that a.s. $\langle F(w_n), x - w_n\rangle \geq 0$ for all $x \in X$. Hence, a.s. $w_n \in S(X,F)$. Noting that $\lambda \in (0,1)$, $\tau \in \left(0, \frac{2(1-\lambda)}{1+\lambda}\right)$, $\sum_{n=1}^{\infty} \frac{1}{\sqrt{S_n}} < \infty$, and $\sum_{n=1}^{\infty} \zeta_n < \infty$ and using Lemma 2.6 and (3.14), one obtains that a.s. $\{\|x_n - z\|^2\}$ converges and

$$\sum_{n=1}^{\infty} \mathbb{E}[\gamma_n^2 \mathscr{R}(w_n)^2 \mid \mathscr{F}_n] < \infty. \tag{3.26}$$

Accordingly, we further obtain that a.s. $\{\|x_n - z\|^2\}$ is bounded. Hence, for every cluster point $\hat{x}$ of $\{x_n\}$, we can construct a subsequence $\{x_{n_l}\} \subset \{x_n\}$ such that a.s. $\lim_{l\to\infty} x_{n_l} = \hat{x}$. By recalling the definition of w_n and ζ_n, one asserts that

$$\|w_n - x_n\| = \eta_n \|x_n - x_{n-1}\| \leq \zeta_n.$$

This together with $\sum_{n=1}^{\infty} \zeta_n < \infty$ yields that $\lim_{n\to\infty} \|w_n - x_n\| = 0$. Noticing that a.s. $\lim_{l\to\infty} x_{n_l} = \hat{x}$, we derives that $\{w_{n_l}\} \subset \{w_n\}$ satisfies that a.s. $\lim_{l\to\infty} w_{n_l} = \hat{x}$. Lemma 3.1 asserts that

$$\mathbb{E}[\gamma_n^2 \mid \mathscr{F}_n] \geq \left(\frac{\min\{\mu, \lambda\nu\}}{|\mathscr{L}(\xi)|_2}\right)^2 > 0. \tag{3.27}$$

Noticing that $w_n \in \mathscr{F}_n$ and combining (3.26) with (3.27), we obtain that

$$\lim_{n\to\infty} \mathbb{E}[\mathscr{R}(w_n)^2 \mid \mathscr{F}_n] = 0. \tag{3.28}$$

By replacing n with n_l in (3.28), one sees that $\lim_{l\to\infty} \mathbb{E}[\mathfrak{R}(w_{n_l})^2] = 0$, i.e.,

$$\lim_{l\to\infty} \mathbb{E}[\|w_{n_l} - \Pi_X(w_{n_l} - G(w_{n_l}))\|^2] = 0.$$

Hence, the continuity of $G(\cdot)$ and $\Pi_X(\cdot)$ asserts that $\mathfrak{R}(\hat{x})^2 = \|\hat{x} - \Pi_X(\hat{x} - G(\hat{x}))\|^2 = 0$. From this, we obtain that every cluster point $\hat{x}$ of $\{x_n\}$ satisfies $\mathfrak{R}(\hat{x})^2 = 0$. From Lemma 2.1 (iv), one obtains that $\hat{x} \in S(X,G)$. The proof is completed. □

4. Complexity Analysis and Rates

In this section, we give explicit estimates on the convergence rate and the oracle complexity of Algorithm 1.

Theorem 4.1. *Let Assumptions 2.1 and 3.1 hold. Let $\{x_n\}$ be a sequence generated by Algorithm 1. Let c_1, c_2, and c_3 be the same as in Lemma 2.5. Let $\chi, p, \varpi, q(z)$, and $\varphi(z)$ be the same as in Proposition 3.2. Let $\sigma_4(\cdot)$ be defined in (2.1). Give two constants $u \in (0,\infty)$ and $v \in (0,\infty)$ such that $0 < v + \varpi u < 1$, and choose $n_0 \in \mathbb{N}$ such that $\sum_{n\geq n_0} \frac{1}{\sqrt{S_n}} \leq u$ and $\sum_{n\geq n_0} \zeta_n \leq v$. Given $z \in M(X,G)$, define, for all $n \in \mathbb{N}$,*

$$\rho := \frac{\chi}{2}\left(\frac{\min\{\mu, \lambda\nu\}}{|\mathscr{L}(\xi)|_2}\right)^2,$$
$$\mathscr{B}(z,u,v,n_0) := \frac{\mathbb{E}[\|x_{n_0} - z\|^2] + v(1+v) + [\varpi v^2 + \varphi(z)]u}{1 - v - \varpi u},$$
$$\mathscr{D}(z,u,v,n_0) := \mathbb{E}[\|x_0 - z\|^2] + (v + \varpi u)\,\mathscr{B}(z,u,v,n_0) + v(1+v) + [\varpi v^2 + \varphi(z)]u,$$
$$\mathscr{U}(z,u,v,n_0) := (9 + 6L^2)v^2 + \frac{3}{\rho}\mathscr{D}(z,u,v,n_0).$$

Then $\sup_{n\geq 0}\mathbb{E}[\|x_n - z\|^2] \leq \mathscr{B}(z,s,t,n_0)$. Furthermore, for any $\varepsilon > 0$, there exists $N_\varepsilon \in \mathbb{N}$ such that $N_\varepsilon = 1$ or $\mathbb{E}[\mathscr{R}(x_{N_\varepsilon})^2] \leq \varepsilon \leq \frac{\mathscr{U}(z,n_0,u,v)}{N_\varepsilon}$.

