J. Appl. Numer. Optim. 4 (2022), No. 3, pp. 341-355
Available online at http://jano.biemdas.com
https://doi.org/10.23952/jano.4.2022.3.03

A STOCHASTIC SELECTIVE PROJECTION METHOD FOR SOLVING A CONSTRAINED STOCHASTIC VARIATIONAL INEQUALITY PROBLEM

SHENGHUA WANG*, RONGGUANG LIN, YITONG SHI

Department of Mathematics and Physics, North China Electric Power University, Baoding 071003, China

Abstract. In this paper, we propose an iterative algorithm by combining the selective projection method and the stochastic approximation method for solving a stochastic pseudomonotone variational inequality problem. Compared with many known related algorithms, the proposed algorithm uses a simpler adaptive rule to deal with the unknown Lipschitz constant of the involved mapping. The convergence is proved and convergence rate is estimated for the proposed algorithm. Finally, a numerical example is given to illustrate the effectiveness of the proposed algorithm.

Keywords. Adaptive rule; Pseudomonotone variational inequality; Stochastic variational inequality; Subgradient extragradient method.

1. INTRODUCTION

Let $X \subset \mathbb{R}^n$ be a nonempty, closed, and convex subset. Let ξ be a random vector defined on a probability space $(\Omega, \mathscr{F}, \mathscr{P})$ and supported on a closed subset $\Xi \subset \mathbb{R}^l$. Let $f(\cdot,\xi) : \mathbb{R}^n \to \mathbb{R}^n$ be a vector-valued function, and assume that $f(x,\cdot)$ is integrable on Ξ for each $x \in \mathbb{R}^n$. The stochastic variational inequality problem (SVIP, for short) is to find a vector $x^* \in X$ such that

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

where $F(x) = \mathbb{E}[f(x,\xi)] \in \mathbb{R}^n$ is the expected value of $f(x,\xi)$. Assume that $F(x)$ is well defined for each $x \in \mathbb{R}^n$. It is easy to see that if the expected value $F(x)$ can be computed explicitly, then the SVIP (1.1) becomes a deterministic variational inequality problem (DVIP, for short), which can be solved by the various methods presented in the literature. However, if the probability distribution of the random variable in (1.1) is unknown, or the expected value $F(x)$ is hard to compute (although the probability distribution is known), then the methods for solving the DVIPs can not be used directly. The stochastic approximation (SA) method is one of the popular methods for solving the SVIPs, which was initially introduced by Jiang and Xu [1] in their seminal work. More precisely, they proposed the following single-projection scheme based on stochastic approximation (SA) technique from [2] for stochastic optimization problems:

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

*Corresponding author.
E-mail address: sheng-huawang@ncepu.edu.cn (S. Wang).
Received April 6, 2022; Accepted August 2, 2022.

where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$, Π_X is the Euclidean projection onto X, ξ_k is a sample of ξ, and $\{\alpha_k\}$ is a nonnegative sequence satisfying the conditions $\sum_{k=0}^{\infty} \alpha_k = \infty$ and $\sum_{k=0}^{\infty} \alpha_k^2 < \infty$. In (1.2), the expected function $F(x)$ of $f(x,\xi)$ is required to be strongly monotone. Under some suitable conditions, they proved that $\{x_k\}$ generated by (1.2) converges almost surely (a.s.) to the unique solution of (1.1). Since the work of Jiang and Xu [1], many authors proposed various methods for solving the SVIPs. Juditsky et al. [3] introduced a stochastic mirror-prox algorithm based on the SA technique for solving the monotone SVIP in which the involved mapping F possesses both smooth and nonsmooth components. Koshal et al. [4] proposed a stochastic Tikhonov regularization method and a proximl-point method for solving a Cartesian monotone SVIP and applied their methods to a stochastic networked rate allocation problem. Kannan and Shanbhag [5] proposed an extragradient method for solving a pseudomonotone SVIP in which the sufficiency conditions for the solvability of SVIP was studied. Recently, Kannan and Shanbhag [6] presented a stochastic mirror-prox scheme for the pseudomonotone (and other class of) SVIPs. Yousefian et al. [7] presented two robust variants of the stochastic extragradient methods for a monotone SVIP where the involved mapping F is required to be bounded over a compact convex set. They also studied a distributed adaptive steplength SA-based algorithm for a strongly monotone Cartesian SVIP in which the mapping F is not necessary to be Lipschitz continuous; see [8]. In 2017, Iusem et al. [9] proposed a dynamic sampled extragradient method for solving a pseudomonotone SVIP that allows the feasible set X to be unbounded, and the variance of the oracle to be nonuniform. Subsequently, Iusem et al. [10] proposed another dynamic sampled extragradient method for solving a pseudomonotone SVIP in which a line search rule was used for the unknown Lipschitz constant of the mapping F. Inspired by Iusem et al. [10], Zhang et al. [11] proposed an infeasible projection algorithm with the line search rule in [10] for a pseudomonotone SVIP in which the same iteration complexity and higher oracle complexity compared with the ones of [10] were presented. Differed with the algorithm of [10], only one projection and one family of samples are needed at each iteration in [11] that is such that the performing of algorithm is easier. Recently, Yang et al. [12] considered a variance-based modified backward-forward algorithm for the SVIP. In [12], the authors also used the line search rule [10] for the unknown Lipschitz constant of the mapping F that has a weaker condition than the pseudomonotonicity.

In many practical problems, such as constrained convex optimization problems, the feasible set X is often reduced to the following form: $X = \cap_{i=1}^{m} X_i$ with

$$X_i = \{x \in \mathbb{R}^n : c_i(x) \le 0\}, \quad i \in \{1, \cdots, m\}, \tag{1.3}$$

where $c_i : \mathbb{R}^n \to \mathbb{R}$ is a convex function for $i = 1, \cdots, m$. The DVIPs with the feasible set X of the form (1.3) has been investigated by some authors; see, e.g., [13]. However, to the best of our knowledge, the SVIPs defined on the feasible set X of the form (1.3) have not been studied in the literature yet. In this paper, we propose a stochastic selective projection algorithm by combining the method present in [13] and the SA method to solve a SVIP defined the feasible set X of the form (1.3). The paper is organized as follows. In Section 2, we introduce some notations and useful lemmas. In Section 3, we present the algorithm for solving the constrained SVIP and prove the convergence of the proposed algorithm. The rate of convergence of the algorithm is also estimated in this section. In Section 4, we give a numerical example to illustrate the effectiveness of the proposed algorithm. Finally, we make the conclusions in Section 5.

2. Preliminaries

Let X be a nonempty, closed, and convex subset of $\mathbb{R}^n$, and let $F : X \to X$ be a mapping. For any $x \in \mathbb{R}^n$, there exists a unique element $z \in X$, denoted by $\Pi_X(x)$, such that $\|z - x\| = \inf_{y \in X} \|y - x\|$. The mapping $\Pi_X : \mathbb{R}^n \to X$ is called a projection from $\mathbb{R}^n$ onto X. The following lemma describes the characterization of Π_X.

Lemma 2.1. *Let Π_X be a projection from $\mathbb{R}^n$ onto X. Then*

(i) $\|\Pi_X(x) - z\|^2 \leq \|x - z\|^2 - \|\Pi_X(x) - x\|^2$ *for all $x \in \mathbb{R}^n$ and $z \in X$.*

(ii) $2\langle v, u - z\rangle \leq \|y - z\|^2 - \|u - z\|^2 - \|u - y\|^2$ *for all $z \in X$, where $u = \Pi_X(y - v)$ with $v \in \mathbb{R}^n$ and $y \in X$.*

(iii) $z = \Pi_X(x)$ *if and only if $\langle x - z, z - y\rangle \geq 0$ for all $y \in X$.*

Lemma 2.2. [14] *For any $x \in \mathbb{R}^n$ and $\beta > 0$, there holds*

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

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

Let $c : \mathbb{R}^n \to \mathbb{R}$ be a convex proper function. A point $v \in \mathbb{R}^n$ is said to be a subgradient of c at $x \in \mathbb{R}^n$ if $c(z) \geq c(x) + \langle v, z - x\rangle$, $\forall z \in \mathbb{R}^n$. The set of all subgradients of c at x is called the subdifferential of c at x and is denoted by $\partial c(x)$. The multivalued mapping $\partial c : x \to \partial c(x)$ is called the subdifferential of c. If $\partial c(x)$ is not empty, then c is said to be subdifferentiable at x. It is well known that if c is a proper function, then $\partial c(x)$ is a closed convex set for each $x \in \mathbb{R}^n$. Moreover, $\partial c(x)$ is bounded for each $x \in D$ if c is finite on a nonempty bounded subset $D \subset \mathbb{R}^n$; see [15] for details.

