Last Iterate Convergence of Popov Method for Non-monotone Stochastic Variational Inequalities

Daniil Vankov DVANKOV@ASU.EDU
Angelia Nedić ANGELIA.NEDICH@ASU.EDU
Lalitha Sankar LSANKAR@ASU.EDU
Arizona State University

Abstract

This paper focuses on non-monotone stochastic variational inequalities (SVIs) that may not have a unique solution. A commonly used efficient algorithm to solve VIs is the Popov method, which is known to have the optimal convergence rate for VIs with Lipschitz continuous and strongly monotone operators. We introduce a broader class of structured non-monotone operators, namely *p-quasi-sharp* operators ($p > 0$), which allows tractably analyzing convergence behavior of algorithms. We show that the stochastic Popov method converges *almost surely* to a solution for all operators from this class under a *linear growth*. In addition, we obtain the last iterate convergence rate (in expectation) for the method under a *linear growth* condition for 2-quasi-sharp operators. Based on our analysis, we refine the results for smooth 2-quasi-sharp and p-quasi-sharp operators (on a compact set), and obtain the optimal convergence rates.

1. Introduction

Recently, the framework of variational inequalities (VIs) has attracted much attention from researchers due to the wide range of its applications. A VI problem results when generalizing a variety of optimization problems, including those involving constraints, min-max optimization, and more general non-zero sum games. The adversarial approach in machine learning (ML) is yet another motivation behind the recent interest in stochastic VIs which allows modeling stochasticity in training.

In ML applications, including optimization of deep neural networks or generative adversarial networks (GANs), a large condition number associated with the operator is a key source of slower convergence rates [13]. When the condition number of the operator, defined as the ratio of the Lipschitz constant to the strong monotonicity constant, is large, it has been observed that the projection method is slower than Popov and EG algorithms [1]. Furthermore, while both Popov and EG methods have the same theoretical upper bound on the number of iterations for monotone operators ([3], [7], the Popov method requires only one oracle call per iteration, while EG requires two. For these reasons, we focus on the Popov method.

Most results on the last iterate convergence of first-order methods for the stochastic VI involve strong monotonicity. Weak sharpness, a weaker condition than strong monotonicity, is widely used to show convergences in optimization and monotone VI problems [11, 17]. But in many real-life applications (e.g., GANs where both the discriminator and generator usually are nonconvex deep neural networks), the resulting VI is not monotone. To address this issue, we present a broad class of structured, non-monotone, constrained VIs with non-unique solutions, called p-quasi-sharp.

Figure 1 visualizes the relationship between our newly introduced class of operators and the existing ones. For this setting, we now summarize our contributions.

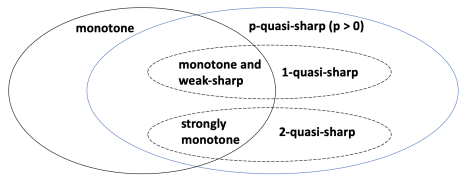

Figure 1: Relations between different operator classes and the new class of p-quasi-sharp operators.

1.1. Our contributions

We summarize our contributions on the last iterate convergence guarantees of stochastic Popov method for *constrained non-monotone stochastic VIs with non-unique solutions* (see also Table 1). A key feature of our analysis is the use of a new type of non-monotone VIs with special operators, termed *quasi-sharp*. The class of monotone and weak-sharp operators is contained in the class of quasi-sharp operators (see Figure 1). Moreover, when the VI solution is not unique, the quasi-sharpness is a weaker condition than that of quasi-strong monotonicity. Also, we presented an example of an operator that satisfies p-quasi sharpness, is not monotone, and does not satisfy previously considered assumptions.

Our main contribution is proving almost sure (*a.s.*) convergence when the operator is assumed to have a *linear growth* and *quasi-sharpness*. The class of linear growth operators includes Lipschitz continuous and bounded operators. *We prove a.s. convergence of the iterates to a solution*, which is a new result for this setting in contrast with the existing results showing only the convergence of the iterate distances to the solution set. To the best of our knowledge, this is the most general result on *a.s.* convergence of the stochastic Popov method.

Our second main contribution is in deriving the *first known* last iterate sublinear convergence rates for linear growth operators under the quasi-sharpness assumption with $p = 2$. Also, leveraging results from [16] we obtained $\mathcal{O}\left(C^2 R_0 \exp\left(-\mu^2 K/C^2\right)/\mu^2 + \sigma^2/\mu^2 K\right)$ when the number K of iterations is known in advance. We then refine the analysis to Lipschitz continuous operators and obtain $\mathcal{O}\left(LR_0 \exp\left(-\mu K/L\right)/\mu + \sigma^2/\mu^2 K\right)$ convergence rates. This rate also holds for the quasi-strongly monotone setting wherein we recover the result in [4] for unconstrained finite-sum VIs.

Finally, we focus on convergence rates under quasi-sharpness with $p \leq 2$. We derive rate bounds for the VIs with a compact constraint set and a continuous operator. In this setting, we can verify a linear growth condition and thus obtain asymptotic convergence. We obtained the last iterate convergence with $\mathcal{O}\left(R_0 \exp\left(-K\right) + (\sigma^2 + D^2) M_U^{2(2-p)}/\mu^2 K\right)$ rates.

1.2. Related Work

There are many works on convergence results of first order method for monotone SVI, including[8–11, 17]. Recently, quasi-strong monotonicity was introduced in [12] to establish last iterate convergence rates of projection and consensus methods for unconstrained SVIs with unique solutions. For

Assumptions on operator $F(\cdot)$	Rates
$\|F(u)\| \leq C\|u\| + D,\ p \in (0,\infty)$	Asymptotic Convergence (Our Thm 1)
$\|F(u)\| \leq C\|u\| + D,\ p = 2$	$\frac{C^2}{\mu^2} R_0 \exp[-\frac{\mu^2}{C^2}K] + \frac{\sigma^2}{\mu^2 K}$ (Our Thm 2)
Lipschitz continuous, $p = 2$	$\frac{\sigma^2}{\mu^2 K}$ [8] $\frac{L}{\mu} R_0 \exp[-\frac{\mu}{L}K] + \frac{\sigma^2}{\mu^2 K}$ [4] $\frac{L}{\mu} R_0 \exp[-\frac{\mu}{L}K] + \frac{\sigma^2}{\mu^2 K}$ (Our Thm 3)
$\|F(u)\| \leq D,\ p \leq 2$	$R_0 \exp[-K] + \frac{(\sigma^2+D^2)M_2^{2(2-p)}}{\mu^2 K}$ (Our Thm 4)

Table 1: Summary of the best known and our results on convergence rates of stochastic Popov method for stochastic VIs under p-quasi sharpness assumption (see Assumption 3). Convergence rates are obtained for the case when the number K of iterations is given and fixed. When the number K of iterations is not given, then the uniform upper bounds on the error after k iterations are $\mathcal{O}(1/k)$ in all cases. Paper [8] provides a convergence rate only for strongly-monotone unconstrained SVIs, while [4] has a convergence rate for finite-sum unconstrained VIs.

the same setting, both [6] and [4] studied EG and Popov methods, respectively, under a quasi-strong monotonicity condition and derived $\mathcal{O}(LR_0 \exp[-\mu K/L]/\mu + \sigma^2/\mu^2 K)$ convergence rates.

We consider constrained SVIs with non-monotone operators under linear growth; we also do not assume uniqueness of the solution. The rest of the paper is organized as follows. In Section 2, we introduce a general VI problem and provide our assumptions. In Section 3, we present our main results on the last iterate convergence and provide convergence rates. We present discussion in Section 4. Also, we present experiments on Popov method for p-quasi sahrp operators with linear growth in Appendix D.

2. Variational Inequality Problem

A variational inequality problem is specified by a (nonempty) set $U \subseteq \mathbb{R}^m$ and an operator $F(\cdot) : U \to \mathbb{R}^m$, and denoted by VI$(U, F)$. For $U = \mathbb{R}^m$, we obtain an unconstrained VI. The variational inequality problem VI(U, F) consists of determining a point $u^* \in U$ such that

$$\langle F(u^*), u - u^* \rangle \geq 0 \qquad \text{for all } u \in U. \tag{1}$$

The solution set for the VI(U, F) is $U^* = \{u^* \in U \mid \langle F(u^*), u - u^* \rangle \geq 0 \ \text{ for all } u \in U\}$.

We focus on a stochastic variational inequality problem (SVI(U, F)), corresponding to the case when the operator $F(u) = \mathbb{E}[\Phi(u, \xi)]$ for all $u \in U$, where ξ is a random vector. For such a problem, we consider a stochastic variant of the Popov method [15] defined by:

$$u_{k+1} = P_U(u_k - \alpha_k \Phi(h_k, \xi_k)), \quad h_{k+1} = P_U(u_{k+1} - \alpha_{k+1} \Phi(h_k, \xi_k)), \tag{2}$$

where $\alpha_k > 0$ is a stepsize, and $u_0, h_0 \in U$ are arbitrary deterministic points[1].

Regarding the stochastic approximation error $\Phi(h_k, \xi_k) - F(h_k)$, we assume that it is unbiased and with a finite variance, formalized as follows.

Assumption 1 *The random sample sequence $\{\xi_k\}$ is such that for some $\sigma > 0$ and for all $k \geq 0$,*

$$\mathbb{E}[\Phi(h_k, \xi_k) - F(h_k) \mid h_k] = 0, \qquad \mathbb{E}[\|\Phi(h_k, \xi_k) - F(h_k)\|^2 \mid h_k] \leq \sigma^2.$$

1. The results easily extend to the case when the initial points are random as long as $\mathbb{E}[\|u_0\|^2]$ and $\mathbb{E}[\|h_0\|^2]$ are finite.

Regarding the VI(U, F), we will assume that the set U is closed and convex. We will also assume that the solution set U^* is nonempty and closed. Our first assumption is a linear growth property defined below.

Assumption 2 *The operator $F(\cdot) : U \to \mathbb{R}^m$ has a linear growth on the set U with $C \geq 0$ and $D \geq 0$:*

$$\|F(u)\| \leq C\|u\| + D \qquad \text{for all } u \in U.$$

An operator $F(\cdot)$ is *bounded on the set U* if the preceding linear growth condition is satisfied with $C = 0$. A continuous operator has a linear growth over a compact set U. Moreover, when operator $F(\cdot)$ is Lipschitz continuous over the set U, then it has a linear growth on U. Additionally, we consider the p-quasi sharpness property which captures the behavior of the operator with respect to the solution set U^*.

Assumption 3 *The operator $F(\cdot) : U \to \mathbb{R}^m$ has a p-quasi sharpness property over U relative to the solution set U^*, i.e., for some $p > 0$, $\mu > 0$, and for all $u \in U$ and all $u^* \in U^*$,*

$$\langle F(u), u - u^* \rangle \geq \mu \, \mathrm{dist}^p(u, U^*). \tag{3}$$

When $p = 1$, quasi-sharpness property is weaker than weak-sharpness and monotonicity considered in [17], [11]. When $p = 2$, the 2-quasi sharpness property includes the quasi-strongly monotone property, which has been used in [12], [6] to analyze the convergence of stochastic gradient and extra-gradient methods. We note that an operator can posses p-quasi sharpness property but need not necessarily be monotone. Leveraging this property, we can show convergence results for non-monotone SVIs (and VIs). Next, we present an example of operator that satisfies p-quasi sharpness and does not satisfy previously considered conditions. We rigorously prove the below observations in Appendix A.

Example 1 *Consider operator* $F(u) = c \begin{bmatrix} \mathrm{sign}(u_1)|u_1|^{p-1} + u_2 \\ \mathrm{sign}(u_2)|u_2|^{p-1} - u_1 \end{bmatrix}$, $c = \begin{cases} 2, & \|u\| \leq 1 \\ 1, & \|u\| \geq 1 \end{cases}$ *with* $p > 0$. *Then F is p-quasi monotone with $\mu = 2^{2-p}$ and has a linear growth for $p \geq 2$. However, F is not monotone and, for any $p \in [1, 2) \cup (2, \infty)$, it does not satisfy positive co-monotonicity or quasi-strong monotonicity conditions. In addition, operator F is not Lipschitz continuous.*

3. Last Iterate Convergence Analysis

In this section, we present our convergence analysis of stochastic Popov method for solving for SVI(U, F).

3.1. Almost Sure and in-Expectation Convergence

In this section, we establish almost sure (*a.s.*) convergence of the stochastic Popov method for SVI(U, F) assuming that the stepsize is diminishing.

Assumption 4 *The positive sequence $\{\alpha_k\}$ is such that*

$$\alpha_k > 0 \text{ for all } k, \qquad \sum_{k=0}^{\infty} \alpha_k = \infty, \qquad \sum_{k=0}^{\infty} \alpha_k^2 < \infty. \tag{4}$$

The following theorem shows *a.s.* convergence of the method when the operator $F(\cdot)$ has a linear growth and the p-quasi sharpness property. Unlike [17] and [11], our result neither requires monotonicity of the operator $F(\cdot)$ nor compactness of the set U. Instead, it relies on the linear growth condition and the p-sharpness property of the operator.

