Monotonicity and Complexity of Multistage Stochastic Variational Inequalities

Jie Jiang [†] Hailin Sun [‡*]

Abstract

In this paper, we consider multistage stochastic variational inequalities (MSVIs). First, we give multistage stochastic programs and multistage multi-player noncooperative game problems as source problems. After that, we derive the monotonicity properties of MSVIs under less restrictive conditions. Finally, the polynomial rate of convergence with respect to sample sizes between the original problem and its sample average approximation counterpart has been established.

Keywords: Stochastic variational inequality; Monotonicity; Sample average approximation; Rate of convergence
MSC 2020: 90C33; 90C15

1 Introduction

As a powerful mathematical paradigm, the variational inequality (VI) has been widely applied to investigate and formulate min-max problems, game problems, equilibria in traffic and economics, etc; see excellent monographs [1, 2] for more examples and details. To deal with the uncertainty in the decision process, the single-stage stochastic VI (SVI) appears. Some well-known methods are developed to tackle the single-stage SVI, such as expected-value (EV) form [3, 4] in which the random mapping is replaced by its expectation counterpart; expected residual minimization (ERM) form [5, 6] in which a residual function is introduced to measure the discrepancy from any point to the solution set and a minimization process of the expected residual function is considered, etc. For an excellent tutorial on applications, analysis and algorithms on single-stage SVI, we refer to [7].

In the seminal work [8], Rockafellar and Wets formally proposed the formulations of two-stage SVIs and multistage SVIs (MSVIs). From then on, two-stage SVIs have received considerable attention. In [9], Chen et al. addressed a general two-stage nonlinear SVI, and the ERM procedure was applied to give a solution. In [10], a strongly monotone two-stage stochastic linear complementarity problem (CP) was studied. A discretization scheme with convergence guarantee was proposed. Also, they first proposed the distributionally robust counterpart for a two-stage stochastic linear CP. In [11], two-stage stochastic generalized equations were considered. Specially, as a specific case, a strongly monotone two-stage stochastic VI-CP problem was detailedly discussed. The sample average approximation (SAA) convergence analysis was extended to the monotonicity case by using a regularization technique in [12]. The quantitative stability of two-stage SVIs was studied in [13, 14]. More recently, a fast algorithm based on the projection and the semismooth Newton method was proposed to solve a class of two-stage SVIs in [15]. As for the application aspect, two-stage SVIs have rich papers, including optimality of

[*]Corresponding Author
[†]College of Mathematics and Statistics, Chongqing University, Chongqing, 401331, China, jiangjiecq@163.com
[‡]Jiangsu Key Laboratory for NSLSCS, School of Mathematical Sciences, Nanjing Normal University, Nanjing, 210023, China, hlsun@njnu.edu.cn

two-stage stochastic programs [9], Cournot-Nash equilibrium in crude oil mark [16], two-player noncooperative game problem [17], water resources allocation problem [18], medical supplies in epidemic management [19], etc. For a comprehensive review of two-stage SVIs, we refer to [20].

Unlike two-stage SVIs, the study of MSVIs to date is in its infancy at best. In [8], Rockafellar and Wets proposed the MSVI model. After that, Rockafellar and Sun applied the progressive hedging algorithm to solve monotone MSVIs when the support set of the random variable is finite [21]. More recently, Cui et al. discussed in [22] the solvability and tractability of multistage pseudomonotone SVIs. Jiang et al. considered the strongly monotone multistage nonlinear CP in [23], and the asymptotic convergence of the SAA with conditional sampling counterpart was established.

According to the state-of-the-art progress of MSVIs, we have the following observations. Firstly, the results on the monotonicity of the general MSVIs are still limited. In [10, 11, 24, 23], the monotonicity assertion heavily relies on the information of subdifferential, which somehow restricts the application range of MSVIs that can establish the monotonicity properties. Secondly, the asymptotic convergence of multistage stochastic nonlinear CPs is discussed in [23], which fails to address how large sample sizes should be to achieve a desired accuracy of SAA estimators. Moreover, [23] focuses only on the multistage nonlinear CP rather than the general MSVI. Keep these observations in mind, and we make the following contributions in this paper.