For any given $z \in \mathbb{R}^n$ and $v \in \partial c(z)$ with $v \neq 0$, define the set $T = \{x \in \mathbb{R}^n : c(z) + \langle v, x - z\rangle \leq 0\}$. Then, for all $y \in \mathbb{R}^n$, the projection $\Pi_T(y)$ is computed by

$$\Pi_T(y) = y - \max\left\{0, \frac{c(z) + \langle v, y - z\rangle}{\|v\|^2}\right\} v. \tag{2.1}$$

(2.1) gives us an explicit manner to compute the projection of any point onto a half-space (see [16, Proposition 2.7]).

Let $(\Omega, \mathscr{F}, \mathscr{P})$ be a probability space, and let ξ be an l-dimensional random vector defined on Ω and supported on a closed subset $\Xi \subset \mathbb{R}^l$. Given a σ-algebra $\mathscr{F}$, the notations $\xi \in \mathscr{F}$ and $\xi \perp\!\!\!\perp \mathscr{F}$ denote that ξ is $\mathscr{F}$-measurable and ξ is independent of $\mathscr{F}$, respectively. For $p \geq 1$, we denote the $\mathscr{L}^p$-norm of ξ by $|\xi|_p$ and the $\mathscr{L}_p$-norm of ξ conditional to $\mathscr{F}$ by $|\xi|\mathscr{F}|_p = \sqrt[p]{\mathbb{E}[|\xi|^p|\mathscr{F}]}$. Let $\{\xi_i\}_{i=1}^k$ be a sequence of l-dimensional random vectors defined on Ω. Define the oracle error for (1.1) by $\varepsilon(x, \xi) = f(x, \xi) - F(x)$, $\forall x \in \mathbb{R}^n$. For $p \in [2, \infty)$, define the p-moment function of $\varepsilon(x, \xi)$ by $\sigma_p(x) = \sqrt[p]{\mathbb{E}[\|\varepsilon(x, \xi)\|^p]}$, which can be used to measure the efficiency of an SA method.

The following assumption is necessary for Lemmas 2.3, 2.4, and 2.5.

Assumption 2.1. (1) For all $x, y \in \mathbb{R}^n$ and almost every $\xi \in \Xi$, $\|f(x, \xi) - f(y, \xi)\| \leq \mathscr{L}(\xi)\|x - y\|$, where $\mathscr{L} : \Xi \to \mathbb{R}^+$ is a measurable function such that $\mathscr{L}(\xi) \geq 1$ for almost every $\xi \in \Xi$. (2) There exist $x_* \in \mathbb{R}^n$ and $p \geq 2$ such that $\mathbb{E}[\|f(x_*, \xi)\|^p] < \infty$ and $\mathbb{E}[\mathscr{L}(\xi)^p] < \infty$.

Lemma 2.3. [10] *Under Assumption 2.1, F and $\sigma_q(\cdot)$ are Lipschitz continuous with modulus L and L_q on $\mathbb{R}^n$, respectively for any $q \in [p, 2p]$ with p from 2.1, where $L = \mathbb{E}[\mathscr{L}(\xi)]$ and $L_q = \sqrt[q]{\mathbb{E}[\mathscr{L}(\xi)^q]} + L$ for any $q > 0$.*

Let $\xi = \{\xi_j\}_{j=1}^N$ be an independent and identically distributed (i.i.d. for short) sample from Ξ and

$$H(x,\xi) = \frac{1}{N}\sum_{j=1}^{N} f(x,\xi_j) \text{ and } \hat{\varepsilon}(x,\xi) = \frac{1}{N}\sum_{j=1}^{N} \varepsilon(x,\xi_j).$$

In the following lemmas (Lemma 2.4 and Lemma 2.5), X^* denotes the set of solutions of the SVIP (1.1) and is assumed to be nonempty.

Lemma 2.4. [10] *Suppose that Assumption 2.1 holds. Then, for any $q \in [p,2p]$ with p from Assumption 2.1, there exists a constant $C_q > 0$ (depending only on q) such that, for any $x \in \mathbb{R}^n, x^* \in X^*$,*

$$\big|\|\hat{\varepsilon}(x,\xi)\|\big|_q \le C_q \frac{\sigma_q(x^*) + L_q\|x - x^*\|}{\sqrt{N}}.$$

Lemma 2.5. [10] *Suppose that Assumption 2.1 holds and $X^* \neq \emptyset$. Let $\lambda_N : \Xi \to [0,\lambda]$ be a random variable for some $0 < \lambda \le 1$. Define $z(x,\lambda_N,\xi) = \Pi_X(x - \lambda_N H(x,\xi))$. Then, for any $p \ge 2$, there exist positive constants $\{c_i\}_{i=1}^4$ (depending on n, p, L_{2p}, and λ)) such that*

$$\big|\|\hat{\varepsilon}(z(x,\lambda_N,\xi),\xi)\|\big|_p \le \frac{c_1\sigma_{2p}(x^*) + \bar{L}_{2p}\|x - x^*\|}{\sqrt{N}}, \ \forall x \in \mathbb{R}^n, \forall x^* \in X^*,$$

where $\bar{L}_{2p} = c_2 L_2 + c_3 L_p + c_4 L_{2p}$, $c_1 > C_p$, and $\sigma_{2p}(x^*) \ge \sigma_p(x^*)$.

Lemma 2.6. [17] *Let $\{V_k\}$, $\{\delta_k\}$, $\{\eta_k\}$, and $\{\theta_k\}$ be sequences of nonnegative random variables adapted to the filtration $\{\mathscr{G}_k\}$ such that a.s. $\sum_{k=1}^\infty \delta_k < \infty$ and $\sum_{k=1}^\infty \theta_k < \infty$ and*

$$\mathbb{E}[V_{k+1}|\mathscr{G}_k] \le (1+\delta_k)V_k - \eta_k + \theta_k, \forall k \in \mathbb{N}.$$

Then a.s. $\{V_k\}$ is convergent and $\sum_{k=1}^\infty \eta_k < \infty$.

3. Stochastic Selective Projection Algorithm

Let $(\Omega,\mathscr{F},\mathscr{P})$ be a probability space, and let ξ be an l-dimensional random vector defined on Ω and supported on a closed subset $\Xi \subset \mathbb{R}^l$. Let $f(\cdot,\xi) : \mathbb{R}^n \to \mathbb{R}^n$ be a nonlinear mapping. Let $F(x) = \mathbb{E}[f(x,\xi)] \in \mathbb{R}^n$ be the expected value of $f(x,\xi)$, and assume that $F(x)$ is well defined for each $x \in \mathbb{R}^n$. Let $c_i : \mathbb{R}^n :\to \mathbb{R}$ be a convex proper function for each $i \in \{1,\cdots,m\}$. In this section, we consider the following the constrained SVIP: find a point $x^* \in X = \cap_{i=1}^m X_i$ such that

$$\langle F(x^*), x - x^*\rangle \ge 0, \ \forall x \in X, \tag{3.1}$$

where

$$X_i = \{x \in \mathbb{R}^n : c_i(x) \le 0\}, \ \forall i = 1,\cdots,m.$$

We denote the set of solutions of (3.1) by X^* and assume that $X^* \neq \emptyset$. For solving the constrained SVIP (3.1), we propose the following algorithm that is a combination of the SA method with the selective projection method used in [13].

