



# Convergence analysis of sample average approximation for a class of stochastic nonlinear complementarity problems: from two-stage to multistage

Jie Jiang<sup>1</sup> · Hailin Sun<sup>2</sup> · Bin Zhou<sup>2</sup>

Received: 7 October 2020 / Accepted: 30 March 2021 / Published online: 27 April 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

## Abstract

In this paper, we consider the sample average approximation (SAA) approach for a class of stochastic nonlinear complementarity problems (SNCPs) and study the corresponding convergence properties. We first investigate the convergence of the SAA counterparts of two-stage SNCPs when the first-stage problem is continuously differentiable and the second-stage problem is locally Lipschitz continuous. After that, we extend the convergence results to a class of multistage SNCPs whose decision variable of each stage is influenced only by the decision variables of adjacent stages. Finally, some preliminary numerical tests are presented to illustrate the convergence results.

**Keywords** Two-stage · Multistage · Stochastic complementarity problems · Sample average approximation · Convergence analysis

**Mathematics Subject Classification (2010)** 90C15 · 90C33

---

✉ Hailin Sun  
hlsun@njnu.edu.cn

Jie Jiang  
jiangjiecq@163.com

Bin Zhou  
318801262@qq.com

<sup>1</sup> College of Mathematics and Statistics, Chongqing University, Chongqing, 401331, China

<sup>2</sup> Jiangsu Key Laboratory for NSLSCS, School of Mathematical Sciences, Nanjing Normal University, Nanjing, 210023, China

## 1 Introduction

Uncertainty is an important ingredient in modern optimization problems and practical applications [1, 24, 28]. The stochastic complementarity problem (SCP) is to consider the complementarity problem under uncertainty, which can be derived from the Karush-Kuhn-Tucker condition for a large class of stochastic programming problems and/or the equilibrium of stochastic multi-player games.

According to the structure (or levels) of uncertainty, SCPs can be roughly divided into single-stage SCPs, two-stage SCPs, and multistage SCPs. Single-stage and two-stage SCPs use random parameters to describe the uncertainty. Multistage SCPs model the uncertainty by stochastic processes. Since the multistage SCP reveals uncertainty more elaborately, it is expected to obtain better solutions compared with the single-stage or two-stage case, especially in a relatively long decision-making period. Of course, it is also challenging in both theoretical analysis and numerical algorithms because of the complexity of uncertainty.

Deterministic variational inequalities and complementarity problems have been extensively discussed; see monographs [11, 12] and the references therein for more details. In the last decade, stochastic variational inequalities (SVIs) have received more and more attention [19, 25]. In [2], by rewriting the single-stage stochastic linear complementarity problem through a nonlinear complementarity problem (NCP) function, the quasi-Monte Carlo method was adopted to solve the reformulated problem. Xu employed in [30] the SAA approach to analyze the discretization of single-stage SVIs. In [7], Chen et al. adopted the expected residual minimization (ERM) approach to formulate the single-stage SVI and a smoothing SAA approach was proposed to solve the ERM formulation. In [3], Chen et al. extended the ERM procedure to general two-stage nonlinear SVIs. In [23], Rockafellar and Wets first proposed the formulation of multistage SVIs and lay the foundation for the further investigation of two-stage or multistage SVIs. Next to [23], Rockafellar and Sun adopted in [22] the progressive hedging method (PHM) to solve the multistage SVIs when the support set is discrete. Inspired by [3, 20, 22, 23], Chen et al. investigated in [6] a discretization scheme for two-stage stochastic linear complementarity problems. In [4], Chen et al. considered the SAA approach for two-stage stochastic generalized equations and its application in two-stage nonlinear SVI-SCP problems. In [17], Jiang et al. adopted the regularized SAA approach to solve a class of two-stage SVIs and applied them to the crude oil market. In [5], Chen et al. modeled crude oil market share under the COVID-19 pandemic using two-stage stochastic equilibrium. In [16], Jiang et al. discussed the quantitative stability as well as the rate of convergence for a class of two-stage linear SVI-SCP problems. For a comprehensive review of recent development in two-stage and multistage SVIs, we refer to [29].

The SAA approach is an important approximation method for single-stage and two-stage stochastic optimization problems [28, 30]. Compared with the single-stage or two-stage case, the convergence analysis results of the SAA of multistage cases are quite limited. The existing results concentrate on multistage stochastic programming problems; see [21, 26, 27]. To the best of our knowledge, there are few results about the SAA approach and corresponding convergence analysis for multistage SCPs. According to these existing results, there are mainly two possible approaches to

sample from a random process. The first one is to view the random process as a whole and sample from its (joint) distribution; the other is the conditional sampling. In this paper, from the viewpoint of practical calculation, we adopt the conditional sampling method and focus on the SAA convergence analysis of multistage SCPs.

There are mainly two contributions of this paper.

- We establish the strong monotonicity assertion for a class of multistage SNCPs, which extends existing results from the two-stage case to the multistage case. Specifically, we weaken the continuously differentiable condition of strong monotonicity assertion in [4, Theorem 3.9] for two-stage nonlinear SVI-SCP problems to the locally Lipschitz continuity (Theorem 2), and then extend it to the multistage case (Theorem 3).
- Under the strong monotonicity assumption, we consider the conditional sampling method for a three-stage SNCP. The asymptotic convergence assertion is derived (Theorem 4). The convergence analysis of the SAA approach for  $T$ -stage SNCPs ( $T > 3$ ) can be derived similarly.

The rest of this paper is organized as follows. In Section 2, we introduce models and some useful properties of multistage SNCPs. In Section 3, we investigate the convergence analysis results of the SAA approach to multistage SNCPs. In Section 4, some preliminary numerical results are given by PHM. Finally, we conclude the whole paper in Section 5.

## 2 Models and properties of multistage SNCPs

In this section, we first introduce models of multistage SNCPs. After that, some useful properties are given for the further discussion.

### 2.1 Models of multistage SNCPs

To describe a  $T$ -stage SNCP model, we need further the following statement. Denote by  $\xi_1, \xi_2, \dots, \xi_T$  the discrete stochastic process with time horizon  $2 \leq T \in \mathbb{N}$ . For simplicity of statement and without loss of generality, we assume that  $\xi_t, t = 1, 2, \dots, T$  are all  $\mathbb{R}^s$ -valued. Furthermore, we assume that  $\xi_1$  is deterministic and random vectors  $\xi_t, t = 2, \dots, T$  are defined on probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , i.e.,  $\xi_t : \Omega \rightarrow \mathcal{E}_t$  with  $\mathcal{E}_t$  being the support set of  $\xi_t$ . Denote  $\xi_{[t]} = (\xi_1, \dots, \xi_t)$  supported on  $\mathcal{E}_{[t]} \subseteq \mathcal{E}_1 \times \dots \times \mathcal{E}_t$ . Then we use  $\mathcal{F}_t$  to denote the  $\sigma$ -field induced by  $\xi_{[t]}$ , i.e.,  $\mathcal{F}_t := \sigma(\xi_{[t]})$ . Thus,  $\xi_{[t]}$  is a random vector on the probability space  $(\Omega, \mathcal{F}_t, \mathbb{P})$ . Specially, without the confusion, we neglect the subscript when  $t = T$ , namely,  $\xi = \xi_{[T]}$ ,  $\mathcal{E} = \mathcal{E}_{[T]}$  and  $\mathcal{F} = \mathcal{F}_T$ . Moreover, we have the inclusion that  $\{\emptyset, \Omega\} = \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots \subseteq \mathcal{F}_T = \mathcal{F}$ , which forms the so-called filtration. Analogously, to simplify the description, we again assume that decision variables  $x_t, t = 1, 2, \dots, T$  are all  $\mathbb{R}^n$ -valued. Of course, the  $t$ th stage decision variable  $x_t$  is assumed to depend only on the available information until time  $t$ , i.e.,  $\xi_{[t]}$ . Or, equivalently,  $x_t = x_t(\xi_{[t]})$ . This is the well-known nonanticipativity condition, which is an essential property in multistage stochastic optimization.

With the above notation, the general  $T$ -stage SNCP model can be written as:

$$\left\{ \begin{array}{l} 0 \in \mathbb{E}_{\xi|\xi_1}[\Phi_1(x_1, x_2(\xi_{[2]}), \dots, x_T(\xi_{[T]}), \xi)] + \mathcal{N}_{\mathbb{R}_+^n}(x_1), \\ 0 \in \mathbb{E}_{\xi|\xi_{[2]}}[\Phi_2(x_1, x_2(\xi_{[2]}), \dots, x_T(\xi_{[T]}), \xi)] + \mathcal{N}_{\mathbb{R}_+^n}(x_2(\xi_{[2]})), \text{ a.e. } \xi_{[2]} \in \mathcal{E}_{[2]}, \\ \vdots \\ 0 \in \mathbb{E}_{\xi|\xi_{[T-1]}}[\Phi_{T-1}(x_1, x_2(\xi_{[2]}), \dots, x_T(\xi_{[T]}), \xi)] + \mathcal{N}_{\mathbb{R}_+^n}(x_{T-1}(\xi_{[T-1]})), \\ \text{a.e. } \xi_{[T-1]} \in \mathcal{E}_{[T-1]}, \\ 0 \in \Phi_T(x_1, x_2(\xi_{[2]}), \dots, x_T(\xi_{[T]}), \xi) + \mathcal{N}_{\mathbb{R}_+^n}(x_T(\xi_{[T]})), \text{ a.e. } \xi_{[T]} \in \mathcal{E}_{[T]}, \end{array} \right. \tag{1}$$

where  $\mathbb{E}_{\xi|\xi_{[t]}}[\cdot]$  means the conditional expectation of  $\xi$  for giving  $\xi_{[t]}$  for  $t = 1, \dots, T - 1$ ;  $\mathcal{N}_{\mathbb{R}_+^n}(x_1)$  is the normal cone of  $\mathbb{R}_+^n$  at  $x_1$  and so on in a similar fashion;  $\Phi_t : \mathbb{R}^{nT} \times \mathcal{E} \rightarrow \mathbb{R}^n$  for  $t = 1, \dots, T$ . Sometimes, for simplification, we also write  $x_2 = x_2(\xi_{[2]}), \dots, x_T = x_T(\xi_{[T]})$  in what follows.

Strategies of multistage SNCP (1) are a series of tuples  $(x_1, x_2(\cdot), \dots, x_T(\cdot))$  satisfying (1). Another important concept for two-stage or multistage SNCPs is the so-called relatively complete recourse condition. We say the multistage SNCP (1) satisfies the relatively complete recourse condition, if

$$\left\{ \begin{array}{l} 0 \in \mathbb{E}_{\xi|\xi_{[t+1]}}[\Phi_{t+1}(x_1, x_2, \dots, x_T, \xi)] + \mathcal{N}_{\mathbb{R}_+^n}(x_{t+1}), \text{ a.e. } \xi_{[t+1]} \in \mathcal{E}_{[t+1]}, \\ \vdots \\ 0 \in \mathbb{E}_{\xi|\xi_{[T-1]}}[\Phi_{T-1}(x_1, x_2, \dots, x_T, \xi)] + \mathcal{N}_{\mathbb{R}_+^n}(x_{T-1}), \text{ a.e. } \xi_{[T-1]} \in \mathcal{E}_{[T-1]}, \\ 0 \in \Phi_T(x_1, x_2, \dots, x_T, \xi) + \mathcal{N}_{\mathbb{R}_+^n}(x_T), \text{ a.e. } \xi_{[T]} \in \mathcal{E}_{[T]} \end{array} \right.$$

has solutions for any given  $\xi_{[t+1]} \in \mathcal{E}_{[t+1]}$  and  $(x_1, x_2, \dots, x_t) \in \mathbb{R}_+^{tn}$ .

Note that model (1) is a very general form of multistage SNCPs. However, each stage may be decided only by the adjoining decision variables, e.g., multistage stochastic inventory problems. Given this, we will focus on the following specific multistage SNCP:

$$\left\{ \begin{array}{l} 0 \leq x_1 \perp \mathbb{E}_{\xi_2}[\Phi_1(x_1, x_2(\xi_{[2]}), \xi_2)] \geq 0, \\ 0 \leq x_2(\xi_{[2]}) \perp \mathbb{E}_{\xi_3|\xi_{[2]}}[\Phi_2(x_1, x_2(\xi_{[2]}), x_3(\xi_{[3]}), \xi_3)] \geq 0, \text{ a.e. } \xi_{[2]} \in \mathcal{E}_{[2]}, \\ \vdots \\ 0 \leq x_{T-1}(\xi_{[T-1]}) \perp \mathbb{E}_{\xi_T|\xi_{[T-1]}}[\Phi_{T-1}(x_{T-2}(\xi_{[T-2]}), x_{T-1}(\xi_{[T-1]}), x_T(\xi_{[T]}), \xi_T)] \geq 0, \\ \text{a.e. } \xi_{[T-1]} \in \mathcal{E}_{[T-1]}, \\ 0 \leq x_T(\xi_{[T]}) \perp \Phi_T(x_{T-1}(\xi_{[T-1]}), x_T(\xi_{[T]}), \xi_T) \geq 0, \text{ a.e. } \xi_{[T]} \in \mathcal{E}_{[T]}. \end{array} \right. \tag{2}$$

In general, it is more interesting to consider the behavior of the first-stage solution  $x_1$ . Intuitively, we only need to determine  $x_1$  at the present stage, and  $x_2, \dots, x_T$  are scenario-dependent (or wait and see) solutions and new information can be employed for more accurate estimate in future stages. In view of this, as well as by convention (see, e.g., [26]), we will focus on the first stage solutions of the multistage SNCP (2). Throughout this paper, we assume that  $\Phi_1(\cdot, \cdot, \xi_2)$  for a.e.  $\xi_{[2]} \in \mathcal{E}_{[2]}$ ,  $\Phi_t(\cdot, \cdot, \cdot, \xi_{t+1})$

for a.e.  $\xi_{[t+1]} \in \mathcal{E}_{[t+1]}$  with  $t = 2, \dots, T - 1$  and  $\Phi_T(\cdot, \cdot, \xi_T)$  for a.e.  $\xi_{[T]} \in \mathcal{E}_{[T]}$  are continuous.

Our analysis is based on strong monotonicity properties of multistage SNCPS. To this end, we first consider strong monotonicity properties of two-stage SNCPS, and then extend them to multistage SNCPS.