Theorem 1 *Let Assumptions 1, 2, 3, and 4 hold. Then, the following statements hold the iterate sequences $\{u_k\}$ and $\{h_k\}$ generated by the stochastic Popov method (2):*

(a) The sequence $\{u_k\}$ and $\{h_k\}$ converge almost surely to some $\bar{u} \in U$, where $\bar{u}$ is a solution a.s.*, i.e. $P(\bar{u} \in U^*) = 1$.*

(b) The sequences $\{\mathbb{E}[\|u_k\|^2]\}$ and $\{\mathbb{E}[\|h_k\|^2]\}$ are bounded.

(c) If the solution set U^* is bounded, then the sequences $\{u_k\}$ and $\{h_k\}$ also converge in expectation, i.e.,

$$\lim_{k\to\infty} \mathbb{E}[\|u_k - \bar{u}\|^2] = 0, \qquad \lim_{k\to\infty} \mathbb{E}[\|h_k - \bar{u}\|^2] = 0,$$

In the existing works on the stochastic first-order methods, such as EG and projection methods for quasi-strongly monotone operators [6], [12], there are no results on *a.s.* convergence of the iterates to a solution, except for the case when the solution set U^* is a singleton. Our Theorem 1 shows that such *a.s.* convergence results are possible even when the solution set U^* is not necessarily a singleton.

3.2. Convergence Rates

Here, we present convergence rate results for the stochastic Popov method when the operator $F(\cdot)$ has p-quasi sharpness property with $p \leq 2$.

Operator $F(\cdot)$ with 2-quasi Sharpness Property

To achieve exponential decay in stochastic part of the convergence rate and sublinear rate in stochastic part we use Lemma 3 of [16] and the stepize choice given in the proof of that lemma, namely, for any given $K \geq 0$, the stepsize α_k, $0 \leq k \leq K$ is given by

$$\begin{aligned} \alpha_k &= \frac{1}{d} \qquad \text{if } K \leq \frac{d}{a} \text{ or } \left(K > \frac{d}{a} \text{ and } k < k_0\right), \\ \alpha_k &= \frac{2}{a\left(\frac{2d}{a} + k - k_0\right)} \qquad \text{if } K > \frac{d}{a} \text{ and } k \geq k_0, \end{aligned} \tag{5}$$

where $k_0 = \lceil \frac{K}{2} \rceil$ and $d \geq a > 0$. Our convergence rate estimate with such a stepsize selection is obtained assuming that the solution set U^* is compact.

Theorem 2 *Let Assumptions 1, 2, and 3 with $p = 2$ hold. For a given $K \geq 1$, let the stepsize α_k be given as in (5) with $a = \frac{\mu}{2}$ and d satisfying $d^{-1} \leq \min\left\{\frac{\mu}{288C^2}, \frac{4}{9\mu}\right\}$, where the constant $C > 0$ is from the linear growth condition (Assumption 2). Then, the following relation holds for the iterate sequence $\{u_k\}$ generated by the stochastic Popov method (2) for all $K \geq 1$,*

$$\mathbb{E}[\text{dist}^2(u_{K+1}, U^*)] \leq \frac{64d}{\mu} r_1 e^{-\frac{\mu(K-1)}{4d}} + \frac{144c}{\mu^2(K-1)},$$

where $r_1 = \mathbb{E}[\text{dist}^2(u_1, U^*) + \|h_0 - u_1\|^2]$, $c = 12\sigma^2 + 2D^2 + 12M_1^2$, and M_1 is an upper bound for the norms of solutions $u^* \in U^*$, i.e., $\|u^*\| \leq M_1$ for all $u^* \in U^*$.

As noted earlier, a Lipschitz continuous operator $F(\cdot)$ on the set U with a Lipschitz constant L satisfies a linear growth condition with $C = L$ and $D = L\|u'\| + \|F(u')\|$ where $u' \in U$ is an arbitrary but fixed point. Thus, Theorem 2 applies with $C = L$ to such an operator. By directly applying Theorem 2 to a Lipschitz continuous operator, we would obtain a convergence rate estimate of the form $\frac{L^2}{\mu^2}\tilde{C}_1 r_1 e^{-\frac{\mu^2}{L^2}K} + \frac{L^2\tilde{C}_2}{\mu^2 K}$ for some positive constants $\tilde{C}_1$ and $\tilde{C}_2$ independent of L and μ. However, a better convergence rate result can be obtained of the form $\frac{L}{\mu}\hat{C}_1 e^{-\frac{\mu}{L}K} + \frac{\sigma^2\hat{C}_2}{\mu^2 K}$, where the positive constants $\hat{C}$ and $\hat{C}$ are independent of L and μ. We establish such an estimate by directly exploiting the Lipschitz continuity of the operator, which also allows us to relax the boundedness assumption for the solution set U^* imposed in Theorem 2, as seen in the following theorem.

Theorem 3 *Let Assumption 1 hold, and assume that the operator $F(\cdot)$ is Lipschitz continuous over U with a constant $L > 0$ and satisfies Assumption 3 with $p = 2$. For any given $K \geq 1$, let the stepsizes α_k be defined by (5) with $a = \mu$ and $d \geq \max\left\{2\sqrt{3}L,\, \mu\right\}$. Then, the iterate sequence $\{u_k\}$ generated by the stochastic Popov method (2) satisfies the following inequality for all $K \geq 1$,*

$$\mathbb{E}\left[\mathrm{dist}^2(u_{K+1}, U^*)\right] \leq \frac{32d}{\mu} r_1 e^{-\frac{\mu(K-1)}{2d}} + \frac{432\sigma^2}{\mu^2(K-1)},$$

where $r_1 = \mathbb{E}\left[\mathrm{dist}^2(u_1, U^*) + \|h_0 - u_1\|^2\right]$.

Operator $F(\cdot)$ with p-quasi Sharpness Property Here, we establish a convergence rate result for the SVI$(U, , F(\cdot))$ with the operator $F(\cdot)$ that has p-quasi-sharpness property with $p \leq 2$. For this result, we assume that the set U is compact.

Theorem 4 *Let U be a compact convex set, and the constants $M_U > 0$ and $D > 0$ be such that $\|u - u'\| \leq M_U$ for all $u, u' \in U$ and $\|F(u)\| \leq D$ for all $u \in U$. Also, let Assumptions 1 and 3 with $p \leq 2$ hold. Then, for the iterate sequence $\{u_k\}$ generated by the stochastic Popov method (2) the following statements are valid. For any $K \geq 1$, let the stepsizes α_k be given by (5) with $a = \frac{\mu}{M_U^{2-p}}$, and $d = \frac{2\mu}{M_U^{2-p}}$. Then, we have for all $K \geq 1$,*

$$\mathbb{E}[\mathrm{dist}^2(u_{K+1}, U^*)] \leq 64\mathbb{E}[\mathrm{dist}^2(u_1, U^*)]e^{-\frac{(K-1)}{4}} + \frac{432(\sigma^2 + 2D^2)M_U^{2(2-p)}}{\mu^2(K-1)}.$$

4. Discussion

We have considered non-monotone SVIs under *linear growth* condition on operators when the solution set is not necessarily a singleton. The class of operators with *linear growth* includes Lipschitz continuous and bounded operators. Focusing on the convergence of the stochastic Popov method, we have proposed a broad class of structured non-monotone VIs called p-quasi-sharp, which generalizes the weak-sharpness condition for monotone VIs. We have proved the *a.s.* last iterate convergence to a solution for the Popov method under p-quasi-sharpness condition for all $p > 0$. Among all existing results on *a.s.* convergence of the Popov method, ours is the most extensive. Moreover, we showed the optimal convergence rate of the Popov method for Lipschitz continuous and 2-quasi-sharp operators.

References

[1] Aleksandr Beznosikov, Boris Polyak, Eduard Gorbunov, Dmitry Kovalev, and Alexander Gasnikov. Smooth monotone stochastic variational inequalities and saddle point problems–survey. *arXiv preprint arXiv:2208.13592*, 2022.

[2] Patrick Billingsley. *Probability and Measure*. Wiley-Interscience Publication, John Wiley & Sons, USA, 3 edition, 1995.

[3] Yang Cai, Argyris Oikonomou, and Weiqiang Zheng. Tight last-iterate convergence of the extragradient method for constrained monotone variational inequalities. *arXiv preprint arXiv:2204.09228*, 2022.

[4] Sayantan Choudhury, Eduard Gorbunov, and Nicolas Loizou. Single-call stochastic extragradient methods for structured non-monotone variational inequalities: Improved analysis under weaker conditions. *arXiv preprint arXiv:2302.14043*, 2023.

[5] Francisco Facchinei and Jong-Shi Pang. *Finite-dimensional variational inequalities and complementarity problems*. Springer, 2003.

[6] Eduard Gorbunov, Hugo Berard, Gauthier Gidel, and Nicolas Loizou. Stochastic extragradient: General analysis and improved rates. *arXiv preprint arXiv:2111.08611*, 2021.

[7] Eduard Gorbunov, Nicolas Loizou, and Gauthier Gidel. Extragradient method: $o(1/k)$ last-iterate convergence for monotone variational inequalities and connections with cocoercivity. *arXiv preprint arXiv:2110.04261*, 2021.

[8] Yu-Guan Hsieh, Franck Iutzeler, Jérôme Malick, and Panayotis Mertikopoulos. On the convergence of single-call stochastic extra-gradient methods. *Advances in Neural Information Processing Systems*, 32, 2019.

[9] Yu-Guan Hsieh, Franck Iutzeler, Jérôme Malick, and Panayotis Mertikopoulos. Explore aggressively, update conservatively: Stochastic extragradient methods with variable stepsize scaling. *Advances in Neural Information Processing Systems*, 33:16223–16234, 2020.

[10] Anatoli Juditsky, Arkadi Nemirovski, and Claire Tauvel. Solving variational inequalities with stochastic mirror-prox algorithm. *Stochastic Systems*, 1(1):17–58, 2011.

[11] Aswin Kannan and Uday V Shanbhag. Optimal stochastic extragradient schemes for pseudomonotone stochastic variational inequality problems and their variants. *Computational Optimization and Applications*, 74(3):779–820, 2019.

[12] Nicolas Loizou, Hugo Berard, Gauthier Gidel, Ioannis Mitliagkas, and Simon Lacoste-Julien. Stochastic gradient descent-ascent and consensus optimization for smooth games: Convergence analysis under expected co-coercivity. *Advances in Neural Information Processing Systems*, 34:19095–19108, 2021.

[13] Eliya Nachmani, Yaron Bachar, Elad Marciano, David Burshtein, and Yair Be'ery. Near maximum likelihood decoding with deep learning. *arXiv preprint arXiv:1801.02726*, 2018.

[14] Boris T Polyak. Introduction to optimization. optimization software. *Inc., Publications Division, New York*, 1:32, 1987.

[15] Leonid Denisovich Popov. A modification of the arrow-hurwicz method for search of saddle points. *Mathematical notes of the Academy of Sciences of the USSR*, 28(5):845–848, 1980.

[16] Sebastian U Stich. Unified optimal analysis of the (stochastic) gradient method. *arXiv preprint arXiv:1907.04232*, 2019.

[17] Farzad Yousefian, Angelia Nedić, and Uday V Shanbhag. Optimal robust smoothing extragradient algorithms for stochastic variational inequality problems. In *53rd IEEE Conference on Decision and Control*, pages 5831–5836. IEEE, 2014.

Appendix A. On p-Quasi Sharpness

We provide proof that operator from the Example 1 is p-quasi sharp and has linear growth. Moreover, such operator does not satisfy assumptions typically studied in the existing literature.

Proof Firstly, we find solution set of variational inequality VI($\mathbb{R}^2, F$). Since $U = \mathbb{R}^2$, a solution u^* of VI($\mathbb{R}^2, F$) must satisfy $F(u^*) = 0$. Let u^* be an arbitrary solution, then $\text{sign}(u_1^*)|u_1^*|^{p-1} + u_2^* = 0$ and $\text{sign}(u_2^*)|u_2^*|^{p-1} - u_1^* = 0$. From the first equality it follows that $\text{sign}(u_1^*) = -\text{sign}(u_2^*)$, while from the second inequality it follows that $\text{sign}(u_2^*) = \text{sign}(u_1^*)$. Hence $u_1^* = u_2^* = 0$, and VI($\mathbb{R}^2, F$) has a unique solution $u^* = (0, 0)$.

