Proceedings of the 2013 Winter Simulation Conference
R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, eds.

A REGULARIZED SMOOTHING STOCHASTIC APPROXIMATION (RSSA) ALGORITHM FOR STOCHASTIC VARIATIONAL INEQUALITY PROBLEMS

Farzad Yousefian

Industrial & Enterprise Systems Engineering
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA

Angelia Nedić

Industrial & Enterprise Systems Engineering
University of Illinois at Urbana-Champaign
Urbana, IL 61801, USA

Uday V. Shanbhag

Industrial & Manufacturing Engineering
Pennsylvania State University
University Park, PA 16802, USA

ABSTRACT

We consider a stochastic variational inequality (SVI) problem with a continuous and monotone mapping over a compact and convex set. Traditionally, stochastic approximation (SA) schemes for SVIs have relied on strong monotonicity and Lipschitzian properties of the underlying map. We present a regularized smoothed SA (RSSA) scheme wherein the stepsize, smoothing, and regularization parameters are diminishing sequences. Under suitable assumptions on the sequences, we show that the algorithm generates iterates that converge to a solution in an almost-sure sense. Additionally, we provide rate estimates that relate iterates to their counterparts derived from the Tikhonov trajectory associated with a deterministic problem.

1 INTRODUCTION

Variational inequalities (VI) represent an immensely important object in applied mathematics and operations research. Variational inequality models find application in capturing a range of optimization and equilibrium problems in engineering, economics, game theory, and finance. Given a set $X \subset \mathbb{R}^n$ and a mapping $F : X \to \mathbb{R}^n$, a VI problem (Facchinei and Pang 2003; Rockafellar and Wets 1998) denoted by VI(X,F), requires a vector $x^* \in X$ such that $F(x^*)^T(x-x^*) \geq 0$, for any $x \in X$. We consider a stochastic generalization of this problem in which the components of the map contain expectations. We are interested in solving VI(X,F) where mapping $F : X \to \mathbb{R}^n$ represents the expected value of a stochastic mapping $\Phi : X \times \Omega \to \mathbb{R}^n$, i.e., $F(x) \triangleq \mathsf{E}[\Phi(x,\xi(\omega))]$ where $\xi : \Omega \to \mathbb{R}^d$ is a $d-$dimensional random variable and with the probability space $(\Omega, \mathscr{F}, \mathbb{P})$. $x^* \in X$ solves VI (X,F) if

$$\mathsf{E}[\Phi(x^*,\xi(\omega))]^T(x-x^*) \geq 0, \qquad \text{for any } x \in X.$$

For purposes of brevity, we let ξ denote $\xi(\omega)$. While SVIs are a natural extension of their deterministic counterparts, generally deterministic schemes cannot be applied directly unless the expectation of the mapping can be efficiently computed. Our interest in this paper is pertaining to finding an exact solution to such problems when the expectations are unavailable in a closed form. Consequently, Monte-Carlo sampling schemes assume relevance. Stochastic approximation methods (SA) and sample average approximation methods (SAA) are amongst the well-known solution approaches in this regime. Moreover, a recent approach for addressing approximate solution of SVI problems is the stochastic mirror-prox algorithm (Juditsky, Nemirovski, and Tauvel 2011). That method allows for both smooth and nonsmooth problems and optimal rate of convergence is attained for a constant choice of the stepsizes. SA methods, first proposed by Robbins and Monro (Robbins and Monro 1951), were motivated by stochastic root-finding problems. The goal in such problems is to find a vector $x \in \mathbb{R}^n$ such that $\mathsf{E}[g(x,\xi)] = 0$, where $\xi : \Omega \to \mathbb{R}^d$ is a random variable, $g(\cdot,\xi) : \mathbb{R}^n \to \mathbb{R}^n$ is a continuous function for any realization of ξ. The SA scheme is based on the iterative scheme $x_{k+1} = x_k - \gamma_k g(x_k,\xi_k)$ for all $k \geq 0$, where $\gamma_k > 0$ is the stepsize and ξ_k is the realization of random variable ξ at k-th iteration .

A comprehensive review on SAA methods in the context of stochastic generalized equations has been provided by Shapiro (Shapiro 2003). Xu investigated the application of SAA methods for the solution of SVIs (Xu 2010). While SA methods have been extensively used in stochastic optimization regime (Ermoliev 1983; Kushner and Yin 2003; Cicek, Broadie, and Zeevi 2011), Jiang and Xu have recently introduced employing SA schemes for solving SVIs (Jiang and Xu 2008). They considered the SVI problem with a strongly monotone and Lipschitz mapping over a closed and convex set and provided global convergence results. In an extension of that work, a regularized SA method is developed for solving SVIs with a merely monotone and Lipschitz mapping (Koshal, Nedić, and Shanbhag 2010). In such a scheme, Lipschitz property of the mapping is still required. The main motivation of this work is addressing ill-posed SVIs where both the strong monotonicity and Lipschitz property of F are either unavailable or cannot be shown.

Before proceeding, we consider the question of nonsmoothness. In a deterministic regime, most of researchers contended with nonsmothness through introducing a sequence of smooth and approximate problems (Facchinei, Jiang, and Qi 1999) or using conjugate and proximal functions (Nesterov 2005). A challenge associated with applying such schemes in stochastic regimes is that they require a closed form of the stochastic functions while such information may not be available. Our work is motivated by a class of averaged functions first introduced by Steklov (Steklov 1907). Several researchers have employed this approach in stochastic programming and optimization (Bertsekas 1973; Norkin 1993) and more recently (Lakshmanan and Farias 2008; Duchi, Bartlett, and Wainwright 2012). It is well-known that given a convex function $f(x) : \mathbb{R}^n \to \mathbb{R}$ and a random variable ω with probability distribution $P(\omega)$, the function $\hat{f}$ defined by $\hat{f}(x) \triangleq \int_{\mathbb{R}^n} f(x+\omega)P(\omega)d\omega = \mathsf{E}[f(x+\omega)]$, is a differentiable function. Employing this technique allowed us to address nonsmoothness in developing adaptive stepsizes SA schemes for stochastic convex optimization problems and Cartesian SVIs in absence or unavailability of a Lipschitz constant (Yousefian, Nedić, and Shanbhag 2012; Yousefian, Nedić, and Shanbhag 2013). A main difference between the present paper and our preceding work is that here we let the smoothing parameter go to zero as the SA algorithm proceeds. This enables us to reach the solution of the original problem rather than an approximate problem. Our main contributions are as follows:

- Addressing nonsmoothness and absence of strong monotonicity: As mentioned earlier, the Lipschitz property of the mapping has been among the main assumptions of much of the previous research. Given an SVI problem, our main goal is to address ill-posed SVI problems by deriving the strong monotonicity and Lipschitzian properties through employing regularization and local smoothing techniques simultaneously.
- Convergence rate analysis: Our second goal lies in analyzing the rate of convergence for the proposed SA method. Suppose $\{x_k\}$ is generated by our proposed SA method and s_k is the solution to the kth regularized and smoothed SVI problem, we derive a bound for the error $\mathsf{E}\left[\|x_{k+1} - s_k\|^2\right]$.

The rest of the paper is organized as follows. Section 2 describes our proposed SA method and the main assumptions of the problem. Section 3 gives the main theoretical results and properties of the proposed SA method. In particular, the almost-sure convergence of the algorithm is provided. In section 4, we focus on analyzing the convergence rate of the algorithm and derive a bound for a particular error of the scheme.

Notation: In this paper, a vector x is assumed to be a column vector, x^T denotes the transpose of a vector x, and $\|x\|$ denotes the Euclidean vector norm, i.e., $\|x\| = \sqrt{x^T x}$. We use $\Pi_X(x)$ to denote the

Euclidean projection of a vector x on a set X, i.e., $\|x-\Pi_X(x)\| = \min_{y\in X}\|x-y\|$. We write *a.s.* as the abbreviation for "almost surely". We use $\mathsf{E}[z]$ to denote the expectation of a random variable z.