### 2.2 Strong monotonicity of two-stage SNCPS

Two-stage SNCP is a special case of  $T$ -stage SNCPS (with  $T = 2$ ) and can be formulated as:

$$\begin{cases} 0 \in \mathbb{E}_\xi[\Phi(x, y(\xi), \xi)] + \mathcal{N}_{\mathbb{R}^n_+}(x), \\ 0 \in \Psi(x, y(\xi), \xi) + \mathcal{N}_{\mathbb{R}^m_+}(y(\xi)), \text{ a.e. } \xi \in \mathcal{E}, \end{cases} \tag{3}$$

where  $\xi$  is a  $s$ -dimensional random vector defined on probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ ,  $\mathbb{E}_\xi[\cdot]$  denotes the expectation operator with respect to (w.r.t.) random vector  $\xi$ ,  $\Phi : \mathbb{R}^n \times \mathbb{R}^m \times \mathcal{E} \rightarrow \mathbb{R}^n$  and  $\Psi : \mathbb{R}^n \times \mathbb{R}^m \times \mathcal{E} \rightarrow \mathbb{R}^m$ ,  $\mathcal{E} \subseteq \mathbb{R}^s$  denotes the support set of random vector  $\xi$ , and  $y(\xi)$  denotes the value of  $y(\cdot)$  at  $\xi$ . We assume that both  $\Phi(\cdot, \cdot, \xi)$  and  $\Psi(\cdot, \cdot, \xi)$  are continuous for a.e.  $\xi \in \mathcal{E}$ .

We first give the definition of the strong monotonicity.

**Definition 1** A mapping  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is strongly monotone with a positive scalar  $\gamma$ , if  $\langle f(x) - f(y), x - y \rangle \geq \gamma \|x - y\|^2$  for any  $x, y \in \mathbb{R}^n$ .

We have from [10, p. 75, Corollary] that the following chain rule property for Clarke generalized Jacobian, which is useful in the following discussion.

**Lemma 1** [10] Let  $G : \mathbb{R}^m \rightarrow \mathbb{R}^k$  and  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ . Assume that  $G$  and  $F$  are locally Lipschitz continuous. Then, for any  $v \in \mathbb{R}^n$ , we have

$$\partial(G \circ F)(x)v \subseteq \text{conv}\{\partial G(F(x))\partial F(x)v\},$$

where  $\partial(G \circ F)(x)$ ,  $\partial G(F(x))$  and  $\partial F(x)$  denote the Clarke generalized Jacobian of  $G \circ F$  at  $x$ ,  $G$  at  $F(x)$  and  $F$  at  $x$ , respectively.

**Proposition 1** Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be locally Lipschitz continuous. Then,  $f$  is strongly monotone with a positive constant  $\gamma$  if and only if, for any  $x \in \mathbb{R}^n$  and  $J(x) \in \partial f(x)$ , it has that  $v^\top J(x)v \geq \gamma \|v\|^2$  for any  $v \in \mathbb{R}^n$ .

Proposition 1 can be considered as a simple corollary of [9, Corollary 3.10], but the Jacobian is defined in the sense of Clarke, rather than Mordukhovich.

**Lemma 2** Consider the following parametric variational inequality:

$$0 \in \Theta(z, p) + \mathcal{N}_C(z), \tag{4}$$

where  $\Theta : \mathbb{R}^k \times \mathbb{R}^l \rightarrow \mathbb{R}^k$  and  $C \subseteq \mathbb{R}^k$  is a closed convex set. Assume that  $\Theta(\cdot, p)$  is strongly monotone and locally Lipschitz continuous for every fixed  $p$ . Then, there

exists a unique solution for problem (4), denoted by  $\hat{z}(p)$ , and it is locally Lipschitz continuous w.r.t.  $p$ .

The result follows by Proposition 1, parametrically CD regularity, and [15, Theorem 4] directly. Thus, we neglect the proof.

To proceed, we define the following inner product in  $\mathbb{R}^n \times \mathcal{Y}$ , where  $\mathcal{Y}$  is the collection of measurable mappings from  $\mathcal{E}$  to  $\mathbb{R}^m$ . Let  $(x, y(\cdot)), (\tilde{x}, \tilde{y}(\cdot)) \in \mathbb{R}^n \times \mathcal{Y}$ . Define  $\langle (x, y(\cdot)), (\tilde{x}, \tilde{y}(\cdot)) \rangle = x^\top \tilde{x} + \mathbb{E}_\xi [y(\xi)^\top \tilde{y}(\xi)]$ .

According to the definition of the inner product above, we can define the strong monotonicity (see [6]) for problem (3): there exists a positive number  $\bar{\kappa}$  such that

$$\begin{aligned} & \left\langle \begin{pmatrix} \mathbb{E}_\xi [\Phi(x, y(\xi), \xi)] - \mathbb{E}_\xi [\Phi(\tilde{x}, \tilde{y}(\xi), \xi)] \\ \Psi(x, y(\cdot), \cdot) - \Psi(\tilde{x}, \tilde{y}(\cdot), \cdot) \end{pmatrix}, \begin{pmatrix} x - \tilde{x} \\ y(\cdot) - \tilde{y}(\cdot) \end{pmatrix} \right\rangle \\ & \geq \bar{\kappa} \langle (x - \tilde{x}, y(\cdot) - \tilde{y}(\cdot)), (x - \tilde{x}, y(\cdot) - \tilde{y}(\cdot)) \rangle \\ & = \bar{\kappa} \left( \|x - \tilde{x}\|^2 + \mathbb{E}_\xi [\|y(\xi) - \tilde{y}(\xi)\|^2] \right) \end{aligned}$$

holds for any  $(x, y(\cdot)), (\tilde{x}, \tilde{y}(\cdot)) \in \mathbb{R}^n \times \mathcal{Y}$ .

We make the following assumption.

**Assumption 1** *There exists a positive continuous function  $\kappa : \mathcal{E} \rightarrow \mathbb{R}_{++}$  with  $\mathbb{E}_\xi [\kappa(\xi)] < +\infty$ , such that*

$$\left\langle \begin{pmatrix} \Phi(u, z, \xi) \\ \Psi(u, z, \xi) \end{pmatrix} - \begin{pmatrix} \Phi(\bar{u}, \bar{z}, \xi) \\ \Psi(\bar{u}, \bar{z}, \xi) \end{pmatrix}, \begin{pmatrix} u - \bar{u} \\ z - \bar{z} \end{pmatrix} \right\rangle \geq \kappa(\xi) \left( \|u - \bar{u}\|^2 + \|z - \bar{z}\|^2 \right)$$

holds for any  $(u, z), (\bar{u}, \bar{z}) \in \mathbb{R}^n \times \mathbb{R}^m$  and a.e.  $\xi \in \mathcal{E}$ , where  $\mathbb{R}_{++} := \{x \in \mathbb{R} : x > 0\}$ .

For any fixed pair  $(x, \xi)$ , under Assumption 1 and the continuity of  $\Psi(x, \cdot, \xi)$ , the second-stage problem of (3) has a unique solution because of the strong monotonicity with constant  $\kappa(\xi)$  [12, Theorem 2.3.3 (b)]. We denote by  $\hat{y}(x, \xi)$  the unique solution. Then, we take it into the first-stage problem and obtain

$$0 \in \mathbb{E}_\xi [\Phi(x, \hat{y}(x, \xi), \xi)] + \mathcal{N}_{\mathbb{R}^n_+}(x). \tag{5}$$

We also use  $\hat{\Phi}(x, \xi)$  to denote  $\Phi(x, \hat{y}(x, \xi), \xi)$ , which stresses that the latter one depends only on  $x$  and  $\xi$ .

Now we are ready to present the main results of this part.

**Theorem 1** *Suppose that  $\Psi(x, \cdot, \xi)$  in the second-stage problem of (3) is strongly monotone for any fixed  $x$  and a.e.  $\xi \in \mathcal{E}$ , and  $\Psi(\cdot, \cdot, \xi)$  is locally Lipschitz continuous for a.e.  $\xi \in \mathcal{E}$ .*

- (i) *For any fixed  $x$  and a.e.  $\xi \in \mathcal{E}$ , the second-stage problem of (3) has a unique solution, denoted by  $\hat{y}(x, \xi)$ , and  $\hat{y}(x, \xi)$  is locally Lipschitz continuous w.r.t.  $x$  for a.e.  $\xi \in \mathcal{E}$ .*
- (ii) *Moreover, if  $\hat{y}(x, \xi)$  is differentiable at  $x$ , then for any  $v \in \mathbb{R}^m$ ,*

$$\nabla_x \hat{y}(x, \xi)v \in \bigcup_{(\mathbf{C}, \mathbf{D}) \in \partial \Psi(x, \hat{y}(x, \xi), \xi)} -(I - \mathcal{D}_\mathcal{L} + \mathcal{D}_\mathcal{L} \mathbf{D})^{-1} \mathcal{D}_\mathcal{L} \mathbf{C}v, \tag{6}$$

where  $\partial\Psi(x, \hat{y}(x, \xi), \xi)$  is the Clarke generalized Jacobian w.r.t.  $(x, y)$ ,  $\mathcal{L}$  is a subset contained in  $\{1, \dots, m\}$  and  $\mathcal{D}_{\mathcal{L}}$  is a diagonal matrix with its diagonal elements defining as  $\mathcal{D}_{ll} = \begin{cases} 1 & l \in \mathcal{L}, \\ 0 & \text{otherwise.} \end{cases}$

*Proof* The first assertion follows from Lemma 2 straightforwardly. In the following proof, we concentrate on the second assertion.

Since  $\hat{y}(x, \xi)$  is locally Lipschitz continuous, it is differentiable almost everywhere. For any fixed  $x$  with  $\hat{y}(\cdot, \xi)$  being differentiable at  $x$ , we consider the following partition of the index set:

$$\begin{aligned} \mathcal{I} &= \{i : (\hat{y}(x, \xi))_i > (\Psi(x, \hat{y}(x, \xi), \xi))_i\}, \\ \mathcal{J} &= \{j : (\hat{y}(x, \xi))_j = (\Psi(x, \hat{y}(x, \xi), \xi))_j\}, \\ \mathcal{K} &= \{k : (\hat{y}(x, \xi))_k < (\Psi(x, \hat{y}(x, \xi), \xi))_k\}. \end{aligned}$$

If index  $j \in \mathcal{J}$ ,  $(\hat{y}(x, \xi))_j = (\Psi(x, \hat{y}(x, \xi), \xi))_j = 0$ . There are two possible cases.

- (a) If there exists a sequence  $\{x^k\}_{k \in \mathbb{N}}$  with  $x^k \neq x$  and  $x^k \rightarrow x$  as  $k \rightarrow \infty$  such that  $(\hat{y}(x^k, \xi))_j = 0$ , based on the definition of differential, we have

$$\|\nabla_x(\hat{y}(x, \xi))_j\| = \lim_{k \rightarrow \infty} \frac{|(\hat{y}(x^k, \xi))_j - (\hat{y}(x, \xi))_j|}{\|x^k - x\|} = \frac{|0 - 0|}{\|x^k - x\|} = 0.$$

- (b) If there exists a sequence  $\{x^k\}_{k \in \mathbb{N}}$  with  $x^k \neq x$  and  $x^k \rightarrow x$  as  $k \rightarrow \infty$  such that  $(\Psi(x^k, \hat{y}(x^k, \xi), \xi))_j = 0$ , we have

$$\|\nabla_x(\Psi(x, \hat{y}(x, \xi), \xi))_j\| = \lim_{k \rightarrow \infty} \frac{|(\Psi(x^k, \hat{y}(x^k, \xi), \xi))_j - (\Psi(x, \hat{y}(x, \xi), \xi))_j|}{\|x^k - x\|} = 0.$$

Thus, we use notation  $\mathcal{J}_1$  and  $\mathcal{J}_2$  to denote these indexes satisfying (a) and (b), respectively. Thus,  $\mathcal{J} = \mathcal{J}_1 \cup \mathcal{J}_2$ . Then, we denote  $\mathcal{L} = \mathcal{I} \cup \mathcal{J}_2$  and  $\mathcal{M} = \mathcal{K} \cup \mathcal{J}_1$ .

If index  $i \in \mathcal{M}$ ,  $(\hat{y}(x, \xi))_i = 0$ , which obviously indicates that  $\nabla_x(\hat{y}(x, \xi))_{\mathcal{M}} = 0$ .

If index  $i \in \mathcal{I}$ , we have  $(\Psi(x, \hat{y}(x, \xi), \xi))_i = 0$ . Actually, we have that

$$(\Psi(x, \hat{y}(x, \xi), \xi))_i = 0$$

in a neighborhood of  $x$  due to  $\hat{y}(x, \xi) > 0$  and the locally Lipschitz continuity of  $\hat{y}(\cdot, \xi)$  w.r.t.  $x$ . This observation together with (b) indicates

$$\nabla_x(\Psi(x, \hat{y}(x, \xi), \xi))_{\mathcal{L}} = 0,$$

where  $(\Psi(x, \hat{y}(x, \xi), \xi))_{\mathcal{L}}$  is the subvector of  $\Psi(x, \hat{y}(x, \xi), \xi)$  with entries in  $\mathcal{L}$ .

For any vector  $v \in \mathbb{R}^m$ , we have

$$\begin{aligned} 0 &= 0v = \nabla_x(\Psi(x, \hat{y}(x, \xi), \xi))_{\mathcal{L}}v \\ &\stackrel{1)}{\in} \text{conv} \left\{ \partial(\Psi(x, \hat{y}(x, \xi), \xi))_{\mathcal{L}} \partial_x \begin{pmatrix} x \\ \hat{y}(x, \xi) \end{pmatrix} v \right\} \\ &= \text{conv} \left\{ \partial(\Psi(x, \hat{y}(x, \xi), \xi))_{\mathcal{L}} \begin{pmatrix} v \\ \nabla_x \hat{y}(x, \xi)v \end{pmatrix} \right\} \\ &\stackrel{2)}{=} \partial(\Psi(x, \hat{y}(x, \xi), \xi))_{\mathcal{L}} \begin{pmatrix} v \\ \nabla_x \hat{y}(x, \xi)v \end{pmatrix}, \end{aligned}$$

where 1) follows from the chain rule of Clarke generalized Jacobian, i.e., Lemma 1; 2) follows from the convexity and compactness of  $\partial(\Psi(x, \hat{y}(x, \xi), \xi))_{\mathcal{L}}$  by [10, Proposition 2.6.2, (a)]. Then, we obtain

$$0 \in \bigcup_{(\mathbf{C}, \mathbf{D}) \in \partial \Psi(x, \hat{y}(x, \xi), \xi)} (\mathbf{C}_{\mathcal{L}}, \mathbf{D}_{\mathcal{L}}) \begin{pmatrix} v \\ \nabla_x \hat{y}(x, \xi)v \end{pmatrix},$$

where  $\mathbf{C}_{\mathcal{L}}$  and  $\mathbf{D}_{\mathcal{L}}$  denote sub-matrices of  $\mathbf{C}$  and  $\mathbf{D}$  with row indexes included in  $\mathcal{L}$ , respectively.