Proof. Since $w_n \in \mathscr{F}_n$, one concludes by using (3.27) obtained in Theorem 3.1 that

$$\mathbb{E}[\mathbb{E}[\gamma_n^2 \mathscr{R}(w_n)^2 \mid \mathscr{F}_n]] = \mathbb{E}[\mathbb{E}[\gamma_n^2 \mid \mathscr{F}_n]\mathscr{R}(w_n)^2] \geq \left(\frac{\min\{\mu, \lambda\nu\}}{|\mathscr{L}(\xi)|_2}\right)^2 \mathbb{E}[\mathscr{R}(w_n)^2].$$

By recalling Proposition 3.2 and taking $\mathbb{E}[\cdot]$ in (3.14), one obtains that

$$\begin{aligned}&\mathbb{E}[\|x_{n+1} - z\|^2]\\ &\leq \left(1 + \zeta_n + \frac{\varpi}{\sqrt{S_n}}\right)\mathbb{E}[\|x_n - z\|^2] + \zeta_n(1+\zeta_n) + \frac{\varpi\zeta_n^2 + \varphi(z)}{\sqrt{S_n}} - \rho\mathbb{E}[\mathscr{R}(w_n)^2].\end{aligned} \tag{4.1}$$

where $\rho := \frac{\chi}{2}\left(\frac{\min\{\mu,\lambda\nu\}}{|\mathscr{L}(\xi)|_2}\right)^2$. Define $\psi(\mathscr{A}) := \inf\{n \geq n_0 + 1 : \mathbb{E}[\|x_n - z\|^2] \geq \mathscr{A}\}$ for any $\mathscr{A} > \mathbb{E}[\|x_{n_0} - z\|^2]$. Now we consider the following two cases

Case 1: Suppose that $\psi(\mathscr{A}) = \infty$. Then $\sup_{n\geq n_0+1}\mathbb{E}[\|x_{n_0} - z\|^2] \leq \mathscr{A} < \infty$.

Case 2: Suppose that $\psi(\mathscr{A}) < \infty$. In view of (4.1) and the definition of $\mathscr{A}$, one sees that

$$\begin{aligned}\mathscr{A} \leq& \mathbb{E}[\|x_{\psi(\mathscr{A})} - z\|^2]\\ \leq& \mathbb{E}[\|x_{n_0} - z\|^2] + \sum_{n=n_0}^{\psi(\mathscr{A})-1}\left(\zeta_n + \frac{\varpi}{\sqrt{S_n}}\right)\mathscr{A} + \sum_{n=n_0}^{\psi(\mathscr{A})-1}\left[\zeta_n(1+\zeta_n) + \frac{\varpi\zeta_n^2 + \varphi(z)}{\sqrt{S_n}}\right]\\ \leq& \mathbb{E}[\|x_{n_0} - z\|^2] + (v + \varpi u)\,\mathscr{A} + v(1+v) + [\varpi v^2 + \varphi(z)]u.\end{aligned}$$

Thus

$$\mathscr{A} \leq \frac{\mathbb{E}[\|x_{n_0} - z\|^2] + v(1+v) + [\varpi v^2 + \varphi(z)]u}{1 - v - \varpi u} := \mathscr{B}(z,u,v,n_0). \tag{4.2}$$

It is noted that if $\mathscr{A} \to \infty$, then (4.2) yields a contradiction. The argument above implies that any threshold Θ, which $\{\mathbb{E}[\|x_n - z\|^2]\}_{n\geq n_0+1}$ eventually exceeds, is bounded above by $\mathscr{B}(z,u,v,n_0)$, that is, $\sup_{n\geq n_0+1}\mathbb{E}[\|x_n - z\|^2] \leq \mathscr{B}(z,u,v,n_0)$.

On the other hand, since $\mathscr{A} > \mathbb{E}[\|x_{n_0} - z\|^2]$ and $v + \mathscr{U}(z)(1+v)u \in (0,1)$, we obtain that $\sup_{1\le i\le n_0} \mathbb{E}[\|x_n - z\|^2] \le \mathscr{B}(z,u,v,n_0)$. Then

$$\sup_{n\ge 1} \mathbb{E}[\|x_n - z\|^2] \le \mathscr{B}(z,u,v,n_0).$$

Next, we estimate the convergence rate for the sequence $\{\mathscr{R}(x_n)^2\}$. In view of (4.1), one deduces that

$$\begin{aligned}
&\rho \sum_{n=1}^{k} \mathbb{E}[\mathscr{R}(w_n)^2] \\
\le& \mathbb{E}[\|x_1 - z\|^2] + \sum_{n=1}^{k} \left(\zeta_n + \frac{\varpi}{\sqrt{S_n}}\right) \mathbb{E}[\|x_n - z\|^2] + \sum_{n=1}^{k} \left[\zeta_n(1+\zeta_n) + \frac{\varpi\zeta_n^2 + \varphi(z)}{\sqrt{S_n}}\right] \\
\le& \mathbb{E}[\|x_1 - z\|^2] + (v + \varpi u)\,\mathscr{B}(z,u,v,n_0) + v(1+v) + [\varpi v^2 + \varphi(z)]u \\
:=& \mathscr{D}(z,u,v,n_0).
\end{aligned}$$

Recalling the Lipschitz continuity of $G(\cdot)$ and Lemma 2.1 (ii), we have that

$$\begin{aligned}
\mathscr{R}(x_n)^2 =& \|x_n - \Pi_X(x_n - G(x_n))\|^2 \\
\le& 3\|x_n - w_n\|^2 + 3\mathscr{R}(w_n)^2 + 3\|\Pi_X(w_n - G(w_n)) - \Pi_X(x_n - G(x_n))\|^2 \\
\le& 3\|x_n - w_n\|^2 + 3\mathscr{R}(w_n)^2 + 3\|(w_n - x_n) - (G(w_n) - G(x_n))\|^2 \\
\le& 9\|x_n - w_n\|^2 + 3\mathscr{R}(w_n)^2 + 6\|G(w_n) - G(x_n)\|^2 \\
\le& (9 + 6L^2)\eta_n^2 \|x_n - x_{n-1}\|^2 + 3\mathscr{R}(w_n)^2 \\
\le& (9 + 6L^2)\zeta_n^2 + 3\mathscr{R}(w_n)^2.
\end{aligned} \tag{4.3}$$