We have the remarks regarding the algorithm.

Algorithm 1 Stochastic Selective Projection Algorithm

Initialization: Select the initial points $x_0, x_1, y_1 \in \mathbb{R}^n$, the parameters $\delta \in (0,1)$, $\alpha_0 > 0$, $\rho \in (0,1)$, and the sample rate $\{N_k\} \subset \mathbb{N}$ with $\sum_{k=0}^{\infty} \frac{1}{\sqrt{N_k}} < \infty$. Set $k = 1$.

Step 1. Take an i.i.d. sample $\xi_{k-1} = \{\xi_{j,k-1}\}_{j=1}^{N_{k-1}}$ from Ξ and compute

$$H(x_{k-1}, \xi_{k-1}) = \frac{1}{N_{k-1}} \sum_{j=1}^{N_{k-1}} f(x_{k-1}, \xi_{j,k-1}) \text{ and } H(y_k, \xi_{k-1}) = \frac{1}{N_{k-1}} \sum_{j=1}^{N_{k-1}} f(y_k, \xi_{j,k-1}).$$

If $\alpha_{k-1}\|H(x_{k-1}, \xi_{k-1}) - H(y_k, \xi_{k-1})\| \leq \rho \|x_{k-1} - y_k\|$, then set $\alpha_k = \alpha_{k-1}$, else set $\alpha_k = \delta\alpha_{k-1}$.

Step 2. For given the current x_k, select $i_k \in \{1, \cdots, m\}$ such that

$$i_k = \arg \max_{1 \leq i \leq m} \{c_i(x_k)\},$$

and construct a half-space by

$$X_k = \{z \in \mathbb{R}^n : c_{i_k}(x_k) + \langle v_k, z - x_k \rangle \leq 0\},$$

where $v_k \in \partial c_{i_k}(x_k)$ with $v_k \neq 0$.

Step 3. Draw an i.i.d. sample $\xi_k = \{\xi_{j,k}\}_{j=1}^{N_k}$ from Ξ and compute

$$\begin{cases} y_{k+1} = \Pi_{X_k}(x_k - \alpha_k H(x_k, \xi_k)) \\ x_{k+1} = \Pi_{X_k}(x_k - \alpha_k H(y_{k+1}, \xi_k)), \end{cases}$$

where

$$\begin{cases} H(x_k, \xi_k) = \dfrac{1}{N_k} \displaystyle\sum_{j=1}^{N_k} f(x_k, \xi_{j,k}), \\ H(y_{k+1}, \xi_k) = \dfrac{1}{N_k} \displaystyle\sum_{j=1}^{N_k} f(y_{k+1}, \xi_{j,k}). \end{cases}$$

Step 4. If $y_{k+1} = x_k$, then the algorithm stops and $x_k \in X^*$; otherwise, set $k = k+1$ and go to *Step 1*.

Remark 3.1. There is no projection involved in our algorithm. For each $k \in \mathbb{N}$, it follows from (2.1) that x_{k+1} and y_{k+1} can be explicitly computed by

$$\begin{cases} y_{k+1} = x_k - \alpha_k H(x_k, \xi_k) - \max\left\{0, \dfrac{c_{i_k}(x_k) - \alpha_k \langle v_k, H(x_k, \xi_k)\rangle}{\|v_k\|}\right\} v_k, \\ x_{k+1} = x_k - \alpha_k H(y_{k+1}, \xi_k) - \max\left\{0, \dfrac{c_{i_k}(x_k) - \alpha_k \langle v_k, H(y_{k+1}, \xi_k)\rangle}{\|v_k\|}\right\} v_k. \end{cases}$$

Remark 3.2. For each $k \in \mathbb{N}$, $x_k \in \mathscr{F}_k$ and $y_k \in \mathscr{F}_k$.

Remark 3.3. For each $k \in \mathbb{N}$, $\mathbb{E}[\varepsilon_k|\mathscr{F}_k] = 0$. In fact, since $x_k \in \mathscr{F}_k$ and $\xi_k \perp\!\!\!\perp \mathscr{F}_k$, by the definition of ε_k, we have

$$\mathbb{E}[\varepsilon_k|\mathscr{F}_k] = \mathbb{E}[H(x_k, \xi_k)|\mathscr{F}_k] - F(x_k) = \mathbb{E}[H(x_k, \xi_k)] - F(x_k) = 0.$$

Remark 3.4. For each $k \in \mathbb{N}_0$, $X \subset X_k$. In fact, for any $z \in X$, i.e., $c_{i_k}(z) \leq 0$, by the definition of ∂c_{i_k} we have

$$c_{i_k}(x_k) + \langle v_k, x - x_k \rangle \leq c_{i_k}(x), \ \forall x \in \mathbb{R}^n, \ \forall k \in \mathbb{N}. \tag{3.2}$$

Replacing x with z in (3.2), we obtain $c_{i_k}(x_k) + \langle v_k, z - x_k \rangle \leq c_{i_k}(z) \leq 0$, $\forall k \in \mathbb{N}_0$. By the definition of c_{i_k}, we have $c_i(z) \leq 0$ for each $i \in \{1, \cdots, m\}$. It follows that $z \subset X_k$. So $X \subset X_k$ for all $k \in \mathbb{N}$.

The following remark shows that the stop criterion can work in Step 3.

Remark 3.5. If $y_{k+1} = x_k$ for some $k \in \mathbb{N}$, then $x_k \in X^*$. Indeed, if $y_{k+1} = x_k$, then it follows from Lemma 2.1 (iii) that $\langle H(x_k, \xi_k), y - x_k \rangle \geq 0, \forall y \in X_k$. That is, $\langle \varepsilon_k, y - x_k \rangle + \langle F(x_k), y - x_k \rangle \geq 0$, $\forall y \in X_k$. From Remark 3.4, it follows that $\langle \varepsilon_k, y - x_k \rangle + \langle F(x_k), y - x_k \rangle \geq 0$, $\forall y \in X$. Taking $\mathbb{E}[\cdot|\mathscr{F}_k]$ and noting that $x_k \in \mathscr{F}_k$, we find from Remark 3.3 that

$$\mathbb{E}[\langle \varepsilon_k, y - x_k \rangle | \mathscr{F}_k] + \langle F(x_k), y - x_k \rangle = \langle F(x_k), y - x_k \rangle \geq 0, \forall y \in X,$$

which implies that $x_k \in X^*$.

To prove the convergence of Algorithm 1 with respect to the filtrations $\{\mathscr{F}_k\}$, we assume that Assumption 2.1 and the following assumption hold.

Assumption 3.1. The mapping F is pseudomonotone on $\mathbb{R}^n$, that is, for $x, y \in \mathbb{R}^n$, if $\langle F(x), y - x \rangle \geq 0$, then $\langle F(y), y - x \rangle \geq 0$.

Under Assumption 2.1, the mapping F is L-Lipschitz continuous, where $L = \mathbb{E}[\mathscr{L}(\xi)]$. Find that the constant L is not used as the input parameter in Algorithm 1. That is, the constant L may be unknown. The control sequence $\{\alpha_k\}$ is used to deal with the unknown constant L. In Algorithm 1, $\{\alpha_k\}$ is generated at each step by an adaptive rule whereas the sequence in [10, 11, 12] for L was obtained by an Armijo line search rule that is a more complex rule than our one.

Lemma 3.1. *For all $x^* \in X^*$, the following conclusion holds:*

$$\left| \|y_{k+1} - x^*\| \|\eta_k\| \middle| \mathscr{F}_k \right|_{\frac{p}{2}} \leq \frac{m_1}{\sqrt{N_k}} + \frac{m_2}{\sqrt{N_k}} \|x_k - x^*\|^2, \ \forall k \in \mathbb{N}_0,$$

where $m_1 = (4 + 2a_0 + 4L\alpha_0)(c_1 \sigma_{2p}(x^*))^2$, $m_2 = 2(1 + L\alpha_0)(1 + 2\bar{L}_{2p}^2) + 2a_0 \bar{L}_{2p}^2$, and $c_1, \bar{L}_{2p}$ with $p \geq 2$ are given in Lemma 2.5.