2 ALGORITHM OUTLINE

We consider the following algorithm where the sequence $\{x_k\}$ is generated by

$$x_{k+1} = \Pi_X\left(x_k - \gamma_k(\Phi(x_k+z_k)+\eta_k x_k)\right), \qquad \text{for all } k\geq 0. \tag{1}$$

where $\{\gamma_k\}$ is the stepsize sequence, $\{\eta_k\}$ is the regularization sequence, $z_k\in\mathbb{R}^n$ is a uniform random variable over the n-dimensional ball centered at the origin with radius ε_k for any $k\geq 0$, and $x_0\in X$ is a random initial vector that is independent of the random variable ξ and such that $\mathsf{E}\left[\|x_0\|^2\right]<\infty$. To have a well defined Φ in algorithm (2), we define the set X^ε as $X^\varepsilon \triangleq X + B_n(0,\varepsilon)$ where the scalar $\varepsilon>0$ is an upper bound of the sequence $\{\varepsilon_k\}$ and $B_n(y,\rho)$ is defined as the ball centered at point y with radius ρ, i.e. $B_n(y,\rho) = \{x\in\mathbb{R}^n \mid \|x-y\|\leq\rho\}$. We let $\mathrm{SOL}(X,F)$ denote the solution set of $\mathrm{VI}(X,F)$ and $\mathscr{F}_k$ denote the history of the method up to time k, i.e., $\mathscr{F}_k = \{x_0,\xi_0,\xi_1,\ldots,\xi_{k-1},z_1,\ldots,z_{k-1}\}$ for $k\geq 1$ and $\mathscr{F}_0=\{x_0\}$. Our first set of assumptions is on the properties of the set X, the mapping F, and random variables.

Assumption 1 Let the following hold:
(a) The set $X\subset\mathbb{R}^n$ is closed, bounded, and convex;
(b) $\Phi(x,\xi)$ is a monotone and continuous mapping over the set X^ε with respect to x for any $\xi\in\Omega$;
(c) $\mathrm{SOL}(X,F)\neq\emptyset$, i.e., there exists an $x^*\in X$ such that $(x-x^*)^T\mathsf{E}[\Phi(x^*,\xi)]\geq 0$, for all $x\in X$;
(d) Random variables z_i and ξ_j are both i.i.d and independent from each other for any $i,j\geq 0$.

Remark 1 Boundedness of the set X implies that there exists $M>0$ for which $\|x\|\leq M$ for any $x\in X$. Moreover, an immediate consequence of continuity of the mapping Φ over the bounded set X^ε is that there exists $C>0$ for which $\|\Phi(x,\xi)\|\leq C$ for any $x\in X^\varepsilon$. Taking expectations on both sides of the preceding inequality and using Jensen's Inequality, we have $\|F(x)\|\leq C$ for any $x\in X^\varepsilon$.

Remark 2 By introducing stochastic errors w_k, algorithm (1) can be rewritten as the following

$$\begin{aligned} x_{k+1} &= \Pi_X\left(x_k-\gamma_k(F(x_k+z_k)+\eta_k x_k + w_k)\right), \qquad \text{for all } k\geq 0\\ w_k &\triangleq \Phi(x_k+z_k,\xi_k) - F(x_k+z_k), \qquad \text{for all } k\geq 0. \end{aligned} \tag{2}$$

Note that the implementation of the algorithm (1) requires evaluation of the mapping Φ.

3 ALMOST-SURE CONVERGENCE

In this section, we present the main results of algorithm (2). After stating the main assumptions on the stepsize, regularization, and smoothing sequences, we establish the convergence result by presenting different properties of the algorithm.

Assumption 2 Let the following hold:
(a) $\{\gamma_k\}$, $\{\eta_k\}$, and $\{\varepsilon_k\}$ are strictly positive sequences for $k\geq 0$ converging to zero;
(b) There exists $K_1\geq 0$ such that $\frac{\gamma_k}{\eta_k\varepsilon_k^2}\leq 0.5\left(\frac{(n-1)!!}{n!!\kappa C}\right)^2$ for any $k\geq K_1$, where n is the dimension of the space and $\kappa=1$ if n is odd and $\kappa=\frac{2}{\pi}$ otherwise;
(c) For any $k\geq 0$, $\varepsilon_k\leq\varepsilon$, where ε is the parameter of the set X^ε;
(d) $\sum_k^\infty \gamma_k\eta_k=\infty$; (e) $\sum_k^\infty\gamma_k^2<\infty$; (f) $\sum_k^\infty \frac{1}{\eta_{k-1}^2\eta_k\gamma_k}\left(1-\frac{\min\{\varepsilon_k,\varepsilon_{k-1}\}}{\max\{\varepsilon_k,\varepsilon_{k-1}\}}\right)^2<\infty$; (g) $\sum_k^\infty\frac{1}{\eta_k\gamma_k}\left(1-\frac{\eta_k}{\eta_{k-1}}\right)^2<\infty$;
(h) $\lim_{k\to\infty}\frac{\gamma_k}{\eta_k}=0$; (i) $\lim_{k\to\infty}\frac{1}{\eta_k^2\gamma_k}\left(1-\frac{\min\{\varepsilon_k,\varepsilon_{k-1}\}}{\max\{\varepsilon_k,\varepsilon_{k-1}\}}\right)=0$; (j) $\lim_{k\to\infty}\frac{1}{\eta_k\gamma_k}\left|1-\frac{\eta_k}{\eta_{k-1}}\right|=0$.

Remark 3 Later in Lemma 5, we provide a suitable choice for the sequences $\{\gamma_k\}$, $\{\eta_k\}$, and $\{\varepsilon_k\}$ that satisfies the conditions of Assumption 2.

The following supermartingale convergence theorem is a key in our analysis in establishing the almost-sure convergence of algorithm (2) and may be found in (Polyak 1987) (cf. Lemma 10, page 49).

Lemma 1 [Robbins Siegmund's Lemma] Let $\{v_k\}$ be a sequence of nonnegative random variables, where $\mathsf{E}[v_0] < \infty$, and let $\{\alpha_k\}$ and $\{\mu_k\}$ be deterministic scalar sequences such that $0 \leq \alpha_k \leq 1$, and $\mu_k \geq 0$ for all $k \geq 0$, $\sum_{k=0}^{\infty} \alpha_k = \infty$, $\sum_{k=0}^{\infty} \mu_k < \infty$, and $\lim_{k\to\infty} \frac{\mu_k}{\alpha_k} = 0$, and $\mathsf{E}[v_{k+1}|v_0,\ldots,v_k] \leq (1-\alpha_k)v_k + \mu_k$ a.s. for all $k \geq 0$. Then, $v_k \to 0$ almost surely as $k \to \infty$.

Lemma 2 [Properties of the stochastic errors w_k defined by (2)] Consider algorithm (2) and suppose Assumptions 1(b) and (d) hold. Then, the stochastic error w_k satisfies the following relations for any $k \geq 0$:

$$\mathsf{E}_\xi[w_k \mid \mathscr{F}_k] = 0 \text{ for any realization of } z_k \text{ and } \mathsf{E}\left[\|w_k\|^2 \mid \mathscr{F}_k\right] \leq C^2.$$

Proof. Let us assume that $k \geq 0$ is fixed. The definition of w_k in (2) implies that

$$\mathsf{E}_\xi[w_k \mid \mathscr{F}_k] = \mathsf{E}_\xi[\Phi(x_k+z_k,\xi_k) \mid \mathscr{F}_k] - F(x_k+z_k) = F(x_k+z_k) - F(x_k+z_k) = 0,$$

where we used the independence of z_k and ξ_k. To show the second inequality, we may write

$$\begin{aligned}\mathsf{E}\left[\|w_k\|^2 \mid \mathscr{F}_k\right] &= \mathsf{E}\left[\|\Phi(x_k+z_k,\xi_k) - F(x_k+z_k)\|^2 \mid \mathscr{F}_k\right] \\ &= \mathsf{E}\left[\|\Phi(x_k+z_k,\xi_k)\|^2 \mid \mathscr{F}_k\right] + \mathsf{E}\left[\|F(x_k+z_k)\|^2 \mid \mathscr{F}_k\right] - 2\mathsf{E}\left[\Phi(x_k+z_k,\xi_k)^T F(x_k+z_k) \mid \mathscr{F}_k\right].\end{aligned}$$

Since z_k and ξ_k are independent random variables (Assumption 1(b)), we can write

$$\mathsf{E}\left[\Phi(x_k+z_k,\xi_k)^T F(x_k+z_k) \mid \mathscr{F}_k\right] = \mathsf{E}_z\left[\mathsf{E}_\xi[\Phi(x_k+z_k,\xi_k) \mid \mathscr{F}_k]^T F(x_k+z_k) \mid \mathscr{F}_k\right] = \mathsf{E}\left[\|F(x_k+z_k)\|^2 \mid \mathscr{F}_k\right].$$

From the two preceding relations and the definition of C in Remark 1, we obtain the desired result. □

Next, we present a Lemma stating that the local smoothing technique preserves the monotonicity property. The proof of this Lemma is straightforward and is omitted.