The fact of $\sum_{n=1}^{\infty} \zeta_n < \infty$ yields that $\lim_{n\to\infty} \zeta_n = 0$. By combining (3.28) with (4.3), we conclude that

$$0 \le \lim_{n\to\infty} \mathbb{E}[\mathscr{R}(x_n)^2] \le (9 + 6L^2) \lim_{n\to\infty} \zeta_n^2 + 3 \lim_{n\to\infty} \mathbb{E}[\mathscr{R}(w_n)^2] = 0. \tag{4.4}$$

By summing (4.3) recursively with n running from 1 to k, one deduces that

$$\sum_{n=1}^{k} \mathscr{R}(x_n)^2 \le (9 + 6L^2) \sum_{n=1}^{k} \zeta_n^2 + 3 \sum_{n=1}^{k} \mathscr{R}(w_n)^2 \le (9 + 6L^2)v^2 + 3 \sum_{n=1}^{k} \mathscr{R}(w_n)^2,$$

which together with (4.3) yields

$$\sum_{n=1}^{k} \mathbb{E}[\mathscr{R}(x_n)^2] \le (9 + 6L^2)v^2 + \frac{3}{\rho}\mathscr{D}(z,u,v,n_0) := \mathscr{U}(z,u,v,n_0). \tag{4.5}$$

Given $\varepsilon > 0$, one defines $N_\varepsilon := \inf\{n \in \mathbb{N} : \mathbb{E}[\mathscr{R}(x_n)^2] \le \varepsilon\}$. This together with (4.4) implies that N_ε is finite for any $\varepsilon > 0$. Suppose that $N_\varepsilon > 1$. In this case, using the definition of N_ε and taking $k := N_\varepsilon$ in (4.5), one deduces that $\varepsilon N_\varepsilon \le \sum_{n=0}^{N_\varepsilon} \mathbb{E}[\mathscr{R}(x_n)^2] \le \mathscr{U}(z,u,v,n_0)$. Hence, one concludes that

$$\mathbb{E}[\mathscr{R}(x_{N_\varepsilon})^2] \le \varepsilon \le \frac{\mathscr{U}(z,u,v,n_0)}{N_\varepsilon}.$$

This completes the proof. □

Theorem 4.2. *Let Assumptions 2.1 and 3.1 hold. Let $z \in M(X,G)$ be chosen arbitrarily. Let $\{x_n\}$ be a sequence generated by Algorithm 1. Let c_1, c_2, and c_3 be the same as in Lemma 2.5. Let $\chi, p, \varpi, q(z)$, and $\varphi(z)$ be the same as in Proposition 3.2. Let ρ be the same as in Theorem 4.1. Let $\sigma_4(\cdot)$ be defined in (2.1). Given $a \in (1,\infty)$ and $N \in \mathbb{N}$, let $S_n := N\lceil (n+2)^{1+a} \rceil$ and $\varsigma_n := \frac{1}{(n+2)^{1+a}}$ for any $n \in \mathbb{N}$. We make use of the following definitions for the sake of clarity:*

$$
\begin{aligned}
&n_0 := \lceil \sqrt[a]{\frac{a(N+\varpi)}{N}} \rceil, \quad \kappa := \frac{a}{N(n_0+1)^a},\\
&\mathscr{B}(z) := \frac{\mathbb{E}[\|x_{n_0} - z\|^2] + \kappa N(1+\kappa N) + [\varpi \kappa^2 N^2 + \varphi(z)]u}{1 - \kappa N - \varpi\kappa},\\
&\mathscr{D}(z) := \mathbb{E}[\|x_0 - z\|^2] + a\left(1 + \frac{\varpi}{N}\right)\mathscr{B}(z) + a(1+a) + [\varpi a^2 + \varphi(z)]\frac{a}{N},\\
&\mathscr{U}(z) := (9 + 6L^2)a^2 + \frac{3}{\rho}\mathscr{D}(z).
\end{aligned}
$$

Given a tolerance level $\varepsilon > 0$, there exists $N_\varepsilon \in \mathbb{N}$ (depending on ε) such that

$$
N_\varepsilon = 1 \text{ or } \mathbb{E}[\mathscr{R}(x_{N_\varepsilon})^2] \le \varepsilon \le \frac{\mathscr{U}(z)}{N_\varepsilon}. \tag{4.6}
$$

After N_ε iterations, a.s. the oracle complexity is bounded above by

$$
\sum_{n=1}^{N_\varepsilon} (1 + \mathbb{E}[k_n]) S_n \le N(1+\sigma)\frac{\mathscr{U}(z)}{\varepsilon}\left[\left(\frac{\mathscr{U}(z)}{\varepsilon} + 2\right)^{1+a} + 1\right].
$$

where k_n is the number of oracle calls used in the line search scheme at iteration n.

Proof. By the definitions of $\{S_n\}$ and $\{\zeta_n\}$, we obtain that, for any $n_0 \in \mathbb{N}$,

$$
\sum_{n \ge n_0} \frac{1}{S_n} \le \frac{a}{N(n_0+1)^a} \quad \text{and} \quad \sum_{n \ge n_0} \zeta_n \le \frac{a}{(n_0+1)^a}. \tag{4.7}
$$

It is enough to choose $n_0 = \lceil \sqrt[a]{\frac{a(N+\varpi)}{N}} \rceil$ such that $0 < v + \varpi u < 1$. Replacing 1 with n_0 in (4.7), one deduces that

$$
\sum_{n \ge 1} \frac{1}{S_n} \le \frac{a}{N} \quad \text{and} \quad \sum_{n \ge 1} \zeta_n \le a. \tag{4.8}
$$