Proof. By Lemma 2.4, we have

$$\begin{aligned} \left| \|\varepsilon_k\| \middle| \mathscr{F}_k \right|_{\frac{p}{2}} \leq \left| \|\varepsilon_k\| \middle| \mathscr{F}_k \right|_p &\leq \frac{C_p \sigma_p(x^*) + C_p L_p \|x_k - x^*\|}{\sqrt{N_k}} \\ &\leq \frac{c_1 \sigma_{2p}(x^*) + \bar{L}_{2p} \|x_k - x^*\|}{\sqrt{N_k}}, \ \forall k \in \mathbb{N}_0. \end{aligned} \tag{3.3}$$

From Lemma 2.5, it follows that

$$\begin{aligned} \left| \|\eta_k\| \middle| \mathscr{F}_k \right|_{\frac{p}{2}} &\leq \left| \|\eta_k\| \middle| \mathscr{F}_k \right|_p \\ &\leq \frac{c_1 \sigma_{2p}(x^*) + \bar{L}_{2p} \|x_k - x^*\|}{\sqrt{N_k}}, \ \forall k \in \mathbb{N}_0. \end{aligned} \tag{3.4}$$

By Remark 3.4, we have $x^* \in X^* \subset X \subset X_k$ and hence $x^* = \Pi_{X_k}(x^* - \alpha_k F(x^*))$ for each $k \in \mathbb{N}_0$. Assumption 2.1 indicates that F is L-Lipschitz continuous. Since $\alpha_k \le \alpha_0$ for all $k \in \mathbb{N}$, we have

$$\begin{aligned}
\|y_{k+1} - x^*\| &= \|\Pi_{X_k}(x_k - \alpha_k H(x_k, \xi_k)) - \Pi_{X_k}(x^* - \alpha_k F(x^*))\| \\
&\le \|x_k - x^* - \alpha_k (H(x_k, \xi_k) - F(x^*))\| \\
&= \|x_k - x^* - \alpha_k (F(x_k) + \varepsilon_k - F(x^*))\| \\
&\le \|x_k - x^*\| + \alpha_k (\|F(x_k) - F(x^*)\| + \|\varepsilon_k\|) \\
&\le \|x_k - x^*\| + \alpha_k L \|x_k - x^*\| + \alpha_k \|\varepsilon_k\| \\
&\le (1 + L\alpha_0)\|x_k - x^*\| + \alpha_0 \|\varepsilon_k\|, \ \forall k \in \mathbb{N}_0.
\end{aligned} \tag{3.5}$$

By the Hölder inequality, (3.3), (3.4), and (3.5), we have

$$\begin{aligned}
&\Big| \|y_{k+1} - x^*\| \|\eta_k\| \Big| \mathscr{F}_k \Big|_{\frac{p}{2}} \\
&\le (1 + L\alpha_0)\|x_k - x^*\| \Big| \|\eta_k\| \Big| \mathscr{F}_k \Big|_{\frac{p}{2}} + \alpha_0 \Big| \|\varepsilon_k\| \|\eta_k\| \Big| \mathscr{F}_k \Big|_{\frac{p}{2}} \\
&\le (1 + L\alpha_0)\|x_k - x^*\| \Big| \|\eta_k\| \Big| \mathscr{F}_k \Big|_{\frac{p}{2}} + a_0 \Big| \|\varepsilon_k\| \Big| \mathscr{F}_k \Big|_p \cdot \Big| \|\eta_k\| \Big| \mathscr{F}_k \Big|_p \\
&\le (1 + L\alpha_0)\|x_k - x^*\| \frac{c_1 \sigma_{2p}(x^*) + \bar{L}_{2p}\|x_k - x^*\|}{\sqrt{N_k}} + \frac{a_0 (c_1 \sigma_{2p}(x^*) + \bar{L}_{2p}\|x_k - x^*\|)^2}{N_k} \\
&\le 2(1 + L\alpha_0) \frac{\|x_k - x^*\|^2 + (c_1 \sigma_{2p}(x^*) + \bar{L}_{2p}\|x_k - x^*\|)^2}{\sqrt{N_k}} + \frac{a_0 (c_1 \sigma_{2p}(x^*) + \bar{L}_{2p}\|x_k - x^*\|)^2}{\sqrt{N_k}} \\
&\le 2(1 + L\alpha_0) \frac{\|x_k - x^*\|^2 + 2(c_1 \sigma_{2p}(x^*))^2 + 2\bar{L}_{2p}^2 \|x_k - x^*\|^2}{\sqrt{N_k}} \\
&\quad + \frac{2a_0 (c_1 \sigma_{2p}(x^*))^2 + 2a_0 \bar{L}_{2p}^2 \|x_k - x^*\|^2}{\sqrt{N_k}} \\
&= \frac{(4 + 2a_0 + 4L\alpha_0)(c_1 \sigma_{2p}(x^*))^2}{\sqrt{N_k}} + \frac{2(1 + L\alpha_0)(1 + 2\bar{L}_{2p}^2) + 2a_0 \bar{L}_{2p}^2}{\sqrt{N_k}} \|x_k - x^*\|^2 \\
&= \frac{m_1}{\sqrt{N_k}} + \frac{m_2}{\sqrt{N_k}} \|x_k - x^*\|^2, \ \forall k \in \mathbb{N}_0.
\end{aligned}$$

This completes the proof. □

Lemma 3.2. *For any $x^* \in X^*$, $\{\|x_k - x^*\|^2\}$ is a.s. convergent.*

Proof. It is easy to see that $\{\alpha_k\}$ is nonnegative and nonincreasing. We demonstrate that $\{\alpha_k\}$ is bounded away from zero. Suppose in contrary that $\alpha_k \to 0$. Then there exists a subsequence $\{\alpha_{k_i}\} \subset \{\alpha_k\}$ satisfying

$$\alpha_{k_i - 1} \|H(x_{k_i - 1}, \xi_{k_i - 1}) - H(y_{k_i}, \xi_{k_i - 1})\| > \rho \|x_{k_i - 1} - y_{k_i}\|.$$

By Assumption 2.1, we have

$$\alpha_{k_i - 1} > \frac{\rho \|x_{k_i - 1} - y_{k_i}\|}{\|H(x_{k_i - 1}, \xi_{k_i - 1}) - H(y_{k_i}, \xi_{k_i - 1})\|} \ge \frac{\rho}{\mathscr{L}(\xi_{k_i - 1})},$$

that is,

$$\rho < \alpha_{k_i - 1} \mathscr{L}(\xi_{k_i - 1}). \tag{3.6}$$

Since $\alpha_{k_i-1} \to 0$ as $i \to \infty$ for any $\lambda > 0$, there exist $i_\lambda \in \mathbb{N}$ such that $\alpha_{k_i-1} < \lambda$ for all $i > i_\lambda$. It follows from (3.6) that $\rho < \lambda \mathscr{L}(\xi_{k_i-1})$, $\forall i > i_\lambda$. Taking $\mathbb{E}[\cdot]$ in the inequality above, we obtain $\rho \le \lambda L$. Since λ is arbitrary, it follows that $\rho \le 0$. It contracts with $\rho > 0$. So there exists a number $k_0 \in \mathbb{N}$ such that

$$\alpha_{k+1} = \alpha_k \text{ and } \alpha_k \|H(x_k, \xi_k) - H(y_{k+1}, \xi_k)\| \le \rho \|x_k - y_{k+1}\|, \ \forall k \ge k_0. \tag{3.7}$$

By the definition of y_{k+1} and Lemma 2.1 (iii), we have $\langle z - y_{k+1}, x_k - \alpha_k H(x_k, \xi_k) - y_{k+1} \rangle \le 0$, $\forall z \in X_k$. It follows that $\langle y_{k+1} - x_k, y_{k+1} - z \rangle \le \alpha_k \langle H(x_k, \xi_k), z - y_{k+1} \rangle$, $\forall z \in X_k$. In particular, from $x_{k+1} \in X_k$, we have

$$\langle y_{k+1} - x_k, y_{k+1} - x_{k+1} \rangle \le \alpha_k \langle H(x_k, \xi_k), x_{k+1} - y_{k+1} \rangle. \tag{3.8}$$