Moreover, this operator has p-quasi sharpness property with $p \geq 1$ and $\mu = 2^{1-p}$. To see this, let $\|u\| > 1$. Then, we have:

$$\begin{aligned}
\langle F(u), u - u^* \rangle &= \left\langle \begin{bmatrix} \text{sign}(u_1)|u_1|^{p-1} + u_2 \\ \text{sign}(u_2)|u_2|^{p-1} - u_1 \end{bmatrix}, \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\rangle \\
&= |u_1|^p + |u_2|^p \\
&\geq 2^{1-p} \left(|u_1| + |u_2|\right)^p \quad \text{Jensen inequality for a convex function } |\cdot|^p \text{ since } p \geq 1 \\
&\geq 2^{1-p} \left(\sqrt{u_1^2 + u_2^2} \right)^p \qquad \text{due to } \|\cdot\|_1 \geq \|\cdot\|_2 \text{ and monotonicity of } |\cdot|^p \\
&= 2^{1-p} \text{dist}^p(u, U^*).
\end{aligned} \tag{6}$$

In case when $\|u\| \leq 1$, the arguments are the same and we get $\langle F(u), u - u^* \rangle \geq 2^{2-p} \text{dist}^p(u, U^*)$. Moreover, it can be shown that operator $F(\cdot)$ is not monotone for $p > 1$. Consider two points $u = (u_1, u_2)'$, where $u_1 = 0, u_2 = 1$, and $v = (v_1, v_2)'$, where $v_1 = 0, v_2 = 1 + \frac{1}{5(p-1)}$. Then, $F(u) = (2, 2)'$, and $F(v) = (1 + \frac{1}{5(p-1)}, (1 + \frac{1}{5(p-1)})^{p-1})'$, and we have

$$\begin{aligned}
\langle F(u) - F(v), u - v \rangle &= \langle \begin{bmatrix} 1 - \frac{1}{5(p-1)} \\ 2 - (1 + \frac{1}{5(p-1)})^{p-1} \end{bmatrix}, \begin{bmatrix} 0 \\ -\frac{1}{5(p-1)} \end{bmatrix} \rangle \\
&= -\frac{1}{5(p-1)} (2 - (1 + \frac{1}{5(p-1)})^{p-1}) \\
&\leq -\frac{1}{5(p-1)} (2 - e^{0.2}) < 0
\end{aligned}$$

where the inequality holds since $(1 + a/x)^x \leq e^a$.

Next, we show that F is discontinuous at $u = (0, 1)'$. Consider $v_k = (0, 1 + 1/k)'$ and notice that as $k \to \infty$, $v_k \to u$, but $\lim_{k \to \infty} \|F(u) - F(v_k)\| = \sqrt{2}$. Hence, F is discontinuous at $u = (0, 1)'$. Now, we show that operator F has linear growth for $p \leq 2$. In case when $\|u\| \leq 1$, $\|F(u)\| = 2\sqrt{(\text{sign}(u_1)|u_1|^{p-1} + u_2)^2 + (\text{sign}(u_2)|u_2|^{p-1} - u_1)^2} \leq 2\sqrt{2^2 + 2^2} = 4\sqrt{2}$. For

$\|u\| > 1$:

$$
\begin{aligned}
\|F(u)\| &= \| \begin{bmatrix} u_2 \\ -u_1 \end{bmatrix} + \begin{bmatrix} \text{sign}(u_1)|u_1|^{p-1} \\ \text{sign}(u_2)|u_2|^{p-1} \end{bmatrix} \| \\
&\leq \|u\| + \| \begin{bmatrix} \text{sign}(u_1)|u_1|^{p-1} \\ \text{sign}(u_2)|u_2|^{p-1} \end{bmatrix} \| \\
&\leq \|u\| + \sqrt{(|u_1|^{p-1})^2 + (|u_2|^{p-1})^2} \\
&\leq \|u\| + \sqrt{u_1^2 + u_2^2 + 2} \\
&\leq 2\|u\| + \sqrt{2}.
\end{aligned}
\tag{7}
$$

Combining these two cases, we obtain that $\|F(u)\| \leq 2\|u\| + 4\sqrt{2}$ for all $u \in \mathbb{R}^2$.

Finally, we show that F does not satisfy quasi-strong monotonicity for any $p \in (0, 2) \cup (2, \infty)$. To arrive at contradiction, we assume that F is quasi-strong monotone with $\mu > 0$. Then, for all $u \in \mathbb{R}^2$

$$
\langle F(u), u - u^* \rangle \geq \mu \|u - u^*\|^2.
$$

Consider $u = (u_1, 0)$. Similar to the derivation in (6), we can see that

$$
\langle F(u), u - u^* \rangle = c(|u_1|^p + |u_2|^p) = c|u_1|^p.
$$

Since $\|u - u^*\|^2 = \|u\|^2 = u_1^2$, the quasi-strong monotonicity would imply that the following inequality holds for $p > 0$ and $p \neq 2$, and for any $u_1 \in \mathbb{R}$,

$$
c|u_1|^p \geq \mu u_1^2,
$$

which is a contradiction. ■

Appendix B. Almost Sure and in-Expectation Convergence

In our analysis of the Popov method (2) we use the properties of the projection operator $P_U(\cdot)$ given in the following lemma.

Lemma 5 *(Theorem 1.5.5 and Lemma 12.1.13 in [5]) Given a convex closed set $U \subset \mathbb{R}^m$, the projection operator $P_U(\cdot)$ has the following properties:*

$$
\langle v - P_U(v), u - P_U(v) \rangle \leq 0 \quad \textit{for all } u \in U, v \in \mathbb{R}^m, \tag{8}
$$

$$
\|u - P_U(v)\|^2 \leq \|u - v\|^2 - \|v - P_U(v)\|^2 \quad \textit{for all } u \in U, v \in \mathbb{R}^m, \tag{9}
$$

$$
\|P_U(u) - P_U(v)\| \leq \|u - v\| \quad \textit{for all } u, v \in \mathbb{R}^m. \tag{10}
$$

B.1. Basis Lemma

We first provide a lemma presenting the main inequality for the iterates of stochastic Popov method (2) without any assumptions on the operator. This lemma is the basis for all the subsequent results. Lemma 6 can be further refined for different types of operators, such as Lipschitz continuous and operators with linear growth.

Lemma 6 *Let U be a closed convex set. Then, for the iterate sequences $\{u_k\}$ and $\{h_k\}$ generated by the stochastic Popov method (2) we surely have for all $y \in U$ and $k \geq 1$,*

$$\begin{aligned}\|u_{k+1}-y\|^2 \leq \|u_k - y\|^2 - \|u_{k+1}-h_k\|^2 - \|u_k - h_k\|^2 \\ - 2\alpha_k\langle e_k + F(h_k), h_k - y\rangle + 6\alpha_k^2\|e_{k-1}\|^2 \\ + 6\alpha_k^2\,\|F(h_k)-F(h_{k-1})\|^2 + 6\alpha_k^2\|e_k\|^2,\end{aligned}$$

where $e_k = \Phi(h_k,\xi_k) - F(h_k)$ for all $k \geq 0$.

Proof Let $k \geq 1$ be arbitrary but fixed. From the definition of u_{k+1} in (2), we have $\|u_{k+1}-y\|^2 = \|P_U(u_k - \alpha_k\Phi(h_k,\xi_k)) - y\|$ for any $y \in U$. Using the inequality (9) of Lemma 5 we obtain for any $y \in U$,

$$\begin{aligned}\|u_{k+1}-y\|^2 &\leq \|u_k - \alpha_k\Phi(h_k,\xi_k) - y\|^2 - \|u_k - \alpha_k\Phi(h_k,\xi_k) - u_{k+1}\|^2 \\ &= \|u_k - y\|^2 - \|u_{k+1}-u_k\|^2 - 2\alpha_k\langle\Phi(h_k,\xi_k), u_{k+1}-y\rangle.\end{aligned} \tag{11}$$

We next consider the term $\|u_{k+1}-u_k\|^2$, where we add and subtract h_k, and thus obtain

$$\begin{aligned}\|u_{k+1}-u_k\|^2 =&\|(u_{k+1}-h_k)-(u_k-h_k)\|^2 \\ =&\|u_{k+1}-h_k\|^2 + \|u_k - h_k\|^2 - 2\langle u_k - h_k, u_{k+1}-h_k\rangle \\ =&\|u_{k+1}-h_k\|^2 + \|u_k - h_k\|^2 - 2\langle u_k - \alpha_k\Phi(h_{k-1},\xi_{k-1}) - h_k, u_{k+1}-h_k\rangle \\ &- 2\alpha_k\langle\Phi(h_{k-1},\xi_{k-1}), u_{k+1}-h_k\rangle,\end{aligned} \tag{12}$$

where the last equality is obtained by adding and subtracting $2\alpha_k\langle\Phi(h_{k-1},\xi_{k-1}), u_{k+1}-h_k\rangle$. Next, we use the projection property (8) of Lemma 5, where we let $v = u_k - 2\alpha_k\Phi(h_{k-1},\xi_{k-1})$, $u = u_{k+1}$, and $h_k = P_U(v)$ (which follows by the definition of h_k in the method (2)). Then, it follows that

$$\langle u_k - \alpha_k\Phi(h_{k-1},\xi_{k-1}) - h_k, u_{k+1}-h_k\rangle \leq 0. \tag{13}$$

Therefore,

$$\|u_{k+1}-u_k\|^2 \geq \|u_{k+1}-h_k\|^2 + \|u_k - h_k\|^2 - 2\alpha_k\langle\Phi(h_{k-1},\xi_{k-1}), u_{k+1}-h_k\rangle \tag{14}$$

Combining (11) and (14) we can see that for any $y \in U$,

$$\begin{aligned}\|u_{k+1}-y\|^2 \leq&\|u_k - y\|^2 - \|u_{k+1}-h_k\|^2 - \|u_k - h_k\|^2 - 2\alpha_k\langle\Phi(h_k,\xi_k), u_{k+1}-h_k\rangle \\ &- 2\alpha_k\langle\Phi(h_k,\xi_k), h_k - y\rangle + 2\alpha_k\langle\Phi(h_{k-1},\xi_{k-1}), u_{k+1}-h_k\rangle \\ =&\|u_k - y\|^2 - \|u_{k+1}-h_k\|^2 - \|u_k - h_k\|^2 - 2\alpha_k\langle\Phi(h_k,\xi_k), h_k - y\rangle \\ &+ 2\alpha_k\langle\Phi(h_{k-1},\xi_{k-1}) - \Phi(h_k,\xi_k), u_{k+1}-h_k\rangle.\end{aligned} \tag{15}$$

To estimate the last inner product in (15), we write

$$\langle \Phi(h_{k-1}, \xi_{k-1}) - \Phi(h_k, \xi_k), u_{k+1} - h_k \rangle \le \|\Phi(h_{k-1}, \xi_{k-1}) - \Phi(h_k, \xi_k)\| \, \|u_{k+1} - h_k\|.$$

From the definitions of u_{k+1} and h_{k+1} in (2), we have $u_{k+1} = P_U(u_k - \alpha_k \Phi(h_k, \xi_k))$ and $h_k = P_U(u_k - \alpha_k \Phi(h_{k-1}, \xi_{k-1}))$. Thus, by using the Lipschitz property of the projection operator (see relation (10) in Lemma 5), we obtain $\|u_{k+1} - h_k\| \le \alpha_k \|\Phi(h_{k-1}, \xi_{k-1}) - \Phi(h_k, \xi_k)\|$, implying that

$$\langle \Phi(h_{k-1}, \xi_{k-1}) - \Phi(h_k, \xi_k), u_{k+1} - h_k \rangle \le \alpha_k \|\Phi(h_{k-1}, \xi_{k-1}) - \Phi(h_k, \xi_k)\|^2.$$

Upon substituting the preceding estimate back in relation (15), we have that

$$\begin{aligned} \|u_{k+1} - y\|^2 \le & \|u_k - y\|^2 - \|u_{k+1} - h_k\|^2 - \|u_k - h_k\|^2 - 2\alpha_k \langle \Phi(h_k, \xi_k), h_k - y \rangle \\ & + 2\alpha_k^2 \|\Phi(h_k, \xi_k) - \Phi(h_{k-1}, \xi_{k-1})\|^2. \end{aligned} \tag{16}$$

In the last term of (16), we add and subtract $F(h_k)$ and $F(h_{k-1})$. Recalling that $e_k = \Phi(h_k, \xi_k) - F(h_k)$, we obtain

$$\begin{aligned} & \|\Phi(h_k, \xi_k) - \Phi(h_{k-1}, \xi_{k-1})\|^2 \\ & \quad = \|(\Phi(h_k, \xi_k) - F(h_k)) + (F(h_k) - F(h_{k-1})) + (F(h_{k-1}) - \Phi(h_{k-1}, \xi_{k-1}))\|^2 \\ & \quad \le 3\|\Phi(h_k, \xi_k) - F(h_k)\|^2 + 3\|F(h_k) - F(h_{k-1})\|^2 + 3\|F(h_{k-1}) - \Phi(h_{k-1}, \xi_{k-1})\|^2 \\ & \quad \le 3\|F(h_k) - F(h_{k-1})\|^2 + 3(\|e_k\|^2 + \|e_{k-1}\|^2), \end{aligned} \tag{17}$$

where the first inequality follows from $(\sum_{i=1}^m a_i)^2 \le m \sum_{i=1}^m a_i^2$, which is valid for any scalars a_i, $i = 1, \ldots, m$, and any integer $m \ge 1$. Combining relations (16) and (17), and using $e_k = \Phi(h_k, \xi_k) - F(h_k)$, we obtain the desired relation:

$$\begin{aligned} \|u_{k+1} - y\|^2 \le & \|u_k - y\|^2 - \|u_{k+1} - h_k\|^2 - \|u_k - h_k\|^2 - 2\alpha_k \langle e_k + F(h_k), h_k - y \rangle \\ & + 6\alpha_k^2 \|F(h_k) - F(h_{k-1})\|^2 + 6\alpha_k^2 (\|e_k\|^2 + \|e_{k-1}\|^2). \end{aligned} \tag{18}$$

■

B.2. Linear Growth Condition

In the following lemma, we refine Lemma 6 for the case when the operator $F(\cdot)$ has a linear growth (see Assumption 2). The part (a) of the following lemma gives a suitable relation for our convergence rate analysis of the method (2), while part (b) is used for establishing almost sure convergence of the method.