- Different from the existing works [10, 11, 24, 23], where the information of subdifferential is leveraged to examine the monotonicity property, we establish the (strong) monotonicity of MSVIs with no need for subdifferential. Note that information of subdifferential requires some restrictive assumptions such as the uniqueness of solutions and the polyhedron [25]. Therefore, our method enlarges the scope of applicability for the corresponding monotonicity properties.

- Complexity analysis of SAA with conditional sampling MSVIs is studied. Compared with the existing work [23], in which the asymptotic convergence of multistage nonlinear CPs is addressed, the polynomial convergence rate between the general MSVI and its SAA counterpart is derived.

The layout of this paper is organized as follows. In Section 2, we first give the model of MSVI and some preliminaries. Then we present multistage stochastic programs (MSPs) and multistage multi-player noncooperative game problems as two source problems. In Section 3, the monotonicity properties of MSVIs are studied. In Section 4, the complexity analysis of MSVIs under SAA with conditional sampling is investigated. Finally, we conclude the paper with a summary in Section 5.

2 Models and Source Problems

Notations. $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$, $\mathbb{R}_+ := \{x \in \mathbb{R} : x \geq 0\}$ and $\mathbb{R}_{++} := \{x \in \mathbb{R} : x > 0\}$; $\mathbb{N}$ denotes the set of natural numbers; $\langle \cdot, \cdot \rangle$ denotes the inner product in the Euclidean space; the domain of $g : X \subseteq \mathbb{R}^n \to \overline{\mathbb{R}}$ is denoted by $\operatorname{dom} g := \{x \in X : g(x) < \infty\}$; $\rightrightarrows$ is used to denote the set-valued mapping; the superscript $\top$ denotes the transpose of vectors or matrices.

2.1 Preliminaries

In this subsection, we first give our models, and then summarize some basic concepts that will be useful in the subsequent sections.

Denote $\xi_1, \xi_2, \cdots, \xi_T$ a discrete stochastic process with time horizon $2 \leq T \in \mathbb{N}$. Random vectors ξ_t, $t = 1, 2 \cdots, T$ are defined on probability space $(\Omega, \mathcal{F}, \mathbb{P})$, i.e., $\xi_t : \Omega \to \Xi_t \subseteq \mathbb{R}^{s_t}$ with Ξ_t being the support set of ξ_t. Denote $\xi_{[t]} = (\xi_1, \cdots, \xi_t)$ supported on $\Xi_{[t]} \subseteq \Xi_1 \times \cdots \times \Xi_t \subseteq \mathbb{R}^{s^t}$

with $s^t := \sum_{i=1}^t s_i$. Then we use $\mathcal{F}_t$ to denote the σ-field induced by $\xi_{[t]}$, i.e., $\mathcal{F}_t := \sigma(\xi_{[t]})$. Thus, $\xi_{[t]}$ is a random vector on the probability space $(\Omega, \mathcal{F}_t, \mathbb{P})$. Moreover, we have the inclusion that $\{\emptyset, \Omega\} = \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}_T = \mathcal{F}$, which forms the so-called filtration. Due to $\mathcal{F}_1 = \{\emptyset, \Omega\}$, we know that ξ_1 is deterministic. Assume that decision variable x_t is $\mathbb{R}^{n_t}$-valued for $t = 1, 2, \cdots, T$. The t-th stage decision variable x_t is assumed to depend only on the available information until time t, i.e., $\xi_{[t]}$. Thus, $x_t = x_t(\xi_{[t]})$. This is the well-known nonanticipativity condition, which is an essential property in multistage stochastic optimization. Finally, denote $n = \sum_{t=1}^T n_t$ and $s = \sum_{t=1}^T s_t$.