Similarly, by the definition of x_{k+1}, we have

$$\langle x_{k+1} - x_k, x_{k+1} - z \rangle \le \alpha_k \langle H(y_{k+1}, \xi_k), z - x_{k+1} \rangle, \ \ z \in X_k. \tag{3.9}$$

On the other hand, we have

$$\begin{aligned} \|x_{k+1} - z\|^2 &= \|x_k - z\|^2 - \|x_{k+1} - y_{k+1}\|^2 - \|y_{k+1} - x_k\|^2 + 2\langle y_{k+1} - x_k, y_{k+1} - z \rangle \\ &\quad + 2\langle x_{k+1} - y_{k+1}, x_{k+1} - z \rangle \\ &= \|x_k - z\|^2 - \|x_{k+1} - y_{k+1}\|^2 - \|y_{k+1} - x_k\|^2 + 2\langle y_{k+1} - x_k, y_{k+1} - x_{k+1} \rangle \\ &\quad + 2\langle x_{k+1} - x_k, x_{k+1} - z \rangle, \ \forall z \in X_k. \end{aligned} \tag{3.10}$$

Inserting (3.8) and (3.9) into (3.10) yields that

$$\begin{aligned} \|x_{k+1} - z\|^2 &\le \|x_k - z\|^2 - \|x_{k+1} - y_{k+1}\|^2 - \|y_{k+1} - x_k\|^2 + 2\alpha_k \langle H(y_{k+1}, \xi_k), z - x_{k+1} \rangle \\ &\quad + 2\alpha_k \langle H(x_k, \xi_k), x_{k+1} - y_{k+1} \rangle \\ &= \|x_k - z\|^2 - \|x_{k+1} - y_{k+1}\|^2 - \|y_{k+1} - x_k\|^2 + 2\alpha_k \langle H(y_{k+1}, \xi_k), z - y_{k+1} \rangle \\ &\quad + 2\alpha_k \langle H(x_k, \xi_k) - H(y_{k+1}, \xi_k), x_{k+1} - y_{k+1} \rangle, \ \forall z \in X_k. \end{aligned}$$

By (3.7), for all $k \ge k_0$, we have

$$\begin{aligned} \|x_{k+1} - z\|^2 &\le \|x_k - z\|^2 - \|x_{k+1} - y_{k+1}\|^2 - \|y_{k+1} - x_k\|^2 + 2\alpha_k \langle H(y_{k+1}, \xi_k), z - y_{k+1} \rangle \\ &\quad + 2\alpha_k \|H(x_k, \xi_k) - H(y_{k+1}, \xi_k)\| \|x_{k+1} - y_{k+1}\| \\ &\le \|x_k - z\|^2 - \|x_{k+1} - y_{k+1}\|^2 - \|y_{k+1} - x_k\|^2 + 2\alpha_k \langle H(y_{k+1}, \xi_k), z - y_{k+1} \rangle \\ &\quad + 2\rho \|x_k - y_{k+1}\| \|x_{k+1} - y_{k+1}\| \\ &\le \|x_k - z\|^2 - (1-\rho)\|x_{k+1} - y_{k+1}\|^2 - (1-\rho)\|y_{k+1} - x_k\|^2 \\ &\quad + 2\alpha_k \langle H(y_{k+1}, \xi_k), z - y_{k+1} \rangle \\ &= \|x_k - z\|^2 - (1-\rho)(\|x_{k+1} - y_{k+1}\|^2 + \|y_{k+1} - x_k\|^2) \\ &\quad + 2\alpha_k \langle F(y_{k+1}), z - y_{k+1} \rangle + 2\alpha_k \langle \eta_k, z - y_{k+1} \rangle, \ \forall z \in X_k. \end{aligned} \tag{3.11}$$

Observe that F is pseudomonotone on $\mathbb{R}^n$. Letting $z = x^* \in X^*$ in (3.11), and noting that $X^* \subset X \subset X_k$, we obtain that

$$\begin{aligned}\|x_{k+1}-x^*\|^2 &\leq \|x_k - x^*\|^2 - (1-\rho)(\|x_{k+1}-y_{k+1}\|^2 + \|y_{k+1}-x_k\|^2) + 2\alpha_k\langle \eta_k, x^* - y_{k+1}\rangle \\ &\leq \|x_k - x^*\|^2 - (1-\rho)(\|x_{k+1}-y_{k+1}\|^2 + \|y_{k+1}-x_k\|^2) + 2\alpha_k\|\eta_k\|\|x^* - y_{k+1}\| \\ &\leq \|x_k - x^*\|^2 + 2\alpha_0\|\eta_k\|\|x^*-y_{k+1}\|, \ \forall k \geq k_0.\end{aligned} \tag{3.12}$$

Taking $\mathbb{E}[\cdot|\mathscr{F}_k]$ in the equality above and using Lemma 3.1 with $p = 2$, we have

$$\begin{aligned}\mathbb{E}[\|x_{k+1}-x^*\|^2|\mathscr{F}_k] &\leq \|x_k - x^*\|^2 + 2\alpha_0\left(\frac{m_1}{\sqrt{N_k}} + \frac{m_2}{\sqrt{N_k}}\|x_k - x^*\|^2\right) \\ &= \left(1 + \frac{2m_2\alpha_0}{\sqrt{N_k}}\right)\|x_k - x^*\|^2 + \frac{2\alpha_0 m_1}{\sqrt{N_k}}, \ \forall k \geq k_0.\end{aligned} \tag{3.13}$$

By the hypothesis on $\{N_k\}$, (3.13), and Lemma 2.6, we obtain that $\{\|x_k - x^*\|^2\}$ is a.s. convergent. Hence $\{x_k\}$ is a.s. bounded. This completes the proof. □

Lemma 3.3. $\sum_{k=0}^{\infty}\mathbb{E}[\|x_{k+1}-y_k\|^2 + \|y_{k+1}-x_k\|^2] < \infty$.

Proof. It follows from (3.12) that

$$\begin{aligned}(1-\rho)(\|x_{k+1}-y_{k+1}\|^2 + \|y_{k+1}-x_k\|^2) \leq \|x_k - x^*\|^2 - \|x_{k+1}-x^*\|^2 \\ + 2\alpha_0\|\eta_k\|\|x^* - y_{k+1}\|, \ \forall k \geq k_0.\end{aligned} \tag{3.14}$$

Lemma 3.2 indicates that $\{\|x_k - x^*\|^2\}$ is a.s. convergent and hence is a.s. bounded. Thus there exists a positive constant M such that a.s. $M \geq \sup_{k\in\mathbb{N}_0}\|x_k - x^*\|^2$. Taking $\mathbb{E}[\cdot]$ in (3.14) and using Lemma 3.1 with $p = 2$, we have

$$\begin{aligned}&(1-\rho)\mathbb{E}[\|x_{k+1}-y_k\|^2 + \|y_{k+1}-x_k\|^2] \\ &\leq \mathbb{E}[\|x_k - x^*\|^2] - \mathbb{E}[\|x_{k+1}-x^*\|^2] + 2\alpha_0\mathbb{E}[\|\eta_k\|\|x^* - y_{k+1}\|] \\ &\leq \mathbb{E}[\|x_k - x^*\|^2] - \mathbb{E}[\|x_{k+1}-x^*\|^2] + 2\alpha_0\mathbb{E}\left[\frac{m_1}{\sqrt{N_k}} + \frac{m_2}{\sqrt{N_k}}\|x_k - x^*\|^2\right] \\ &\leq \mathbb{E}[\|x_k - x^*\|^2] - \mathbb{E}[\|x_{k+1}-x^*\|^2] + 2\alpha_0\left(\frac{m_1}{\sqrt{N_k}} + \frac{m_2 M}{\sqrt{N_k}}\right), \ \forall k \geq k_0,\end{aligned} \tag{3.15}$$