Using (4.7) and (4.8), we obtain (4.6) immediately. Let k_n be the number of line search in iteration n. It follows from Lemma 3.1 that a.s. $\gamma_n = \mu \nu^{k_n} \ge \left(\frac{\lambda\nu}{\mathscr{L}_n}\right) \wedge \mu$. From this, it yields that a.s.

$$
k_n \le \log_{\frac{1}{\nu}} \frac{\mathscr{L}_n \mu}{(\lambda\nu) \wedge \nu}.
$$

By using Jensen's inequality and the concavity of $x \mapsto \log_{\frac{1}{\nu}}$, we give that

$$
\mathbb{E}[k_n] \le \mathbb{E}\left[\log_{\frac{1}{\nu}} \frac{\mu\mathscr{L}_n}{(\lambda\nu)\wedge\nu}\right] \le \log_{\frac{1}{\nu}} \frac{\mu\mathbb{E}[\mathscr{L}_n]}{(\lambda\nu)\wedge\nu} \le \log_{\frac{1}{\nu}} \frac{\mu L}{(\lambda\nu)\wedge\nu} := \sigma. \tag{4.9}
$$

Indeed, (4.6) gives that $N_\varepsilon \leq \frac{\mathscr{U}(z)}{\varepsilon}$. Using the definition of S_n and (4.9), we see that the oracle complexity is upper bounded by

$$\begin{aligned}\sum_{n=1}^{N_\varepsilon}(1+\mathbb{E}[k_n])S_n \leq & \sum_{n=1}^{N_\varepsilon}(1+\mathbb{E}[k_n])N\left[(n+2)^{1+a}+1\right]\\ \leq & N(1+\sigma)\frac{\mathscr{U}(z)}{\varepsilon}\left[\left(\frac{\mathscr{U}(z)}{\varepsilon}+2\right)^{1+a}+1\right].\end{aligned}$$

The proof is completed. □

5. Numerical Examples

In this section, we provide computational experiments and compare the convergence performance between our proposed algorithm and some other existing algorithms in [16, 17, 34]. All the programs are written in Python 3.9 on a PC Desktop Intel(R) Core(TM) i5-11300H @ 3.10 GHz(8 CPUs), 3.1 GHz, RAM 16 384 MB. In all the numerical implementations, the batch size sequence is chosen as $S_n = 2\lceil(n+3)(\ln(n+3)^2\rceil$, for any $n \in \mathbb{N}$.

Example 5.1. Consider the saddle-point problem of the form

$$\max_{x\in X_1}\min_{y\in X_2} g(x,y,\xi) = x^TAy + a(\xi)^Tx + b(\xi)^Ty + \frac{c}{2}\|x\|^2 - \frac{c}{2}\|y\|^2,$$

where $X_1 \subseteq \mathbb{R}^m, X_2 \subseteq \mathbb{R}^m$, A is an $m\times m$ deterministic positive definite matrix, $a(\xi)\in\mathbb{R}^m$ and $b(\xi)\in\mathbb{R}^m$ are random vectors with their i.i.d. entries generated uniformly from $(-1,1)$, and $c>0$. The first order optimality condition can be written as SVI (1.1) with $X = X_1\times X_2$ and $F: X\times\Xi$ given by

$$F(z,\xi) = \begin{pmatrix}\nabla_x g(z,\xi)\\ -\nabla_y g(z,\xi)\end{pmatrix} = \begin{pmatrix}Ay + a(\xi) + cx\\ A^Tx + b(\xi) - cy\end{pmatrix}, \forall z = \begin{pmatrix}x\\ y\end{pmatrix} \in X.$$

It is known that $\mathbb{E}[F(\cdot,\xi)]$ is Lipschitz continuous with $L = \sqrt{\|A\|^2+c^2}$. It is clear that $(0,0)$ is a weak solution to the corresponding SVI. It is noted that $\mathbb{E}[F(\cdot,\xi)]$ is nonmonotone rather than pseudomonotone. z belongs to the constraint set $X = [-10,10]^{2m}$. We use our algorithm to solve this problem. In all tests, the parameters ν and χ are generated randomly in $(0,1)$. Meanwhile, we set $\theta := 3$, $\varsigma_n := \frac{1}{(n+3)^2}$, and τ is chosen randomly in $(0, \frac{2(1-\lambda)}{1+\lambda})$. We use $res(z) := \|x - \Pi_X(z-\hat{F}(z,\xi))\|$ to measure the error of the convergence behavior of our algorithm. In all tests, we choose the maximum iteration number of 200 or $res(z)\leq 0.0005$ as a common termination criterion. Now we test two different dimensions.

Case 1: Set $m = 1$. For this case, we choose $c = 2$ and A is randomly generated from $(2,4)$. The i.i.d. entries of random vectors $a(\xi)$ and $b(\xi)$ are drawn randomly from the standard Gaussian distribution. Two initial points of z are randomly generated on $(0,10)^2$. Figure 1 demonstrates the convergence behavior of the proposed algorithm. To illustrate the computational performance, Figure 2 depicts the isometric view of $g(x,y,\xi)$ in a 3-D space. Obviously, the convergence of $(x,y)^T$ to $(0,0)^T$ implies that the convergence point $(0,0)^T$ is the optimal solution to this saddle-point problem over the set of $[-10,10]^2$.