With the above notations, the T-stage SVI can be written as (see more motivations in subsections 2.2 and 2.3):

$$\begin{cases} 0 \in \mathbb{E}_{\xi_2}[\Phi_1(x_1, x_2(\xi_{[2]}), \xi_2)] + \mathcal{N}_{\mathcal{C}_1}(x_1), \\ 0 \in \mathbb{E}_{\xi_3|\xi_{[2]}}\left[\Phi_2(x_1, x_2(\xi_{[2]}), x_3(\xi_{[3]}), \xi_3)\right] + \mathcal{N}_{\mathcal{C}_2(\xi_{[2]})}(x_2(\xi_{[2]})), \text{ a.e. } \xi_{[2]} \in \Xi_{[2]}, \\ \qquad\qquad\qquad \vdots \\ 0 \in \mathbb{E}_{\xi_T|\xi_{[T-1]}}\left[\Phi_{T-1}(x_{T-2}(\xi_{[T-2]}), x_{T-1}(\xi_{[T-1]}), x_T(\xi_{[T]}), \xi_T)\right] + \mathcal{N}_{\mathcal{C}_{T-1}(\xi_{[T-1]})}(x_{T-1}(\xi_{[T-1]})), \\ \text{a.e. } \xi_{[T-1]} \in \Xi_{[T-1]}, \\ 0 \in \Phi_T(x_{T-1}(\xi_{[T-1]}), x_T(\xi_{[T]}), \xi_T) + \mathcal{N}_{\mathcal{C}_T(\xi_{[T]})}(x_T(\xi_{[T]})), \text{ a.e. } \xi_{[T]} \in \Xi_{[T]}, \end{cases} \tag{1}$$

where a.e. is the abbreviation for almost everywhere; $\mathbb{E}_{\xi_2}[\cdot]$ denotes the expectation with respect to (w.r.t.) ξ_2, $\mathbb{E}_{\xi_{t+1}|\xi_{[t]}}[\cdot]$ denotes the conditional expectation of ξ_{t+1} for giving $\xi_{[t]}$ for $t = 2, \cdots, T-1$; $\Phi_1 : \mathbb{R}^{n_1} \times \mathbb{R}^{n_2} \times \mathbb{R}^{s_2} \to \mathbb{R}^{n_1}$, $\Phi_t : \mathbb{R}^{n_{t-1}} \times \mathbb{R}^{n_t} \times \mathbb{R}^{n_{t+1}} \times \mathbb{R}^{s_{t+1}} \to \mathbb{R}^{n_t}$ for $t = 2, \cdots, T-1$, $\Phi_T : \mathbb{R}^{n_{T-1}} \times \mathbb{R}^{n_T} \times \mathbb{R}^{s_T} \to \mathbb{R}^{n_T}$; $\mathcal{C}_1 \subseteq \mathbb{R}^{s_1}$ is a closed and convex set and $\mathcal{C}_t : \Xi_{[t]} \rightrightarrows \mathbb{R}^{n_t}$ for $t = 2, \cdots, T$ are closed and convex set-valued mappings. Without confusion, we sometimes also simply write $x_2 = x_2(\xi_{[2]}), \cdots, x_T = x_T(\xi_{[T]})$ in the sequel.

Before proceeding further, we give some useful concepts and definitions, which can be found in, for instance, [1, 26].

Definition 2.1. *A set $\mathcal{K} \subseteq \mathbb{R}^n$ is called a* cone *if $0 \in \mathcal{K}$ and $\lambda x \in \mathcal{K}$ for all $x \in \mathcal{K}$ and $\lambda > 0$. Then the* dual cone *(or negative polar cone) of $\mathcal{K}$, denoted by $\mathcal{K}^\circ$, is*

$$\mathcal{K}^\circ := \{v : \langle v, z \rangle \geq 0, \forall z \in \mathcal{K}\}.$$

Definition 2.2. *Let $F : \mathcal{C} \subseteq \mathbb{R}^n \to \mathbb{R}^n$.*

(i) F is said to be monotone *on $\mathcal{C}$, if $\langle F(u) - F(v), u - v \rangle \geq 0$ for any $u, v \in \mathcal{C}$.*