where m_1 and m_2 are given as in Lemma 3.1. Taking a sum recursively for k from k_0 to K with $K \geq k_0$ in 3.15 yields that

$$\begin{aligned}&(1-\rho)\sum_{k=k_0}^{K}\mathbb{E}[\|x_{k+1}-y_k\|^2 + \|y_{k+1}-x_k\|^2] \\ &\leq \mathbb{E}[\|x_{k_0} - x^*\|^2] - \mathbb{E}[\|x_{K+1}-x^*\|^2] + 2\alpha_0\sum_{k=k_0}^{K}\left(\frac{m_1}{\sqrt{N_k}} + \frac{m_2 M}{\sqrt{N_k}}\right) \\ &\leq \mathbb{E}[\|x_{k_0} - x^*\|^2] + 2\alpha_0\sum_{k=k_0}^{\infty}\left(\frac{m_1}{\sqrt{N_k}} + \frac{m_2 M}{\sqrt{N_k}}\right) \\ &< M + 2\alpha_0\sum_{k=0}^{\infty}\left(\frac{m_1}{\sqrt{N_k}} + \frac{m_2 M}{\sqrt{N_k}}\right) < \infty.\end{aligned}$$

Letting $K \to \infty$, we obtain the desired conclusion immediately. This completes the proof. $\square$

Theorem 3.1. *The sequence $\{x_k\}$ generated by Algorithm 1 a.s. converges to some point $\bar{x} \in X^*$.*

Proof. In the proof Lemma 3.2, we have proved that $\{\alpha_k\}$ is nonincreasing and is bounded away from zero. Hence there exists some $\hat{\alpha} > 0$ such that $\lim_{k\to\infty} \alpha_k = \hat{\alpha}$. Lemmas 2.1 and 2.2 yield

$$
\begin{aligned}
(\min\{1,\hat{\alpha}\})^2(r_1(x_k))^2 &\le (\min\{1,\alpha_k\})^2(r_1(x_k))^2 \le (r_{\alpha_k}(x_k))^2 \\
&\le 2\|x_k - y_{k+1}\|^2 + 2\|y_{k+1} - \Pi_{X_k}(x_k - \alpha_k F(x_k))\|^2 \\
&= 2\|x_k - y_{k+1}\|^2 + 2\|\Pi_{X_k}(x_k - \alpha_k(F(x_k)+\varepsilon_k)) - \Pi_{X_k}(x_k - \alpha_k F(x_k))\|^2 \\
&\le 2\|x_k - y_{k+1}\|^2 + 2\alpha_k^2\|\varepsilon_k\|^2 \\
&\le 2\|x_k - y_{k+1}\|^2 + 2\alpha_0^2\|\varepsilon_k\|^2,\ \forall k \in \mathbb{N}_0.
\end{aligned}
$$

Taking $\mathbb{E}[\cdot|\mathscr{F}_k]$ in the inequality above and using $\left|\|\varepsilon_k\|^2\middle|\mathscr{F}_k\right|_{\frac{p}{2}} = \left|\|\varepsilon_k\|\middle|\mathscr{F}_k\right|_p^2$, we obtain by (3.3) with $p=2$ that

$$
\begin{aligned}
(\min\{1,\hat{\alpha}\})^2(r_1(x_k))^2 &\le 2\mathbb{E}[\|x_k - y_{k+1}\|^2|\mathscr{F}_k] + 2\alpha_0{}^2\mathbb{E}[\|\varepsilon_k\|^2|\mathscr{F}_k] \\
&\le 2\mathbb{E}[\|x_k - y_{k+1}\|^2] + 4\alpha_0^2\frac{c_1^2(\sigma_4(x^*))^2 + \bar{L}_4^2\|x_k - x^*\|^2}{\sqrt{N_k}} \\
&\le 2\mathbb{E}[\|x_k - y_{k+1}\|^2] + 4\alpha_0^2\frac{c_1^2(\sigma_4(x^*))^2 + \bar{L}_4^2 M}{\sqrt{N_k}},\ \forall k \in \mathbb{N}_0.
\end{aligned}
$$

Letting $k \to \infty$ in the inequality above and using Lemma 3.3, we have

$$\lim_{k\to\infty}(r_1(x_k))^2 = \lim_{k\to\infty}\|x_k - \Pi_{X_k}(x_k - F(x_k))\|^2 = 0.$$

For convenience, we put $u_k = \Pi_{X_k}(x_k - F(x_k))$ for each $k \in \mathbb{N}_0$. It follows that

$$\lim_{k\to\infty}\|x_k - u_k\|^2 = 0. \tag{3.16}$$

On the other hand, by Lemma 2.1 (iii) and the definition of u_k, we have $\langle u_k - z, x_k - F(x_k) - u_k\rangle \ge 0$, $\forall z \in X_k$. Since $X \subset X_k$ for each $k \in \mathbb{N}_0$, we have

$$\langle u_k - z, x_k - F(x_k) - u_k\rangle \ge 0,\ \forall z \in X. \tag{3.17}$$

From Lemma 3.2, it follows that there exists a subsequence $\{x_{k_j}\}$ of $\{x_k\}$ that a.s. converges to a point $\bar{x}$. From (3.16), it follows that $\{u_{k_j}\}$ converges to $\bar{x}$ as $j \to \infty$. Note that $\bar{x} \in X$. In fact, from $y_{k_j+1} \in X_{k_j}$ and the definition of X_{k_j} we have

$$c_{i_{k_j}}(x_{k_j}) + \langle v_{k_j}, y_{k_j+1} - x_{k_j}\rangle \le 0,\ \forall j \in \mathbb{N}. \tag{3.18}$$

Note that $\{x_k\}$ is a.s. bounded and each ∂c_i $(i \in \{1,\cdots,m\})$ is bounded on bounded set. Thus there exists a constant $s > 0$ such that $\|v_{i_{k_j}}\| \le s$ for all $j \in \mathbb{N}$. Note that Lemma 3.3 implies that $\lim_{j\to\infty}\|x_{k_j} - y_{k_j+1}\| = 0$. Hence it follows from ((3.18))that

$$c_{i_{k_j}}(x_{k_j}) \le s\|y_{k_j+1} - x_{k_j}\| \to 0 \text{ as } j \to \infty. \tag{3.19}$$

The *w-lsc* of c_i and (3.19) imply that, for each $i \in \{1,\cdots,m\}$,

$$c_i(\bar{x}) \le \liminf_{j\to\infty} c_i(x_{k_j}) \le \liminf_{j\to\infty} c_{i_{k_j}}(x_{k_j}) \le 0.$$

It follows that $\bar{x} \in X = \cap_{i=1}^{m} X_i$.

Next we prove that $\bar{x} \in X^*$. To end this, replacing k with k_j in (3.17) and letting $j \to \infty$, we obtain from (3.16) that $\langle z - \bar{x}, F(\bar{x}) \rangle \geq 0, \forall z \in X$, which with $\bar{x} \in X$ implies that $\bar{x} \in X^*$.

Finally, we prove that $\{x_k\}$ a.s. converges to $\bar{x}$. Note that we have proved that the limit of $\|x_k - \bar{x}\|$ a.s. exists by Lemma 3.2. Thus $\lim_{k\to\infty} \|x_k - \bar{x}\|^2 = \lim_{j\to\infty} \|x_{k_j} - \bar{x}\|^2 = 0$. This completes the proof. □

In the final part of this section, we discuss the convergence rate of Algorithm 1. Remark 3.5 demonstrates that $x_k \in X^*$ if $x_k = y_{k+1}$ for some $k \in \mathbb{N}$ and Lemma 3.3 implies that $\lim_{k\to\infty} \|x_k - y_{k+1}\| = 0$. Hence we can show the convergence rate of Algorithm 1 by considering the convergence rate of $\{\|x_k - y_{k+1}\|^2\}$. On the convergence rate of Algorithm 1, we have the following conclusion.