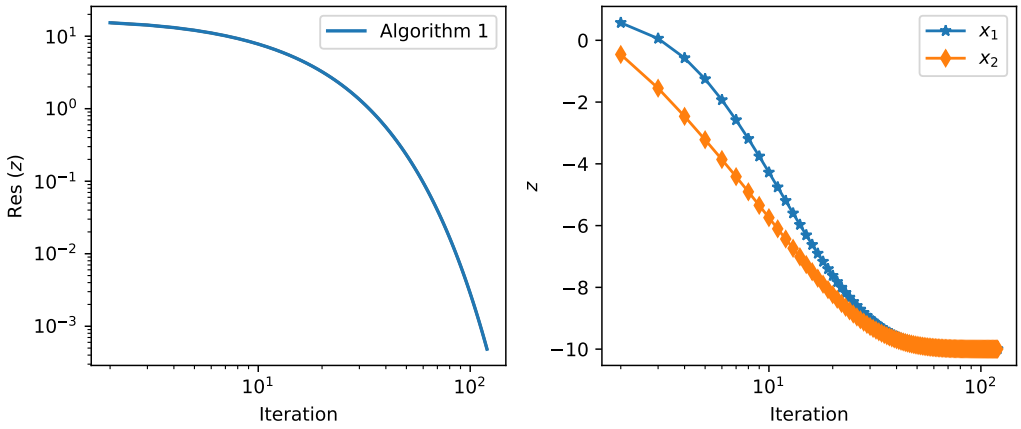

FIGURE 1. Numerical Results for Example 5.1.

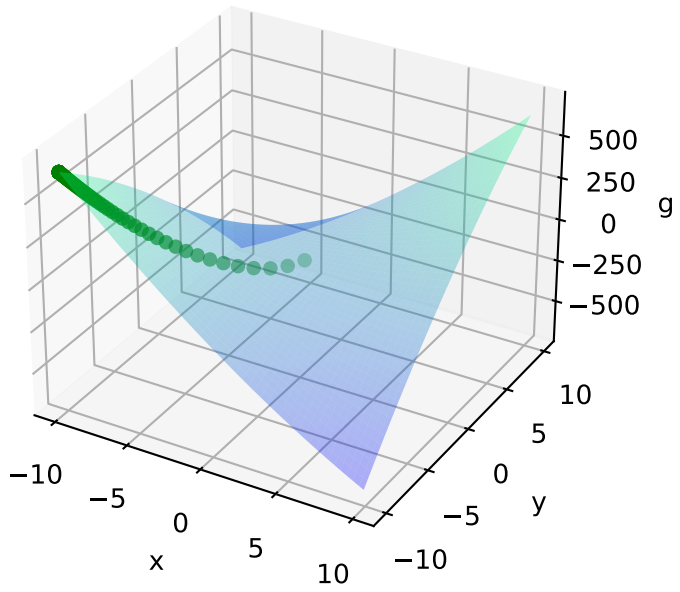

FIGURE 2. The Behavior of (x, y, g) During the Iteration Process.

Case 2: Set $m = 2$. For this experiment, we choose $c = 2$ and all the entries of A are randomly generated from $(2, 16)$. The i.i.d. entries of random vectors $a(\xi)$ and $b(\xi)$ are generated at random from the standard Gaussian distribution. We randomly choose two starting points of z in the range of $(0, 1)^8$. Figures 3 and 4 demonstrate the numerical behaviors of our algorithm with two different choices of starting points.

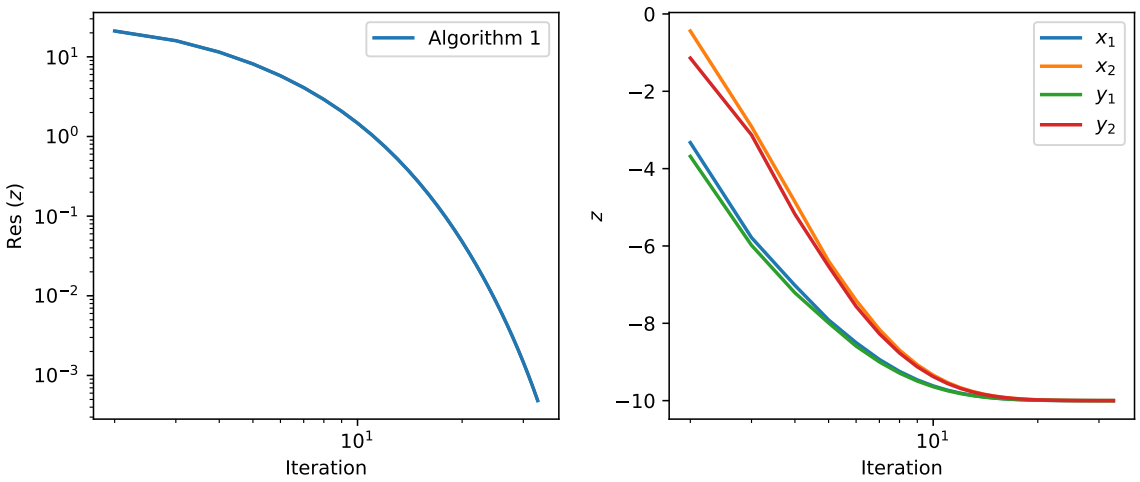

FIGURE 3. Numerical Results for Example 5.1.

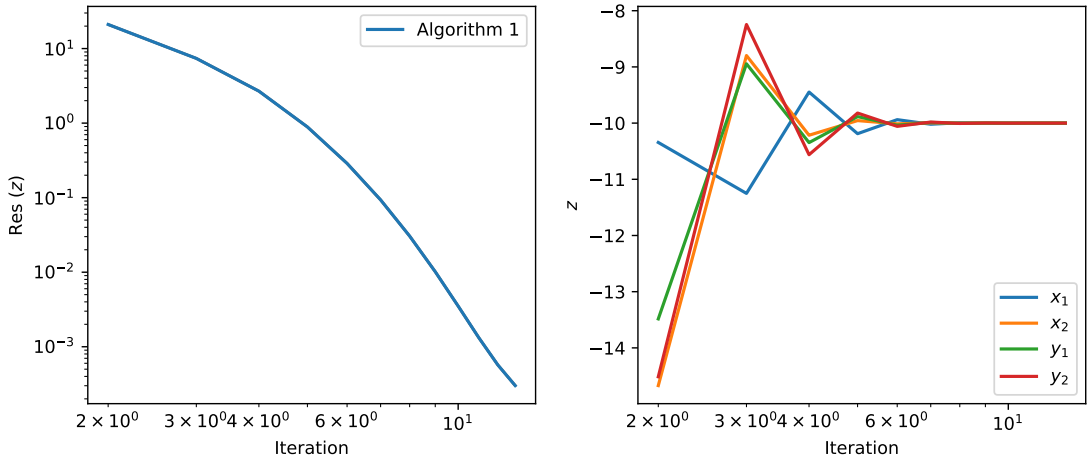

FIGURE 4. Numerical Results for Example 5.1.