Note that  $(\nabla_x \hat{y}(x, \xi))_{\mathcal{M}=0}$ , which implies that

$$\mathbf{D}_{\mathcal{L}}(\nabla_x \hat{y}(x, \xi)v) = \mathbf{D}_{\mathcal{L}\mathcal{L}}(\nabla_x \hat{y}(x, \xi)v)_{\mathcal{L}},$$

where  $\mathbf{D}_{\mathcal{L}\mathcal{L}}$  denotes the sub-matrix of  $\mathbf{D}$  with row index and column index contained in  $\mathcal{L}$ . Then we have

$$0 \in \bigcup_{(\mathbf{C}, \mathbf{D}) \in \partial \Psi(x, \hat{y}(x, \xi), \xi)} \mathbf{C}_{\mathcal{L}}v + \mathbf{D}_{\mathcal{L}\mathcal{L}}(\nabla_x \hat{y}(x, \xi)v)_{\mathcal{L}}.$$

According to Proposition 1, the strong monotonicity of  $\Psi(x, \cdot, \xi)$  implies the strongly positive definiteness of  $\mathbf{D}_{\mathcal{L}\mathcal{L}}$ . Thus, the inverse of  $\mathbf{D}_{\mathcal{L}\mathcal{L}}$  is well-defined. Finally, we obtain

$$(\nabla_x \hat{y}(x, \xi)v)_{\mathcal{L}} \in \bigcup_{(\mathbf{C}, \mathbf{D}) \in \partial \Psi(x, \hat{y}(x, \xi), \xi)} \mathbf{D}_{\mathcal{L}\mathcal{L}}^{-1} \mathbf{C}_{\mathcal{L}}v.$$

Combining  $(\nabla_x \hat{y}(x, \xi)v)_{\mathcal{M}}$  with  $(\nabla_x \hat{y}(x, \xi)v)_{\mathcal{L}}$  together, we have (6) by using the same reformulation techniques in [6, 8]. □

*Remark 1* In Theorem 1, we weaken the conditions in existing results in [18, Corollary 2.1] from continuous differentiability of  $\Psi(\cdot, \cdot, \xi)$  to locally Lipschitz continuity of  $\Psi(\cdot, \cdot, \xi)$ .

With the above preparation, we now give the strong monotonicity assertion for the two-stage SNCP where the second-stage problem is only locally Lipschitz continuous. This will lay the foundation for the multistage case.

**Theorem 2** *Suppose that: (i) Assumption 1 holds; (ii)  $\Phi(\cdot, \cdot, \xi)$  is continuously differentiable and  $\Psi(\cdot, \cdot, \xi)$  is locally Lipschitz continuous for a.e.  $\xi \in \mathcal{E}$ . Then  $\hat{\Phi}(\cdot, \xi)$*

is strongly monotone for a.e.  $\xi \in \mathcal{E}$  with constant  $\kappa(\xi)$  and thus  $\mathbb{E}_\xi[\hat{\Phi}(\cdot, \xi)]$  is strongly monotone with constant  $\mathbb{E}_\xi[\kappa(\xi)]$ .

*Proof* According to Lemma 2,  $\hat{y}(x, \xi)$  is locally Lipschitz continuous w.r.t.  $x$  and thus  $\Phi(x, \hat{y}(x, \xi), \xi)$  is locally Lipschitz continuous w.r.t.  $x$ . Then, its Clarke generalized Jacobian w.r.t.  $x$  is well-defined. We first verify that  $\hat{\Phi}(x, \xi) := \Phi(x, \hat{y}(x, \xi), \xi)$  is strongly monotone w.r.t.  $x$  for a.e.  $\xi \in \mathcal{E}$ . Equivalently, we show that any  $J(\xi) \in \partial_x \hat{\Phi}(x, \xi)$  is strongly positive definite (see Proposition 1). Based on the definition of Clarke generalized Jacobian, we have that, for any  $v \in \mathbb{R}^n$ , that

$$\begin{aligned} \partial_x \hat{\Phi}(x, \xi)v &= \partial_x \Phi(x, \hat{y}(x, \xi), \xi)v \\ &\stackrel{(a)}{=} \text{conv} \left\{ \lim_{\substack{x' \rightarrow x \\ x' \in \mathcal{D}_{\hat{y}(\cdot, \xi)}}} (\nabla_1 \Phi(x', \hat{y}(x', \xi), \xi) + \nabla_2 \Phi(x', \hat{y}(x', \xi), \xi) \nabla_x \hat{y}(x', \xi)) v \right\} \\ &\stackrel{(b)}{\subseteq} \text{conv} \left\{ \lim_{\substack{x' \rightarrow x \\ x' \in \mathcal{D}_{\hat{y}(\cdot, \xi)}}} \left( \bigcup_{(\mathbf{C}(x', \xi), \mathbf{D}(x', \xi)) \in \partial \Psi(x', \hat{y}(x', \xi), \xi)} \mathbf{A}(x', \xi) \right. \right. \\ &\quad \left. \left. - \mathbf{B}(x', \xi)(I - \mathcal{D}_{\mathcal{L}} + \mathcal{D}_{\mathcal{L}} \mathbf{D}(x', \xi))^{-1} \mathcal{D}_{\mathcal{L}} \mathbf{C}(x', \xi) \right) v \right\}, \end{aligned}$$

where  $\mathcal{D}_{\hat{y}(\cdot, \xi)}$  denotes the collection of differentiable points of  $\hat{y}(\cdot, \xi)$ ;  $\mathbf{A}(x', \xi) := \nabla_1 \Phi(x', \hat{y}(x', \xi), \xi)$ ,  $\mathbf{B}(x', \xi) := \nabla_2 \Phi(x', \hat{y}(x', \xi), \xi)$ ,  $\nabla_1$  is the differential at the first position and  $\nabla_2$  is the differential at the second position. Note that (a) follows from the definition of Clarke generalized Jacobian, and (b) follows from (6).

It knows from the strong monotonicity in Assumption 1 and Proposition 1 that any

$$\begin{pmatrix} \mathbf{A}(x', \xi) & \mathbf{B}(x', \xi) \\ \mathbf{C}(x', \xi) & \mathbf{D}(x', \xi) \end{pmatrix} \in \partial \begin{pmatrix} \Phi(x', \hat{y}(x', \xi), \xi) \\ \Psi(x', \hat{y}(x', \xi), \xi) \end{pmatrix}$$

is strongly positive definite with constant  $\kappa(\xi)$ . Then by [6, Lemma 2.1] and Assumption 1,

$$z^\top (\mathbf{A}(x', \xi) - \mathbf{B}(x', \xi)(I - \mathcal{D}_{\mathcal{L}} + \mathcal{D}_{\mathcal{L}} \mathbf{D}(x', \xi))^{-1} \mathcal{D}_{\mathcal{L}} \mathbf{C}(x', \xi)) z \geq \kappa(\xi) \|z\|^2$$

for any  $z \in \mathbb{R}^n$  and  $x' \in \mathcal{D}_{\hat{y}(\cdot, \xi)}$ . This implies that each element in

$$\begin{aligned} &\text{conv} \left\{ \lim_{\substack{x' \rightarrow x \\ x' \in \mathcal{D}_{\hat{y}(\cdot, \xi)}}} \left( \bigcup_{(\mathbf{C}(x', \xi), \mathbf{D}(x', \xi)) \in \partial \Psi(x', \hat{y}(x', \xi), \xi)} \mathbf{A}(x', \xi) \right. \right. \\ &\quad \left. \left. - \mathbf{B}(x', \xi)(I - \mathcal{D}_{\mathcal{L}} + \mathcal{D}_{\mathcal{L}} \mathbf{D}(x', \xi))^{-1} \mathcal{D}_{\mathcal{L}} \mathbf{C}(x', \xi) \right) v \right\} \end{aligned}$$

is strongly positive definite with constant  $\kappa(\xi)$ . Then, the strong monotonicity of  $\hat{\Phi}(\cdot, \xi)$  with constant  $\kappa(\xi)$  follows from Proposition 1.

Furthermore, we have  $\mathbb{E}_\xi[\hat{\Phi}(\cdot, \xi)]$  is strongly monotone with constant  $\mathbb{E}_\xi[\kappa(\xi)]$ . □

### 2.3 Strong monotonicity of multistage SNCPs

In this part, we consider the strong monotonicity of the  $T$ -stage SNCP (2). It can be viewed as an extension of [6] from two-stage to multistage and linear case to nonlinear case. Analogously, we need to define the following inner product in  $\mathbb{R}^n \times \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_T$ , where  $\mathcal{Y}_t$ ,  $t = 2, \dots, T$  stand for the set of all  $n$ -dimensional measurable mappings on  $\mathcal{E}_{[t]}$ . For

$$(x_1, x_2(\cdot), \dots, x_T(\cdot)), (y_1, y_2(\cdot), \dots, y_T(\cdot)) \in \mathbb{R}^n \times \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_T,$$

we define

$$\begin{aligned} &\langle (x_1, x_2(\cdot), \dots, x_T(\cdot)), (y_1, y_2(\cdot), \dots, y_T(\cdot)) \rangle \\ &= x_1^\top y_1 + \mathbb{E}_{\xi_{[2]}}[x_2(\xi_{[2]})^\top y_2(\xi_{[2]})] + \cdots + \mathbb{E}_{\xi_{[T]}}[x_T(\xi_{[T]})^\top y_T(\xi_{[T]})]. \end{aligned}$$

Then, we have the corresponding norm

$$\| (x_1, x_2(\cdot), \dots, x_T(\cdot)) \|^2 = \langle (x_1, x_2(\cdot), \dots, x_T(\cdot)), (x_1, x_2(\cdot), \dots, x_T(\cdot)) \rangle.$$

Denote

$$\Lambda(x_1, \dots, x_T, \xi) = \begin{pmatrix} \Phi_1(x_1, x_2, \xi_2) \\ \Phi_2(x_1, x_2, x_3, \xi_3) \\ \vdots \\ \Phi_{T-1}(x_{T-2}, x_{T-1}, x_T, \xi_T) \\ \Phi_T(x_{T-1}, x_T, \xi_T) \end{pmatrix}$$

and

$$\Gamma(x_1, \dots, x_T(\cdot), \xi) = \begin{pmatrix} \mathbb{E}_{\xi_2}[\Phi_1(x_1, x_2(\xi_{[2]}), \xi_2)] \\ \mathbb{E}_{\xi_3|\xi_{[2]}}[\Phi_2(x_1, x_2(\xi_{[2]}), x_3(\xi_{[3]}), \xi_3)] \\ \vdots \\ \mathbb{E}_{\xi_T|\xi_{[T-1]}}[\Phi_{T-1}(x_{T-2}(\xi_{[T-2]}), x_{T-1}(\xi_{[T-1]}), x_T(\xi_{[T]}), \xi_T)] \\ \Phi_T(x_{T-1}(\xi_{[T-1]}), x_T(\xi_{[T]}), \xi_T) \end{pmatrix}.$$

Then, we can give the definition of the strong monotonicity for  $\mathcal{G} : \mathbb{R}^n \times \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_T \rightarrow \mathbb{R}^n \times \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_T$  as follows: there exists a positive number  $\bar{\eta}$  such that

$$\begin{aligned} &\langle \mathcal{G}(x_1, \dots, x_T(\cdot)) - \mathcal{G}(y_1, \dots, y_T(\cdot)), (x_1 - y_1, \dots, x_T(\cdot) - y_T(\cdot)) \rangle \\ &\geq \bar{\eta} \| (x_1 - y_1, x_2(\cdot) - y_2(\cdot), \dots, x_T(\cdot) - y_T(\cdot)) \|^2 \end{aligned}$$

for any  $(x_1, x_2(\cdot), \dots, x_T(\cdot)), (y_1, y_2(\cdot), \dots, y_T(\cdot)) \in \mathbb{R}^n \times \mathcal{Y}_2 \times \cdots \times \mathcal{Y}_T$ .

**Assumption 2** For any  $(u_1, u_2, \dots, u_T), (z_1, z_2, \dots, z_T) \in \mathbb{R}^{Tn}$ , there exists  $\eta : \mathcal{E} \rightarrow \mathbb{R}_{++}$  satisfying  $\mathbb{E}_\xi[\eta(\xi)] < +\infty$ , such that

$$\begin{aligned} &\langle \Lambda(u_1, u_2, \dots, u_T, \xi) - \Lambda(z_1, z_2, \dots, z_T, \xi), (u_1 - z_1, u_2 - z_2, \dots, u_T - z_T) \rangle \\ &\geq \eta(\xi) \left( \|u_1 - z_1\|^2 + \|u_2 - z_2\|^2 + \cdots + \|u_T - z_T\|^2 \right) \end{aligned} \tag{7}$$

holds for a.e.  $\xi \in \mathcal{E}$ .

It is easy to know from (7) that

$$\langle \Phi_t(u_{t-1}, u_t, u_{t+1}, \xi_{t+1}) - \Phi_t(u_{t-1}, z_t, u_{t+1}, \xi_{t+1}), (u_t - z_t) \rangle \geq \eta(\xi) \|u_t - z_t\|^2 \tag{8}$$

for  $t = 1, 2, \dots, T$ . Then,

$$\begin{aligned} & \langle \Gamma(x_1, \dots, x_T(\cdot), \xi) - \Gamma(y_1, \dots, y_T(\cdot), \xi), (x_1 - y_1, \dots, x_T(\cdot) - y_T(\cdot)) \rangle \\ & \geq \mathbb{E}_\xi[\eta(\xi)] \left( \|x_1 - y_1\|^2 + \dots + \|x_T(\cdot) - y_T(\cdot)\|^2 \right), \end{aligned}$$

which implies  $\Gamma(x_1, x_2, \dots, x_T, \xi)$  is strongly monotone.

**Proposition 2** *Let Assumption 2 hold. For any given  $x_{[t]} \in \mathbb{R}^{tn}$  and a.e.  $\xi_{[t+1]} \in \mathcal{E}_{[t+1]}$ , the last  $(T - t)$ -stage problem*

$$\left\{ \begin{aligned} & 0 \leq x_{t+1} \perp \mathbb{E}_{\xi_{t+2}|\xi_{[t+1]}} [\Phi_{t+1}(x_t, x_{t+1}, x_{t+2}, \xi_{t+2})] \geq 0, \\ & 0 \leq x_{t+2} \perp \mathbb{E}_{\xi_{t+3}|\xi_{[t+2]}} [\Phi_{t+2}(x_{t+1}, x_{t+2}, x_{t+3}, \xi_{t+3})] \geq 0, \text{ a.e. } \xi_{[t+2]} \in \mathcal{E}_{[t+2]}, \\ & \vdots \\ & 0 \leq x_T \perp \Phi_T(x_{T-1}, x_T, \xi_T) \geq 0, \text{ a.e. } \xi_{[T]} \in \mathcal{E}_{[T]} \end{aligned} \right. \tag{9}$$

has a unique solution  $(\hat{x}_{t+1}, \dots, \hat{x}_T)$ .

Similar to the previous discussion, we know from [14, Theorem 12.2 and Lemma 12.2] that (9) has a unique solution. Specifically, when  $t = 0$ , problem (2) has a unique solution under Assumption 2.