Lemma 3 Suppose mapping $F : X^\varepsilon \to \mathbb{R}^n$ is monotone over the set X^ε. For $k \geq 0$, consider mappings $F_k : X \to \mathbb{R}$ where $F_k(x) = \mathsf{E}[F(x+z_k)]$ and $z_k \in \mathbb{R}^n$ is a uniform random variable defined on an n-dimensional ball with radius $\varepsilon_k > 0$ where $\varepsilon_k \leq \varepsilon$ for $k \geq 0$. Then, the mapping F_k is monotone over the set X.

Remark 4 Lemma 3 implies that the mapping $F_k + \eta_k\mathbf{I}$ is strongly monotone. When the set X is closed and convex, Theorem 2.3.3 of (Facchinei and Pang 2003) indicates that $\mathrm{VI}(X, F_k + \eta_k\mathbf{I})$ has a unique solution. Throughout this paper, we let the sequence $\{s_k\}$ be defined such that s_k is the unique solution of $\mathrm{VI}(X, F_k + \eta_k\mathbf{I})$ for $k \geq 0$, where $F_k : X \to \mathbb{R}^n$ is defined by $F_k(x) = \mathsf{E}[F(x+z_k)]$.

The following proposition, presents a bound on the rate $\|s_k - s_{k-1}\|$, convergence of $\{s_k\}$, and the Lipschitzian property of the approximate mapping F_k.

Proposition 1 [Convergence of $\{s_k\}$ and Lipschitzian property of F_k] Suppose Assumption 1 holds. Consider the sequence $\{s_k\}$ such that $s_k \in \mathrm{SOL}(X, F_k + \eta_k\mathbf{I})$ for $k \geq 0$, where $\varepsilon_k \leq \varepsilon$ for any $k \geq 0$. Then,

(a) For any $k \geq 1$, $\|s_k - s_{k-1}\| \leq \frac{2nC}{\eta_{k-1}}\left(1 - \frac{\min\{\varepsilon_k,\varepsilon_{k-1}\}}{\max\{\varepsilon_k,\varepsilon_{k-1}\}}\right) + M\left|1 - \frac{\eta_k}{\eta_{k-1}}\right|$, where M and C are the norm bounds on the set X and the mapping F respectively (Remark 1).

(b) Suppose $\mathrm{SOL}(X,F) \neq \emptyset$ and let the sequences $\{\eta_k\}$ and $\{\varepsilon_k\}$ go to zero. Then $\lim_{k\to\infty} s_k = x^*$, where x^* is a solution of $\mathrm{VI}(X,F)$.

(c) For any $k \geq 0$, the mapping F_k is Lipschitz over the set X with the parameter $\kappa\frac{n!!}{(n-1)!!}\frac{C}{\varepsilon}$, where $\kappa = 1$ if n is odd and $\kappa = \frac{2}{\pi}$ otherwise.

Proof. (a) Suppose $k \geq 1$ is fixed. Since $s_k \in \mathrm{SOL}(X, F_k + \eta_k\mathbf{I})$ and $s_{k-1} \in \mathrm{SOL}(X, F_{k-1} + \eta_{k-1}\mathbf{I})$,

$$(s_{k-1} - s_k)^T(F_k(s_k) + \eta_k s_k) \geq 0 \text{ and } (s_k - s_{k-1})^T(F_{k-1}(s_{k-1}) + \eta_{k-1}s_{k-1}) \geq 0.$$

Adding the preceding relations, yields $(s_{k-1}-s_k)^T(F_k(s_k)-F_{k-1}(s_{k-1})+\eta_k s_k-\eta_{k-1}s_{k-1})\geq 0$. By adding and subtracting $F_{k-1}(s_k)+\eta_{k-1}s_k$, we obtain that

$$(s_{k-1}-s_k)^T(F_k(s_k)-F_{k-1}(s_k))+(s_{k-1}-s_k)^T(F_{k-1}(s_k)-F_{k-1}(s_{k-1}))$$
$$+(\eta_k-\eta_{k-1})(s_{k-1}-s_k)^T s_k-\eta_{k-1}\|s_k-s_{k-1}\|^2\geq 0.$$

By monotonicity of F_{k-1}, $\eta_{k-1}\|s_k-s_{k-1}\|^2\leq(s_{k-1}-s_k)^T(F_k(s_k)-F_{k-1}(s_k))+(\eta_k-\eta_{k-1})(s_{k-1}-s_k)^T s_k$. By the Cauchy-Schwartz inequality and the definition of M, we obtain

$$\eta_{k-1}\|s_k-s_{k-1}\|\leq\|F_k(s_k)-F_{k-1}(s_k)\|+M|\eta_{k-1}-\eta_k|. \tag{3}$$

Let p_u denote the probability density function of the random vector z and suppose it is given by $p_u(z)\triangleq\frac{1}{c_n\varepsilon^n}$ fro any $z\in B_n(0,\varepsilon)$, where $c_n\triangleq\frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)}$. In the following, we estimate the term $\|F_k(s_k)-F_{k-1}(s_k)\|$. First, let us consider the case $\varepsilon_k\leq\varepsilon_{k-1}$.

$$\begin{aligned}
&\|F_k(s_k)-F_{k-1}(s_k)\|=\left\|\int_{\mathbb{R}^n}F(s_k+z_k)p_u(z_k)dz_k-\int_{\mathbb{R}^n}F(s_k+z_{k-1})p_u(z_{k-1})dz_{k-1}\right\|\\
&=\left\|\int_{\|z\|<\varepsilon_k}F(s_k+z)\frac{1}{c_n\varepsilon_k^n}dz-\int_{\|z\|<\varepsilon_{k-1}}F(s_k+z)\frac{1}{c_n\varepsilon_{k-1}^n}dz\right\|\\
&=\left\|\int_{\|z\|<\varepsilon_k}F(s_k+z)\frac{1}{c_n\varepsilon_k^n}dz-\left(\int_{\|z\|<\varepsilon_k}F(s_k+z)\frac{1}{c_n\varepsilon_{k-1}^n}dz+\int_{\varepsilon_k\leq\|z\|<\varepsilon_{k-1}}F(s_k+z)\frac{1}{c_n\varepsilon_{k-1}^n}dz\right)\right\|\\
&\leq\left\|\int_{\|z\|<\varepsilon_k}F(s_k+z)\left(\frac{1}{c_n\varepsilon_k^n}-\frac{1}{c_n\varepsilon_{k-1}^n}\right)dz\right\|+\left\|\int_{\varepsilon_k\leq\|z\|<\varepsilon_{k-1}}F(s_k+z)\frac{1}{c_n\varepsilon_{k-1}^n}dz\right\|\\
&\leq\int_{\|z\|<\varepsilon_k}\|F(s_k+z)\|\left|\frac{1}{c_n\varepsilon_k^n}-\frac{1}{c_n\varepsilon_{k-1}^n}\right|dz+\int_{\varepsilon_k\leq\|z\|<\varepsilon_{k-1}}\|F(s_k+z)\|\frac{1}{c_n\varepsilon_{k-1}^n}dz,
\end{aligned}$$

where in the third equality we used $\{z\in\mathbb{R}^n\mid\|z\|<\varepsilon_{k-1}\}=\{z\in\mathbb{R}^n\mid\|z\|<\varepsilon_k\}\cup\{z\in\mathbb{R}^n\mid\varepsilon_k\leq\|z\|<\varepsilon_{k-1}\}$ when $\varepsilon_k\leq\varepsilon_{k-1}$, and in the last two inequalities we made use of the triangle inequality and the Jensen's inequality respectively. By the definition of C in Remark 1, we obtain

$$\begin{aligned}
\|F_k(s_k)-F_{k-1}(s_k)\|&\leq C\int_{\|z\|<\varepsilon_k}\left|\frac{1}{c_n\varepsilon_k^n}-\frac{1}{c_n\varepsilon_{k-1}^n}\right|dz+C\int_{\varepsilon_k\leq\|z\|<\varepsilon_{k-1}}\frac{1}{c_n\varepsilon_{k-1}^n}dz\\
&=C(c_n\varepsilon_k^n)\left(\frac{1}{c_n\varepsilon_k^n}-\frac{1}{c_n\varepsilon_{k-1}^n}\right)+C(c_n\varepsilon_{k-1}^n-c_n\varepsilon_k^n)\frac{1}{c_n\varepsilon_{k-1}^n}=2C\left(1-\left(\frac{\varepsilon_k}{\varepsilon_{k-1}}\right)^n\right).
\end{aligned}$$

Now, using relation (3), we obtain

$$\|s_k-s_{k-1}\|\leq\frac{2C}{\eta_{k-1}}\left(1-\left(\frac{\varepsilon_k}{\varepsilon_{k-1}}\right)^n\right)+M\left|1-\frac{\eta_k}{\eta_{k-1}}\right|. \tag{4}$$

Since we assumed that $\varepsilon_k\leq\varepsilon_{k-1}$, we may write

$$1-\left(\frac{\varepsilon_k}{\varepsilon_{k-1}}\right)^n=\left(1-\frac{\varepsilon_k}{\varepsilon_{k-1}}\right)\left(1+\left(\frac{\varepsilon_k}{\varepsilon_{k-1}}\right)+\ldots+\left(\frac{\varepsilon_k}{\varepsilon_{k-1}}\right)^{n-1}\right)\leq n\left(1-\frac{\varepsilon_k}{\varepsilon_{k-1}}\right). \tag{5}$$

Therefore when $\varepsilon_k\leq\varepsilon_{k-1}$, from (5) and (4), the desired inequality holds for all $k\geq 1$. Now, suppose $\varepsilon_k\geq\varepsilon_{k-1}$. Following the similar steps above, one can check that if $\varepsilon_k\geq\varepsilon_{k-1}$, then $\|F_k(s_k)-F_{k-1}(s_k)\|\leq 2C\left(1-(\frac{\varepsilon_{k-1}}{\varepsilon_k})^n\right)$. Therefore, $\|F_k(s_k)-F_{k-1}(s_k)\|\leq 2nC\left(1-\frac{\varepsilon_{k-1}}{\varepsilon_k}\right)$ implyingthe desired inequality.