Example 5.2. Consider the following stochastic linear complementarity problem

$$x \in \mathbb{R}^m_+, \ \mathbb{E}[F(x,\xi)] \geq 0, \text{ and } x^T \mathbb{E}[F(x,\xi)] = 0.$$

Define a random operator $F : \mathbb{R}^2_+ \times \Xi$ by

$$F(x,\xi) = \begin{bmatrix} \xi_1 & \xi_2 \\ \xi_3 & \xi_4 \end{bmatrix} x + \begin{pmatrix} \xi_5 \\ \xi_6 \end{pmatrix} - \begin{pmatrix} 15 \\ 30 \end{pmatrix}, \tag{5.1}$$

where $\xi_1, \cdots, \xi_6$ is a random sample of 6 realizations of ξ. It is easy to verify that $\mathbb{E}[F(\cdot,\xi)]$ is nonmonotone and Lipschitz continuous with $L = \mathbb{E}[\|A(\xi)\|]$. Suppose that ξ follows the uniform distribution over the box $\{\xi \in \mathbb{R}^6 : (0,0,0,0,-6,-6) \leq \xi \leq (2,1,4,6,6,6)\}$. Note that the problem above can be cast as SVI (1.1) with $X = \mathbb{R}^m_+$ and $F(\cdot,\cdot) : \mathbb{R}^2 \times \Xi \to \mathbb{R}^2$ defined as (5.1).

In this example, we compare the proposed algorithm with several previously known algorithms, including the variance-based subgradient extragradient method proposed by Yang et

al. in [17, Algorithm 1] (shortly, Method SEM), the SA based subgradient extragradient algorithm with variance reduction suggested by Long and He in [16, Algorithm 1] (shortly, Method ASEA), and the stochastic forward backward forward method presented by Boţ et al. in [34, Algorithm 1] (shortly, Method FBF). The parameters of all algorithms are set as follows

- For Algorithm 1, the three parameters μ, ν, and λ are randomly generated in $(0,1)$, τ is chosen randomly in $(0, \frac{2(1-\lambda)}{1+\lambda})$, $\theta = 5$, and $\zeta_n := \frac{1}{(n+3)^2}$ for all $n \in \mathbb{N}$.
- For Method SEM, $\hat{\alpha}$ and θ are randomly chosen in $(0,1)$, and λ is randomly chosen in $(0, 1/\sqrt{3})$.
- For Method ASEA, μ, γ, and θ are randomly generated in $(0,1)$.
- For Method FBF, α_n is generated in $(0, \frac{1}{\sqrt{2}L})$, $\forall n \in \mathbb{N}$.

We define the residual distance function $res(x_n) = \|x_n - \Pi_X(x_n - F(x_n))\|$, which is equal to zero if and only if x_n is a solution to the SVI. Since we do not know the exact solution of the problem, we use $\{res(x_n)\}$ to measure the n-th iteration error of all algorithms. The maximum number of iterations of 150 is as the common stopping criterion in the following experiments. For comparison on the convergence speed between all algorithms, we use the same random starting points.

Now, we show the numerical results of all algorithms for two different types of the starting points.

Case 1: x_0 and x_1, the fixed starting points, are taken randomly from $(0,1)^2$ by using *numpy.random.rand*(2) in Python.

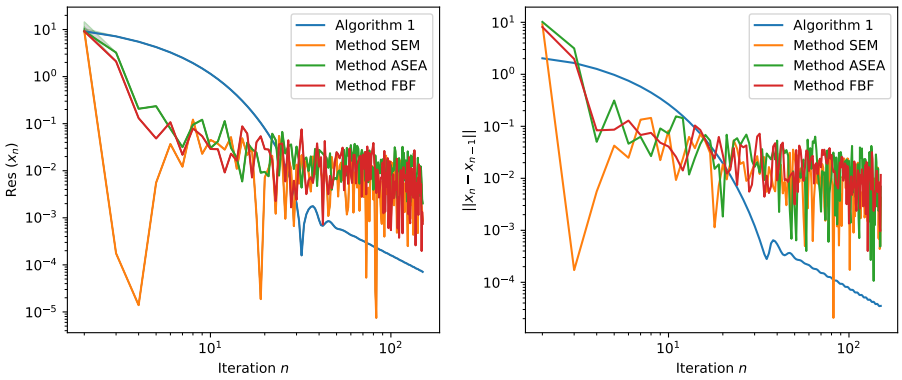

FIGURE 5. Numerical Results of All Algorithms.

TABLE 1. Comparison Results between Algorithm 1, Method SEM, Method ASEA, and Method FBF.

Res(x_n) \ Method / Iter.	Algorithm 1	Method SEM	Method ASEA	Method FBF
50	**0.00057**	0.01272	0.01017	0.01837
100	**0.00015**	0.00356	0.01123	0.00291
150	**7.10160E-5**	0.00086	0.00202	0.00074

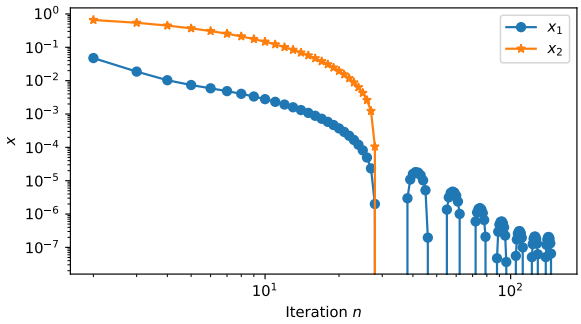

FIGURE 6. The Behavior of $x = (x_1, x_2)^T$ During the Iteration Process.