**Theorem 3** *Suppose that: (i) Assumption 2 holds; (ii)  $\Phi_1(\cdot, \cdot, \xi_2)$ ,  $\Phi_t(\cdot, \cdot, \cdot, \xi_{t+1})$  for  $t = 2, \dots, T - 1$  and  $\Phi_T(\cdot, \cdot, \xi_T)$  are continuously differentiable almost surely. Then*

$$0 \leq x_1 \perp \mathbb{E}_{\xi_2} [\Phi_1(x_1, \hat{x}_2(x_1, \xi_{[2]}), \xi_2)] \geq 0 \tag{10}$$

and the following problem

$$\left\{ \begin{aligned} & 0 \leq x_1 \perp \mathbb{E}_{\xi_2} [\Phi_1(x_1, x_2(\xi_{[2]}), \xi_2)] \geq 0, \\ & 0 \leq x_2 \perp \mathbb{E}_{\xi_3|\xi_{[2]}} [\Phi_2(x_1, x_2(\xi_{[2]}), x_3(\xi_{[3]}), \xi_3)] \geq 0, \text{ a.e. } \xi_{[2]} \in \mathcal{E}_{[2]}, \\ & \vdots \\ & 0 \leq x_t \perp \mathbb{E}_{\xi_{t+1}|\xi_{[t]}} [\Phi_t(x_{t-1}(\xi_{[t-1]}), x_t(\xi_t), \hat{x}_{t+1}(x_{[t]}, \xi_{[t+1]}), \xi_{t+1})] \geq 0, \\ & \text{a.e. } \xi_{[t]} \in \mathcal{E}_{[t]} \end{aligned} \right. \tag{11}$$

for  $t = 2, \dots, T - 1$  are strongly monotone with constant  $\mathbb{E}_\xi[\eta(\xi)]$ , where  $\hat{x}_{t+1}(x_t, \xi_{[t+1]})$  is the unique solution of the  $(t + 1)$ th stage for given  $x_{[t]}$  and  $\xi_{[t+1]}$  in Proposition 2.

*Proof* We give the proof recursively. Denote

$$\tilde{\Lambda}_t(x_1, \dots, x_t, \xi_{[t+1]}) = \begin{pmatrix} \Phi_1(x_1, x_2, \xi_2) \\ \Phi_2(x_1, x_2, x_3, \xi_3) \\ \vdots \\ \Phi_t(x_{t-1}, x_t, \hat{x}_{t+1}(x_t, \xi), \xi_{t+1}) \end{pmatrix}$$

and

$$\tilde{\Gamma}_t(x_1, \dots, x_t, \xi_{[t]}) = \begin{pmatrix} \mathbb{E}_{\xi_2}[\Phi_1(x_1, x_2, \xi_2)] \\ \mathbb{E}_{\xi_3|\xi_{[2]}}[\Phi_2(x_1, x_2, x_3, \xi_3)] \\ \vdots \\ \mathbb{E}_{\xi_{t+1}|\xi_{[t]}}[\Phi_t(x_{t-1}, x_t, \hat{x}_{t+1}(x_t, \xi_{[t+1]}), \xi_{t+1})] \end{pmatrix}$$

for  $t = 2, \dots, T - 1$ , where  $\hat{x}_{t+1}(x_t, \xi_{[t+1]})$  is the unique  $(t + 1)$ th stage solution in Proposition 2.

Then, we first verify  $t = T - 1$  for problem (11). According to Proposition 2, we have for given  $x_{T-1}$  and a.e.  $\xi \in \Xi$  that the last-stage problem

$$0 \leq x_T \perp \Phi_T(x_{T-1}, x_T, \xi_T) \geq 0$$

has a unique solution, denoted by  $\hat{x}_T(x_{T-1}, \xi)$ . Moreover, by Lemma 2, we know that  $\hat{x}_T(x_{T-1}, \xi)$  is locally Lipschitz continuous w.r.t.  $x_{T-1}$ . By viewing

$$\left( \begin{pmatrix} \Phi_1(x_1, x_2, \xi_2) \\ \vdots \\ \Phi_{T-1}(x_{T-2}, x_{T-1}, x_T, \xi_T) \\ \Phi_T(x_{T-1}, x_T, \xi_T) \end{pmatrix} \right) = \begin{pmatrix} \tilde{\Lambda}_{T-1}(x_1, \dots, x_{T-1}, \xi_{[T]}) \\ \Phi_T(x_{T-1}, x_T, \xi_T) \end{pmatrix}$$

as two parts, we can obtain from Theorem 2 that  $\tilde{\Lambda}_{T-1}(x_1, \dots, x_{T-1}, \xi_{[T]})$  satisfies Assumption 2 in  $\mathbb{R}^{(T-1)n}$  with constant  $\eta(\xi)$ . Moreover, by a similar procedure, we can obtain

$$\begin{aligned} & \left| \tilde{\Gamma}_{T-1}(x_1, x_2(\cdot), \dots, x_{T-1}(\cdot), \xi_{[T-1]}) - \tilde{\Gamma}_{T-1}(y_1, y_2(\cdot), \dots, y_{T-1}(\cdot), \xi_{[T-1]}) \right. \\ & \left. (x_1 - y_1, x_2(\cdot) - y_2(\cdot), \dots, x_{T-1}(\cdot) - y_{T-1}(\cdot)) \right| \\ & \geq \mathbb{E}_\xi[\eta(\xi)] \left( \|x_1 - y_1\|^2 + \|x_2(\cdot) - y_2(\cdot)\|^2 + \dots + \|x_{T-1}(\cdot) - y_{T-1}(\cdot)\|^2 \right) \end{aligned}$$

for any  $(x_1, x_2(\cdot), \dots, x_{T-1}(\cdot)), (y_1, y_2(\cdot), \dots, y_{T-1}(\cdot)) \in \mathbb{R}^n \times \mathcal{Y}_2 \times \dots \times \mathcal{Y}_{T-1}$ . Thus, problem (11) is strongly monotone with constant  $\mathbb{E}_\xi[\eta(\xi)]$  for  $t = T - 1$ .

Now, we assume that for  $T - 1, T - 2, \dots, t + 1$  ( $t \geq 2$ ) that problem (11) is strongly monotone with constant  $\mathbb{E}_\xi[\eta(\xi)]$  and  $\hat{x}_{t+1}(x_t, \xi_{[t+1]})$  is locally Lipschitz continuous w.r.t.  $x_t$ . Analogously, we view  $\tilde{\Lambda}_t(x_1, \dots, x_t, \xi_{[t+1]})$  as the following two parts:

$$\left( \begin{pmatrix} \Phi_1(x_1, x_2, \xi_2) \\ \vdots \\ \Phi_{t-1}(x_{t-2}, x_{t-1}, x_t, \xi_t) \\ \Phi_t(x_{t-1}, x_t, \hat{x}_{t+1}(x_t, \xi_{[t]}), \xi_{t+1}) \end{pmatrix} \right) = \begin{pmatrix} \tilde{\Lambda}_{t-1}(x_1, \dots, x_{t-1}, \xi_{[t]}) \\ \Phi_t(x_{t-1}, x_t, \hat{x}_{t+1}(x_t, \xi_{[t]}), \xi_{t+1}) \end{pmatrix}.$$

Then, by Lemma 2 and Theorem 2, we have that  $\hat{x}_t(x_{t-1}, \xi_{[t]})$  is locally Lipschitz continuous w.r.t.  $x_{t-1}$  and  $\tilde{\Gamma}_t(x_1, \dots, x_t, \xi_{[t]})$  is strongly monotone with constant  $\mathbb{E}_\xi[\eta(\xi)]$ .

Finally, we obtain

$$\begin{cases} 0 \leq x_1 \perp \mathbb{E}_{\xi_2}[\Phi_1(x_1, x_2(\xi_{[2]}), \xi_2)] \geq 0; \\ 0 \leq x_2 \perp \mathbb{E}_{\xi_3|\xi_{[2]}}[\Phi_2(x_1, x_2(\xi_{[2]}), \hat{x}_3(x_2, \xi_{[3]}), \xi_3)] \geq 0, \text{ a.e. } \xi_{[2]} \in \mathcal{E}_{[2]} \end{cases}$$

is strongly monotone with constant  $\mathbb{E}_\xi[\eta(\xi)]$ , and  $\hat{x}_3(\cdot, \xi_{[3]})$  is locally Lipschitz continuous. By directly using Theorem 2, we verify (10). Then we complete the proof.  $\square$

Theorem 3 can be viewed as a multistage extension of Theorem 2. The following corollary directly follows from the proof of Theorem 3.

**Corollary 1** *Let assumptions in Theorem 3 hold. For  $t = 2, \dots, T$ ,  $\hat{x}_t(x_{t-1}, \xi_{[t]})$ , which is the  $t$ th stage solution in Proposition 2, is locally Lipschitz continuous w.r.t.  $x_{t-1}$  for a.e.  $\xi_{[t]} \in \mathcal{E}_{[t]}$ .*

### 3 Conditional sampling for multistage SNCPs

To simplify the presentation as well as by convention [26, 27], we only discuss the case of the SAA approach to solve problem (2) when  $T = 3$ . That is

$$\begin{cases} 0 \leq x_1 \perp \mathbb{E}_{\xi_2}[\Phi_1(x_1, x_2(\xi_{[2]}), \xi_2)] \geq 0, \\ 0 \leq x_2(\xi_{[2]}) \perp \mathbb{E}_{\xi_3|\xi_{[2]}}[\Phi_2(x_1, x_2(\xi_{[2]}), x_3(\xi_{[3]}), \xi_3)] \geq 0, \text{ a.e. } \xi_{[2]} \in \mathcal{E}_{[2]}, \\ 0 \leq x_3(\xi_{[3]}) \perp \Phi_3(x_2(\xi_{[2]}), x_3(\xi_{[3]}), \xi_3) \geq 0, \text{ a.e. } \xi_{[3]} \in \mathcal{E}_{[3]} \end{cases} \quad (12)$$

or equivalently

$$\begin{cases} 0 \leq x_1 \perp \mathbb{E}_{\xi_2}[\Phi_1(x_1, x_2(\xi_{[2]}), \xi_2)] \geq 0, \\ 0 \leq x_2(\xi_{[2]}) \perp \mathbb{E}_{\xi_3|\xi_{[2]}}[\Phi_2(x_1, x_2(\xi_{[2]}), \hat{x}_3(x_2(\xi_{[2]}), \xi_{[3]}), \xi_3)] \geq 0, \text{ a.e. } \xi_{[2]} \in \mathcal{E}_{[2]}, \end{cases} \quad (13)$$

where  $\hat{x}_3(x_2(\xi_{[2]}), \xi_{[3]})$  is the unique solution of the third-stage problem for given  $x_2(\xi_{[2]})$  and  $\xi_{[3]}$ .

Let  $\{\xi_2^1, \xi_2^2, \dots, \xi_2^{N_1}\}$  be  $N_1$  i.i.d. random samples according to random vector  $\xi_2$ . Further, we have according to  $\xi_2^i$  for  $1 \leq i \leq N_1$  the following conditional  $N_2$  i.i.d.

random samples  $\{\xi_3^{i1}, \xi_3^{i2}, \dots, \xi_3^{iN_2}\}$  of  $\xi_3$ . Thus, we have total  $N := N_1 N_2$  scenarios or sample paths. Then, we obtain the conditional sampling SAA problem of (13):

$$\begin{cases} 0 \leq x_1 \perp \frac{1}{N_1} \sum_{i=1}^{N_1} \Phi_1(x_1, x_2(\xi_2^i), \xi_2^i) \geq 0, \\ 0 \leq x_2(\xi_2^j) \perp \frac{1}{N_2} \sum_{k=1}^{N_2} \Phi_2(x_1, x_2(\xi_2^j), \hat{x}_3(x_2(\xi_2^j), \xi_{[3]}^{jk}), \xi_3^{jk}) \geq 0, \\ j = 1, 2, \dots, N_1 \text{ and } k = 1, 2, \dots, N_2, \end{cases} \tag{14}$$

where  $\xi_{[3]}^{jk} := (\xi_1, \xi_2^j, \xi_3^{jk})$  for  $j = 1, 2, \dots, N_1$  and  $k = 1, 2, \dots, N_2$ . Since  $\xi_1$  is deterministic, we also neglect  $\xi_1$  and write  $\xi_{[2]}^j$  as  $\xi_2^j$ . Similar to the notation in the preceding context, we use  $\hat{x}_2(x_1, \xi_2)$  and  $x_1^*$  to denote the optimal solution of the second-stage problem of (13) for given  $x_1$  and  $\xi_2$  and the optimal solution of the first-stage problem of (13), respectively. Moreover, according to Theorem 3, we have that the SAA problem (14) has the unique solution by viewing that the underlying probability distributions are discrete. We use  $\tilde{x}_1^{N_1, N_2}$  to denote the first-stage solution of problem (14) and  $\tilde{x}_2^{N_2}(x_1, \xi_2^j)$  to denote the second-stage solution of problem (14) for given  $x_1$  and  $\xi_2^j$ .

It should be mentioned that the convergence analysis of SAA of three-stage SNCs (between problems (13) and (14)) is not a trivial extension from SAA of two-stage SNCs. The main reason is that the second-stage problem in (13) changes when the SAA procedure is conducted (see (14)), which makes the SAA convergence of the first-stage problem of (14) to that of (13) harsh. To derive the convergence result, we need the following standard assumption.

**Assumption 3** *Let the following assertions hold:*

(A1) *There exist two compact and convex sets  $X_1, X_2 \subseteq \mathbb{R}^n$  such that  $x_1^* \cup \tilde{x}_1^{N_1, N_2} \subseteq X_1$  for any  $N_1, N_2 \in \mathbb{N}$  and  $\hat{x}_2(x_1, \xi_2) \cup \{\tilde{x}_2^{N_2}(x_1, \xi_2^j)\}_{j=1}^{N_1} \subseteq X_2$  for any  $x_1 \in X_1, \xi_2, \xi_2^j \in \mathcal{E}_2$  and  $N_1, N_2 \in \mathbb{N}, j = 1, 2, \dots, N_1$ ;*

(A2) *For each  $\xi_2 \in \mathcal{E}_2$ , the following conditional expectation*

$$\sup_{x_1 \in X_1, x_2 \in X_2} \mathbb{E}_{\xi_3 | \xi_2} [\|\Phi_2(x_1, x_2, \hat{x}_3(x_2, \xi_{[3]}), \xi_3)\|]$$

*is finite.*

The following lemma is a uniform convergence result.

**Lemma 3** *Let  $f : X \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a continuous mapping with  $X$  being a compact set and  $f_k : \mathbb{R}^n \times \Omega \rightarrow \mathbb{R}^m$  for  $k \in \mathbb{N}$ . Suppose that: (i)  $\|f_k(x, \omega) - f_k(y, \omega)\| \leq L_k(\omega) \|x - y\|$  for any  $x, y \in X$  and  $k \in \mathbb{N}$  with  $L_k(\omega) \rightarrow L$  with probability 1 (w.p.1) as  $k \rightarrow \infty$ , where  $L$  is a positive constant; (ii) for any fixed  $x \in X$ ,*

$f_k(x, \omega) \rightarrow f(x)$  w.p.1 as  $k \rightarrow \infty$ . Then  $f_k(x, \omega) \rightarrow f(x)$  w.p.1 uniformly w.r.t.  $x$  over  $X$  as  $k \rightarrow \infty$ , i.e.,

$$\lim_{k \rightarrow \infty} \sup_{x \in X} |f_k(x, \omega) - f(x)| = 0 \text{ w.p.1.}$$

*Proof* Choose now a point  $\bar{x} \in X$ . For any  $\epsilon > 0$ , due to assumption (i), there exists a neighborhood  $W$  of  $\bar{x}$  such that w.p.1 for sufficiently large  $k$ ,

$$\sup_{x \in W \cap X} \|f_k(x, \omega) - f_k(\bar{x}, \omega)\| \leq L_k(\omega) \cdot \text{diam}(W) \leq \epsilon, \tag{15}$$

where  $\text{diam}(W) := \sup_{w_1, w_2 \in W} \|w_1 - w_2\|$  is the diameter of  $W$ .