(b) The proof is similar to the proof of Proposition 9 in (Yousefian, Nedić, and Shanbhag 2013).

(c) The proof can be done similar to the proof of Lemma 8 of (Yousefian, Nedić, and Shanbhag 2012). □

Next, we construct a recursive relation for the error between the iterate x_{k+1} and its counterpart s_k. Such a relation is a key for the convergence analysis of the proposed algorithm.

Lemma 4 [A recursive bound for the error] Consider algorithm (2) where $\{\gamma_k\}$, $\{\eta_k\}$, and $\{\varepsilon_k\}$ are strictly positive sequences. Let Assumptions 1, 2(b), and 2(c) hold and suppose there exists $K_2 \geq 0$ such that for any $k \geq K_2$, we have $\eta_k\gamma_k < 1$. Then, the following relation holds a.s. for any $k \geq \max\{K_1, K_2\}$:

$$\begin{aligned}\mathsf{E}\left[\|x_{k+1}-s_k\|^2 \mid \mathscr{F}_k\right] \leq \left(1-\frac{1}{2}\eta_k\gamma_k\right)\|x_k-s_{k-1}\|^2 + 2C^2\gamma_k^2 + 4M^2\eta_k^2\gamma_k^2 \\ + 16n^2C^2\left(1-\frac{\min\{\varepsilon_k,\varepsilon_{k-1}\}}{\max\{\varepsilon_k,\varepsilon_{k-1}\}}\right)^2\frac{1}{\eta_{k-1}^2\eta_k\gamma_k} + 4M^2\left(1-\frac{\eta_k}{\eta_{k-1}}\right)^2\frac{1}{\eta_k\gamma_k}.\end{aligned}$$

where K_1 is given by Assumptions 2(b).

Proof. Using the fixed point property of the projection operator at the solution $s_k \in \mathrm{SOL}(X, F_k + \eta_k\mathbf{I})$, we can write $s_k = \Pi_X(s_k - \gamma_k(F_k(s_k) + \eta_k s_k))$. Employing the nonexpansiveness property of the projection operator, the preceding relation, and algorithm (2), we obtain

$$\begin{aligned}&\|x_{k+1}-s_k\|^2 \leq \|x_k - \gamma_k(F(x_k+z_k)+\eta_k x_k + w_k) - s_k + \gamma_k(F_k(s_k)+\eta_k s_k)\|^2 \\ &= \|(1-\eta_k\gamma_k)(x_k-s_k) - \gamma_k(F(x_k+z_k)-F_k(s_k)) - \gamma_k w_k\|^2 \\ &= (1-\eta_k\gamma_k)^2\|x_k-s_k\|^2 + \gamma_k^2\|F(x_k+z_k)-F_k(s_k)\|^2 + \gamma_k^2\|w_k\|^2 \\ &- 2\gamma_k(1-\eta_k\gamma_k)(x_k-s_k)^T(F(x_k+z_k)-F_k(s_k)) - 2\Big((1-\eta_k\gamma_k)(x_k-s_k) - \gamma_k(F(x_k+z_k)-F_k(s_k))\Big)^T w_k.\end{aligned}$$

Adding and subtracting $F_k(x_k)$, we obtain

$$\begin{aligned}&\|x_{k+1}-s_k\|^2 \leq (1-\eta_k\gamma_k)^2\|x_k-s_k\|^2 + \gamma_k^2\|F(x_k+z_k)-F_k(x_k)\|^2 + \gamma_k^2\|F_k(x_k)-F_k(s_k)\|^2 \\ &+ 2\gamma_k^2(F(x_k+z_k)-F_k(x_k))^T(F_k(x_k)-F_k(s_k)) + \gamma_k^2\|w_k\|^2 - 2\gamma_k(1-\eta_k\gamma_k)(x_k-s_k)^T(F(x_k+z_k)-F_k(x_k)) \\ &- 2\gamma_k(1-\eta_k\gamma_k)(x_k-s_k)^T(F_k(x_k)-F_k(s_k)) - 2\Big((1-\eta_k\gamma_k)(x_k-s_k) - \gamma_k(F(x_k+z_k)-F_k(s_k))\Big)^T w_k.\end{aligned}$$

Taking the expectation in the preceding result conditioned on $\mathscr{F}_k$, using $\|F(x_k+z_k)\| \leq C$, and F_k is Lipschitz with constant $\kappa\frac{n!!}{(n-1)!!}\frac{C}{\varepsilon_k}$ (Proposition 1(c)), we obtain

$$\begin{aligned}&\mathsf{E}\left[\|x_{k+1}-s_k\|^2 \mid \mathscr{F}_k\right] \leq (1-\eta_k\gamma_k)^2\|x_k-s_k\|^2 + \gamma_k^2C^2 + \gamma_k^2\|F_k(x_k)\|^2 - 2\gamma_k^2\mathsf{E}[F(x_k+z_k)\mid\mathscr{F}_k]^T F_k(x_k) \\ &+ \gamma_k^2\left(\kappa\frac{n!!}{(n-1)!!}\frac{C}{\varepsilon_k}\right)^2\|x_k-s_k\|^2 + 2\gamma_k^2(\mathsf{E}[F(x_k+z_k)\mid\mathscr{F}_k] - F_k(x_k))^T(F_k(x_k)-F_k(s_k)) \\ &+ \gamma_k^2\mathsf{E}\left[\|w_k\|^2 \mid \mathscr{F}_k\right] - 2\gamma_k(1-\eta_k\gamma_k)(x_k-s_k)^T(\mathsf{E}[F(x_k+z_k\mid\mathscr{F}_k)] - F_k(x_k)) \\ &- 2\gamma_k(1-\eta_k\gamma_k)(x_k-s_k)^T(F_k(x_k)-F_k(s_k)) - 2\mathsf{E}\left[\Big((1-\eta_k\gamma_k)(x_k-s_k) - \gamma_k(F(x_k+z_k)-F_k(s_k))\Big)^T w_k \mid \mathscr{F}_k\right].\end{aligned}$$

Note that $\mathsf{E}_\xi[w_k \mid \mathscr{F}_k] = \mathsf{E}[w_k \mid \mathscr{F}_k] = 0$ by Lemma 2 implying that the last term is zero. Therefore, Term $1 = 0$. Using the preceding result and $\mathsf{E}[F(x_k+z_k)\mid\mathscr{F}_k] = F_k(x_k)$, we obtain that

$$\begin{aligned}\mathsf{E}\left[\|x_{k+1}-s_k\|^2 \mid \mathscr{F}_k\right] \leq (1-\eta_k\gamma_k)^2\|x_k-s_k\|^2 + \gamma_k^2C^2 - \gamma_k^2\|F_k(x_k)\|^2 + \gamma_k^2\left(\kappa\frac{n!!}{(n-1)!!}\frac{C}{\varepsilon_k}\right)^2\|x_k-s_k\|^2 \\ + \gamma_k^2\mathsf{E}\left[\|w_k\|^2 \mid \mathscr{F}_k\right] - 2\gamma_k(1-\eta_k\gamma_k)(x_k-s_k)^T(F_k(x_k)-F_k(s_k)).\end{aligned}$$