To illustrate the convergence performance of all algorithms, we display the values of $\{res(x_n)\}$ and $\{\|x_n - x_{n-1}\|\}$ with the number of iterations in Figure 5. One can check that our algorithm achieves a more stable and higher precision with the number of iterations, in contrast with other methods. The test results summarized in Table 1 indicate that our algorithm enjoys a fast convergence speed, which outperforms the convergence speed of other methods. Furthermore, Figure 6 plots the convergence behavior of entries of $\{x_n\}$ for all algorithms. From this, one sees that $(0,0)^T$ is the convergence point.

Case 2: x_0 and x_1, the fixed starting points, are taken randomly from $(0,10)^2$ by using $numpy.random.uniform(0,10,2)$ in Python.

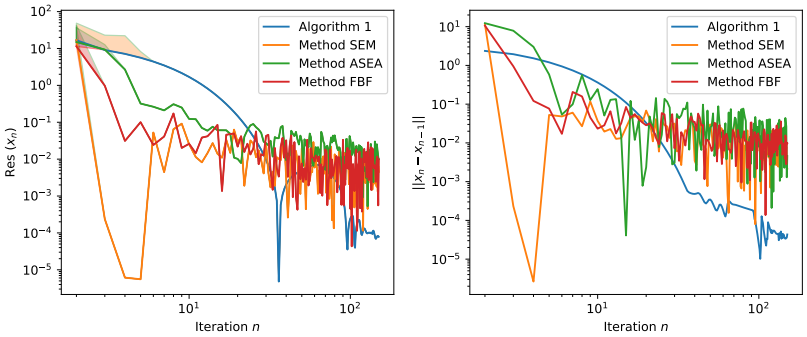

FIGURE 7. Numerical Results of All Algorithms.

TABLE 2. Comparison Results between Algorithm 1, Method SEM, Method ASEA, and Method FBF.

$\|x_n - x_{n-1}\|$ Method / Iter.	Algorithm 1	Method SEM	Method ASEA	Method FBF
50	**0.00721**	0.02604	0.00336	0.00617
100	**6.29144E-5**	0.00879	0.00562	0.01617
150	**3.28920E-5**	0.00786	0.00245	0.02166

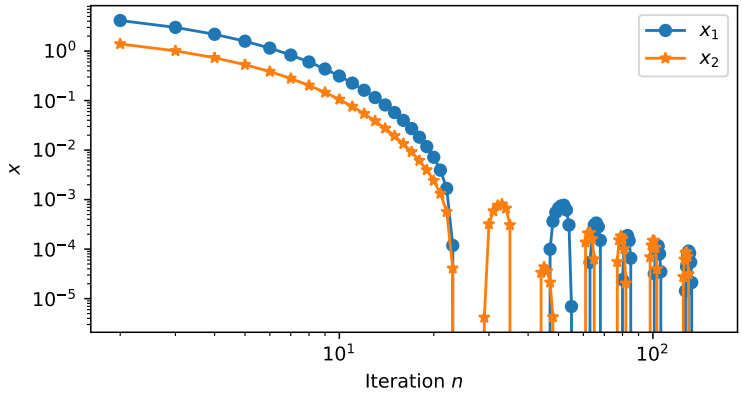

FIGURE 8. The Behavior of $x = (x_1, x_2)^T$ During the Iteration Process.

Figure 7 demonstrates the comparison on the convergence speed between all algorithms. The shaded area therein indicates the deviation. Meanwhile, we code the test results in Table 1. It can be seen that our proposed algorithm has a better convergence behavior of $\{res(x_n)\}$ and $\{\|x_n - x_{n-1}\|\}$. This illustrates that the speed of our algorithm with inertial effects is more efficient than the methods without inertial effects. Besides, the convergence of $res(x_n)$ to 0 implies that $\{x_n\}$ converges to the solution of the SVI. Furthermore, from the results reported in Figure 8, we observe that the convergent point is $(0,0)^T$.

Remark 5.1. We have the following observations for Examples 5.1 and 5.2.

(i) Figures 5 and 7 demonstrate that Algorithm 1 converges faster than the methods proposed in [16,17,34]. This illustrates the fact that the presence of the initial extrapolation term plays a key role in the acceleration process.

(ii) The involved operator F is nonmonotone rather than pseudomonotone, which implies that our algorithm have a broader scope of applications.

(iii) The numerical results are independent of the size of the dimensions and the choices of starting points. It entails that our algorithm is efficient and robust.

(iv) Our algorithm is efficient in solving a large spectrum of problems including saddle-point problems and linear complementarity problems.

6. CONCLUSION

In this paper, we devised an SA-based projection and contraction algorithm for approximating to a solution of SVIs without assuming any monotonicity on the operator involved. The algorithm embedded inertial terms leads to the fast convergence result. The step sizes are produced by employing the SA version of Armijo's line search rule. Primary numerical experiments were presented to demonstrate the effectiveness and the computational performance of our algorithm in contrast with some related methods in the literature. The proposed algorithm improved and extended some earlier results from theoretical as well as practical point of view.

Acknowledgments

This work was supported by National Natural Science Foundation of China (Grant No.12201517), the Natural Science Foundation of Chongqing (Grant No. CSTB2022NSCQ-MSX0580), and the Fundamental Research Funds for the Central Universities (Grant No. SWU-KQ22014).