Lemma 7 *Assume that U is a closed convex set and that the operator $F(\cdot) : U \to \mathbb{R}^m$ has a linear growth on the set U. Then, the iterates u_k and h_k of the stochastic Popov method (2) satisfy the following relations:*

(a) For all $y \in U$ and for all $k \ge 1$,

$$\begin{aligned} \|u_{k+1} - y\|^2 \le & \|u_k - y\|^2 - \|u_{k+1} - h_k\|^2 - \|u_k - h_k\|^2 - 2\alpha_k \langle e_k + F(h_k), h_k - y \rangle \\ & + 24\alpha_k^2 C^2 (\|h_k\|^2 + \|h_{k-1}\|^2) + 6\alpha_k^2 (\|e_k\|^2 + \|e_{k-1}\|^2 + 4D^2); \end{aligned} \tag{19}$$

(b) For all $y \in U$, $z \in \mathbb{R}^m$, and for all $k \geq 1$,

$$\begin{aligned}\|u_{k+1} - y\|^2 \leq &\|u_k - y\|^2 - \|u_{k+1} - h_k\|^2 - \|u_k - h_k\|^2 - 2\alpha_k\langle e_k + F(h_k), h_k - y\rangle \\ &+ 72\alpha_k^2 C^2(\|h_k - u_k\|^2 + \|u_k - h_{k-1}\|^2 + 2\|u_k - z\|^2) \\ &+ 6\alpha_k^2(\|e_k\|^2 + \|e_{k-1}\|^2 + 4D^2 + 24\|z\|^2);\end{aligned} \tag{20}$$

where $e_k = \Phi(h_k, \xi_k) - F(h_k)$ for all $k \geq 0$.

Proof Let $k \geq 1$ and $y \in U$ be arbitrary. By Lemma 6, we have

$$\begin{aligned}\|u_{k+1} - y\|^2 \leq &\|u_k - y\|^2 - \|u_{k+1} - h_k\|^2 - \|u_k - h_k\|^2 - 2\alpha_k\langle e_k + F(h_k), h_k - y\rangle \\ &+ 6\alpha_k^2\|F(h_k) - F(h_{k-1})\|^2 + 6\alpha_k^2(\|e_k\|^2 + \|e_{k-1}\|^2).\end{aligned} \tag{21}$$

Using $(\sum_{i=1}^m a_i)^2 \leq m\sum_{i=1}^m a_i^2$, which is valid for any scalars a_i, $i = 1, \ldots, m$, and any integer $m \geq 1$, to estimate $\|F(h_k) - F(h_{k-1})\|^2$, and the linear growth of operator $F(\cdot)$ we obtain

$$\begin{aligned}\|F(h_k) - F(h_{k-1})\|^2 &\leq 2\|F(h_k)\|^2 + 2\|F(h_{k-1})\|^2 \\ &\leq 4C^2(\|h_k\|^2 + \|h_{k-1}\|^2) + 4D^2.\end{aligned} \tag{22}$$

By substituting the preceding estimate back in relation (21) we arrive at the relation in part (a).

To obtain the relation in part (b), for $\|h_k\|^2$ we write

$$\|h_k\|^2 = \|(h_k - u_k) + (u_k - z) + z\|^2.$$

where $z \in \mathbb{R}^m$ is arbitrary. Using $(\sum_{i=1}^m a_i)^2 \leq m\sum_{i=1}^m a_i^2$, with $m = 3$, we find that

$$\|h_k\|^2 \leq 3\left(\|h_k - u_k\|^2 + \|u_k - z\|^2 + \|z\|^2\right).$$

Similarly, we can see that

$$\|h_{k-1}\|^2 \leq 3\left(\|h_{k-1} - u_k\|^2 + \|u_k - z\|^2 + \|z\|^2\right).$$

Therefore,

$$\|h_k\|^2 + |h_{k-1}\|^2 \leq 3\left(\|h_k - u_k\|^2 + \|h_{k-1} - u_k\|^2 + 2\|u_k - z\|^2 + 2\|z\|^2\right).$$

Upon substituting the preceding estimate back in relation (22) we obtain

$$\|F(h_k) - F(h_{k-1})\|^2 \leq 12C^2(\|h_k - u_k\|^2 + \|h_{k-1} - u_k\|^2 + 2\|u_k - z\|^2 + 2\|z\|^2) + 4D^2. \tag{23}$$

Combining the estimate in (23) with relation (21) we obtain the following relation

$$\begin{aligned}\|u_{k+1} - y\|^2 \leq &\|u_k - y\|^2 - \|u_{k+1} - h_k\|^2 - \|u_k - h_k\|^2 - 2\alpha_k\langle e_k + F(h_k), h_k - y\rangle \\ &+ 72\alpha_k^2 C^2(\|h_k - u_k\|^2 + \|u_k - h_{k-1}\|^2 + 2\|u_k - z\|^2) \\ &+ 6\alpha_k^2(\|e_k\|^2 + \|e_{k-1}\|^2 + 4D^2 + 24\|z\|^2),\end{aligned} \tag{24}$$

which is the relation stated in part (b). ■

In the forthcoming analysis, we use Lemma 11 [14], which is stated below.

Lemma 8 *[Lemma 11 [14]] Let* $\{v_k\}, \{z_k\}, \{a_k\}, \{b_k\}$ *be nonnegative random scalar sequences such that almost surely for all* $k \geq 0$,

$$\mathbb{E}[v_{k+1} \mid \mathcal{F}_k] \leq (1 + a_k)v_k - z_k + b_k, \tag{25}$$

where $\mathcal{F}_k = \{v_0, \dots, v_k, z_0, \dots, z_k, a_0, \dots, a_k, b_0, \dots, b_k\}$, *and* a.s. $\sum_{k=0}^{\infty} a_k < \infty$, $\sum_{k=0}^{\infty} b_k < \infty$. *Then, almost surely,* $\lim_{k\to\infty} v_k = v$ *for some nonnegative random variable* v *and* $\sum_{k=0}^{\infty} z_k < \infty$.

As a direct consequence of Lemma 8, when the sequences $\{v_k\}, \{z_k\}, \{a_k\}, \{b_k\}$ are deterministic, we obtain the following result.

Lemma 9 *Let* $\{\bar{v}_k\}, \{\bar{z}_k\}, \{\bar{a}_k\}, \{\bar{b}_k\}$ *be nonnegative scalar sequences such that for all* $k \geq 0$,

$$\bar{v}_{k+1} \leq (1 + \bar{a}_k)\bar{v}_k - \bar{z}_k + \bar{b}_k, \tag{26}$$

where $\sum_{k=0}^{\infty} \bar{a}_k < \infty$ *and* $\sum_{k=0}^{\infty} \bar{b}_k < \infty$. *Then,* $\lim_{k\to\infty} \bar{v}_k = \bar{v}$ *for some scalar* $\bar{v} \geq 0$ *and* $\sum_{k=0}^{\infty} \bar{z}_k < \infty$.

We also use Lebesgue Dominated Convergence Theorem, which is stated below and can be found, for example, in the textbook [2], Theorem 16.4, page 209.

Theorem 10 (Lebesgue Dominated Convergence Theorem) *Let* $\{f_k\}$ *be a sequence of functions and* g *be a function in some measure space with a measure* ν, *and let* $|f_k| \leq g$ *almost everywhere. If* g *is integrable and* $f_k \to f$ *almost everywhere, then* $\int f_k d\nu \to \int f d\nu$.

B.3. Proof of Theorem 1

We use Lemmas 8 and 9 to establish parts (a) and (b), respectively, while we use Theorem 10 to prove part (c). (a) Using Lemma 7(b), where we set $y = z = u^*$ for an arbitrary $u^* \in U^*$, after re-arranging the terms, we obtain for all $u^* \in U^*$ and for all $k \geq 1$,

$$\begin{aligned} \|u_{k+1} - u^*\|^2 + \|u_{k+1} - h_k\|^2 \leq{}& (1 + 144\alpha_k^2 C^2)\|u_k - u^*\|^2 - (1 - 72\alpha_k^2 C^2)\|u_k - h_k\|^2 \\ &- 2\alpha_k \langle F(h_k), h_k - u^*\rangle + 2\alpha_k \langle e_k, u^* - h_k\rangle + 72\alpha_k^2 C^2 \|u_k - h_{k-1}\|^2 \\ &+ 6\alpha_k^2(\|e_{k-1}\|^2 + \|e_k\|^2 + 24\|u^*\|^2 + 4D^2). \end{aligned} \tag{27}$$

Under Assumption 3, the following relation is valid for all $k \geq 0$,

$$\langle F(h_k), h_k - u^*\rangle \geq \mu \text{dist}^p(h_k, U^*), \tag{28}$$

with $p > 0$ and $\mu > 0$. Combining (28) with (27) we get for all $u^* \in U^*$ and for all $k \geq 1$,

$$\begin{aligned} \|u_{k+1} - u^*\|^2 + \|u_{k+1} - h_k\|^2 \leq{}& (1 + 144\alpha_k^2 C^2)\|u_k - u^*\|^2 - 2\alpha_k \mu \text{dist}^p(h_k, U^*) \\ &-(1 - 72\alpha_k^2 C^2)\|u_k - h_k\|^2 + 2\alpha_k \langle e_k, u^* - h_k\rangle \\ &+72\alpha_k^2 C^2 \|u_k - h_{k-1}\|^2 + 6\alpha_k^2(\|e_{k-1}\|^2 + \|e_k\|^2) \\ &+6\alpha_k^2(24\|u^*\|^2 + 4D^2). \end{aligned} \tag{29}$$

By writing

$$72\alpha_k^2 C^2 \|u_k - h_{k-1}\|^2 \leq (1 + 144\alpha_k^2 C^2)\|u_k - h_{k-1}\|^2 - \|u_k - h_{k-1}\|^2,$$

and regrouping some of the terms in (29), we have for all $u^* \in U^*$ and for all $k \geq 1$,

$$\begin{aligned}\|u_{k+1} - u^*\|^2 + \|u_{k+1} - h_k\|^2 \leq &(1 + 144\alpha_k^2 C^2)(\|u_k - u^*\|^2 + \|u_k - h_{k-1}\|^2) \\ &-2\alpha_k \mu \text{dist}^p(h_k, U^*) - (1 - 72\alpha_k^2 C^2)\|u_k - h_k\|^2 \\ &+2\alpha_k \langle e_k, u^* - h_k \rangle - \|u_k - h_{k-1}\|^2 \\ &+6\alpha_k^2(\|e_{k-1}\|^2 + \|e_k\|^2 + 24\|u^*\|^2 + 4D^2).\end{aligned}$$

Next, we add $7\alpha_{k+1}^2\|e_k\|^2$ to both sides of the preceding relation and, after slightly re-arranging the terms, we obtain for all $u^* \in U^*$ and for all $k \geq 1$,

$$\begin{aligned}\|u_{k+1} - u^*\|^2 + &\|u_{k+1} - h_k\|^2 + 7\alpha_{k+1}^2\|e_k\|^2 \\ \leq &(1 + 144\alpha_k^2 C^2)(\|u_k - u^*\|^2 + \|u_k - h_{k-1}\|^2) + 6\alpha_k^2\|e_{k-1}\|^2 \\ &- 2\alpha_k \mu \text{dist}^p(h_k, U^*) - (1 - 72\alpha_k^2 C^2)\|u_k - h_k\|^2 \\ &+ 2\alpha_k \langle e_k, u^* - h_k \rangle - \|u_k - h_{k-1}\|^2 \\ &+ 7\alpha_{k+1}^2\|e_k\|^2 + 6\alpha_k^2(\|e_k\|^2 + 24\|u^*\|^2 + 4D^2).\end{aligned} \tag{30}$$