Since $\eta_k\gamma_k < 1$ for any $k \geq K_2$, the term $1-\eta_k\gamma_k$ is positive. On the other hand, monotonicity of F_k implies that the term $(x_k - s_k)^T(F_k(x_k) - F_k(s_k))$ is nonnegative. Therefore, $2\gamma_k(1-\eta_k\gamma_k)(x_k-s_k)^T(F_k(x_k)-F_k(s_k)) \geq 0$ for any $k \geq K_2$. Using this, the preceding relation, and $\gamma_k^2\|F_k(x_k)\|^2 \geq 0$, for any $k \geq K_2$,

$$\mathsf{E}\left[\|x_{k+1}-s_k\|^2 \mid \mathscr{F}_k\right] \leq \left(1 - 2\eta_k\gamma_k + \eta_k^2\gamma_k^2 + \gamma_k^2\left(\kappa\frac{n!!}{(n-1)!!}\frac{C}{\varepsilon_k}\right)^2\right)\|x_k - s_k\|^2 + \gamma_k^2\left(\mathsf{E}\left[\|w_k\|^2 \mid \mathscr{F}_k\right] + C^2\right).$$

Using the definition of M in Remark 1 and the triangle inequality, we can write $\|y-z\| \leq \|y\| + \|z\| \leq 2M$. Taking this to account and using $\mathsf{E}\left[\|w_k\|^2 \mid \mathscr{F}_k\right] \leq C^2$ from Lemma 2, the preceding inequality yields

$$\mathsf{E}\left[\|x_{k+1}-s_k\|^2 \mid \mathscr{F}_k\right] \leq \left(1 - 2\eta_k\gamma_k + \gamma_k^2\left(\kappa\frac{n!!}{(n-1)!!}\frac{C}{\varepsilon_k}\right)^2\right)\|x_k - s_k\|^2 + 2C^2\gamma_k^2 + 4M^2\eta_k^2\gamma_k^2. \tag{6}$$

Note that the above inequality is not yet a recursive relation. To obtain a recursive relation, we need to estimate the term $\|x_k - s_k\|$ in terms of $\|x_k - s_{k-1}\|$. Using the triangle inequality, we can write $\|x_k - s_k\| \leq \|x_k - s_{k-1}\| + \|s_k - s_{k-1}\|$.Therefore, we obtain

$$\|x_k - s_k\|^2 \leq \|x_k - s_{k-1}\|^2 + \|s_k - s_{k-1}\|^2 + 2\|s_k - s_{k-1}\|\|x_k - s_{k-1}\|. \tag{7}$$

Using the relation $2ab \leq a^2 + b^2$, for $a, b \in R$, we obtain that

$$2\|s_k - s_{k-1}\|\|x_k - s_{k-1}\| = 2\left(\sqrt{\eta_k\gamma_k}\|x_k - s_{k-1}\|\right)\left(\frac{\|s_k - s_{k-1}\|}{\sqrt{\eta_k\gamma_k}}\right) \leq \eta_k\gamma_k\|x_k - s_{k-1}\|^2 + \frac{\|s_k - s_{k-1}\|^2}{\eta_k\gamma_k}.$$

Combining this result, Proposition 1(a), and (7), we obtain

$$\begin{aligned}\|x_k - s_k\|^2 &\leq (1+\eta_k\gamma_k)\|x_k - s_{k-1}\|^2 + \left(\frac{2nC}{\eta_{k-1}}\left(1 - \frac{\min\{\varepsilon_k, \varepsilon_{k-1}\}}{\max\{\varepsilon_k, \varepsilon_{k-1}\}}\right) + M\left|1 - \frac{\eta_k}{\eta_{k-1}}\right|\right)^2\left(1 + \frac{1}{\eta_k\gamma_k}\right) \\ &\leq (1+\eta_k\gamma_k)\|x_k - s_{k-1}\|^2 + 2\left(\frac{2nC}{\eta_{k-1}}\left(1 - \frac{\min\{\varepsilon_k, \varepsilon_{k-1}\}}{\max\{\varepsilon_k, \varepsilon_{k-1}\}}\right) + M\left|1 - \frac{\eta_k}{\eta_{k-1}}\right|\right)^2\frac{1}{\eta_k\gamma_k},\end{aligned} \tag{8}$$

where in the last inequality we used $1 + \frac{1}{\eta_k\gamma_k} < \frac{2}{\eta_k\gamma_k}$ as a consequence of $\gamma_k\eta_k < 1$. Let us define $q_k \triangleq 1 - 2\eta_k\gamma_k + \gamma_k^2\left(\kappa\frac{n!!}{(n-1)!!}\frac{C}{\varepsilon_k}\right)^2$. Now, inequalities (6) and (8) imply that for $k \geq K_2$

$$\begin{aligned}\mathsf{E}\left[\|x_{k+1}-s_k\|^2 \mid \mathscr{F}_k\right] &\leq q_k(1+\eta_k\gamma_k)\|x_k - s_{k-1}\|^2 + 2C^2\gamma_k^2 + 4M^2\eta_k^2\gamma_k^2 \\ &\quad + 2q_k\left(\underbrace{\frac{2nC}{\eta_{k-1}}\left(1 - \frac{\min\{\varepsilon_k, \varepsilon_{k-1}\}}{\max\{\varepsilon_k, \varepsilon_{k-1}\}}\right)}_{a} + \underbrace{M\left|1 - \frac{\eta_k}{\eta_{k-1}}\right|}_{b}\right)^2\frac{1}{\eta_k\gamma_k}.\end{aligned} \tag{9}$$

By Assumption 2(b), we can write for $k \geq K_1$,

$$\frac{\gamma_k}{\eta_k\varepsilon_k^2} \leq 0.5\left(\frac{(n-1)!!}{n!!\kappa C}\right)^2 \Rightarrow \gamma_k^2\left(\kappa\frac{n!!}{(n-1)!!}\frac{C}{\varepsilon_k}\right)^2 \leq \frac{\eta_k\gamma_k}{2} \Rightarrow -2\eta_k\gamma_k + \gamma_k^2\left(\kappa\frac{n!!}{(n-1)!!}\frac{C}{\varepsilon_k}\right)^2 \leq -\frac{3}{2}\eta_k\gamma_k.$$

Therefore, $q_k \leq 1 - \frac{3}{2}\eta_k\gamma_k$. This implies that $q_k(1+\eta_k\gamma_k) \leq (1 - \frac{3}{2}\eta\gamma_k)(1+\eta_k\gamma_k) = 1 - \frac{1}{2}\eta_k\gamma_k - \frac{3}{2}\eta_k^2\gamma_k^2 \leq 1 - \frac{1}{2}\eta_k\gamma_k$. Using relation (9) and $q_k \leq 1$ (which follows by $q_k \leq 1 - \frac{3}{2}\eta_k\gamma_k$), and the fact that for real numbers a and b, $(a+b)^2 \leq 2a^2 + 2b^2$, we conclude that the desired relation holds. □

We are now ready to present the almost-sure convergence result.

Proposition 2 [Almost-sure convergence] Let Assumptions 1 and 2 hold. Suppose $\{x_k\}$ is given by algorithm (2). Then $\{x_k\}$ converges to a solution of $\text{VI}(X,F)$ almost surely.

Proof. From Assumption 2(a), γ_k and η_k go to zero. Thus, there exists a constant $K_2 \geq 0$ such that $\gamma_k \eta_k < 1$ for any $k \geq K_2$. Let use define define sequences $\{v_k\}$, $\{\alpha_k\}$, and $\{\mu_k\}$ for $k \geq \max\{K_1, K_2\}$ given by $v_k \triangleq \|x_k - s_{k-1}\|$, $\quad \alpha_k \triangleq \frac{1}{2}\gamma_k\eta_k$ and

$$\mu_k \triangleq 2C^2\gamma_k^2 + 4M^2\eta_k^2\gamma_k^2 + 16n^2C^2 \overbrace{\left(1 - \frac{\min\{\varepsilon_k, \varepsilon_{k-1}\}}{\max\{\varepsilon_k, \varepsilon_{k-1}\}}\right)^2 \frac{1}{\eta_{k-1}^2\eta_k\gamma_k}}^{\text{Term 1}} + 4M^2 \overbrace{\left(1 - \frac{\eta_k}{\eta_{k-1}}\right)^2 \frac{1}{\eta_k\gamma_k}}^{\text{Term 2}}.$$