(ii) F is said to be strongly monotone *on $\mathcal{C}$, if $\langle F(u) - F(v), u - v \rangle \geq \kappa \|u - v\|^2$ for some $\kappa > 0$ and any $u, v \in \mathcal{C}$.*

Definition 2.3. *Let $\mathcal{C}$ be a convex set. Then the* normal cone *of $\mathcal{C}$ at $x \in \mathcal{C}$, denoted by $\mathcal{N}_{\mathcal{C}}(x)$, is*

$$\mathcal{N}_{\mathcal{C}}(x) := \{v : \langle v, y - x \rangle \leq 0, \forall y \in \mathcal{C}\}.$$

Recall that we say a mapping $G : \mathbb{R}^n \to \mathbb{R}^\ell$ is *convex w.r.t. a convex cone* $\mathcal{K}$ [27, Section 2.3.2] if

$$G(\lambda x + (1-\lambda)y) \preceq_{\mathcal{K}} \lambda G(x) + (1-\lambda)G(y), \forall x, y \in \mathbb{R}^n, \forall \lambda \in [0, 1],$$

where $a \preceq_{\mathcal{K}} b$ means that $b - a \in \mathcal{K}$ and the sum of two sets is the Minkowski sum.

Next, we consider two source problems of MSVI (1) that arise from MSPs and multistage N-player ($N \geq 2$) noncooperative game problems, respectively. Under some standard settings, they both can be equivalently reformulated as the model (1), which therefore provide a wide range of potential applications for MSVI (1).

2.2 Multistage Stochastic Programs

Consider a MSP as follows (see e.g. [27, Chapter 3]):

$$\begin{aligned}\min_{x_1\in\mathcal{M}_1} f_1(x_1,\xi_1)+\mathbb{E}\Bigg[\min_{x_2\in\mathcal{M}_2(x_1,\xi_2)} f_2(x_2,\xi_2)+\mathbb{E}\Big[\min_{x_3\in\mathcal{M}_3(x_2,\xi_3)} f_3(x_3,\xi_3)+\cdots \\ +\mathbb{E}\big[\min_{x_T\in\mathcal{M}_T(x_{T-1},\xi_T)} f_T(x_T,\xi_T)\big]\Big]\Bigg],\end{aligned} \tag{2}$$

where the discrete stochastic process $\xi_1,\xi_2,\cdots,\xi_T$ is defined in subsection 2.1, $f_t:\mathbb{R}^{n_t}\times\mathbb{R}^{s_t}\to\mathbb{R}$ for $t=1,2,\cdots,T$, $\mathcal{M}_1\subseteq\mathbb{R}^{n_1}$ is a closed and convex subset and $\mathcal{M}_t(x_{t-1},\xi_t)$ for $t=2,\cdots,T$ are defined as

$$\mathcal{M}_t(x_{t-1},\xi_t):=\left\{x_t\in\mathcal{X}_t(\xi_{[t]}):W_t(x_t,\xi_t)+H_t(x_{t-1},\xi_t)\in-\mathcal{D}_t(\xi_{[t]})\right\}$$

with $W_t:\mathbb{R}^{n_t}\times\mathbb{R}^{s_t}\to\mathbb{R}^{r_t}$, $H_t:\mathbb{R}^{n_{t-1}}\times\mathbb{R}^{s_t}\to\mathbb{R}^{r_t}$, $\mathcal{X}_t:\mathbb{R}^{s^t}\rightrightarrows\mathbb{R}^{n_t}$, $\mathcal{D}_t:\mathbb{R}^{s^t}\rightrightarrows\mathbb{R}^{r_t}$ being vector-valued or set-valued mappings. Further, we assume that, for fixed $\xi_{[t]}$, $\mathcal{X}_t(\xi_{[t]})$ is a closed convex set and $\mathcal{D}_t(\xi_{[t]})$ is a closed convex cone.