We next consider the term $6\alpha_k^2\|e_{k-1}\|^2$ for which we write

$$6\alpha_k^2\|e_{k-1}\|^2 = 7\alpha_k^2\|e_{k-1}\|^2 - \alpha_k^2\|e_{k-1}\|^2 \leq 7(1 + 144\alpha_k^2 C^2)\alpha_k^2\|e_{k-1}\|^2 - \alpha_k^2\|e_{k-1}\|^2.$$

Upon substituting the preceding estimate back in (30) we have that for all $u^* \in U^*$ and for all $k \geq 1$,

$$\begin{aligned}\|u_{k+1} - u^*\|^2 + &\|u_{k+1} - h_k\|^2 + 7\alpha_{k+1}^2\|e_k\|^2 \\ \leq &(1 + 144\alpha_k^2 C^2)(\|u_k - u^*\|^2 + \|u_k - h_{k-1}\|^2 + 7\alpha_k^2\|e_{k-1}\|^2) \\ &- \alpha_k^2\|e_{k-1}\|^2 - 2\alpha_k \mu \text{dist}^p(h_k, U^*) - (1 - 72\alpha_k^2 C^2)\|u_k - h_k\|^2 \\ &+ 2\alpha_k \langle e_k, u^* - h_k \rangle - \|u_k - h_{k-1}\|^2 \\ &+ 7\alpha_{k+1}^2\|e_k\|^2 + 6\alpha_k^2(\|e_k\|^2 + 24\|u^*\|^2 + 4D^2).\end{aligned} \tag{31}$$

Since $\sum_{k=0}^{\infty} \alpha_k^2 < \infty$, it follows that $\alpha_k \to 0$, so there exists an index $N \geq 1$ such that the stepsize satisfies $1 - 72\alpha_k^2 C^2 \geq 1/2$ for all $k \geq N$. Thus, by defining

$$v_k = \|u_k - u^*\|^2 + \|u_k - h_{k-1}\|^2 + 6\alpha_k^2\|e_{k-1}\|^2 \qquad \text{for all } k \geq 1, \tag{32}$$

from relation (31) we obtain for all $u^* \in U^*$ and $k \geq N$,

$$\begin{aligned}v_{k+1} \leq &(1 + 144\alpha_k^2 C^2)v_k - \alpha_k^2\|e_{k-1}\|^2 - 2\alpha_k \mu \text{dist}^p(h_k, U^*) - \frac{1}{2}\|u_k - h_k\|^2 \\ &+2\alpha_k \langle e_k, u^* - h_k \rangle - \|u_k - h_{k-1}\|^2 + 7\alpha_{k+1}^2\|e_k\|^2 + 6\alpha_k^2(\|e_k\|^2 + 24\|u^*\|^2 + 4D^2).\end{aligned} \tag{33}$$

Recalling that $e_k = \Phi(h_k, \xi_k) - F(h_k)$ and using the stochastic properties of ξ_k imposed by Assumption 1, we have $\mathbb{E}[\langle e_k, h_k - u^* \rangle | \mathcal{F}_{k-1}] = 0$ and $\mathbb{E}[\|e_k\|^2 | \mathcal{F}_{k-1}] \leq \sigma^2$ for all $k \geq 1$. Thus, by taking the conditional expectation on $\mathcal{F}_{k-1} = \{\xi_0, \ldots, \xi_{k-1}\}$ in relation (33), we obtain for all $u^* \in U^*$ and for all $k \geq N$,

$$\begin{aligned}\mathbb{E}[v_{k+1} \mid \mathcal{F}_{k-1}] \leq &(1 + 144\alpha_k^2 C^2)v_k - \alpha_k^2\|e_{k-1}\|^2 - 2\alpha_k \mu \text{dist}^p(h_k, U^*) - \frac{1}{2}\|u_k - h_k\|^2 \\ &-\|u_k - h_{k-1}\|^2 + 7\alpha_{k+1}^2\sigma^2 + 6\alpha_k^2(\sigma^2 + 24\|u^*\|^2 + 4D^2).\end{aligned} \tag{34}$$

Notice that when $u^* \in U^*$ is a fixed solution, then $\|u^*\|$ is a constant.

Since $\sum_{k=0}^{\infty} \alpha_k^2 < \infty$, the inequality in (34) satisfies the conditions of Lemma 8 for all $k \geq N$, with

$$z_k = \alpha_k^2 \|e_{k-1}\|^2 + 2\alpha_k \mu \mathrm{dist}^p(h_k, U^*) + \frac{1}{2}\|u_k - h_k\|^2 + \|u_k - h_{k-1}\|^2,$$
$$a_k = 144\alpha_k^2 C^2, \qquad b_k = 7\alpha_{k+1}^2\sigma^2 + 6\alpha_k^2(\sigma^2 + 24\|u^*\|^2 + 4D^2).$$

By Lemma 8 (where we shift the indices to start with $k = N$), it follows that the sequence $\{v_k\}$ converges *a.s.* to a non-negative scalar for any $u^* \in U^*$, and almost surely we have

$$\sum_{k=N}^{\infty} \alpha_k^2 \|e_{k-1}\|^2 < \infty, \quad \sum_{k=N}^{\infty} \alpha_k \mathrm{dist}^p(h_k, U^*) < \infty, \quad \sum_{k=N}^{\infty} (\|u_k - h_k\|^2 + \|u_k - h_{k-1}\|^2) < \infty.$$

Thus, it follows that

$$\lim_{k\to\infty} \alpha_k^2 \|e_{k-1}\|^2 = 0 \qquad a.s. \tag{35}$$

$$\lim_{k\to\infty} \|u_k - h_k\| = 0 \qquad a.s. \tag{36}$$

$$\lim_{k\to\infty} \|u_k - h_{k-1}\| = 0 \qquad a.s. \tag{37}$$

Moreover, since $\sum_{k=0}^{\infty} \alpha_k = \infty$, it follows that

$$\liminf_{k\to\infty} \mathrm{dist}^p(h_k, U^*) = 0 \quad a.s.$$

Since the sequence $\{v_k\}$ converges *a.s.* for any given $u^* \in U^*$, in view of the definition of v_k in (32) combined with relations(35) and (37), it follows that the sequence $\{\|u_k - u^*\|^2\}$ converges *a.s.* for all $u^* \in U^*$. Thus, the sequence $\{u_k\}$ is bounded *a.s.* and, consequently, it has accumulation points *a.s.* In view of relation (36), the sequences $\{u_k\}$ and $\{h_k\}$ have the same accumulation points.

Now, let $\{k_i \mid i \geq 1\}$ be a (random) index sequence such that

$$\lim_{i\to\infty} \mathrm{dist}^p(h_{k_i}, U^*) = \liminf_{k\to\infty} \mathrm{dist}^p(h_k, U^*) = 0 \quad a.s. \tag{38}$$

Without loss of generality we may assume that $\{u_{k_i}\}$ is a convergent sequence (for otherwise we will select a convergent subsequence), and let $\bar{u}$ be its (random) limit point, i.e.,

$$\lim_{i\to\infty} \|u_{k_i} - \bar{u}\| = 0 \qquad a.s. \tag{39}$$

By relation (36), it follows that

$$\lim_{i\to\infty} \|h_{k_i} - \bar{u}\| = 0 \qquad a.s.$$

By continuity of the distance function $\mathrm{dist}(\cdot, U^*)$, from relation (38) we conclude that $\mathrm{dist}(\bar{u}, U^*) = 0$ *a.s.*, which implies that $\bar{u} \in U^*$ almost surely since the set U^* is closed. Since the sequence $\{\|u_k - u^*\|^2\}$ converges *a.s.* for any $u^* \in U^*$, it follows that $\{\|u_k - \bar{u}\|^2\}$ converges *a.s.*, and by relation (39) we conclude that $\lim_{k\to\infty} \|u_k - \bar{u}\|^2 = 0$.

(b) Taking the total expectation in (34), we obtain for all $u^* \in U^*$ and all $k \geq N$,

$$\begin{aligned}\mathbb{E}[v_{k+1}] \leq &(1+144\alpha_k^2 C^2)\mathbb{E}[v_k] - \alpha_k^2\mathbb{E}[\|e_{k-1}\|^2] - 2\alpha_k\mu\mathbb{E}[\text{dist}^p(h_k, U^*)] - \frac{1}{2}\mathbb{E}[\|u_k - h_k\|^2] \\ &- \mathbb{E}[\|u_k - h_{k-1}\|^2] + 7\alpha_{k+1}^2\sigma^2 + 6\alpha_k^2(\sigma^2 + 24\|u^*\|^2 + 4D^2).\end{aligned} \tag{40}$$

We can now apply Lemma 9 for $k \geq N$ (instead of $k \geq 0$), with

$$\bar{v}_k = \mathbb{E}[v_k], \quad \bar{z}_k = \alpha_k^2\mathbb{E}[\|e_{k-1}\|^2] + 2\alpha_k\mu\mathbb{E}[\text{dist}^p(h_k, U^*)] + \frac{1}{2}\mathbb{E}[\|u_k - h_k\|^2] + \mathbb{E}[\|u_k - h_{k-1}\|^2],$$

$$\bar{a}_k = 144\alpha_k^2 C^2, \qquad \bar{b}_k = 7\alpha_{k+1}^2\sigma^2 + 6\alpha_k^2(\sigma^2 + 24\|u^*\|^2 + 4D^2).$$

Since $\sum_{k=0}^{\infty}\alpha_k^2 < \infty$, the inequality (40) satisfies the conditions of Lemma 9 for all $k \geq N$. By Lemma 9 (where the indices are shifted to start with $k = N$ instead of $k = 0$), and the definitions of $\bar{v}_k$, v_k in (32), and $\bar{z}_k$, it follows that

$$\lim_{k\to\infty} \mathbb{E}[\|u_k - u^*\|^2 + \|u_k - h_{k-1}\|^2 + 6\alpha_k^2\|e_{k-1}\|^2] \quad \text{exist for every } u^* \in U^*, \tag{41}$$

$$\sum_{k=N}^{\infty}(\alpha_k^2\mathbb{E}[\|e_{k-1}\|^2] + 2\alpha_k\mu\mathbb{E}[\text{dist}^p(h_k, U^*)] + \frac{1}{2}\mathbb{E}[\|u_k - h_k\|^2] + \mathbb{E}[\|u_k - h_{k-1}\|^2]) < \infty.$$

Therefore, it follows that

$$\lim_{k\to\infty}(\alpha_k^2\mathbb{E}[\|e_{k-1}\|^2] + \mathbb{E}[\|u_k - h_{k-1}\|^2) = 0, \tag{42}$$

$$\lim_{k\to\infty}\mathbb{E}[\|u_k - h_k\|^2] = 0. \tag{43}$$

From relations (41) and (42) we conclude that

$$\lim_{k\to\infty}\mathbb{E}[\|u_k - u^*\|^2] \quad \text{exist for every } u^* \in U^*, \tag{44}$$

which implies that $\{\mathbb{E}[\|u_k - u^*\|^2]\}$ is bounded. Hence, for a fixed $u^* \in U^*$ and all $k \geq 0$,

$$\mathbb{E}[\|u_k\|^2] = \mathbb{E}[\|(u_k - u^*) + u^*\|^2] \leq \mathbb{E}[(\|u_k - u^*\| + \|u^*\|)^2 \leq 2\mathbb{E}[\|u_k - u^*\|^2] + 2\|u^*\|^2,$$

implying that the sequence $\{\mathbb{E}[\|u_k\|^2]\}$ is bounded. Moreover, we have that all $k \geq 0$,

$$\mathbb{E}[\|h_k\|^2] = \mathbb{E}[\|(h_k - u_k) + u_k\|^2] \leq \mathbb{E}[(\|h_k - u_k\| + \|u_k\|)^2 \leq 2\mathbb{E}[\|h_k - u_k\|^2] + 2\mathbb{E}[\|u_k\|^2],$$

thus implying that the sequence $\{\mathbb{E}[\|h_k\|^2]\}$ is bounded due to relation (43) and the boundedness of $\{\mathbb{E}[\|u_k\|^2]\}$.