Due to the compactness of  $X$ , by finite covering theorem, we can select finite points  $x_1, \dots, x_q$  and corresponding neighborhoods  $W_1, \dots, W_q$  covering  $X$  such that w.p.1 for large enough  $k$ , we have

$$\sup_{x \in W_j \cap X} \|f_k(x, \omega) - f_k(x_j, \omega)\| \leq L_k(\omega) \cdot \text{diam}(W_j) \leq \epsilon, \quad j = 1, \dots, q. \tag{16}$$

Since  $f$  is continuous, neighborhoods of  $x_j, j = 1, \dots, q$  can be chosen in such a way that

$$\sup_{x \in W_j \cap X} \|f(x) - f(x_j)\| \leq \epsilon, \quad j = 1, \dots, q. \tag{17}$$

We also know from assumption (ii) that w.p.1 for sufficiently large  $k$ ,

$$\max_{1 \leq j \leq q} \|f_k(x_j, \omega) - f(x_j)\| \leq \epsilon, \quad j = 1, \dots, q. \tag{18}$$

For any  $x \in X$ , denote by  $j(x) \in \{1, \dots, q\}$  the index such that  $x \in W_{j(x)}$ . Then we have from (16) to (18) that w.p.1 for sufficiently large  $k$ ,

$$\begin{aligned} & \|f_k(x, \omega) - f(x)\| \\ &= \|f_k(x, \omega) - f(x) + f_k(x_{j(x)}, \omega) - f_k(x_{j(x)}, \omega) + f(x_{j(x)}) - f(x_{j(x)})\| \\ &\leq \|f_k(x, \omega) - f_k(x_{j(x)}, \omega)\| + \|f(x) - f(x_{j(x)})\| + \|f_k(x_{j(x)}, \omega) - f(x_{j(x)})\| \\ &\leq 3\epsilon. \end{aligned}$$

Due to the arbitrariness of  $x$  and  $\epsilon$ , we complete the proof. □

**Proposition 3** *Suppose that: (i) Assumptions 2 and 3 hold; (ii)  $\Phi_2(\cdot, \cdot, \cdot, \xi_3)$  and  $\Phi_3(\cdot, \cdot, \xi_3)$  are continuously differentiable and Lipschitz continuous with Lipschitz moduli  $L_2(\xi_3)$  and  $L_3(\xi_3)$  respectively for a.e.  $\xi_3 \in \Xi_3$ .*

(a) For each  $\xi_2^j \in \Xi_2$ ,

$$\begin{aligned} & \sup_{x_1 \in X_1, x_2 \in X_2} \left( \frac{1}{N_2} \sum_{k=1}^{N_2} \Phi_2(x_1, x_2, \hat{x}_3(x_2, \xi_{[3]}^{jk}), \xi_3^{jk}) \right. \\ & \quad \left. - \mathbb{E}_{\xi_3 | \xi_2^j} [\Phi_2(x_1, x_2, \hat{x}_3(x_2, \xi_{[3]}), \xi_3)] \right) \rightarrow 0 \end{aligned}$$

w.p.1 as  $N_2 \rightarrow \infty$ ,

(b) If, in addition, (iii) for every  $\xi_2^j \in \mathcal{E}_2$ ,

$$\mathbb{E}_{\xi_3|\xi_2^j}[L_2(\xi_3)] < \infty, \mathbb{E}_{\xi_3|\xi_2^j}[\eta(\xi_{[3]})] < \infty \text{ and } \mathbb{E}_{\xi_3|\xi_2^j} \left[ \frac{L_2(\xi_3)L_3(\xi_3)}{\eta(\xi_{[3]})} \right] < \infty,$$

then for every  $\xi_2^j \in \mathcal{E}_2$ , we have

$$\sup_{x_1 \in X_1} \left\| \tilde{x}_2^{N_2}(x_1, \xi_2^j) - \hat{x}_2(x_1, \xi_2^j) \right\| \rightarrow 0$$

w.p.1 as  $N_2 \rightarrow \infty$ .

*Proof* We know from Corollary 1 that  $\hat{x}_3(\cdot, \xi_3)$  is locally Lipschitz continuous. This together with the locally Lipschitz continuity of  $\Phi_2(\cdot, \cdot, \cdot, \xi_3)$  and the compactness of  $X_1$  and  $X_2$  result in (a) by directly using [28, Theorem 7.53].

In the rest of the proof, we focus on part (b). We prove it based on Lemma 3. Thus, we first verify the pointwise convergence. Let  $\{\tilde{x}_2^{N_2}(x_1, \xi_2^j)\}$  be the solution sequence of the second-stage problem of (14) with given  $x_1, \xi_2^j$  and samples  $\{\xi_3^{jk}\}_{k=1}^{N_2}$  and  $\bar{x}_2$  be any accumulation point of  $\{\tilde{x}_2^{N_2}(x_1, \xi_2^j)\}$  due to (A1) of Assumption 3. According to part (a), we have, for fixed  $x_1 \in X_1$  and  $\xi_2^j \in \mathcal{E}_2$ , that

$$\sup_{x_2 \in X_2} \left( \frac{1}{N_2} \sum_{k=1}^{N_2} \Phi_2(x_1, x_2, \hat{x}_3(x_2, \xi_3^{jk}), \xi_3^{jk}) - \mathbb{E}_{\xi_3|\xi_2^j} [\Phi_2(x_1, x_2, \hat{x}_3(x_2, \xi_{[3]}), \xi_3)] \right) \rightarrow 0$$

w.p.1 as  $N_2 \rightarrow \infty$ . This implies, for fixed  $x_1 \in X_1$  and  $\xi_2^j \in \mathcal{E}_2$ , that  $\tilde{x}_2^{N_2}(x_1, \xi_2^j) \rightarrow \hat{x}_2(x_1, \xi_2^j)$  w.p.1 as  $N_2 \rightarrow \infty$  (see, e.g., [30, Proposition 2.1]). Next, we verify this pointwise convergence is uniform w.r.t.  $x_1$  over  $X_1$ .

To this end, we first consider the Lipschitz modulus of  $\hat{x}_3(\cdot, \xi_3^{jk})$ , which is the upper bound of the Clarke generalized Jacobian  $\partial_{x_2} \hat{x}_3(x_2, \xi_3^{jk})$ .

By directly using (6) in Theorem 1, we have

$$\partial_{x_2} \hat{x}_3(x_2, \xi_3^{jk})v \subseteq \bigcup_{(\mathbf{C}, \mathbf{D}) \in \partial \Phi_3(x_2, \hat{x}_3(x_2, \xi_3^{jk}), \xi_3^{jk})} -(I - \mathcal{D}_{\mathcal{L}} + \mathcal{D}_{\mathcal{L}}\mathbf{D})^{-1} \mathcal{D}_{\mathcal{L}}\mathbf{C}v$$

for any  $v \in \mathbb{R}^n$ . It means that

$$\begin{aligned} \sup_{\Delta \in \partial_{x_2} \hat{x}_3(x_2, \xi_3^{jk})} \|\Delta\| &\leq \sup_{(\mathbf{C}, \mathbf{D}) \in \partial \Phi_3(x_2, \hat{x}_3(x_2, \xi_3^{jk}), \xi_3^{jk})} \|(I - \mathcal{D}_{\mathcal{L}} + \mathcal{D}_{\mathcal{L}}\mathbf{D})^{-1} \mathcal{D}_{\mathcal{L}}\mathbf{C}\| \\ &\leq \sup_{(\mathbf{C}, \mathbf{D}) \in \partial \Phi_3(x_2, \hat{x}_3(x_2, \xi_3^{jk}), \xi_3^{jk})} \|(I - \mathcal{D}_{\mathcal{L}} + \mathcal{D}_{\mathcal{L}}\mathbf{D})^{-1} \mathcal{D}_{\mathcal{L}}\| \|\mathbf{C}\|. \end{aligned}$$

We know from Proposition 1 and Assumption 2 that  $\mathbf{D}$  is strongly monotone (positive definite) w.r.t.  $\eta(\xi_{[3]}^{jk})$ . Then we obtain from [6, Lemma 2.1] that

$$\|(I - \mathcal{D}_{\mathcal{L}} + \mathcal{D}_{\mathcal{L}}\mathbf{D})^{-1} \mathcal{D}_{\mathcal{L}}\| \leq \frac{1}{\eta(\xi_{[3]}^{jk})}.$$

Note also that  $\|C\|$  is bounded from above by the Lipschitz modulus of  $\Phi_3(\cdot, \cdot, \xi_3^{jk})$  based on the definition of Clarke generalized Jacobian. Then  $\hat{x}_3(\cdot, \xi_3^{jk})$  is Lipschitz continuous with modulus  $\frac{L_3(\xi_3^{jk})}{\eta(\xi_{[3]}^{jk})}$ . Since  $X_2$  is convex and compact, we have  $\hat{x}_3(x_2, \xi_3^{jk})$  is Lipschitz continuous over  $X_2$  with modulus  $\frac{L_3(\xi_3^{jk})}{\eta(\xi_{[3]}^{jk})}$ .

According to the Lipschitz continuity of  $\Phi_2(\cdot, \cdot, \cdot, \xi_{[3]}^{jk})$ , a Lipschitz modulus of term

$$\frac{1}{N_2} \sum_{k=1}^{N_2} \Phi_2(\cdot, \cdot, \hat{x}_3(\cdot, \xi_{[3]}^{jk}), \xi_3^{jk})$$

is  $\frac{1}{N_2} \sum_{k=1}^{N_2} L_2(\xi_3^{jk}) \left(1 + \frac{L_3(\xi_3^{jk})}{\eta(\xi_{[3]}^{jk})}\right)$ .

Moreover, by Assumption 2,  $\frac{1}{N_2} \sum_{k=1}^{N_2} \Phi_2(x_1, x_2, \hat{x}_3(x_2, \xi_3^{jk}), \xi_3^{jk})$  is strongly monotone with positive constant  $\frac{1}{N_2} \sum_{k=1}^{N_2} \eta(\xi_{[3]}^{jk})$ . Finally, by the same procedure as above, we obtain a Lipschitz modulus of  $\tilde{x}_2^{N_2}(\cdot, \xi_2^j)$  is

$$\frac{\frac{1}{N_2} \sum_{k=1}^{N_2} L_2(\xi_3^{jk}) \left(1 + \frac{L_3(\xi_3^{jk})}{\eta(\xi_{[3]}^{jk})}\right)}{\frac{1}{N_2} \sum_{k=1}^{N_2} \eta(\xi_{[3]}^{jk})},$$

which can be bounded from above by

$$\frac{\mathbb{E}_{\xi_3|\xi_2^j} \left[ L_2(\xi_3) \left(1 + \frac{L_3(\xi_3)}{\eta(\xi_{[3]})}\right) \right] + 1}{\mathbb{E}_{\xi_3|\xi_2^j} [\eta(\xi_{[3]})]}$$

w.p.1 as  $N_2 \rightarrow \infty$ . Then by directly using Lemma 3, we conclude (b). □

**Proposition 4** *Suppose that: (i) Assumptions 2 and 3 hold; (ii)  $\Phi_1(\cdot, \cdot, \xi_2)$ ,  $\Phi_2(\cdot, \cdot, \cdot, \xi_3)$  and  $\Phi_3(\cdot, \cdot, \xi_3)$  are continuously differentiable and Lipschitz continuous for a.e.  $\xi_2 \in \Xi_3$  and  $\xi_3 \in \Xi_3$  with Lipschitz moduli  $L_1(\xi_2)$ ,  $L_2(\xi_3)$  and  $L_3(\xi_3)$  respectively, and*

$$\mathbb{E}_{\xi_2} \left[ L_1(\xi_2) \mathbb{E}_{\xi_3|\xi_2} \left[ \frac{L_2(\xi_3) \left(1 + \frac{L_3(\xi_3)}{\eta(\xi_{[3]})}\right)}{\eta(\xi_{[3]}^{ik})} + 1 \right] \right] < \infty.$$

Then we have

$$\sup_{x_1 \in X_1} \left\| \frac{1}{N_1} \sum_{i=1}^{N_1} \Phi_1(x_1, \tilde{x}_2^{N_2}(x_1, \xi_2^i), \xi_2^i) - \mathbb{E}_{\xi_2} [\Phi_1(x_1, \hat{x}_2(x_1, \xi_{[2]}), \xi_2)] \right\| \rightarrow 0 \quad (19)$$

w.p.1 as  $N_2 \rightarrow \infty$  and  $N_1 \rightarrow \infty$ .