Therefore, Lemma 4 implies that $\mathsf{E}[v_{k+1} \mid \mathscr{F}_k] \leq (1-\alpha_k)v_k + \mu_k \quad$ for $k \geq \max\{K_1, K_2\}$. To claim convergence of the sequence $\{x_k\}$, we show that conditions of Lemma 1 hold. The nonnegativity of v_k, α_k, and μ_k for $k \geq \max\{K_1, K_2\}$ is trivial. Assumption 2(d) indicates that the condition $\sum_k \alpha_k = \infty$ is satisfied. On the other hand, positivity of γ_k and η_k indicates that $\alpha_k \leq 1$ holds for $k \geq \max\{K_1, K_2\}$. Since η_k goes to zero, there exists a bound $\bar{\eta}$ such that $\eta_k \leq \bar{\eta}$. Therefore, $\mu_k \leq (2C^2 + 4M^2\bar{\eta}^2)\gamma_k^2 + 16n^2C^2(\text{Term 1}) + 4M^2(\text{Term 2})$. Assumptions 2(e), (f), and (g) show that γ_k^2, Terms 1 and 2 are summable. Therefore, we conclude that μ_k is summable too. It remains to show that $\lim_{k\to\infty} \frac{\mu_k}{\alpha_k} = 0$. It suffices to show that $\lim_{k\to\infty} \frac{\gamma_k^2}{\alpha_k} = 0$, $\lim_{k\to\infty} \frac{\text{Term 1}}{\alpha_k} = 0$, and $\lim_{k\to\infty} \frac{\text{Term 2}}{\alpha_k} = 0$. These three conditions hold due to Assumptions 2(h), (i), and (j) respectively. In conclusion, all of the conditions of Lemma 1 hold and thus $\|x_k - s_k\|$ goes to zero almost surely. Moreover, since η_k and ε_k go to zero, Proposition 1(b) implies that the sequence $\{s_k\}$ converges to a solution of $\text{VI}(X,F)$. Hence, we conclude that the sequence $\{x_k\}$ generated by algorithm (2) converges to a solution of $\text{VI}(X,F)$ almost surely. □

This section is ended by providing a class of the stepsize, regularization, and smoothing sequences that guarantees almost-sure convergence.

Lemma 5 Suppose sequences $\{\gamma_k\}$, $\{\eta_k\}$, and $\{\varepsilon_k\}$ are given by $\gamma_k = \gamma_0(k+1)^{-a}$, $\eta_k = \eta_0(k+1)^{-b}$, and $\varepsilon_k = \varepsilon_0(k+1)^{-c}$ where $a,b,c > 0$, $\quad a+3b < 1$, $\quad a > b+2c$, $\quad a > 0.5$, and γ_0, η_0, ε_0 are strictly positive scalars and $\varepsilon_0 = \varepsilon$. Then, sequences $\{\gamma_k\}$, $\{\eta_k\}$, and $\{\varepsilon_k\}$ satisfy Assumption 2.

Proof. We show that each part of Assumption 2 holds as follows:

(a) Assumption 2(a) holds since a, b, c, γ_0, η_0, and ε_0 are strictly positive.

(b) To show that part (b) holds, we write $\frac{\gamma_k}{\eta_k\varepsilon_k^2} = \frac{\gamma_0(k+1)^{-a}}{\eta_0(k+1)^{-b}\varepsilon_0^2(k+1)^{-2c}} = (k+1)^{-(a-b-2c)}\frac{\gamma_0}{\eta_0\varepsilon_0^2}$. Since $a > b+2c$, then $(k+1)^{-(a-b-2c)} \to 0$. Therefore, $\frac{\gamma_k}{\eta_k\varepsilon_k^2} \to 0$ implying that there exists $K_1 \geq 0$ such that $\frac{\gamma_k}{\eta_k\varepsilon_k^2} \leq 0.5\left(\frac{(n-1)!!}{n!!\kappa C}\right)^2$ for any $k \geq K_1$. This indicates that Assumption 2(b) holds.

(c) Part (c) holds because $\varepsilon_k \leq \varepsilon_0$ for any $k \geq 0$ and $\varepsilon_0 = \varepsilon$.

(d) Let us now check part (d) to see if it holds. We have $\sum_{k=0}^{\infty} \eta_k\gamma_k = \eta_0\gamma_0\sum_{k=0}^{\infty}\frac{1}{(k+1)^{a+b}}$. Since $a,b > 0$ and $a+3b < 1$, then $a+b < 1$. Thus, $\sum_{k=0}^{\infty}\frac{1}{(k+1)^{a+b}} = \infty$. Therefore, Assumption 2(d) is met.

(e) To show that part (e) holds we need to show that γ_k^2 is summable. We have $\gamma_k^2 = \gamma_0^2(k+1)^{-2a}$ and $2a > 1$ since $a > 0.5$. Therefore, γ_k^2 is summable which means that condition (e) is satisfied.

(f) Note that sequences $\{\eta_k\}$ and $\{\varepsilon_k\}$ are both decreasing. Therefore,

$$\frac{1}{\eta_{k-1}^2\eta_k\gamma_k}\left(1 - \frac{\min\{\varepsilon_k, \varepsilon_{k-1}\}}{\max\{\varepsilon_k, \varepsilon_{k-1}\}}\right)^2 = \frac{1}{\eta_{k-1}^2\eta_k\gamma_k}\left(1 - \frac{\varepsilon_k}{\varepsilon_{k-1}}\right)^2 < \frac{1}{\eta_k^3\gamma_k}\left(1 - \frac{\varepsilon_k}{\varepsilon_{k-1}}\right)^2 \triangleq \text{Term 1}.$$

It suffices to show that Term 1 is summable. First, we estimate $1-\frac{\varepsilon_k}{\varepsilon_{k-1}}$. We have $1-\frac{\varepsilon_k}{\varepsilon_{k-1}}=1-\frac{\varepsilon_0(k+1)^{-c}}{\varepsilon_0 k^{-c}}=1-\left(\frac{k}{k+1}\right)^c=1-\left(1-\frac{1}{k+1}\right)^c$. Recall that the Taylor expansion of $(1-x)^p$ for $|x|<1$ and any scalar p is given by $(1-x)^p=\sum_{j=0}^{\infty}(-1)^j\left(\begin{array}{c}p\\ j\end{array}\right)x^j=1-px+\frac{p(p-1)}{2}x^2-\frac{p(p-1)(p-2)}{6}x^3+\cdots$. Using this expansion for $x=\frac{1}{k+1}$ and $p=c$, we have $1-\frac{\varepsilon_k}{\varepsilon_{k-1}}=1-\left(1-c\frac{1}{k+1}+\frac{c(c-1)}{2}\frac{1}{(k+1)^2}+\cdots\right)=O(k^{-1})$. Therefore, from the preceding relation, we obtain Term $1=\frac{O(k^{-2})}{\eta_0^3\gamma_0(k+1)^{-3b-a}}=O(k^{-(2-a-3b)})$. To have Term 1 summable, we need to have $2-a-3b>1$ or equivalently $a+3b<1$. This holds by our assumptions.

(g) In a similar fashion that we used in part (f), we can show that $1-\frac{\eta_k}{\eta_{k-1}}=O(k^{-1})$. Therefore, Term $3\triangleq\frac{1}{\eta_k\gamma_k}\left(1-\frac{\eta_k}{\eta_{k-1}}\right)^2\frac{O(k^{-2})}{\eta_0\gamma_0(k+1)^{-(a+b)}}=O(k^{-(2-a-b)})$. To show that condition (g) is satisfied, we need to show that Term 3 is summable. From the preceding relation, we need to show that $2-a-b>1$ or equivalently $a+b<1$. We assumed that $a+3b<1$ and $b>0$. Thus, we have $a+b=a+3b-2b<1-2b<1$. Therefore, $O(k^{-(2-a-b)})$ is summable and we conclude that condition (g) is met.

(h) We have $\frac{\gamma_k}{\eta_k}=\frac{\gamma_0(k+1)^{-a}}{\eta_0(k+1)^{-b}}=\frac{\gamma_0}{\eta_0}(k+1)^{-(a-b)}$. To show that $\frac{\gamma_k}{\eta_k}$ goes to zero when k goes to infinity, we only need to show that $a>b$. We assumed that $a+3b<1$. Therefore, $b<\frac{1}{3}(1-a)$. Since $a>0.5$, the preceding relation yields $b<\frac{1}{3}0.5$. Thus, $b<0.5<a$, implying that condition (h) holds.

(i) From the discussion in part (f), we have $1-\frac{\varepsilon_k}{\varepsilon_{k-1}}=O(k^{-1})$. To show the condition (i), we write

$$\text{Term } 4\triangleq\frac{1}{\eta_k^2\gamma_k}\left(1-\frac{\min\{\varepsilon_k,\varepsilon_{k-1}\}}{\max\{\varepsilon_k,\varepsilon_{k-1}\}}\right)=\frac{1}{\eta_0^2\gamma_0(k+1)^{-a-2b}}O(k^{-1})=O(k^{-(1-a-2b)}).$$

Thus, it suffices to show that $a+2b<1$. This is true since $a+3b<1$ and $b>0$. Hence, Term 4 goes to zero implying that part (i) holds.