(c) By part (a), we have that almost surely

$$\lim_{k\to\infty}\|u_k - \bar{u}\|^2 = 0, \qquad \lim_{k\to\infty}\|h_k - \bar{u}\|^2 = 0,$$

for some random solution $\bar{u} \in U^*$. When the set U^* is bounded, we further have that

$$\|u_k - \bar{u}\|^2 \leq (\|u_k\| + \|\bar{u}\|)^2 \leq 2\|u_k\|^2 + 2\|u^*\|^2 \leq 2\|u_k\|^2 + 2M_0^2,$$

where $M_0 = \max_{u^* \in U^*} \|u^*\|$. Similarly, we have

$$\|h_k - \bar{u}\|^2 \le 2\|h_k\|^2 + 2M_0^2.$$

We note that by part (b), the sequences $\{\mathbb{E}[\|u_k\|^2]\}$ and $\{\mathbb{E}[\|h_k\|^2]\}$ are bounded. By applying the Lebesgue Dominated Convergence Theorem, with $f_k = \|u_k - \bar{u}\|^2$ and $g = 2\|u_k\|^2 + 2M_0^2$, we conclude that

$$\lim_{k\to\infty} \mathbb{E}[\|u_k - \bar{u}\|^2] = 0.$$

Similarly, applying the Lebesgue Dominated Convergence Theorem, with $f_k = \|h_k - \bar{u}\|^2$ and $g = 2\|h_k\|^2 + 2M_0^2$, we obtain that

$$\lim_{k\to\infty} \mathbb{E}[\|h_k - \bar{u}\|^2] = 0.$$

Appendix C. Convergence Rates

C.1. Auxiliary Results

In our analysis we make use of Lemma 3 and Lemma 7 from [16], as well as the sequences provided in the proofs in [16].

Lemma 11 *Let $\{r_k\}$ and $\{s_k\}$ be nonnegative scalar sequences that satisfy the following relation*

$$r_{k+1} \le (1 - a\gamma_k)r_k - b\gamma_k s_k + c\gamma_k^2 \qquad \text{for all } k \ge 0,$$

where $a > 0$, $b > 0$, $c \ge 0$, and

$$\gamma_k = \frac{2}{a\left(\frac{2d}{a} + k\right)} \qquad \text{for all } k \ge 0,$$

where $d \ge a$. Then, for any given $K \ge 0$, the following relation holds:

$$\frac{b}{W_K}\sum_{k=0}^{K} w_k s_k + a r_{K+1} \le \frac{8d^2}{aK^2} r_0 + \frac{2c}{aK},$$

where $w_k = 2d/a + k$, $0 \le k \le K$, and $W_K = \sum_{k=0}^{K} w_k$.

Lemma 12 *Let $\{r_k\}$, $\{s_k\}$, and $\{\gamma_k\}$ be nonnegative scalar sequences that satisfy the following relation*

$$r_{k+1} \le (1 - a\gamma_k)r_k - b\gamma_k s_k + c\gamma_k^2 \qquad \text{for all } k \ge 0,$$

where $a > 0$, $b > 0$, $c \ge 0$, and $\gamma_k \le d^{-1}$ for some $d \ge a$ and for all $k \ge 0$. Then, for any given $K \ge 0$, we can choose the stapsizes γ_k and the weights $w_k \ge 0$, $0 \le k \le K$, such that the following relation holds:

$$\frac{b}{W_K}\sum_{k=0}^{K} w_k s_k + a r_{K+1} \le 32 d r_0 e^{-\frac{aK}{2d}} + \frac{36c}{aK},$$

where $W_K = \sum_{k=0}^{K} w_k$.

A specific choice of the stepsize and the weights for which the preceding lemma holds is as follows:

$$\begin{aligned} \gamma_k &= \tfrac{1}{d}, \qquad w_k = \left(1-\tfrac{a}{d}\right)^{-(k+1)} \qquad \text{if } K \le \tfrac{d}{a}, \\ \gamma_k &= \tfrac{1}{d}, \qquad w_k = 0 \qquad \text{if } K > \tfrac{d}{a} \ \text{ and } \ k < k_0, \\ \gamma_k &= \tfrac{2}{a\left(\frac{2d}{a}+k-k_0\right)}, \qquad w_k = \left(\tfrac{2d}{a}+k-k_0\right)^2 \qquad \text{if } K > \tfrac{d}{a} \ \text{ and } \ k \ge k_0, \end{aligned} \tag{45}$$

where $k_0 = \left\lceil \frac{K}{2} \right\rceil$.

C.2. Proof of Theorem 2

Proof By equation (29) in the proof of Theorem 1, the following relation holds for all $u^* \in U^*$ and for all $k \ge 1$,

$$\begin{aligned} \|u_{k+1}-u^*\|^2 + \|u_{k+1}-h_k\|^2 \le & (1+144\alpha_k^2C^2)\|u_k-u^*\|^2 - 2\alpha_k\mu \mathrm{dist}^2(h_k,U^*) \\ & -(1-72\alpha_k^2C^2)\|u_k-h_k\|^2 + 2\alpha_k\langle e_k, u^*-h_k\rangle \\ & +72\alpha_k^2C^2\|u_k-h_{k-1}\|^2 \\ & +6\alpha_k^2(\|e_{k-1}\|^2+\|e_k\|^2+24\|u^*\|^2+4D^2). \end{aligned} \tag{46}$$

The solution set U^* is closed, so there exists a projection u_k^* of the iterate u_k on the optimal set U^*, i.e., there is a point $u_k^* \in U^*$ such that $\|u_k - u_k^*\| = \mathrm{dist}(u_k, U^*)$. Thus, by letting $u^* = u_k^*$ in relation (46), and noting that $\mathrm{dist}(u_{k+1}, U^*) \le \|u_{k+1} - u_k^*\|$ we obtain for all $k \ge 1$,

$$\begin{aligned} \mathrm{dist}^2(u_{k+1},U^*) + \|u_{k+1}-h_k\|^2 \le & (1+144\alpha_k^2C^2)\mathrm{dist}^2(u_k,U^*) - 2\alpha_k\mu \mathrm{dist}^2(h_k,U^*) \\ & -(1-72\alpha_k^2C^2)\|u_k-h_k\|^2 + 2\alpha_k\langle e_k, u_k^*-h_k\rangle \\ & +72\alpha_k^2C^2\|u_k-h_{k-1}\|^2 \\ & +6\alpha_k^2(\|e_{k-1}\|^2+\|e_k\|^2+24\|u_k^*\|^2+4D^2). \end{aligned} \tag{47}$$

We next estimate the term $-2\mathrm{dist}^2(h_k, U^*)$ in (47) by using the relation shown in (**??**), i.e.,

$$-2\mathrm{dist}^2(h_k,U^*) \le 2\|u_k-h_k\|^2 - \mathrm{dist}^2(u_k,U^*).$$

By substituting the preceding estimate in relation (47), we obtain for all $k \ge 1$,

$$\begin{aligned} \mathrm{dist}^2(u_{k+1},U^*) + \|u_{k+1}-h_k\|^2 \le & (1+144\alpha_k^2C^2-\mu\alpha_k)\mathrm{dist}^2(u_k,U^*) \\ & +72\alpha_k^2C^2\|u_k-h_{k-1}\|^2 \\ & -(1-2\mu\alpha_k-72\alpha_k^2C^2)\|u_k-h_k\|^2 + 2\alpha_k\langle e_k, u_k^*-h_k\rangle \\ & +6\alpha_k^2(\|e_{k-1}\|^2+\|e_k\|^2) + 6\alpha_k^2(24\|u_k^*\|^2+4D^2). \end{aligned} \tag{48}$$

By Assumption 1, we have that $\mathbb{E}[\|e_k\|^2 \mid h_k] \le \sigma^2$ and $\mathbb{E}[e_k \mid h_k] = 0$ for all $k \ge 1$, implying that $\mathbb{E}[\|e_k\|^2] \le \sigma^2$ for all $k \ge 1$, and

$$\mathbb{E}[\langle e_k, h_k-u_k^*\rangle] = \mathbb{E}\left[\mathbb{E}[\langle e_k, h_k-u_k^*\rangle \mid h_k, u_k^*]\right] = 0 \qquad \text{for all } k \ge 1.$$

Therefore, by taking the expectation in relation (48) and using the assumption that the set U^* is bounded, we obtain for all $k \geq 1$,

$$\begin{aligned}\mathbb{E}[\text{dist}^2(u_{k+1}, U^*) + \|u_{k+1} - h_k\|^2] \leq &(1 + 144\alpha_k^2 C^2 - \mu\alpha_k)\mathbb{E}[\text{dist}^2(u_k, U^*)] \\ &+72\alpha_k^2 C^2 \mathbb{E}[\|h_{k-1} - u_k\|^2] \\ &-(1 - 2\mu\alpha_k - 72\alpha_k^2 C^2)\mathbb{E}[\|u_k - h_k\|^2] \\ &+12\alpha_k^2(\sigma^2 + 2D^2 + 12M_1^2),\end{aligned} \tag{49}$$

where $M_1 > 0$ is such that $\|u^*\| \leq M_1$ for all $u^* \in U^*$.

By the stepsize choice we have that $\alpha_k \leq d^{-1}$ with $d^{-1} \leq \frac{\mu}{288C^2}$ for all $k \geq 0$, implying that $144\alpha_k C^2 \leq \mu/2$, and consequently

$$1 + 144\alpha_k^2 C^2 - \mu\alpha_k \leq 1 + \frac{\mu}{2}\alpha_k - \mu\alpha_k = 1 - \frac{\mu}{2}\alpha_k \qquad \text{for all } k \geq 0.$$

Thus, it follows that

$$\begin{aligned}\mathbb{E}[\text{dist}^2(u_{k+1}, U^*) + \|u_{k+1} - h_k\|^2] \leq &\left(1 - \frac{\mu}{2}\alpha_k\right)\mathbb{E}[\text{dist}^2(u_k, U^*)] + 72\alpha_k^2 C^2 \mathbb{E}[\|h_{k-1} - u_k\|^2] \\ &-(1 - 2\mu\alpha_k - 72\alpha_k^2 C^2)\mathbb{E}[\|u_k - h_k\|^2] \\ &+12\alpha_k^2(\sigma^2 + 2D^2 + 12M_1^2).\end{aligned} \tag{50}$$

Since $\alpha_k \leq \frac{\mu}{288C^2}$ for all $k \geq 0$, we also have

$$72\alpha_k C^2 \leq \frac{\mu}{4} \quad \Longrightarrow \quad 72\alpha_k^2 C^2 \leq \frac{\mu}{4}\alpha_k.$$

Therefore

$$1 - 2\mu\alpha_k - 72\alpha_k^2 C^2 \geq 1 - 2\mu\alpha_k - \frac{\mu}{4}\alpha_k = 1 - \frac{9}{4}\mu\alpha_k \geq 0, \tag{51}$$

where the last inequality follows from the stepsize choice so that $\alpha_k \leq d^{-1}$, and our assumption that $d^{-1} \leq \frac{4}{9\mu}$ for all k. Moreover, from $1 - 2\mu\alpha_k - 72\alpha_k^2 C^2 \geq 0$, it follows that

$$72\alpha_k^2 C^2 \leq 1 - 2\mu\alpha_k < 1 - \frac{\mu}{2}\alpha_k \qquad \text{for all } k \geq 0. \tag{52}$$

By using the estimates (51) and (52) in relation (50) we obtain that for all $k \geq 1$,

$$\begin{aligned}\mathbb{E}[\text{dist}^2(u_{k+1}, U^*) + \|u_{k+1} - h_k\|^2] \leq &\left(1 - \frac{\mu}{2}\alpha_k\right)\left(\mathbb{E}[\text{dist}^2(u_k, U^*)] + \mathbb{E}[\|h_{k-1} - u_k\|^2]\right) \\ &+ 12\alpha_k^2(\sigma^2 + 2D^2 + 12M_1^2).\end{aligned} \tag{53}$$

The equation (53) satisfies the conditions of Lemma 12 with the following identification

$$r_k = \mathbb{E}[\text{dist}^2(u_{k+1}, U^*) + \|u_{k+1} - h_k\|^2], \qquad s_k = 0, \qquad \gamma_k = \alpha_k, \qquad a = \frac{\mu}{2},$$

$$d \geq \frac{1}{\min\{\frac{\mu}{288C^2}, \frac{4}{9\mu}\}}, \qquad c = 12(\sigma^2 + 2D^2 + 12M_1^2).$$

By applying Lemma 12 with a time shift to start with $k = 1$ instead of $k = 0$, we obtain that for all $K \geq 1$,

$$a\mathbb{E}[\text{dist}^2(u_{K+1}, U^*) + \|u_{K+1} - h_K\|^2] \leq 32d(\mathbb{E}[\text{dist}^2(u_1, U^*) + \|h_0 - u_1\|^2])e^{-\frac{a(K-1)}{2d}} + \frac{36c}{a(K-1)}. \tag{54}$$

Upon dividing by $a = \frac{\mu}{2}$ and omitting the term $\|u_{k+1} - h_k\|^2$, we arrive at

$$\mathbb{E}[\text{dist}^2(u_{K+1}, U^*)] \leq \frac{64d}{\mu}\left(\mathbb{E}[\text{dist}^2(u_1, U^*) + \|h_0 - u_1\|^2]\right) e^{-\frac{\mu(K-1)}{4d}} + \frac{144c}{\mu^2(K-1)}.$$