*Proof* We prove pointwise convergence firstly. For any fixed  $x_1 \in X_1$ ,

$$\begin{aligned} & \left\| \frac{1}{N_1} \sum_{i=1}^{N_1} \Phi_1(x_1, \tilde{x}_2^{N_2}(x_1, \xi_2^i), \xi_2^i) - \mathbb{E}_{\xi_2}[\Phi_1(x_1, \hat{x}_2(x_1, \xi_{[2]}), \xi_2)] \right\| \\ & \leq \left\| \frac{1}{N_1} \sum_{i=1}^{N_1} \left( \Phi_1(x_1, \tilde{x}_2^{N_2}(x_1, \xi_2^i), \xi_2^i) - \Phi_1(x_1, \hat{x}_2(x_1, \xi_2^i), \xi_2^i) \right) \right\| \tag{20} \\ & \quad + \left\| \frac{1}{N_1} \sum_{i=1}^{N_1} \Phi_1(x_1, \hat{x}_2(x_1, \xi_2^i), \xi_2^i) - \mathbb{E}_{\xi_2}[\Phi_1(x_1, \hat{x}_2(x_1, \xi_{[2]}), \xi_2)] \right\|. \end{aligned}$$

For the first term of the right-hand side of (20), we have

$$\begin{aligned} & \left\| \frac{1}{N_1} \sum_{i=1}^{N_1} \left( \Phi_1(x_1, \tilde{x}_2^{N_2}(x_1, \xi_2^i), \xi_2^i) - \Phi_1(x_1, \hat{x}_2(x_1, \xi_2^i), \xi_2^i) \right) \right\| \\ & \leq \frac{1}{N_1} \sum_{i=1}^{N_1} L_1(\xi_2^i) \left\| \tilde{x}_2^{N_2}(x_1, \xi_2^i) - \hat{x}_2(x_1, \xi_2^i) \right\|. \end{aligned}$$

It knows from Proposition 3 that

$$L_1(\xi_2^i) \left\| \tilde{x}_2^{N_2}(x_1, \xi_2^i) - \hat{x}_2(x_1, \xi_2^i) \right\| \rightarrow 0$$

w.p.1 as  $N_2 \rightarrow \infty$ , and

$$L_1(\xi_2^i) \left\| \tilde{x}_2^{N_2}(x_1, \xi_2^i) - \hat{x}_2(x_1, \xi_2^i) \right\| \leq L_1(\xi_2^i) \cdot \text{diam}(X_2),$$

where  $\text{diam}(X_2)$  denotes the diameter of  $X_2$ . Then, by Lebesgue’s dominated convergence theorem, we have

$$\lim_{N_2 \rightarrow \infty} \mathbb{E}_{\xi_2^i} \left[ L_1(\xi_2^i) \left\| \tilde{x}_2^{N_2}(x_1, \xi_2^i) - \hat{x}_2(x_1, \xi_2^i) \right\| \right] = 0.$$

Thus, by law of large numbers (LLN), as  $N_1 \rightarrow \infty$  and  $N_2 \rightarrow \infty$ , it holds pointwisely that

$$\frac{1}{N_1} \sum_{i=1}^{N_1} L_1(\xi_2^i) \left\| \tilde{x}_2^{N_2}(x_1, \xi_2^i) - \hat{x}_2(x_1, \xi_2^i) \right\| \rightarrow 0.$$

Analogously, by LLN,

$$\left\| \frac{1}{N_1} \sum_{i=1}^{N_1} \Phi_1(x_1, \hat{x}_2(x_1, \xi_2^i), \xi_2^i) - \mathbb{E}_{\xi_2}[\Phi_1(x_1, \hat{x}_2(x_1, \xi_{[2]}), \xi_2)] \right\| \rightarrow 0$$

as  $N_1 \rightarrow \infty$  pointwisely. Then for any fixed  $x_1 \in X_1$ ,

$$\left\| \frac{1}{N_1} \sum_{i=1}^{N_1} \Phi_1(x_1, \tilde{x}_2^{N_2}(x_1, \xi_2^i), \xi_2^i) - \mathbb{E}_{\xi_2}[\Phi_1(x_1, \hat{x}_2(x_1, \xi_{[2]}), \xi_2)] \right\| \rightarrow 0$$

as  $N_1 \rightarrow \infty$  and  $N_2 \rightarrow \infty$ .

Note that  $\frac{1}{N_1} \sum_{i=1}^{N_1} \Phi_1(\cdot, \tilde{x}_2^{N_2}(\cdot, \xi_2^i), \xi_2^i)$  is Lipschitz continuous with Lipschitz modulus being

$$\frac{1}{N_1} \sum_{i=1}^{N_1} L_1(\xi_2^i) \left( \frac{1}{N_2} \sum_{k=1}^{N_2} \frac{L_2(\xi_3^{ik}) \left( 1 + \frac{L_3(\xi_3^{ik})}{\eta(\xi_{[3]}^{ik})} \right)}{\eta(\xi_{[3]}^{ik})} + 1 \right),$$

which converges to a finite number w.p.1 due to LLN. Then we complete the proof by quoting Lemma 3. □

Immediately, we can derive the following asymptotic convergence result by using Proposition 4 (see, e.g., [30, Proposition 2.1]).

**Theorem 4** *Under the same conditions of Proposition 4, we have*

$$\tilde{x}_1^{N_1, N_2} \rightarrow x_1^*$$

w.p.1 as  $N_1 \rightarrow \infty$  and  $N_2 \rightarrow \infty$ .

We make the following remarks.

*Remark 2* From the above discussion, it can be observed that the extension of convergence analysis of two-stage SNCs to that of three-stage SNCs is not trivial. The key reason is that the second-stage problem in (13) changes when the conditional SAA procedure is conducted (see (14)), which makes the SAA convergence of the first-stage problem of (14) to that of (13) unusual.

We consider the convergence analysis of the SAA approach of three-stage SNC for the following reasons: (1) Similar to multistage stochastic optimization [26], the extension of the convergence analysis from two-stage SNCs to three-stage SNCs already demonstrates the main difficulties of extending the convergence analysis from two-stage SNCs to multistage SNCs. It is not difficult to extend our results to the general  $T$ -stage SNCs for  $T > 3$  by a similar procedure. (2) The discussion of three-stage SNCs simplifies the notation.

*Remark 3* In this section, we concentrate on convergence properties of the SAA approach for a class of multistage SNCs (three-stage SNCs whose decision variable of each stage is influenced only by the decision variables of adjoining stages). This is mainly motivated by some practical problems, like inventory problems, and multistage stochastic optimization problems (see, e.g., [28, Chapter 3]). It should be pointed out that there are some challenges to extend our results to the case of the general multistage SNCs (1).

We use the following general three-stage SNC (by letting  $T = 3$  in (1)) to explain these challenges:

$$\begin{cases} 0 \in \mathbb{E}_{\xi_{[1]}} [\Phi_1(x_1, x_2(\xi_{[2]}), x_3(\xi_{[3]}), \xi_{[3]})] + \mathcal{N}_{\mathbb{R}_+^n}(x_1), \\ 0 \in \mathbb{E}_{\xi_{[1]}\xi_{[2]}} [\Phi_2(x_1, x_2(\xi_{[2]}), x_3(\xi_{[3]}), \xi_{[3]})] + \mathcal{N}_{\mathbb{R}_+^n}(x_2(\xi_{[2]})), \text{ a.e. } \xi_{[2]} \in \mathcal{E}_{[2]}, \\ 0 \in \Phi_3(x_1, x_2(\xi_{[2]}), x_3(\xi_{[3]}), \xi_{[3]}) + \mathcal{N}_{\mathbb{R}_+^n}(x_3(\xi_{[3]})), \text{ a.e. } \xi_{[3]} \in \mathcal{E}_{[3]}. \end{cases} \quad (21)$$

Under certain conditions (see, e.g., Lemma 2), we have the third-stage problem of (21) with given  $(x_1, x_2, \xi_{[3]})$  has a unique and locally Lipschitz continuous solution  $\hat{x}_3(x_1, x_2, \xi_{[3]})$ . Taking it into the first and second stages, we rewrite problem (21) as

$$\begin{cases} 0 \in \mathbb{E}_{\xi_1|\xi_1}[\Phi_1(x_1, x_2(\xi_{[2]}), \hat{x}_3(x_1, x_2(\xi_{[2]}), \xi_{[3]}), \xi_{[3]})] + \mathcal{N}_{\mathbb{R}_+^q}(x_1), \\ 0 \in \mathbb{E}_{\xi_1|\xi_{[2]}}[\Phi_2(x_1, x_2(\xi_{[2]}), \hat{x}_3(x_1, x_2(\xi_{[2]}), \xi_{[3]}), \xi_{[3]})] + \mathcal{N}_{\mathbb{R}_+^q}(x_2(\xi_{[2]})), \\ \text{a.e. } \xi_{[2]} \in \mathcal{E}_{[2]}. \end{cases} \quad (22)$$

By a similar procedure as that in Theorem 1, the second-stage problem of (22) can have a unique and locally Lipschitz continuous solution  $\hat{x}_2(x_1, \xi_{[2]})$ . Submitting it into the first stage of (22), we obtain

$$0 \in \mathbb{E}_{\xi_1|\xi_1}[\Phi_1(x_1, \hat{x}_2(x_1, \xi_{[2]}), \hat{x}_3(x_1, \hat{x}_2(x_1, \xi_{[2]}), \xi_{[3]}), \xi_{[3]})] + \mathcal{N}_{\mathbb{R}_+^q}(x_1). \quad (23)$$

It can observe from problem (23) that the mapping in problem (23) is highly composite w.r.t.  $x_1$ . Moreover, our analysis can not apply to this problem directly, and it will become more complicated as the number of stages increases.

### 4 Numerical results

In this section, we extend the two-stage stochastic non-cooperative game problem [6, 16, 20, 31] to a three-stage case, and apply PHM to solve it. The numerical results are used to illustrate the theoretical results.

Let  $\xi_2 : \Omega_2 \rightarrow \mathcal{E}_2 \subseteq \mathbb{R}^2$  and  $\xi_3 : \Omega_3 \rightarrow \mathcal{E}_3 \subseteq \mathbb{R}^2$  be two independent random vectors. For  $i = 1, 2$ ,  $x_i \in \mathbb{R}^{n_i}$ ,  $y_i : \mathcal{E}_2 \rightarrow \mathbb{R}^{m_i}$  and  $z_i : \mathcal{E}_2 \times \mathcal{E}_3 \rightarrow \mathbb{R}^{l_i}$  be the decision variables of the  $i$ th player at the first, second, and third stages, respectively, with  $n_i = m_i = l_i = 3$ . Denote  $n = n_1 + n_2$ ,  $m = m_1 + m_2$  and  $l = l_1 + l_2$ . By convention, we also use  $x_{-i}$ ,  $y_{-i}$  and  $z_{-i}$  to denote the rival’s decisions for  $i = 1, 2$ . In this game, the  $i$ th ( $i = 1, 2$ ) player solves the following optimization problem:

$$\min_{x_i \geq 0} \theta_i(x_i, x_{-i}) + \mathbb{E}_{\xi_2}[\psi_i(x_i, y_{-i}(\xi_2), \xi_2)], \quad (24)$$

where  $\theta_i : \mathbb{R}^{n_i} \times \mathbb{R}^{n-i} \rightarrow \mathbb{R}$ ,  $\psi_i : \mathbb{R}^{n_i} \times \mathbb{R}^{m-i} \times \mathcal{E}_2 \rightarrow \mathbb{R}$  is defined as

$$\begin{aligned} \psi_i(x_i, y_{-i}(\xi_2), \xi_2) := & \min_{y_i \geq 0} \phi_i(y_i, x_i, y_{-i}(\xi_2), \xi_2) \\ & + \mathbb{E}_{\xi_3|\xi_2}[\varphi_i(y_i, z_{-i}(\xi_2, \xi_3), \xi_2, \xi_3)], \end{aligned} \quad (25)$$

here  $\phi_i : \mathbb{R}^{m_i} \times \mathbb{R}^{n_i} \times \mathbb{R}^{m-i} \times \mathcal{E}_2 \rightarrow \mathbb{R}$  and  $\varphi_i : \mathbb{R}^{m_i} \times \mathbb{R}^{l-i} \times \mathcal{E}_2 \times \mathcal{E}_3 \rightarrow \mathbb{R}$  is the optimal value function of the third-stage problem, that is

$$\varphi_i(y_i(\xi_2), z_{-i}(\xi_2, \xi_3), \xi_2, \xi_3) := \min_{z_i \geq 0} \varsigma_i(z_i, y_i(\xi_2), z_{-i}(\xi_2, \xi_3), \xi_2, \xi_3), \quad (26)$$

with  $\varsigma_i : \mathbb{R}^{l_i} \times \mathbb{R}^{m_i} \times \mathbb{R}^{l-i} \times \mathcal{E}_2 \times \mathcal{E}_3 \rightarrow \mathbb{R}$ .

To proceed, we consider the strongly convex quadratic game (see also [4, 31]), that is

$$\begin{aligned} \theta_i(x_i, x_{-i}) &= \frac{1}{2}x_i^\top H_i x_i + q_i^\top x_i + x_i^\top P_i x_{-i}, \\ \phi_i(y_i, x_i, y_{-i}(\xi_2), \xi_2) &= \frac{1}{2}y_i^\top Q_i^1(\xi_2)y_i + c_i^1(\xi_2)^\top y_i \\ &\quad + \sum_{j=1}^2 y_i^\top S_{ij}^1(\xi_2)x_j + y_i^\top O_i^1(\xi_2)y_{-i}(\xi_2), \\ \varsigma_i(z_i, y_i(\xi_2), z_{-i}(\xi_2, \xi_3), \xi_2, \xi_3) &= \frac{1}{2}z_i^\top Q_i^2(\xi_2, \xi_3)z_i + c_i^2(\xi_2, \xi_3)^\top z_i \\ &\quad + \sum_{j=1}^2 z_i^\top S_{ij}^2(\xi_2, \xi_3)y_j(\xi_2) + z_i^\top O_i^2(\xi_2, \xi_3)z_{-i}(\xi_2, \xi_3), \end{aligned}$$

where  $H_i \in \mathbb{R}^{n_i \times n_i}$ ,  $Q_i^1(\xi_2) \in \mathbb{R}^{m_i \times m_i}$  and  $Q_i^2(\xi_2, \xi_3) \in \mathbb{R}^{l_i \times l_i}$  are symmetric positive definite matrices,  $q_i \in \mathbb{R}^{n_i}$ ,  $P_i \in \mathbb{R}^{n_i \times n_{-i}}$ ,  $c_i^1(\xi_2) \in \mathbb{R}^{m_i}$ ,  $S_{ij}^1(\xi_2) \in \mathbb{R}^{m_i \times n_j}$ ,  $O_i^1(\xi_2) \in \mathbb{R}^{m_i \times m_{-i}}$ ,  $c_i^2(\xi_2, \xi_3) \in \mathbb{R}^{l_i}$ ,  $S_{ij}^2 \in \mathbb{R}^{l_i \times m_j}$ ,  $S_{ij}^2 \in \mathbb{R}^{l_i \times m_j}$  and  $O_i^2(\xi_2, \xi_3) \in \mathbb{R}^{l_i \times l_{-i}}$ .

Denote  $x = (x_1, x_2)$ ,  $y(\cdot) = (y_1(\cdot), y_2(\cdot))$  and  $z(\cdot, \cdot) = (z_1(\cdot, \cdot), z_2(\cdot, \cdot))$ . Due to the strong convexity, problems (25) and (26) have unique solutions, denoted by  $\bar{y}_i(\xi_2)$  and  $\bar{z}_i(\xi_2, \xi_3)$  respectively.