(j) We have Term $5\triangleq\frac{1}{\eta_k\gamma_k}\left|1-\frac{\eta_k}{\eta_{k-1}}\right|=\frac{1}{\eta_0\gamma_0(k+1)^{-a-b}}O(k^{-1})=O(k^{-(1-a-b)})$. Since $a+3b<1$ and $b>0$, we have $a+b<1$, showing that Term 5 converges to zero. □

4 A BOUND FOR THE ERROR OF THE APPROXIMATE PROBLEM

In the second part of this paper, we focus on the rate analysis of algorithm (2). We begin the discussion by a family of assumptions on the sequences. This set of assumptions are essential to derive a particular rate.

Assumption 3 Let the following hold:

(a) There exist $0<\delta<0.5$ and $K_3\geq 0$ such that $\frac{\gamma_k}{\eta_k\varepsilon_k^2}\leq\frac{\gamma_{k+1}}{\eta_{k+1}\varepsilon_{k+1}^2}(1+\delta\eta_{k+1}\gamma_{k+1})$ for any $k\geq K_3$;

(b) There exists a constant $B_1>0$ such that $\frac{\varepsilon_k^2}{\eta_{k-1}^2\eta_k\gamma_k^3}\left(1-\frac{\min\{\varepsilon_k,\varepsilon_{k-1}\}}{\max\{\varepsilon_k,\varepsilon_{k-1}\}}\right)^2\leq B_1$ for any $k\geq 0$;

(c) There exists a constant $B_2>0$ such that $\frac{\varepsilon_k^2}{\eta_k\gamma_k^3}\left(1-\frac{\eta_k}{\eta_{k-1}}\right)^2\leq B_2$ for any $k\geq 0$.

Remark 5 Similar to the result of Lemma 5, one can provide a feasible choice of the sequences that satisfy Assumption 3. We omitted this result due to space limitations.

The following result, provides a bound on the error that relates the iterates $\{x_k\}$ and the approximate sequence $\{s_k\}$. This result provides us an estimate of the performance of our algorithm with respect to the iterates of the solutions to the approximated problems $\text{VI}(X,F_k+\eta_k\mathbf{I})$.

Proposition 3 [An upper bound for $\mathsf{E}\left[\|x_{k+1}-s_k\|^2\right]$] Consider algorithm (2) where $\{\gamma_k\}$, $\{\eta_k\}$, and $\{\varepsilon_k\}$ are strictly positive sequences. Let Assumptions 1, 2(b), 2(c), and 3 hold. Suppose $\{\eta_k\}$ is bounded by $\bar{\eta}$ and there exists some scalar $K_2\geq 0$ such that for any $k\geq K_2$ we have $\eta_k\gamma_k<1$. Then,

$$\mathsf{E}\left[\|x_{k+1}-s_k\|^2\right]\leq\theta\frac{\gamma_k}{\eta_k\varepsilon_k^2},\qquad\text{for any } k\geq\bar{K},\tag{10}$$

where $\bar{K} \triangleq \max\{K_1, K_2, K_3\}$, s_k is the unique solution of $\text{VI}(X, F_k + \eta_k \mathbf{I})$, K_1 and K_3 are given by Assumptions 2(b), and 3(d) respectively. More precisely, relation (10) holds if

$$\theta = \max\left\{4M^2 \frac{\eta_{\bar{K}} \varepsilon_{\bar{K}}^2}{\gamma_{\bar{K}}}, \frac{2C^2\varepsilon^2 + 4M^2\bar{\eta}^2\varepsilon^2 + 16n^2C^2B_1 + 4M^2B_2}{0.5-\delta}\right\}. \tag{11}$$

Proof. We begin the proof by employing Lemma 4. Let us define $e_k \triangleq \mathsf{E}\left[\|x_k - s_{k-1}\|^2\right]$ for $k \geq \bar{K}+1$.Taking expectation in the relation of Lemma 4, we obtain a recursive inequality in terms of the mean squared error between x_{k+1} and s_k. For any $k \geq \bar{K}+1$ we have

$$e_{k+1} \leq \left(1 - \frac{1}{2}\eta_k\gamma_k\right) e_k + 2C^2\gamma_k^2 + 4M^2\eta_k^2\gamma_k^2 + 16n^2C^2 \frac{\left(1 - \frac{\min\{\varepsilon_k, \varepsilon_{k-1}\}}{\max\{\varepsilon_k, \varepsilon_{k-1}\}}\right)^2}{\eta_{k-1}^2\eta_k\gamma_k} + 4M^2 \frac{\left(1 - \frac{\eta_k}{\eta_{k-1}}\right)^2}{\eta_k\gamma_k}. \tag{12}$$

To show the main result, we use induction on k. The first step is to show that the result holds for $k = \bar{K}$. Using the definition of M in Remark 1 and the Cauchy-Schwartz inequality, we can write

$$\begin{aligned} e_{\bar{K}+1} &= \mathsf{E}\left[\|x_{\bar{K}+1} - s_{\bar{K}}\|^2\right] = \mathsf{E}\left[\|x_{\bar{K}+1}\|^2 - 2x_{\bar{K}+1}^T s_{\bar{K}} + \|s_{\bar{K}}\|^2\right] \leq \mathsf{E}\left[\|x_{\bar{K}+1}\|^2 + 2\|x_{\bar{K}+1}\|\|s_{\bar{K}}\| + \|s_{\bar{K}}\|^2\right] \\ &\leq M^2 + 2M^2 + M^2\left(4M^2 \frac{\eta_{\bar{K}}\varepsilon_{\bar{K}}^2}{\gamma_{\bar{K}}}\right)\frac{\gamma_{\bar{K}}}{\eta_{\bar{K}}\varepsilon_{\bar{K}}^2}. \end{aligned}$$

Let us define $\theta_{\bar{K}} \triangleq 4M^2 \frac{\eta_{\bar{K}}\varepsilon_{\bar{K}}^2}{\gamma_{\bar{K}}}$. Thus, the preceding relation implies that the main result holds for $k = \bar{K}$ with $\theta = \theta_{\bar{K}}$. Now, suppose $e_{t+1} \leq \theta \frac{\eta_t\gamma_k}{\varepsilon_t^2}$ for $\bar{K} < t \leq k-1$ for some finite constant $\theta > 0$. We will show that $e_{k+1} \leq \theta \frac{\gamma_k}{\eta_k\varepsilon_k^2}$. Using the induction hypothesis, (12), and Assumptions 3(e) and (f) we obtain

$$e_{k+1} \leq \left(1 - \frac{1}{2}\eta_k\gamma_k\right)\theta \frac{\gamma_{k-1}}{\eta_{k-1}\varepsilon_{k-1}^2} + 2C^2\gamma_k^2 + 4M^2\eta_k^2\gamma_k^2 + 16n^2C^2 \frac{\gamma_k^2}{\varepsilon_k^2}B_1 + 4M^2 \frac{\gamma_k^2}{\varepsilon_k^2}B_2.$$

Using the Assumption 3(d) we obtain

$$e_{k+1} \leq \left(1 - \frac{1}{2}\eta_k\gamma_k\right)(1 + \delta\eta_k\gamma_k)\theta \frac{\gamma_k}{\eta_k\varepsilon_k^2} + 2C^2\gamma_k^2 + 4M^2\eta_k^2\gamma_k^2 + 16n^2C^2 \frac{\gamma_k^2}{\varepsilon_k^2}B_1 + 4M^2 \frac{\gamma_k^2}{\varepsilon_k^2}B_2. \tag{13}$$

Note that we have

$$\left(1 - \frac{1}{2}\eta_k\gamma_k\right)(1 + \delta\eta_k\gamma_k)\theta \frac{\gamma_k}{\eta_k\varepsilon_k^2} = \theta \frac{\gamma_k}{\eta_k\varepsilon_k^2} - \theta\left(\frac{\delta}{2}\right)\frac{\eta_k\gamma_k^3}{\varepsilon_k^2} + \theta\eta_k\gamma_k\left(-\frac{1}{2} + \delta\right)\frac{\gamma_k}{\eta_k\varepsilon_k^2} + 2C^2\gamma_k^2. \tag{14}$$

Using nonpositivity of $-\theta\left(\frac{\delta}{2}\right)\frac{\eta_k\gamma_k^3}{\varepsilon_k^2}$, (13), (14) and by taking out the factor $\frac{\gamma_k^2}{\varepsilon_k^2}$, it follows that

$$e_{k+1} \leq \theta \frac{\gamma_k}{\eta_k\varepsilon_k^2} + \frac{\gamma_k^2}{\varepsilon_k^2}\overbrace{\left[-\theta\left(\frac{1}{2} - \delta\right) + 2C^2\varepsilon^2 + 4M^2\bar{\eta}^2\varepsilon^2 + 16n^2C^2B_1 + 4M^2B_2\right]}^{\text{Term 1}}. \tag{15}$$