■

C.3. Proof of Theorem 3

Proof By Lemma 6 we have that surely for all $y \in U$ and $k \geq 1$,

$$\|u_{k+1} - y\|^2 \leq \|u_k - y\|^2 - \|u_{k+1} - h_k\|^2 - \|u_k - h_k\|^2 - 2\alpha_k\langle F(h_k), h_k - y\rangle - 2\alpha_k\langle e_k, h_k - y\rangle + 6\alpha_k^2\|F(h_{k-1}) - F(h_k)\|^2 + 6\alpha_k^2\left(\|e_{k-1}\|^2 + \|e_k\|^2\right). \tag{55}$$

Using the Lipschitz continuity of the operator $F(\cdot)$ we can bound the term $\|F(h_k) - F(h_{k-1})\|^2$, as follows

$$\begin{aligned}\|F(h_k) - F(h_{k-1})\|^2 &\leq L^2\|h_{k-1} - h_k\|^2 \\ &\leq L^2\left(\|h_{k-1} - u_k\| + \|u_k - h_k\|\right)^2 \\ &\leq 2L^2(\|h_{k-1} - u_k\|^2 + \|u_k - h_k\|^2),\end{aligned} \tag{56}$$

where the last inequality follows the inequality $(\sum_{i=1}^m a_i)^2 \leq m\sum_{i=1}^m a_i^2$, which is valid for any scalars a_i, $i = 1, \dots, m$, and any integer $m \geq 1$. Combining relations (55) and (56), and letting $y = u^* \in U^*$, we surely obtain for all $u^* \in U^*$ and $k \geq 1$,

$$\begin{aligned}\|u_{k+1} - u^*\|^2 + \|u_{k+1} - h_k\|^2 \leq& \|u_k - u^*\|^2 - (1 - 12\alpha_k^2L^2)\|u_k - h_k\|^2 \\ &-2\alpha_k\langle F(h_k), h_k - u^*\rangle - 2\alpha_k\langle e_k, h_k - u^*\rangle \\ &+12\alpha_k^2L^2\|h_{k-1} - u_k\|^2 + 6\alpha_k^2\left(\|e_{k-1}\|^2 + \|e_k\|^2\right).\end{aligned} \tag{57}$$

By the 2-quasi sharpness property of $F(\cdot)$ (Assumption 3, with $p = 2$), we have that $\langle F(h_k), h_k - u^*\rangle \geq \text{dist}^2(h_k, U^*)$, thus implying that

$$\begin{aligned}\|u_{k+1} - u^*\|^2 + \|u_{k+1} - h_k\|^2 \leq& \|u_k - u^*\|^2 - 2\alpha_k\mu\text{dist}^2(h_k, U^*) \\ &-(1 - 12\alpha_k^2L^2)\|u_k - h_k\|^2 - 2\alpha_k\langle e_k, h_k - u^*\rangle \\ &+12\alpha_k^2L^2\|h_{k-1} - u_k\|^2 + 6\alpha_k^2\left(\|e_{k-1}\|^2 + \|e_k\|^2\right).\end{aligned} \tag{58}$$

Using the relation (see (**??**))

$$-2\text{dist}^2(h_k, U^*) \leq 2\|u_k - h_k\|^2 - \text{dist}^2(u_k, U^*),$$

we obtain surely for all $u^* \in U^*$ and $k \geq 1$,

$$\begin{aligned}\|u_{k+1}-u^*\|^2+\|u_{k+1}-h_k\|^2 \leq &(1-\mu\alpha_k)\|u_k-u^*\|^2-(1-2\mu\alpha_k-12\alpha_k^2L^2)\|u_k-h_k\|^2\\ &-2\alpha_k\langle e_k, h_k-u^*\rangle+12\alpha_k^2L^2\|h_{k-1}-u_k\|^2\\ &+6\alpha_k^2\left(\|e_{k-1}\|^2+\|e_k\|^2\right).\end{aligned} \tag{59}$$

Since U^* is closed, there is a projection of u_k^* of the iterate u_k on the solution set U^* such that $\|u_k-u_k^*\| = \text{dist}(u_k, U^*)$. Thus, by letting $u^* = u_k^*$ and by noting that $\text{dist}(u_{k+1}, U^*) \leq \|u_{k+1}-u_k^*\|$, we can see that for all $k \geq 1$,

$$\begin{aligned}\text{dist}^2(u_{k+1}, U^*)+\|u_{k+1}-h_k\|^2| \leq &(1-\mu\alpha_k)\text{dist}^2(u_k, U^*)\\ &-(1-2\mu\alpha_k-12\alpha_k^2L^2)\|u_k-h_k\|^2\\ &-2\alpha_k\langle e_k, h_k-u_k^*\rangle+12\alpha_k^2L^2\|h_{k-1}-u_k\|^2\\ &+6\alpha_k^2\left(\|e_{k-1}\|^2+\|e_k\|^2\right).\end{aligned} \tag{60}$$

We next consider the coefficient $1-2\mu\alpha_k-12\alpha_k^2L^2$. We note that the polynomial $p(s) = 1 - 2\mu s - 12L^2s^2$, $s \in \mathbb{R}$, has two real roots $s_1/2 = \frac{\mu \pm \sqrt{\mu^2+12L^2}}{12L^2}$. Thus, since the stepsize is selecetd so that $0 < \alpha_k \leq \frac{1}{2\sqrt{3}L}$ for all k and since we have

$$\frac{1}{2\sqrt{3}L} = \frac{\sqrt{12L^2}}{12L^2} \leq \frac{\mu+\sqrt{\mu^2+12L^2}}{12L^2},$$

it follows that the stepsize α_k satisfies

$$1-2\mu\alpha_k-12\alpha_k^2L^2 \geq 0 \qquad \text{for all } k \geq 0.$$

Subsequently, we have that $12\alpha_k^2L^2 \leq 1-2\mu\alpha_k < 1-\mu\alpha_k$ for all k. Thus, from (60) we obtain surely for all $k \geq 1$,

$$\begin{aligned}\text{dist}^2(u_{k+1}, U^*)+\|u_{k+1}-h_k\|^2| \leq &(1-\mu\alpha_k)\left(\text{dist}^2(u_k, U^*)+\|h_{k-1}-u_k\|^2\right)\\ &-2\alpha_k\langle e_k, h_k-u_k^*\rangle+6\alpha_k^2\left(\|e_{k-1}\|^2+\|e_k\|^2\right).\end{aligned} \tag{61}$$

By Assumption 1, we have that $\mathbb{E}[\|e_k\|^2 \mid h_k] \leq \sigma^2$ and $\mathbb{E}[e_k \mid h_k] = 0$ for all $k \geq 1$. Therefore, $\mathbb{E}[\|e_k\|^2] \leq \sigma^2$ for all $k \geq 1$, and

$$\mathbb{E}[\langle e_k, h_k-u_k^*\rangle] = \mathbb{E}\left[\mathbb{E}[\langle e_k, h_k-u_k^*\rangle \mid h_k, u_k^*]\right] = 0 \qquad \text{for all } k \geq 1.$$

Hence, by taking the expectation in relation (61) we obtain for all $k \geq 1$,

$$\begin{aligned}\mathbb{E}\left[\text{dist}^2(u_{k+1}, U^*)+\|u_{k+1}-h_k\|^2\right] \leq &(1-\mu\alpha_k)\mathbb{E}\left[\text{dist}^2(u_k, U^*)+\|h_{k-1}-u_k\|^2\right]\\ &+12\alpha_k^2\sigma^2.\end{aligned} \tag{62}$$

Relation (62) satisfies the conditions of Lemma 12 with the following identification

$$r_k = \mathbb{E}[\text{dist}^2(u_k, U^*)+\|u_k-h_{k-1}\|^2], \qquad s_k = 0, \qquad \gamma_k = \alpha_k, \qquad a = \mu,$$

$$d \geq \max\left\{2\sqrt{3}L,\, \mu\right\}, \qquad c = 12\sigma^2.$$

Thus, by using Lemma 12 with a time shift to start with $k = 1$ instead of $k = 0$, we obtain for all $K \geq 1$,

$$\begin{aligned}\mu\mathbb{E}[\text{dist}^2(u_{K+1}, U^*) + \|u_{K+1} - h_K\|^2] \leq & 32d(\mathbb{E}[\text{dist}^2(u_1, U^*) + \|h_0 - u_1\|^2])e^{-\frac{\mu(K-1)}{2d}} \\ & + \frac{36c}{\mu(K-1)}.\end{aligned} \tag{63}$$

Upon dividing by μ and substituting $c = 12\sigma^2$, we find that for all $K \geq 1$,

$$\begin{aligned}\mathbb{E}\left[\text{dist}^2(u_{K+1}, U^*) + \|u_{K+1} - h_K\|^2\right] \leq & \frac{32d}{\mu}\left(\mathbb{E}[\text{dist}^2(u_1, U^*) + \|h_0 - u_1\|^2]\right)e^{-\frac{\mu(K-1)}{2d}} \\ & + \frac{432\sigma^2}{\mu^2(K-1)},\end{aligned} \tag{64}$$

which implies the stated relation. ■

C.4. Proof of Theorem 4

Proof By Lemma 6 we surely have for all $y \in U$ and all $k \geq 1$,

$$\begin{aligned}\|u_{k+1} - y\|^2 \leq & \|u_k - y\|^2 - \|u_{k+1} - h_k\|^2 - \|u_k - h_k\|^2 - 2\alpha_k\langle e_k + F(h_k), h_k - y\rangle \\ & + 6\alpha_k^2\,\|F(h_k) - F(h_{k-1})\|^2 + 6\alpha_k^2(\|e_k\| + \|e_{k-1}\|^2).\end{aligned}$$

Since the set U is compact and the operator $F(\cdot)$ is continuous, it follows by Corollary 2.2.5 in [5] that the solution set U^* of the SVI(U, F) is a nonempty and compact set. Therefore, by letting $u = u^*$ with $u^* \in U^*$ in the preceding relation, we surely obtain for all $u^* \in U^*$ and all $k \geq 1$,

$$\begin{aligned}\|u_{k+1} - u^*\|^2 \leq & \|u_k - u^*\|^2 - \|u_{k+1} - h_k\|^2 - \|u_k - h_k\|^2 - 2\alpha_k\langle e_k + F(h_k), h_k - u^*\rangle \\ & + 6\alpha_k^2\,\|F(h_k) - F(h_{k-1})\|^2 + 6\alpha_k^2(\|e_k\| + \|e_{k-1}\|^2).\end{aligned}$$

The set U is compact and the operator $F(\cdot)$ is continuous, so there is a constant $D > 0$ such that $\|F(u)\| \leq D$ for all $u \in U$. Moreover, we have that $\{h_k\} \subset U$, implying that surely for all $k \geq 1$,

$$\|F(h_k) - F(h_{k-1})\|^2 \leq (\|F(h_k)\| + \|F(h_{k-1})\|)^2 \leq 4D^2.$$

By combining the preceding two relations, and using the p-quasi sharpness property of the operator, we obtain that surely for all $u^* \in U^*$ and all $k \geq 1$,

$$\begin{aligned}\|u_{k+1} - u^*\|^2 \leq & \|u_k - u^*\|^2 - \|u_{k+1} - h_k\|^2 - \|u_k - h_k\|^2 - 2\alpha_k\mu\,\text{dist}^p(h_k, U^*) \\ & -2\alpha_k\langle e_k, h_k - u^*\rangle + 24\alpha_k^2\,D^2 + 6\alpha_k^2(\|e_k\| + \|e_{k-1}\|^2).\end{aligned} \tag{65}$$

Next, we estimate $\text{dist}(h_k, U^*)$. Since U is a compact set, there is an $M_U > 0$ such that

$$\|u - u'\| \leq M_U \qquad \text{for all } u, u' \in U.$$

Therefore, for any $u^* \in U^* \subseteq U$, we have

$$\mathrm{dist}(h_k, U^*) \le \|h_k - u^*\| \le M_U,$$

which implies that

$$\mathrm{dist}^2(h_k, U^*) \le M_U^{2-p}\mathrm{dist}^p(h_k, U^*) \qquad \Longrightarrow \qquad -\mathrm{dist}^p(h_k, U^*) \le -\frac{1}{M_U^{2-p}}\mathrm{dist}^2(h_k, U^*). \tag{66}$$

Moreover, by using the following relation (see (**??**))

$$-2\mathrm{dist}^2(h_k, U^*) \le 2\|u_k - h_k\|^2 - \mathrm{dist}^2(u_k, U^*),$$

from (66) we obtain that

$$-2\mathrm{dist}(h_k, U^*) \le \frac{1}{M_U^{2-p}}\left(2\|u_k - h_k\|^2 - \mathrm{dist}^2(u_k, U^*)\right). \tag{67}$$

Bu using (67) in relation (65) we surely obtain for all $u^* \in U^*$ and $k \ge 1$,

$$\begin{aligned}\|u_{k+1} - u^*\|^2 \le & \|u_k - u^*\|^2 - \|u_{k+1} - h_k\|^2 - \|u_k - h_k\|^2 \\ & + \frac{\alpha_k \mu}{M_U^{2-p}}\left(2\|u_k - h_k\|^2 - \mathrm{dist}^2(u_k, U^*)\right) \\ & - 2\alpha_k\langle e_k, h_k - u^*\rangle + 24\alpha_k^2 D^2 + 6\alpha_k^2(\|e_k\| + \|e_{k-1}\|^2).\end{aligned} \tag{68}$$