REFERENCES

[1] G. Gürkan, A. Yonca Özge, S.M. Robinson, Sample-path solution of stochastic variational inequalities, Math. Program. 84 (1999), 313-333.
[2] M, Limosani, M, Milasi, D, Scopelliti, Deregulated electricity market, a stochastic variational approach, Energy Econ. 103 (2021), 105493.
[3] F. Yousefian, A. Nedić, U.V. Shanbhag, On smoothing, regularization, and averaging in stochastic approximation methods for stochastic variational inequality problems, Math. Program. 165 (2017), 391-431.
[4] 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, Commun. Nonlinear Sci. Numer. Simul. 111 (2022), 106406.
[5] D. Banholzer, J. Fliege, R. Werner, On rates of convergence for sample average approximations in the almost sure sense and in mean, Math. Program. 191 (2022), 307-345.
[6] F. Sohrabi, M. Rohaninejad, M.R. Hesamzadeh, J. Bemš, Coordinated bidding of multi-product charging station in electricity markets using rolling planning and sample average approximation, Int. J. Elec. Power Energy Syst. 146 (2023), 108786, 2023.
[7] D. Bertsimas, V. Gupta, N. Kallus, Robust sample average approximation, Math. Program. 171 (2018), 217-282.
[8] H. Robbins, S. Monro, A stochastic approximation method, Ann. Math. Statist. 22 (1951), 400-407.
[9] S. Ghadimi, A. Ruszczynski, M. Wang, A single timescale stochastic approximation method for nested stochastic optimization, SIAM J. Optim. 30 (2020), 960-979.
[10] H.D. Kaushik, F. Yousefian, A method with convergence rates for optimization problems with variational inequality constraints, SIAM J. Optim. 31 (2021), 2171-2198.
[11] C. Geiersbach, E. Loayza-Romero, K. Welker, Stochastic approximation for optimization in shape spaces, SIAM J. Optim. 31 (2021), 348-376.
[12] H. Jiang, H. Xu, Stochastic approximation approaches to the stochastic variational inequality problem, IEEE Trans. Auto. Contr. 53 (2008), 1462-1475.
[13] F. Yousefian, A. Nedić, U.V. Shanbhag, Distributed adaptive steplength stochastic approximation schemes for cartesian stochastic variational inequality problems, arXiv preprint arXiv:1301.1711, 2013.
[14] J. Koshal, A. Nedic, U.V. Shanbhag, Regularized iterative stochastic approximation methods for stochastic variational inequality problems, IEEE Trans. Automat. Contr. 58 (2012), 594-609.
[15] A.N. Iusem, A. Jofré, R.I. Oliveira, P. Thompson, Extragradient method with variance reduction for stochastic variational inequalities, SIAM J. Optim. 27 (2017), 686-724.
[16] X. Long, Y. He, A fast stochastic approximation-based subgradient extragradient algorithm with variance reduction for solving stochastic variational inequality problems, J. Comput. Appl. Math. 420 (2023), 114786.
[17] Z. Yang, J. Zhang, Y. Wang, G. Lin, Variance-based subgradient extragradient method for stochastic variational inequality problems, J. Sci. Comput. 89 (2021), 4.
[18] 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 (2019), 175-206.
[19] X. Zhang, X. Du, Z. Yang, G. Lin, An infeasible stochastic approximation and projection algorithm for stochastic variational inequalities, J. Optim. Theory Appl. 183 (2019), 1053-1076.
[20] B.T. Polyak, Some methods of speeding up the convergence of iteration methods, USSR. Comput. Math. Math. Phys. 4 (1964), 1-17.
[21] M.A. Olona, T.O. Alakoya, A.O.-E. Owolabi, O.T. Mewomo, Inertial algorithm for solving equilibrium, variational inclusion and fixed point problems for an infinite family of strict pseudocontractive mappings, J. Nonlinear Funct. Anal. 2021 (2021), 10.

[22] B. Tan, S.Y. Cho, Strong convergence of inertial forward–backward methods for solving monotone inclusions, Appl. Anal. 101 (2022), 5386-5414.
[23] F.U. Ogbuisi, The projection method with inertial extrapolation for solving split equilibrium problems in Hilbert spaces, Appl. Set-Valued Anal. Optim. 3 (2021), 239-255.
[24] C. Izuchukwu, S. Reich, Y. Shehu, Relaxed inertial methods for solving the split monotone variational inclusion problem beyond co-coerciveness, Optimization 72 (2023), 607-646.
[25] Y. Shehu, O.S. Iyiola, J.C. Yao, New projection methods with inertial steps for variational inequalities, Optimization 71 (2022), 4731-4762.
[26] F. Facchinei, J.S. Pang, Finite-Dimensional Variational Inequalities and Complementarity Problems, Springer, New York, 2003.
[27] S. Karamardian, S. Schaible, Seven kinds of monotone maps, J. Optim. Theory Appl. 66 (1990), 37-46.
[28] D. Aussel, N. Hadjisavvas, On quasimonotone variational inequalities, J. Optim. Theory Appl. 121 (2004), 445-450.
[29] S. Komlósi, On the stampacchia and minty variational inequalities, Generalized Convexity and Optimization for Economic and Financial Decisions, pp. 231-260, 1999.
[30] D. Kinderlehrer, G. Stampacchia, An Introduction to Variational Inequalities and Their Applications, SIAM, New York, 2000.
[31] C. Song, Z. Zhou, Y. Zhou, Y. Jiang, Y. Ma, Optimistic dual extrapolation for coherent non-monotone variational inequalities, Advances in Neural Information Processing Systems, 33 (2020), 14303-14314.
[32] B. Davis, R. Song, Integral inequalities for convex functions of operators on martingales, In: Selected Works of Donald L. Burkholder, pp. 181-198, Springer, New York, 2011.
[33] H. Robbins, D. Siegmund, A convergence theorem for non negative almost supermartingales and some applications, In: Optimizing Methods in Statistics, pp. 233-257, Elsevier, New York, 1971.
[34] R.I. Boţ, P. Mertikopoulos, M. Staudigl, P.T. Vuong, Minibatch forward-backward-forward methods for solving stochastic variational inequalities, Stoch. Syst. 11 (2021), 112-139.