By [13, Theorem 5.3 and Corollary 5.4],

$$\psi_i(x_i, y_{-i}(\xi_2), \xi_2) \text{ and } \varphi_i(y_i, z_{-i}(\xi_2, \xi_3), \xi_2, \xi_3)$$

are continuously differentiable w.r.t.  $x_i$  and  $y_i$ , respectively, and their gradients are

$$\nabla_{x_i} \psi_i(x_i, y_{-i}(\xi_2), \xi_2) = (S_{ii}^1(\xi_2))^\top \bar{y}_i(\xi_2)$$

and

$$\nabla_{y_i} \varphi_i(y_i(\xi_2), z_{-i}(\xi_2, \xi_3), \xi_2, \xi_3) = (S_{ii}^2(\xi_2, \xi_3))^\top \bar{z}_i(\xi_2, \xi_3).$$

Hence, the three-stage stochastic game problem can be reformulated as a three-stage SVI (it can be rewritten as a three-stage stochastic linear complementarity problem): for  $i = 1, 2$ ,

$$\begin{aligned} 0 &\in A^1 x + \mathbb{E}_{\xi_2}[B^1(\xi_2)y(\xi_2)] + h_1 + \mathcal{N}_{\mathbb{R}_+^m}(x), \\ 0 &\in A^2(\xi_2)y(\xi_2) + L^2(\xi_2)x + \mathbb{E}_{\xi_3}[B^2(\xi_3)z(\xi_2, \xi_3)] + h_2(\xi_2) + \mathcal{N}_{\mathbb{R}_+^m}(y(\xi_2)) \\ &\quad \text{for a.e. } \xi_2 \in \bar{\mathcal{E}}_2, \\ 0 &\in A^3(\xi_2, \xi_3)z(\xi_2, \xi_3) + L^3(\xi_2, \xi_3)y(\xi_2) + h_3(\xi_2, \xi_3) + \mathcal{N}_{\mathbb{R}_+^m}(z(\xi_2, \xi_3)) \\ &\quad \text{for a.e. } (\xi_2, \xi_3) \in \bar{\mathcal{E}}_2 \times \bar{\mathcal{E}}_3, \end{aligned} \tag{27}$$

where

$$\begin{aligned} A^1 &= \begin{pmatrix} H_1 & P_1 \\ P_2 & H_2 \end{pmatrix}, \quad B^1(\xi) = \begin{pmatrix} (S_{11}^1)^\top(\xi_2) & 0 \\ 0 & (S_{22}^1)^\top(\xi_2) \end{pmatrix}, \quad h_1 = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}, \\ A^2(\xi_2) &= \begin{pmatrix} Q_1^1(\xi_2) & O_1^1(\xi_2) \\ O_2^1(\xi_2) & Q_2^1(\xi_2) \end{pmatrix}, \quad L^2(\xi_2) = \begin{pmatrix} S_{11}^1(\xi_2) & S_{12}^1(\xi_2) \\ S_{21}^1(\xi_2) & S_{22}^1(\xi_2) \end{pmatrix}, \quad h_3(\xi_2, \xi_3) = \begin{pmatrix} c_1^2(\xi_2, \xi_3) \\ c_2^2(\xi_2, \xi_3) \end{pmatrix}, \end{aligned}$$

$$B^2(\xi_2, \xi_3) = \begin{pmatrix} (S_{11}^2)^\top(\xi_2, \xi_3) & 0 \\ 0 & (S_{22}^2)^\top(\xi_2, \xi_3) \end{pmatrix}, \quad h_2(\xi_1) = \begin{pmatrix} c_1^1(\xi_2) \\ c_2^1(\xi_2) \end{pmatrix},$$

$$A^3(\xi_2, \xi_3) = \begin{pmatrix} Q_1^2(\xi_2, \xi_3) & O_1^2(\xi_2, \xi_3) \\ O_2^2(\xi_2, \xi_3) & Q_2^2(\xi_2, \xi_3) \end{pmatrix} \text{ and } L^3(\xi_2, \xi_3) = \begin{pmatrix} S_{11}^2(\xi_2, \xi_3) & S_{12}^2(\xi_2, \xi_3) \\ S_{21}^2(\xi_2, \xi_3) & S_{22}^2(\xi_2, \xi_3) \end{pmatrix}.$$

Moreover, if there exists a positive continuous function  $\kappa(\xi_2, \xi_3)$  such that  $\mathbb{E}[\kappa(\xi_2, \xi_3)] < \infty$  and for a.e.  $(\xi_2, \xi_3) \in \mathcal{E}_2 \times \mathcal{E}_3$ ,

$$(\tau^\top, u^\top, k^\top) \begin{pmatrix} A^1 & B^1(\xi_2) \\ L^2(\xi_2) & A^2(\xi_2) & B^2(\xi_2) \\ L^3(\xi_2, \xi_3) & A^3(\xi_2, \xi_3) \end{pmatrix} \begin{pmatrix} \tau \\ u \\ k \end{pmatrix} \geq \kappa(\xi_2, \xi_3)(\|\tau\|^2 + \|u\|^2 + \|k\|^2), \tag{28}$$

then for each  $(\tau^\top, u^\top, k^\top)^\top \in \mathbb{R}^{m+n+l}$ , the three-stage SVI (27) satisfies Assumption 2. Moreover, if  $\mathcal{E}_2 \times \mathcal{E}_3$  is compact, by Assumption 2, the conditions in Theorem 4 hold.

In this example, we consider a stage-wise independent case, that is  $\xi_2$  and  $\xi_3$  are independent. Let  $\{\xi_2^i\}_{i=1}^{N_1}$  be an i.i.d sample of the second-stage random variable  $\xi_2$  and  $\{\xi_3^j\}_{j=1}^{N_2}$  be an i.i.d sample of the third stage of  $\xi_3$  when given the value  $\xi_2^i$ . Then the SAA problem of problem (27) is

$$\begin{aligned} 0 &\in A^1x + \frac{1}{N_1} \sum_{i=1}^{N_1} B^1(\xi_2^i)y(\xi_2^i) + h_1 + \mathcal{N}_{\mathbb{R}_+^n}(x), \\ 0 &\in A^2(\xi_2^i)y(\xi_2^j) + L^2(\xi_2^i)x + \frac{1}{N_2} \sum_{j=1}^{N_2} B^2(\xi_3^j)z(\xi_2^i, \xi_3^j) + h_2(\xi_2^i) + \mathcal{N}_{\mathbb{R}_+^m}(y(\xi_2^j)), \\ &\text{for } i = 1 \dots N_1 \\ 0 &\in A^3(\xi_2^i, \xi_3^j)z(\xi_2^i, \xi_3^j) + L^3(\xi_2^i, \xi_3^j)y(\xi_2^i) + h_3(\xi_2^i, \xi_3^j) + \mathcal{N}_{\mathbb{R}_+^l}(z(\xi_2^i, \xi_3^j)) \\ &\text{for } i = 1 \dots N_1, j = 1 \dots N_2. \end{aligned} \tag{29}$$

### 4.1 Generation of matrices satisfying condition (28)

We first generate i.i.d. samples  $\{\xi_2^i\}_{i=1}^{N_1} \subset [0, 1]^2$  of random variable  $\xi_2 \in \mathbb{R}^2$  and  $\{\xi_3^j\}_{j=1}^{N_2} \subset [0, 1]^2$  of random variable  $\xi_3 \in \mathbb{R}^2$  from uniform distributions over  $\mathcal{E}_2 = [0, 1]^2$  and  $\mathcal{E}_3 = [0, 1]^2$ , respectively.

Note that PHM converges to a solution of (29) if condition (28) holds. To this end, for any given  $(\xi_2, \xi_3)$ , we generate matrices  $A^1, B^1(\xi_2), L^2(\xi_2), A^2(\xi_2), B^2(\xi_2), L^3(\xi_2, \xi_3), A^3(\xi_2, \xi_3)$  by the following procedure. To simplify notation, we use  $A^1, B^1, L^2, A^2, B^2, L^3, A^3$  to denote them. Let

$$E := \begin{pmatrix} I & 0 & 0 \\ 0 & 0 & I \\ -(B^1 + (L^2)^\top)^\top(A^1 + (A^1)^\top)^{-1} & I & -((B^2)^\top + L^3)(A^3 + (A^3)^\top)^{-1} \end{pmatrix}$$

and

$$U_1 := A^2 + (A^2)^\top - ((B^2)^\top + L^3)(A^3 + (A^3)^\top)^{-1}((B^2)^\top + L^3)^\top - ((B^1)^\top + L^2)(A^1 + (A^1)^\top)^{-1}((B^1)^\top + L^2)^\top.$$

Then

$$K := \begin{pmatrix} A^1 + (A^1)^\top & 0 & 0 \\ 0 & A^3 + (A^3)^\top & 0 \\ 0 & 0 & U_1 \end{pmatrix} = E \begin{pmatrix} A^1 + (A^1)^\top & B^1 + (L^2)^\top \\ (B^1)^\top + L^2 & A^2 + (A^2)^\top & B^2 + (L^3)^\top \\ (B^2)^\top + L^3 & A^3 + (A^3)^\top & \end{pmatrix} E^\top.$$

Since  $E$  is nonsingular when  $K$  is positive definite, condition (28) holds.

Let  $U_2 := 2H_2 - (P_2 + P_1^\top)^\top \frac{1}{2} H_1^{-1} (P_2 + P_1^\top)$ . Then

$$(A^1 + (A^1)^\top) = \begin{pmatrix} 2H_1 & P_1 + P_2^\top \\ P_2 + P_1^\top & 2H_2 \end{pmatrix} = \begin{pmatrix} I & 0 \\ (P_2 + P_1^\top)^\top \frac{1}{2} H_1^{-1} & I \end{pmatrix} \begin{pmatrix} 2H_1 & 0 \\ 0 & U_2 \end{pmatrix} \begin{pmatrix} I & 0 \\ (P_2 + P_1^\top)^\top \frac{1}{2} H_1^{-1} & I \end{pmatrix}^\top.$$

We first randomly generate a symmetric positive definite matrix  $H_1 \in \mathbb{R}^{n_1 \times n_1}$  and matrices  $P_1 \in \mathbb{R}^{n_1 \times n_2}$ ,  $P_2 \in \mathbb{R}^{n_2 \times n_1}$  with entries within  $[-1, 1]$ . Set  $H_2 = \frac{1}{4} H_1 (P_1^\top + P_2) H_1^{-1} (P_1 + P_2^\top) + \alpha I_{n_2}$ , where  $\alpha$  is a positive number (we set  $\alpha = 1$  in the numerical tests). Then  $\begin{pmatrix} 2H_1 & 0 \\ 0 & U_2 \end{pmatrix}$  is positive definite and then  $A^1 + (A^1)^\top$  is positive definite.

We then randomly generate two matrices  $O_1^2 \in \mathbb{R}^{l_1 \times l_2}$  and  $O_2^2 \in \mathbb{R}^{l_2 \times l_1}$  with entries within  $[-1, 1]$  and a symmetric matrix  $\bar{Q}_1^2 \in \mathbb{R}^{m_2 \times m_2}$  with diagonal entries greater than  $\max(m - 1 + \alpha, l - 1 + \alpha)$  and off-diagonal entries belonging to  $[-1, 1]$ . Let

$$Q_1^2 = \xi_2^{i(1)} \xi_3^{j(1)} \bar{Q}_1^2 + \alpha I_{l_1}, \quad Q_2^2 = \frac{1}{4} ((O_1^2)^\top + O_2^2) (Q_1^2)^{-1} (O_1^2 + (O_2^2)^\top) + \alpha I_{l_1}$$

and

$$U_3 = 2Q_2^2 - ((O_1^2)^\top + O_2^2) \frac{1}{2} (Q_1^2)^{-1} (O_1^2 + (O_2^2)^\top) = 2\alpha I_{l_1},$$

where  $\xi_2^{i(k)}$  and  $\xi_3^{j(k)}$  denote the  $k$ th component of vector  $\xi_2^i$  and vector  $\xi_3^j$ , respectively. Then

$$(A^3 + (A^3)^\top) = \begin{pmatrix} 2Q_1^2 & O_1^2 + (Q_2^2)^\top \\ O_2^2 + (O_1^2)^\top & 2Q_2^2 \end{pmatrix} = \begin{pmatrix} I & 0 \\ (O_2^2 + (O_1^2)^\top)^\top \frac{1}{2} (Q_1^2)^{-1} & I \end{pmatrix} \begin{pmatrix} 2Q_1^2 & 0 \\ 0 & U_3 \end{pmatrix} \begin{pmatrix} I & 0 \\ (O_2^2 + (O_1^2)^\top)^\top \frac{1}{2} (Q_1^2)^{-1} & I \end{pmatrix}^\top$$

is positive definite.

Finally, we generate a positive definite  $U_1$ . We randomly generate matrices

$$O_1^1 \in \mathbb{R}^{m_1 \times m_2}, O_2^1 \in \mathbb{R}^{m_2 \times m_1}, S_{11}^1 \in \mathbb{R}^{m_1 \times n_1}, S_{12}^1 \in \mathbb{R}^{m_1 \times n_2}, S_{21}^1 \in \mathbb{R}^{m_2 \times n_1},$$

$$S_{22}^1 \in \mathbb{R}^{m_2 \times n_2}, S_{11}^2 \in \mathbb{R}^{l_1 \times m_1}, S_{12}^2 \in \mathbb{R}^{l_1 \times m_2}, S_{21}^2 \in \mathbb{R}^{l_2 \times m_1}, S_{22}^2 \in \mathbb{R}^{l_2 \times m_2}$$

with entries belonging to  $[-1, 1]$ , and two symmetric matrices  $\bar{Q}_1^1 \in \mathbb{R}^{m_1 \times m_1}$  and  $\bar{Q}_2^1 \in \mathbb{R}^{m_2 \times m_2}$  with diagonal entries greater than  $\max(m - 1 + \alpha, l - 1 + \alpha)$  and off-diagonal entries belonging to  $[-1, 1]$ , respectively. Set

$$Q_1^1 = \bar{Q}_1^1 + \left( \xi_2^{i(1)} + \frac{(n+m)^2}{\lambda_{\min}(A^1 + (A^1)^\top)} + \frac{(m+l)^2}{\lambda_{\min}(A^3 + (A^3)^\top)} \right) I_{m_1}$$

and

$$Q_2^1 = \bar{Q}_2^1 + \left( \xi_2^{i(2)} + \frac{(n+m)^2}{\lambda_{\min}(A^1 + (A^1)^\top)} + \frac{(m+l)^2}{\lambda_{\min}(A^3 + (A^3)^\top)} \right) I_{m_2}.$$

Then

$$A^2 + (A^2)^\top = \begin{pmatrix} 2\bar{Q}_1^1 & O_1^1 + (O_1^1)^\top \\ O_2^1 + (O_1^1)^\top & 2\bar{Q}_2^1 \end{pmatrix} + \begin{pmatrix} 2(Q_1^1 - \bar{Q}_1^1) & 0 \\ 0 & 2(Q_2^1 - \bar{Q}_2^1) \end{pmatrix}$$

and  $\begin{pmatrix} 2\bar{Q}_1^1 & O_1^1 + (O_1^1)^\top \\ O_2^1 + (O_1^1)^\top & 2\bar{Q}_2^1 \end{pmatrix}$  is diagonally dominant and positive definite.