Theorem 3.2. *For any $K \in \mathbb{N}$ with $K > k_0$, the following holds:*

$$\min_{k=k_0,\cdots,K} \mathbb{E}[\|y_{k+1} - x_k\|^2] \leq \frac{\overline{M}}{1+K-k_0},$$

where $\overline{M}$ is a constant, and

$$k_0 = \max\{k \in \mathbb{N} : \alpha_k \|H(x_k, \xi_k) - H(y_{k+1}, \xi_k)\| > \rho \|x_k - y_{k+1}\|\}.$$

Proof. From (3.7), we see that there exists a positive integer k_0 such that

$$\alpha_k \|H(x_k, \xi_k) - H(y_{k+1}, \xi_k)\| \leq \rho \|x_k - y_{k+1}\|, \ \forall k \geq k_0.$$

Hence

$$k_0 = \max\{k \in \mathbb{N} : \alpha_k \|H(x_k, \xi_k) - H(y_{k+1}, \xi_k)\| > \rho \|x_k - y_{k+1}\|\}.$$

On the other hand, set $\overline{M} = M + 2\alpha_0 \sum_{k=0}^{\infty} \left(\frac{m_1}{\sqrt{N_k}} + \frac{m_2 M}{\sqrt{N_k}} \right)$, where M is defined as in the proof line of Lemma 3.3. It follows from (3.1) that

$$(1-\rho) \sum_{k=k_0}^{K} \mathbb{E}[\|x_{k+1} - y_k\|^2 + \|y_{k+1} - x_k\|^2] < \overline{M}.$$

Finally, by the fact that

$$\min_{k=k_0,\cdots,K} \mathbb{E}[\|y_{k+1} - x_k\|^2] \leq \frac{1}{1+K-k_0} \sum_{k=k_0}^{K} \mathbb{E}[\|y_{k+1} - x_k\|^2],$$

we obtain the desired conclusion. This completes the proof. □

4. Numerical Experiment

In this section, we illustrate the effectiveness of Algorithm 1 by a synthetic stochastic optimization problem with the constraint, which is a stochastic modification of the example in [13]. The code of performing Algorithm 1 was written by Matlab R2016b running on a MacBook air Desktop with Core(TM) i5-4260U CPU @1.40GHz 2.00GHz, RAM 4GB.

Consider the following stochastic optimization problem:

$$\min_{x \in X} \mathbb{E}[g(x, \xi)], \tag{4.1}$$

where $g(\cdot, \xi) : \mathbb{R}^3 \to \mathbb{R}$ is a continuously differentiable convex function for each realization value of a random vector ξ, and $E[g(x, \xi)]$ denotes the expected value of $g(x, \xi)$. Let $F(x) =$

$\mathbb{E}[g(x,\xi)]$ for all $x \in \mathbb{R}^m$. Since $g(x,\xi)$ is convex in x, it follows that $\nabla F(x) = \mathbb{E}[\nabla g(x,\xi)]$ for each x. It is equivalent to the stochastic variational inequality problem of finding a point $x^* \in X$ such that $\langle \mathbb{E}[\nabla g(x^*,\xi)], y - x^* \rangle \geq 0$, $\forall y \in X$.

Let $\xi = (\xi_1, \xi_2, \xi_3)$ be a random vector, where ξ_1 is a normal random variable with mean $\mu = 1$ and variance $\sigma^2 = 5$, ξ_2 is a standard normal random variable, and ξ_3 is a random variable obeying exponential distribution with parameter $\theta = 1$. Let the feasible set $X = \cap_{i=1}^4 X_i$, where

$$X_1 = \left\{(x_1,x_2,x_3) \in \mathbb{R}^3 : \frac{x_1^2}{31} + \frac{x_2^2}{36} + \frac{x_3^2}{16} \leq 1\right\},$$
$$X_2 = \left\{(x_1,x_2,x_3) \in \mathbb{R}^3 : \frac{(x_1+1)^2}{25} + \frac{(x_2+1)^2}{9} + \frac{(x_3-1)^2}{36} \leq 1\right\},$$
$$X_3 = \left\{(x_1,x_2,x_3) \in \mathbb{R}^3 : (x_1-1)^2 + x_2^2 + (x_3+2)^2 \leq 8\right\},$$
$$X_4 = \left\{(x_1,x_2,x_3) \in \mathbb{R}^3 : \frac{(x_1+1)^2}{15} + \frac{(x_2+2)^2}{10} + \frac{(x_3+0.5)^2}{14} \leq 1\right\}.$$

Consider the convex optimization problem (4.1) with

$$g(x,\xi) = \frac{1}{2}x^T Q(\xi)x + q(\xi)^T x.$$

Since $g(x,\xi)$ is convex in $x \in \mathbb{R}^3$, it follows that $\nabla \mathbb{E}[g(x,\xi)] = \mathbb{E}[\nabla g(x,\xi)]$ for all $x \in \mathbb{R}^3$. So the above stochastic optimization problem is equivalent to the following stochastic variational inequality problem: find a point $x^* \in X$ such that

$$\langle \mathbb{E}[Q(\xi)]x^* + \mathbb{E}[q(\xi)], y - x^* \rangle \geq 0, \ \forall y \in X. \tag{4.2}$$

Test 1. Set

$$Q(\xi) = \begin{pmatrix} \xi_1 & 0 & 0 \\ 0 & 10+\xi_2 & 0 \\ 0 & 0 & 30\xi_3 \end{pmatrix} \quad \text{and} \quad q(\xi) = \begin{pmatrix} \xi_1 \\ 2 \\ 3\xi_3 \end{pmatrix}.$$

In this case,

$$\mathbb{E}[Q(\xi)] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 30 \end{pmatrix} \text{ and } \mathbb{E}[q(\xi)] = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Let

$$F(x) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 30 \end{pmatrix} x + \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \ \forall x \in C.$$

Then stochastic inequality problem (4.2) becomes the problem of finding $x^* \in X$ such that

$$\langle F(x^*), y - x^* \rangle \geq 0, \ \forall y \in X. \tag{4.3}$$

It is easy to check that $F(x)$ is pseudomonotone and Lipschitz continuous on $\mathbb{R}^3$. The unique solution of (4.3) is $x^* = (-1, -0.2, -0.1)^T$. Since the exact solution is known, we use $\|x_k - x^*\|$ to measure the distance between the n-th iterate x_k and x^*. The maximum iteration of 1000 as the stopping criterion is used for Algorithm 1.

Table 1 and 2 give the numerical results of Algorithm 1 in this test with different initial points, sample rate, and input parameters.

Table 1 Numerical results of Algorithm 1 with sample rate $N_k = \lceil 0.991^{-k} \rceil$, initial points $x_0 = x_1 = y_1$ generated randomly from (0,1)

	$\delta = 0.1, \alpha_0 = 1, \rho = 0.2$	$\delta = 0.5, \alpha_0 = 2, \rho = 0.8$	$\delta = 0.8, \alpha_0 = 4, \rho = 0.5$
Iteration	$\|x_k - x^*\|$	$\|x_k - x^*\|$	$\|x_k - x^*\|$
100	2.6083	0.0573	2.6938×10^{11}
200	2.7157	0.0784	0.0670
300	2.1841	0.0327	0.0306
400	2.1235	0.0055	0.0126
500	1.8713	0.0030	0.0062
600	1.6412	0.0014	0.0027
700	1.4632	6.0380×10^{-4}	0.0012
800	1.3113	2.5793×10^{-4}	6.0154×10^{-4}
900	0.9003	8.9288×10^{-5}	2.8114×10^{-4}
970	0.8433	6.4037×10^{-5}	1.6561×10^{-4}
980	0.8350	5.9032×10^{-5}	1.5320×10^{-4}
990	0.8269	4.5105×10^{-5}	1.4328×10^{-4}
1000	0.8193	4.0899×10^{-5}	1.2818×10^{-4}
CPU time (s)	107.4430	105.2757	83.1931

Table 2 Numerical results of Algorithm 1 with sample rate $N_k = \lceil (k/100)^3 \rceil$, initial points $x_0 = x_1 = y_1 = (-20, 10, 5)$