It is known that we can rewrite the nested formulation MSP (2) as the following dynamic programming form (see e.g. [27, Chapter 3]):

$$\min_{x_1\in\mathcal{M}_1} f_1(x_1,\xi_1)+\mathbb{E}_{\xi_{[2]}}[Q_2(x_1,\xi_{[2]})], \tag{3}$$

where $Q_t:\mathbb{R}^{n_{t-1}}\times\mathbb{R}^{s^t}\to\mathbb{R}$ for $t=2,\cdots,T$ are so-called cost-to-go functions, which are defined as

$$Q_t(x_{t-1},\xi_{[t]}):=\min_{x_t\in\mathcal{M}_t(x_{t-1},\xi_t)} f_t(x_t,\xi_t)+\mathbb{E}_{\xi_{[t+1]}|\xi_{[t]}}[Q_{t+1}(x_t,\xi_{[t+1]})] \tag{4}$$

for $t=2,\cdots,T-1$, and

$$Q_T(x_{T-1},\xi_{[T]}):=\min_{x_T\in\mathcal{M}_T(x_{T-1},\xi_T)} f_T(x_T,\xi_T). \tag{5}$$

The Lagrangian functions of problems (4)-(5) are

$$\mathcal{L}_t(x_t,\pi_t):=f_t(x_t,\xi_t)+\mathbb{E}_{\xi_{[t+1]}|\xi_{[t]}}[Q_{t+1}(x_t,\xi_{[t+1]})]+\pi_t^\top(W_t(x_t,\xi_t)+H_t(x_{t-1},\xi_t))$$

for $t=2,\cdots,T-1$, and

$$\mathcal{L}_T(x_t,\pi_t):=f_T(x_T,\xi_T)+\pi_T^\top(W_T(x_T,\xi_T)+H_T(x_{T-1},\xi_T)).$$

Then the Lagrangian dual problems of (4)-(5) can be written as

$$\max_{\pi_t\in\mathcal{D}_t^\circ(\xi_{[t]})}\left\{\min_{x_t\in\mathcal{X}_t(\xi_{[t]})} f_t(x_t,\xi_t)+\mathbb{E}_{\xi_{[t+1]}|\xi_{[t]}}[Q_{t+1}(x_t,\xi_{[t+1]})]+\pi_t^\top(W_t(x_t,\xi_t)+H_t(x_{t-1},\xi_t))\right\} \tag{6}$$

for $t=2,\cdots,T-1$, and

$$\max_{\pi_T\in\mathcal{D}_T^\circ(\xi_{[T]})}\left\{\min_{x_T\in\mathcal{X}_T(\xi_{[T]})} f_T(x_T,\xi_T)+\pi_T^\top(W_T(x_T,\xi_T)+H_T(x_{T-1},\xi_T))\right\}, \tag{7}$$

where $\mathcal{D}_t^\circ(\xi_{[t]})$ is the dual cone of $\mathcal{D}_t(\xi_{[t]})$ for $t=2,\cdots,T$, i.e.,

$$\mathcal{D}_t^\circ(\xi_{[t]}):=\left\{v:\langle v,z\rangle\geq 0,\forall z\in\mathcal{D}_t(\xi_{[t]})\right\}.$$

For $t=2,\cdots,T$ and given x_{t-1} and $\xi_{[t]}$, we use $\Pi_t(x_{t-1},\xi_{[t]})$ to denote the sets of optimal solutions of the dual problems (6)-(7).

Recall that $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is *proper* if $f(x) > -\infty$ and $\mathrm{dom} f$ is nonempty. Let $F : \mathbb{R}^n \times \Xi \to \mathbb{R}$, where Ξ is the support set of random vector ξ. $F(x, \xi)$ is *random lower semicontinuous* (in this case, F is also called a normal integrand, see [27, Definition 7.40]) if the associated epigraphical multifunction $\xi \mapsto \mathrm{epi} F(\cdot, \xi)$ is closed valued and measurable. A large class of random lower semicontinuous functions is given by the so-called Carathéodory functions, i.e., real-valued functions such that $F(x, \cdot)$ is measurable for every x and $F(\cdot, \xi)$ is continuous for a.e. ξ.