Moreover, for any  $y \in \mathbb{R}^{m_1+m_2}$ ,

$$y^\top (A^2 + (A^2)^\top) y \geq \left( 2\alpha + \frac{(n+m)^2}{\lambda_{\min}(A^1 + (A^1)^\top)} + \frac{(m+l)^2}{\lambda_{\min}(A^3 + (A^3)^\top)} \right) \|y\|^2.$$

Let  $U_4 = A^2 + (A^2)^\top - U_1$ . Note that

$$\|(B^2)^\top + L^3\|_2^2 \leq \|(B^2)^\top + L^3\|_1 \|(B^2)^\top + L^3\|_\infty \leq (m+n)^2$$

and

$$\|(B^1)^\top + L^2\|^2 \leq \|(B^1)^\top + L^2\|_1 \|(B^1)^\top + L^2\|_\infty \leq (m+l)^2.$$

We have

$$U_4 = ((B^2)^\top + L^3)(A^3 + (A^3)^\top)^{-1}((B^2)^\top + L^3)^\top + (B^1 + (L^2)^\top)^\top (A^1 + (A^1)^\top)^{-1} (B^1 + (L^2)^\top)$$

and

$$\begin{aligned} y^\top U_4 y &\leq \left( \frac{\|(B^2)^\top + L^3\|^2}{\lambda_{\min}(A^3 + (A^3)^\top)} + \frac{\|(B^1)^\top + L^2\|^2}{\lambda_{\min}(A^1 + (A^1)^\top)} \right) \|y\|^2 \\ &\leq \left( \frac{(n+m)^2}{\lambda_{\min}(A^1 + (A^1)^\top)} + \frac{(m+l)^2}{\lambda_{\min}(A^3 + (A^3)^\top)} \right) \|y\|^2, \end{aligned}$$

which implies

$$y^\top U_1 y = y^\top (A^2 + (A^2)^\top - U_4) y \geq 2\alpha \|y\|_2^2.$$

Thus,  $U_1$  is positive definite. Combining the above analysis,  $K$  is positive definite and condition (28) holds.

We also set  $h_1 = (-1, -1, -1, -3, -3, -3)^\top$ ,  $h_2(\xi^2) = (-1, -1, -1, -1, -1, -1)^\top$  and  $h_3(\xi^2, \xi^3) = (-1, -1, -1, -1, -1, -1)^\top$  for all  $\xi^2 \in \mathcal{E}_2$  and  $\xi^3 \in \mathcal{E}_3$ .

### 4.2 Numerical results

For each sample size  $(N_1, N_2) = (25, 25), (35, 35), (50, 50), (100, 100), (200, 200)$ , we randomly generate 30 test problems and solve problem (29) by PHM as follows (see [22] for more details).

---

#### Algorithm 1

---

(PHM) Given initial points  $\bar{x}^0 \in \mathbb{R}^n, x_{ij}^0 = x^0 \in \mathbb{R}^n, \bar{y}_i^0 \in \mathbb{R}^m, y_{ij}^0 = y_i^0, z_{ij}^0 \in \mathbb{R}^l$ , and  $w_{ij}^0 \in \mathbb{R}^n, s_{ij}^0 \in \mathbb{R}^m$ , such that  $\frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} w_{ij}^0 = 0$  and  $\frac{1}{N_2} \sum_{j=1}^{N_2} s_{ij}^0 = 0, i = 1, \dots, N_1$ . Let  $v = 0$ .

**Step 1.** For  $i = 1, \dots, N_1, j = 1, \dots, N_2$ , solve the LCP

$$\begin{aligned} 0 &\in A^1 x_{ij} + B^1(\xi_2^i) y_{ij} + h_1 + w_{ij}^v + r(x_{ij} - x_{ij}^v) + \mathcal{N}_{\mathbb{R}_+^n}(x_{ij}), \\ 0 &\in A^2(\xi_2^i) y_{ij} + L^2(\xi_2^j) x_{ij} + B^2(\xi_3^j) z_{ij} + h_2(\xi_2^i) + s_{ij}^v + r(y_{ij} - y_{ij}^v) + \mathcal{N}_{\mathbb{R}_+^m}(y_{ij}), \\ 0 &\in A^3(\xi_2^i, \xi_3^j) z_{ij} + L^3(\xi_2^i, \xi_3^j) y_{ij} + h_3(\xi_2^i, \xi_3^j) + r(z_{ij} - z_{ij}^v) \mathcal{N}_{\mathbb{R}_+^l}(z_{ij}) \end{aligned} \tag{30}$$

and obtain a solution  $(\hat{x}_{ij}^v, \hat{y}_{ij}^v, \hat{z}_{ij}^v), i = 1, \dots, N_1, j = 1, \dots, N_2$ .

**Step 2.** Let  $\bar{x}^{v+1} = \frac{1}{N_1 N_2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} \hat{x}_{ij}^v, \bar{y}_i^{v+1} = \frac{1}{N_2} \sum_{j=1}^{N_2} \hat{y}_{ij}^v, i = 1, \dots, N_1$ . For  $i = 1, \dots, N_1, j = 1, \dots, N_2$ , set

$$x_{ij}^{v+1} = \bar{x}^{v+1}, y_{ij}^{v+1} = \bar{y}_i^v, z_{ij}^{v+1} = \hat{z}_{ij}^v$$

and

$$w_{ij}^{v+1} = w_{ij}^v + r(\hat{x}_{ij}^v - x_{ij}^{v+1}), s_{ij}^{v+1} = s_{ij}^v + r(\hat{y}_{ij}^v - y_{ij}^{v+1}).$$

Set  $v = v + 1$ , go to Step 1.

---

We set the parameter  $r = 4$  in Algorithm 1, run the algorithm using Matlab 2018b on a desktop computer with two Intel Core i7 3.40GHz CPUs and 8GB RAM, and stop the iteration when the residual

$$\text{res} = \left\| x - \min \left( x - A^1 x - \frac{1}{N_1} \sum_{i=1}^{N_1} B^1(\xi_2^i) y(x, \xi_2^i), 0 \right) \right\| \leq 10^{-3},$$

where  $y(x, \xi_2^i)$  is the solution of the second-stage problem. Figure 1 shows the convergence tendency of  $x_1, x_2, x_3, x_4, x_5$  and  $x_6$ , respectively.

### 5 Concluding remarks

In this paper, the strong monotonicity of two-stage and multistage SNCs has been investigated under weaker conditions compared with the existing works, such as [4, 6]. Moreover, the conditional sampling approach has been adopted to discretize the multistage SNCs and the asymptotic convergence results have been established. Finally, some numerical tests have been presented through PHM to illustrate our theoretical results.

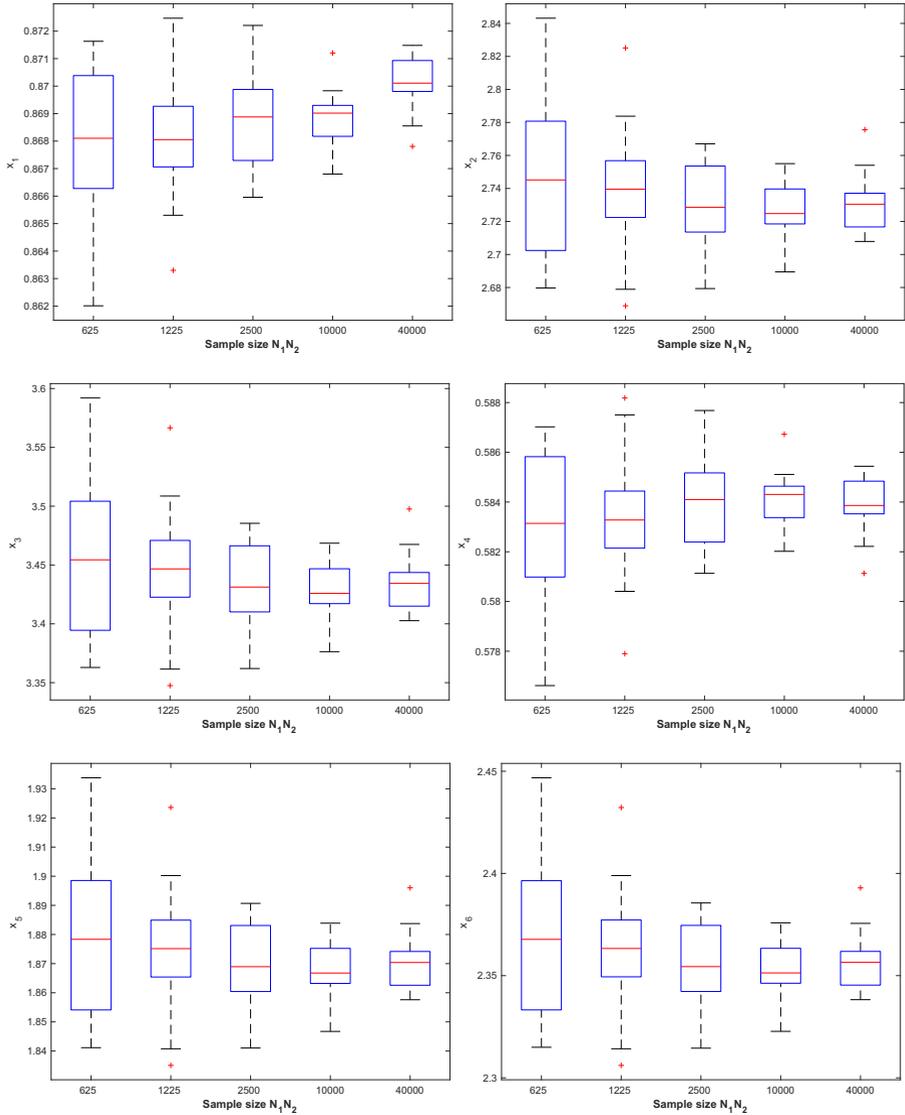


Fig. 1 The box plots from  $x_1$  to  $x_6$

There are still several questions that need to be addressed. We just analyze the asymptotic convergence of the conditional sampling approach for multistage SNCs. What is the rate of convergence of the conditional sampling approach for multistage SNCs? We leave this as our further research. Moreover, based on our numerical experience, the cost of solving three-stage SNCs in PHM is relatively expensive with large sample sizes.

However, large sample size and multistage  $T \geq 3$  may be necessary for some real applications. How to improve the performance of PHM or propose more effective algorithms is our another further research direction.

**Acknowledgements** The authors would like to thank Professor Alexander Shapiro for his enlightening discussion of this topic. They also thank two anonymous referees for their insightful and detailed comments, which improve this paper significantly.

**Funding** This work was supported by Postdoctoral Science Foundation of China (Grant No. 2020M673117) and National Natural Science Foundation of China (Grant No. 11871276).

## References

1. Birge, J.R., Louveaux, F.: Introduction to Stochastic Programming. Springer, Berlin (2011)
2. Chen, X., Fukushima, M.: Expected residual minimization method for stochastic linear complementarity problems. *Math. Oper. Res.* **30**(4), 1022–1038 (2005)
3. Chen, X., Pong, T.K., Wets, R.J.B.: Two-stage stochastic variational inequalities: an erm-solution procedure. *Math. Program.* **165**(1), 71–111 (2017)
4. Chen, X., Shapiro, A., Sun, H.: Convergence analysis of sample average approximation of two-stage stochastic generalized equations. *SIAM J. Optim.* **29**(1), 135–161 (2019)
5. Chen, X., Shi, Y., Wang, X.: Equilibrium oil market share under the COVID-19 pandemic To preprint on. arXiv (2020)
6. Chen, X., Sun, H., Xu, H.: Discrete approximation of two-stage stochastic and distributionally robust linear complementarity problems. *Math. Program.* **177**(1-2), 255–289 (2019)
7. Chen, X., Wets, R.J.B., Zhang, Y.: Stochastic variational inequalities: residual minimization smoothing sample average approximations. *SIAM J. Optim.* **22**(2), 649–673 (2012)
8. Chen, X., Xiang, S.: Newton iterations in implicit time-stepping scheme for differential linear complementarity systems. *Math. Program.* **138**(1-2), 579–606 (2013)
9. Chieu, N.H., Trang, N.: Coderivative and monotonicity of continuous mappings. *Taiwan. J. Math.* **16**(1), 353–365 (2012)
10. Clarke, F.H.: Optimization and Nonsmooth Analysis, Volume 5 of Classics in Applied Mathematics. SIAM, Philadelphia (1990)
11. Cottle, R.W., Pang, J.S., Stone, R.E.: The linear complementarity problem. SIAM, Philadelphia (2009)
12. Facchinei, F., Pang, J.S.: Finite-dimensional variational inequalities and complementarity problems. Springer Science & Business Media, New York (2007)
13. Gauvin, J., Dubeau, F.: Differential properties of the marginal function in mathematical programming. *Math. Programm. Stud.* **21**, 199–212 (1984)
14. Hadjisavvas, N., Komlósi, S., Schaible, S.S.: Handbook of Generalized Convexity and Generalized Monotonicity, vol. 76. Springer, New York (2006)
15. Izmailov, A.F.: Strongly regular nonsmooth generalized equations. *Math. Program.* **147**(1-2), 581–590 (2014)
16. Jiang, J., Chen, X., Chen, Z.: Quantitative analysis for a class of two-stage stochastic linear variational inequality problems. *Comput. Optim. Appl.* **76**(2), 431–460 (2020)
17. Jiang, J., Shi, Y., Wang, X., Chen, X.: Regularized two-stage stochastic variational inequalities for Cournot-Nash equilibrium under uncertainty. *J. Comput. Math.* **37**(6), 813–842 (2019)
18. Kyparisis, J.: Solution differentiability for variational inequalities. *Math. Program.* **48**(1-3), 285–301 (1990)
19. Lin, G., Fukushima, M.: Stochastic equilibrium problems and stochastic mathematical programs with equilibrium constraints: a survey. *Pac. J. Optim.* **6**, 455–482 (2010)
20. Pang, J.S., Sen, S., Shanbhag, U.V.: Two-stage non-cooperative games with risk-averse players. *Math. Program.* **165**(1), 235–290 (2017)
21. Reaiche, M.: A note on sample complexity of multistage stochastic programs. *Oper. Res. Lett.* **44**(4), 430–435 (2016)

22. Rockafellar, R.T., Sun, J.: Solving monotone stochastic variational inequalities and complementarity problems by progressive hedging. *Math. Program.* **174**(1-2), 453–471 (2019)
23. Rockafellar, R.T., Wets, R.J.B.: Stochastic variational inequalities: single-stage to multistage. *Math. Program.* **165**(1), 331–360 (2017)
24. Ruszczyński, A., Shapiro, A.: Stochastic programming models. *Handb. Oper. Res. Manag. Sci.* **10**, 1–64 (2003)
25. Shanbhag, U.V.: Stochastic variational inequality problems: applications, analysis, and algorithms. *INFORMS Tutor. Oper. Res.* 71–107 (2013)
26. Shapiro, A.: Inference of statistical bounds for multistage stochastic programming problems. *Math. Methods Oper. Res.* **58**(1), 57–68 (2003)
27. Shapiro, A.: On complexity of multistage stochastic programs. *Oper. Res. Lett.* **34**(1), 1–8 (2006)
28. Shapiro, A., Dentcheva, D., Ruszczyński, A. *Lectures on Stochastic Programming: Modeling and Theory*, 2nd edn. SIAM, Philadelphia (2014)
29. Sun, H., Chen, X.: Two-stage stochastic variational inequalities: theory, algorithms and applications. *J. Oper. Res. Soc. China* **9**, 1–32 (2021)
30. Xu, H.: Sample average approximation methods for a class of stochastic variational inequality problems. *Asia-Pac. J. Oper. Res.* **27**(01), 103–119 (2010)
31. Zhang, M., Sun, J., Xu, H.: Two-stage quadratic games under uncertainty and their solution by progressive hedging algorithms. *SIAM J. Optim.* **29**(3), 1799–1818 (2019)

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.