If we show that the multiplier of the term $\frac{\gamma_k^2}{\varepsilon_k^2}$ in the brackets is nonpositive for some $\theta > 0$, we obtain the desired result. Note that $\{\eta_k\}$ is bounded by $\bar{\eta}$ and Assumption 3(c) implies that $\varepsilon_k \leq \varepsilon$. By Assumption 3(d), we have $\left(\frac{1}{2} - \delta\right) > 0$. Therefore, if $\theta \geq \frac{2C^2\varepsilon^2 + 4M^2\bar{\eta}^2\varepsilon^2 + 16n^2C^2B_1 + 4M^2B_2}{0.5-\delta}$, then Term 1 is nonpositive. This implies that $e_{k+1} \leq \theta \frac{\gamma_k}{\eta_k\varepsilon_k^2}$ and therefore the induction argument is done. In conclusion, if θ satisfies relation (11), then relation (10) holds for any $k \geq \bar{K}$. □

Remark 6 Proposition 3 provides an upper bound for the MSE between iterates of the algorithm (1) and solutions of the approximate problems. However, in order to obtain the rate of convergence of algorithm (1), we need an estimate of the error $\mathsf{E}\left[\|s_k - x^*\|^2\right]$. This question is not addressed in this paper and it is a future direction to our work.

5 CONCLUDING REMARKS

We consider a stochastic variational inequality problem with monotone and possibly non-Lipschitzian maps over a closed, convex, and compact set. Such problems may arise from stochastic nonsmooth convex optimization problems as well as from stochastic nonsmooth Nash games. A regularized smoothing stochastic approximation (SA) scheme is presented wherein the map is simultaneously regularized and smoothed. A Tikhonov-based regularization ensures that the map is strongly monotone at every step with a constant given by the regularization constant. Similarly, a convolution-based smoothing allows for claiming that the map is Lipschitz continuous with a prescribed constant. In the resulting SA scheme, the steplength, regularization parameter, and the smoothing parameter are all diminishing. By suitable choices of such sequences, almost sure convergence of the scheme can be recovered. Additionally, an error bound is provided that relates the error in the generated iterates and a suitably defined approximate solution.

REFERENCES

Bertsekas, D. P. 1973. "Stochastic Optimization Problems with Nondifferentiable Cost Functionals". *Journal of Optimization Theory and Applications* 12 (2): 218–231.

Cicek, D., M. Broadie, and A. Zeevi. 2011. "General Bounds and Finite-time Performance Improvement for the Kiefer-Wolfowitz Stochastic Approximation Algorithm". *Operations Research* 59:1211–1224.

Duchi, J. C., P. L. Bartlett, and M. J. Wainwright. 2012. "Randomized Smoothing for Stochastic Optimization". *SIAM Journal on Optimization (SIOPT)* 22 (2): 674–701.

Ermoliev, Y. M. 1983. "Stochastic Quasigradient Methods and Their Application to System Optimization". *Stochastics* 9:1–36.

Facchinei, F., H. Jiang, and L. Qi. 1999. "A Smoothing Method for Mathematical Programs with Equilibrium Constraints (English Summary)". *SIAM Journal on Optimization* 85 (1): 107–134.

Facchinei, F., and J.-S. Pang. 2003. *Finite-dimensional Variational Inequalities and Complementarity Problems. Vols. I,II*. Springer Series in Operations Research. New York: Springer-Verlag.

Jiang, H., and H. Xu. 2008. "Stochastic Approximation Approaches to the Stochastic Variational Inequality Problem". *IEEE Transactions in Automatic Control* 53 (6): 1462–1475.

Juditsky, A., A. Nemirovski, and C. Tauvel. 2011. "Solving Variational Inequalities with Stochastic Mirror-prox Algorithm". *Stochastic Systems* 1 (1): 17–58.

Koshal, J., A. Nedić, and U. V. Shanbhag. 2010,. "Single Timescale Regularized Stochastic Approximation Schemes for Monotone Nash games under Uncertainty". In *Proceedings of the IEEE Conference on Decision and Control (CDC)*, 231–236.

Kushner, H. J., and G. G. Yin. 2003. *Stochastic Approximation and Recursive Algorithms and Applications*. New York: Springer.

Lakshmanan, H., and D. Farias. 2008. "Decentralized Recourse Allocation In Dynamic Networks of Agents". *SIAM Journal on Optimization* 19 (2): 911–940.

Nesterov, Y. 2005. "Smooth Minimization of Non-smooth Functions". *Math. Prog.* 103:127–152.

Norkin, V. I. 1993. "The Analysis and Optimization of Probability Functions". Technical report, International Institute for Applied Systems Analysis technical report. WP-93-6.

Polyak, B. 1987. *Introduction to Optimization*. New York: Optimization Software, Inc.

Robbins, H., and S. Monro. 1951. "A stochastic Approximation Method". *Ann. Math. Statistics* 22:400–407.

Rockafellar, R., and R.-B. Wets. 1998. *Variational Analysis*. Berlin: Springer.

Shapiro, A. 2003. "Monte Carlo Sampling Methods". In *Handbook in Operations Research and Management Science*, Volume 10, 353–426. Amsterdam: Elsevier Science.

Steklov, V. A. 1907. "Sur les expressions asymptotiques decertaines fonctions dfinies par les quations diffrentielles du second ordre et leers applications au problme du dvelopement d'une fonction arbitraire en sries procdant suivant les diverses fonctions". *Comm. Charkov Math. Soc.* 2 (10): 97–199.

Xu, H. 2010. "Sample Average Approximation Methods for a Class of Stochastic Variational Inequality problems". *Asia-Pacific Journal of Operational Research* 27 (1): 103–119.

Yousefian, F., A. Nedić, and U. V. Shanbhag. 2012. "On Stochastic Gradient and Subgradient Methods with adaptive steplength sequences". *Automatica* 48 (1): 56–67. An extended version of the paper available at: http://arxiv.org/abs/1105.4549.

Yousefian, F., A. Nedić, and U. V. Shanbhag. 2013. "Distributed Adaptive Steplength Stochastic Approximation Schemes for Cartesian Stochastic Variational Inequality Problems". Submitted, http://arxiv.org/abs/1301.1711.

AUTHOR BIOGRAPHIES

Farzad Yousefian is a Ph.D. candidate at the Department of Industrial and Enterprise Systems Engineering, University of Illinois at Urbana-Champaign. His general interest is in optimization including models, algorithms, and applications. His current research interests lie in the development of algorithms in the regime of variational inequalities and optimization in uncertain and nonsmooth settings. His email and web addresses are yousefi1@illinois.edu and https://sites.google.com/site/farzad1yousefian, respectively.

Angelia Nedić received her B.S. degree from the University of Montenegro (1987) and M.S. degree from the University of Belgrade (1990), both in Mathematics. She received her Ph.D. degrees from Moscow State University (1994) in Mathematics and Mathematical Physics, and from Massachusetts Institute of Technology in Electrical Engineering and Computer Science (2002). She has been at the BAE Systems Advanced Information Technology from 2002-2006. In Fall 2006, as Assistant Professor, she has joined the Department of Industrial and Enterprise Systems Engineering at the University of Illinois at Urbana-Champaign, USA. She has recently received the Donald Biggar Willett Scholar of Engineering award from the College of Engineering at the University of Illinois at Urbana-Champaign. Her general interest is in optimization and dynamics including fundamental theory, models, algorithms, and applications. Her current research interest is focused on large-scale convex optimization, distributed multi-agent optimization and equilibrium problems, stochastic approximations, and network aggregation-dynamics with applications in signal processing, machine learning, and decentralized control. Her email address is angelia@illinois.edu and her web page is http://www.ifp.illinois.edu/~angelia/.

Uday V. Shanbhag is an Associate Professor in the Harold and Inge Marcus Department of Industrial and Manufacturing Engineering at Penn. State University. He received his Ph.D. degree in operations research from the Department of Management Science and Engineering, Stanford University, Stanford, CA, in 2006. His interests lies in the development of analytical and algorithmic tools in the context of optimization and variational problems, in regimes complicated by uncertainty, dynamics and nonsmoothness. Dr. Shanbhag received the triennial A.W. Tucker Prize for his dissertation from the mathematical programming society (MPS) in 2006, the Computational Optimization and Applications (COAP) Best Paper Award in 2007 (With Walter Murray), and the NSF Career Award (Operations Research) in 2012 . His email address is udaybag@psu.edu and his web page is http://www2.ie.psu.edu/shanbhag/personal/index.htm.