	$\delta = 0.1, \alpha_0 = 1, \rho = 0.2$	$\delta = 0.5, \alpha_0 = 2, \rho = 0.8$	$\delta = 0.8, \alpha_0 = 4, \rho = 0.5$
Iteration	$\|x_k - x^*\|$	$\|x_k - x^*\|$	$\|x_k - x^*\|$
100	3.8193×10^{35}	7.2443×10^{93}	6.3604×10^{130}
200	3.0129×10^{5}	5.9946×10^{69}	5.4290×10^{100}
300	0.3809	4.7289×10^{39}	4.2827×10^{70}
400	0.1337	3.7304×10^{9}	3.3784×10^{40}
500	0.0932	0.0953	2.6651×10^{10}
600	0.0601	0.0625	0.0764
700	0.0559	0.0420	0.0361
800	0.0514	0.0284	0.0160
900	0.0471	0.0192	0.0072
970	0.0441	0.0147	0.0042
980	0.0438	0.0142	0.0039
990	0.0434	0.0136	0.0036
1000	0.0430	0.0131	0.0033
CPU time (s)	28.1855	26.8902	37.7395

Test 2. Set

$$Q(\xi)=\begin{pmatrix} \xi_1\xi_2 & 0 & 0 \\ 0 & 10+\xi_2 & 0 \\ 0 & 0 & 30\xi_1\xi_3 \end{pmatrix} \quad \text{and} \quad q(\xi)=\begin{pmatrix} \xi_1\xi_2 \\ 2 \\ 3\xi_1\xi_3 \end{pmatrix}.$$

Note that we do not assume that ξ_1 and ξ_2, and ξ_1 and ξ_3 are independent. Hence the joint density function of ξ_1 and ξ_2, and ξ_1 and ξ_3 are unknown, which results in that the values of $\mathbb{E}[Q(\xi)]$ and $\mathbb{E}[q(\xi)]$ can not be computed explicitly. In this case, the exact solution of (4.2) is unknown. Remark 3.5 shown that if $y_{k+1}=x_k$, then x_k is the solution to (4.2). Hence we may use the convergence of $\{\|y_{k+1}-x_k\|\}$ to illustrate the convergence of Algorithm 1 in this test. Figure 1 and 2 give the numerical results of Algorithm 1 with different initial points, sample rate, and input parameters.

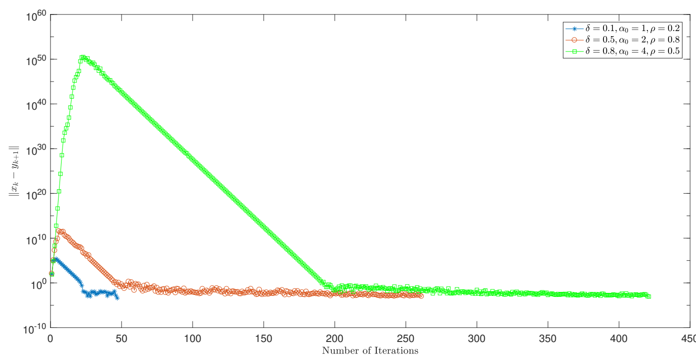

Figure 1 CPU times for Algorithm 1 and ESA with different dimension

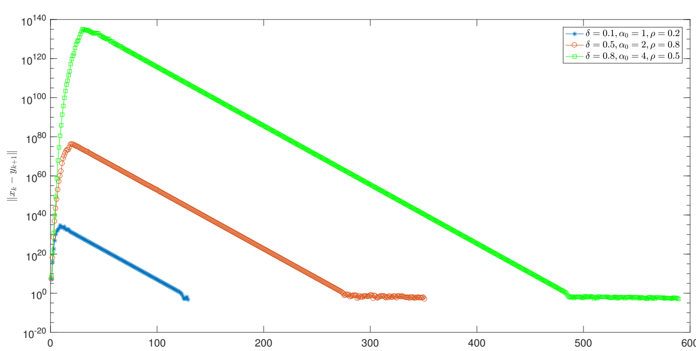

Figure 2 CPU times for Algorithm 1 and ESA with different dimension

From Table 1, Table 2, Figure 1, and Figure 2, we see that the numerical results of Algorithm 1 are sensitive to the initial points, sample rate and input parameters. In particular, in Table 2, the value of $\|x_k-x^*\|$ is larger for the input parameters $\delta=0.8,\alpha_0=4$, and $\rho=0.5$ when the iterative step $k\geq 500$.

5. Conclusion

In this paper, we investigated a SVIP with constraint functions and proposed a method by combining the SA method and the selective projection method in [13] for solving the problem, in which a simple adaptive rule was used to deal with the unknown Lipschitz constant of involved mapping. Since the projections are onto the half-spaces, they can be computed explicitly and in fact there is no projection is included in our method. The convergence and convergence rate of the proposed algorithm were proved and discussed. Finally, a numerical example was given to illustrate the convergence of the proposed method. However, a disadvantage of our method is that the selection process of i_k is a computational workload, especially when the number of constrained functions is very large. How to overcome this weakness is one of our future topics.

References

[1] H. Jiang, H. Xu, Stochastic approximation approaches to the stochastic variational inequality problem, IEEE Trans. Automat. Control 53 (2008), 1462–1475.
[2] H. Robbins, S. Monro, A stochastic approximation method, Ann. Math. Stat. 22 (1951), 400-407.
[3] A. Juditsky, A. Nemirovski, C. Tauvel, Solving variational inequalities with stochastic mirror-prox algorithm, Stoch. Syst. 1 (2011), 17-58.
[4] J. Koshal, A. Nedić, U.V. Shanbhag, Regularized iterative stochastic approximation methods for stochastic variational inequality problems, IEEE Trans. Automat. Control 58 (2013), 594-609.
[5] A. Kannan, U.V. Shanbhag, The pseudomonotone stochastic variational inequality problem: analytical statements and stochastic extragradient schemes, In: The 2014 Conference on American Control Conference, pp. 2930–2935, 2014.
[6] A. Kannan, U.V. Shanbhag, Optimal stochastic extragradient schemes for pseudomonotone stochastic variational inequality problems and their variants, Comput. Optim. Appl. 74 (2019), 779–820.
[7] F. Yousefian, A. Nedic, U.V. Shanbhag, Optimal robust smoothing extragradient algorithms for stochastic variational inequality problems, In: The 2014 IEEE Annual Conference on Decision and Control, pp. 5831–5836, 2014.
[8] F. Yousefian, A. Nedić, U.V. Shanbhag, Distributed adaptive steplength stochastic approximation schemes for Cartesian stochastic variational inequality problems, IEEE Trans. Automat. Control 61 (2016), 1753–1766.
[9] 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.
[10] 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.
[11] X.J. Zhang, X.W. Du, Z.P. Yang, G.H. Lin, An infeasible stochastic approximation and projection algorithm for stochastic variational inequalities, J. Optim. Theory Appl. 183 (2019), 1053-1076.
[12] Z.P. Yang, Y.L. Wang, G.H. Lin, Variance-based modified backward-forward algorithm with line search for stochastic variational inequality problems and its applications, Pacific J. Oper. Res. 37 (2020), 2050011.
[13] S.N. He, H.L. Tian, Selective projection methods for solving a class of variational inequalities, Numer. Algo. 80 (2019), 617-634.
[14] C. Fang, S. Chen, Some extragradient algorithms for variational inequalities, In: W. Han, S. Migórski, M. Sofonea, M. (eds) Advances in Variational and Hemivariational Inequalities, Advances in Mechanics and Mathematics, vol 33. Springer, Cham, 2015.
[15] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, 1970.
[16] J.Y. Bello Cruz, R.D. Millán, A relaxed-projection splitting algorithm for variational inequalities in Hilbert spaces, J. Glob. Optim. 65 (2016), 597–614.
[17] H. Robbins, D. Siegmund, A convergence theorem for nonnegative almost supermartingales and some applications, Optim. Meth. Statist. (1971), 233–257.