Next, we let $u^* = u_k^*$, where u_k^* is a projection of u_k on the solution set U^*, which exists since U^* is a closed set. We also use $\|u_k - u_k^*\| = \mathrm{dist}(u_k, U^*)$ and $\mathrm{dist}(u_{k+1}, U^*) \le \|u_{k+1} - u_k^*\|$ and, thus, after re-arranging the terms in (68) we obtain that surely for all $k \ge 1$,

$$\begin{aligned}\mathrm{dist}^2(u_{k+1}, U^*) + \|u_{k+1} - h_k\|^2 \le & \left(1 - \frac{\alpha_k \mu}{M_U^{2-p}}\right)\mathrm{dist}^2(u_k, U^*) - \left(1 - \frac{2\alpha_k \mu}{M_U^{2-p}}\right)\|u_k - h_k\|^2 \\ & - 2\alpha_k\langle e_k, h_k - u^*\rangle + 24\alpha_k^2 D^2 + 6\alpha_k^2(\|e_k\| + \|e_{k-1}\|^2).\end{aligned} \tag{69}$$

By Assumption 1, we have that $\mathbb{E}[\|e_k\|^2 \mid h_k] \le \sigma^2$ and $\mathbb{E}[e_k \mid h_k] = 0$ for all $k \ge 1$, implying that $\mathbb{E}[\|e_k\|^2] \le \sigma^2$ for all $k \ge 1$,and

$$\mathbb{E}[\langle e_k, h_k - u_k^*\rangle] = \mathbb{E}\left[\mathbb{E}[\langle e_k, h_k - u_k^*\rangle \mid h_k, u_k^*]\right] = 0 \qquad \text{for all } k \ge 1.$$

Therefore, by taking the expectation in relation (69) (and omitting the term $\|u_{k+1} - h_k\|^2$), we obtain for all $k \ge 1$,

$$\begin{aligned}\mathbb{E}[\mathrm{dist}^2(u_{k+1}, U^*)] \le & \left(1 - \frac{\alpha_k \mu}{M_U^{2-p}}\right)\mathbb{E}[\mathrm{dist}^2(u_k, U^*)] - \left(1 - \frac{2\alpha_k \mu}{M_U^{2-p}}\right)\mathbb{E}[\|u_k - h_k\|^2] \\ & + 12\alpha_k^2(\sigma^2 + 2D^2).\end{aligned} \tag{70}$$

We now consider the two stepsize choices in parts (a) and (b) separately.

(a) Since the stepsize is given by $\alpha_k = \frac{2M_U^{2-p}}{\mu(3+k)}$ for all $k \geq 0$, it follows that $\alpha_k \leq \frac{M_U^{2-p}}{2\mu}$ for all $k \geq 1$. Hence, $1 - 2\mu\alpha_k/M_U^{2-p} \geq 0$, implying that

$$\mathbb{E}[\text{dist}^2(u_{k+1}, U^*)] \leq \left(1 - \frac{\alpha_k \mu}{M_U^{2-p}}\right) \mathbb{E}[\text{dist}^2(u_k, U^*)] + 12\alpha_k^2(\sigma^2 + 2D^2). \tag{71}$$

The equation (71) satisfies the conditions of Lemma 11, where the recursive relation is starting with $k = 1$ and with the following identification

$$r_k = \mathbb{E}[\text{dist}^2(u_k, U^*)], \quad s_k = 0, \quad \gamma_k = \alpha_k, \quad a = \frac{\mu}{M_U^{2-p}}, \quad d = \frac{2\mu}{M_U^{2-p}}, \quad c = 12(\sigma^2 + 2D^2). \tag{72}$$

By applying Lemma 11, with a time shift to start with $k = 1$ instead of $k = 0$, we find that for all $K \geq 1$,

$$a\mathbb{E}[\text{dist}^2(u_{K+1}, U^*)] \leq \frac{8d^2}{a(K-1)^2}(\mathbb{E}[\text{dist}^2(u_1, U^*)]) + \frac{2c}{a(K-1)}.$$

Upon dividing by $a = \frac{\mu}{M_2}$, and substituting the corresponding values for d and c, we obtain

$$\mathbb{E}[\text{dist}^2(u_{K+1}, U^*)] \leq \frac{32}{(K-1)^2}\mathbb{E}[\text{dist}^2(u_1, U^*)] + \frac{24(\sigma^2 + 2D^2)M_U^2}{\mu^2(K-1)}.$$

(b) The stepsizes given by relations in (45), with $a = \frac{\mu}{M_U^{2-p}}$ and $d = \frac{2\mu}{M_U^{2-p}}$, satisfies $\alpha_k \leq d^{-1}$ for all $k = 0, 1, \ldots, K-1$, for any $K \geq 1$. Hence, we have $1 - 2\mu\alpha_k/M_U^{2-p} \geq 0$, implying that relation (71) is valid for all $k \geq 1$. Thus, Lemma 12 applies with the same identification as in (72). Thus, by applying Lemma 12, with a time shift to start with $k = 1$ instead of $k = 0$, we obtain the following result for all $K \geq 1$,

$$a\mathbb{E}[\text{dist}^2(u_{K+1}, U^*)] \leq 32d\mathbb{E}[\text{dist}^2(u_1, U^*)]e^{\frac{a(K-1)}{2d}} + \frac{36c}{a(K-1)},$$

and after dividing by $a = \frac{\mu}{M_U^{2-p}}$, we obtain

$$\mathbb{E}[\text{dist}^2(u_{K+1}, U^*)] \leq 64\mathbb{E}[\text{dist}^2(u_1, U^*)]e^{-\frac{(K-1)}{4}} + \frac{432(\sigma^2 + 2D^2)M_U^{2(2-p)}}{\mu^2(K-1)}.$$

■

Appendix D. Experiment Details

Firstly, we verify that the operator $F(\cdot)$ defined in (**??**) is Lipschitz continuous, so it satisfies Assumption 2. For this, we define the matrix J, as follows

$$J = \begin{bmatrix} A & B \\ -B' & C \end{bmatrix}.$$

Then, for all $u, v \in U$ we have that

$$\|F(u) - F(v)\| = \|J(u - v)\| = \sqrt{\langle J'J(u - v), u - v\rangle},$$

where

$$J'J = \begin{bmatrix} A' & -B \\ B' & C' \end{bmatrix} \begin{bmatrix} A & B \\ -B' & C \end{bmatrix} \begin{bmatrix} A'A + BB' & A'B - BC \\ B'A - C'B' & B'B + C'C \end{bmatrix}.$$

Therefore, the Lipschitz constant L for the operator $F(\cdot)$ is the square root of the largest eigenvalue of the matrix $J'J$, i.e., $L = \sqrt{\lambda_{\max}(J'J)}$.

Now, we show that the operator $F(\cdot)$ satisfies Assumption 3 with $p = 2$. In particular, we show that the operator $F(\cdot)$ is strongly monotone and identify the unique solution to the VI(U, F), which implies that Assumption 3 holds with $p = 2$. To show that $F(\cdot)$ is strongly monotone, we let $u, v \in \mathbb{R}^{m+s}$ be arbitrary, and note that we have

$$\langle F(u) - F(v), u - v\rangle = \langle J(u - v), u - v\rangle$$

Let $z = u - v \in \mathbb{R}^{m+s}$, and note that $z = [z_1, z_2]'$ with $z_1 \in \mathbb{R}^m$ and $z_2 \in \mathbb{R}^s$. Thus, we have

$$\begin{aligned} \langle z, J'z\rangle &= [z_1, z_2] \begin{bmatrix} A' & -B \\ B' & C' \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \\ &= \langle z_1, Az_1\rangle + \langle z_2, B'z_1\rangle - \langle z_1, Bz_2\rangle + \langle z_2, Cz_2\rangle \\ &= \langle z_1, Az_1\rangle + \langle z_2, Cz_2\rangle \\ &\geq \mu_A\langle z_1, z_1\rangle + \mu_C\langle z_2, z_2\rangle \\ &\geq \min\{\mu_A, \mu_C\}\|z\|^2, \end{aligned}$$

where the first inequality in the preceding relation holds since A and C are symmetric positive definite matrices. Hence, the operator $F(\cdot)$ is strongly monotone with the constant $\mu = \min\{\mu_A, \mu_C\}$.

D.1. Experiments

We consider the following finite-sum min-max game with quadratic pay-off function as in [12]:

$$\min_{u_1 \in \mathbb{R}^m} \max_{u_2 \in \mathbb{R}^s} \frac{1}{n} \sum_{i=1}^{n} f_i(u_1, u_2),$$

where for each $i = 1, \ldots, n$, the function $f_i(\cdot)$ is given by

$$f_i(u_1, u_2) = \langle u_1, A_i u_1\rangle + \langle u_1, B_i u_2\rangle - \langle u_2, C_i u_2\rangle + \langle \mathbf{a}_i, u_1\rangle - \langle \mathbf{c}_i, u_2\rangle.$$

To formulate the preceding min-max problem as a finite-sum VI problem, we define $u = [u_1, u_2]'$ and the operator for every $i = 1, \ldots, n$

$$F_i(u) = \begin{bmatrix} A_i & B_i \\ -B_i' & C_i \end{bmatrix} u + \begin{bmatrix} \mathbf{a}_i \\ \mathbf{c}_i \end{bmatrix} \quad \text{for all } u \in \mathbb{R}^{m+s}, \tag{73}$$

and we let

$$F(u) = \frac{1}{n} \sum_{i=1}^{n} F_i(u). \tag{74}$$

In this notation, the corresponding VI(U, F) for the min-max problem consists of determining a point $u^* \in \mathbb{R}^{m+s}$ such that

$$\langle F(u^*), u - u^* \rangle \geq 0 \text{ for all } u \in U, \text{ with } U = \mathbb{R}^{m+s}. \tag{75}$$

We view the preceding VI(U, F) problem as a stochastic VI where

$$F(u) = \mathbb{E}[\Phi(u, \xi)],$$

where ξ is a uniform random variable taking values in the set $\{1, 2, \ldots, n\}$, with

$$\Phi(u, i) = f_i(u) \qquad \text{when } \xi = i \text{ for } i \in \{1, 2, \ldots, n\}.$$

Since the constraint set is $U = \mathbb{R}^{m+s}$, the SVI(U, F) in (75) reduces to the problem of determining a point $u^* \in \mathbb{R}^{m+s}$ such that $F(u^*) = 0$, i.e. $\mathbb{E}[\Phi(u^*, \xi)] = 0$.

In our experiments, the number n of random realizations of the uniform random variable ξ is $n = 20$. For every $i = 1, \ldots, n$, we generate positive definite symmetric matrices A_i and C_i with smallest eigenvalues $\mu_A > 0$ and $\mu_C > 0$, respectively. For all $i = 1, \ldots, n$, to generate symmetric matrix A_i, firstly, we generate eigenvalues uniformly from on $[\mu_A, L_A]$ such that μ_A, L_A are always generated. Then we generate a square random matrix $S_i \in \mathbb{R}^m$, do QR decomposition $S_i = Q_i R_i$, and get matrix A_i as $A_i = Q_i \Lambda_i Q_i'$, where Λ_i is a diagonal matrix with generated eigenvalues. We follow the same generation process for C_i matrices. For every $i = 1, \ldots, n$, the matrix B_i and vectors $\mathbf{a_i}, \mathbf{c_i}$ are sampled from a zero mean normal distribution with variances $\sigma_B^2 = 1/(m+s)^2$, $\sigma_{bias}^2 = 1/(m+s)$, respectively. We set threshold for both methods $k_0 = 200$, and the same values for four parameters μ_A, μ_C, L_A, L_C as in Section **??**. The results are presented in Figure 2.

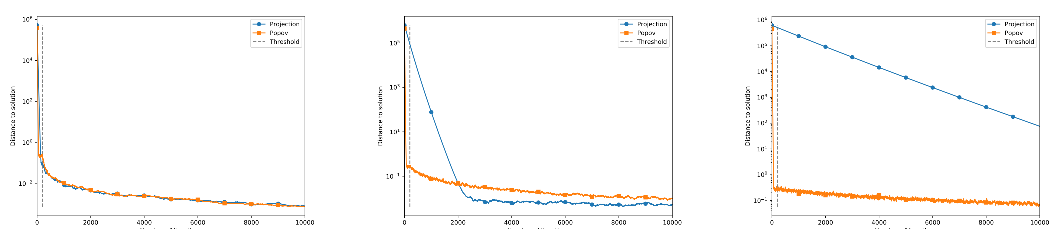

Figure 2: Comparison of stochastic Popov method and stochastic projection method with stepsize rule given in (5) and $k_0 = 200$ for a finite-sum VI.