The following lemma gives a description of the subdifferential of $\mathbb{E}[F(x, \xi)]$.

Lemma 2.1 ([27, Theorem 7.52])**.** *Suppose that: (i) the function $F(x, \xi)$ is random lower semicontinuous; (ii) for a.e. ξ the function $F(\cdot, \xi)$ is convex; (iii) the expectation function $f(x) = \mathbb{E}[F(x, \xi)]$ is proper; (iv) the domain of f has a nonempty interior. Then for any $x_0 \in \mathrm{dom} f$,*

$$\partial f(x_0) = \mathbb{E}[\partial F(x_0, \xi)] + \mathcal{N}_{\mathrm{dom} f}(x_0).$$

To describe the necessary optimality conditions of problems (3)-(5), we first consider the following minimization problem with fixed $x \in \mathbb{R}^n$:

$$\begin{aligned} \min_{y \in Y} \quad & q(y) \\ \text{s.t.} \quad & W(y) + H(x) \in -C, \end{aligned} \tag{8}$$

where $Y \subseteq \mathbb{R}^m$ is a nonempty convex closed subset, $q : \mathbb{R}^m \to \mathbb{R}$ is convex, $C \subseteq \mathbb{R}^\ell$ is a closed convex cone, $W : \mathbb{R}^m \to \mathbb{R}^\ell$ and $H : \mathbb{R}^n \to \mathbb{R}^\ell$ are convex w.r.t. C and differentiable. Denote the Lagrangian dual problem of (8) by

$$\max_{\pi \in C^\circ} \left\{ \min_{y \in Y} \left(q(y) + \pi^\top (H(x) + W(y)) \right) \right\}, \tag{9}$$

where C° is the dual cone of C. Since W is convex w.r.t. C, for any $\pi \in C^\circ$, $\pi^\top W(y)$ is a convex function of y. Therefore, the inner minimization problem w.r.t. y is a convex optimization problem.

Denote the optimal value of problem (8) by $\vartheta(x)$. Denote the optimal solution set of (9) by $\Pi(x)$.

The following two lemmas follow from [27, Propositions 2.21-2.22] and [28, Proposition 3.3] directly, and thus we omit the proof here.

Lemma 2.2. *Suppose that: (i) for every $x \in \mathbb{R}^n$, problem (8) is feasible; (ii) problems (8) and (9) have finite and equal optimal values. Then $\vartheta(x)$ is convex and $\partial \vartheta(x) = \nabla H(x)^\top \Pi(x)$.*

Lemma 2.3. *Let x be fixed and consider the optimality conditions of problem (8):*

$$\begin{cases} 0 \in \partial q(y) + \nabla W(y)^\top \pi + \mathcal{N}_Y(y), \\ 0 \in -W(y) - H(x) + \mathcal{N}_{C^\circ}(\pi). \end{cases} \tag{10}$$

If there is no duality gap between (8) and (9), y^* is an optimal solution of problem (8) and π^* is an optimal solution of problem (9), then the tuple (y^*, π^*) satisfies (10). Conversely, if the tuple $(\bar{y}, \bar{\pi})$ satisfies (10), then $\bar{y}$ is an optimal solution of (8), $\bar{\pi}$ is an optimal solution of (9), and there is no duality gap between (8) and (9).

We then make the following standard assumptions for the dynamic form of MSP (3)-(5).

Assumption 1. *Let the following assertions hold.*

(i) (differentiability and convexity) $f_t(x_t, \xi_t)$ is continuously differentiable and convex w.r.t. x_t for a.e. ξ_t for $t = 1, 2, \cdots, T$. $W_t(\cdot, \xi_t)$ is continuously differentiable and convex w.r.t. $\mathcal{D}_t(\xi_{[t]})$ for a.e. $\xi_{[t]}$ and $H_t(\cdot, \xi_t)$ is continuously differentiable and convex w.r.t. $\mathcal{D}_t(\xi_{[t]})$ for a.e. $\xi_{[t]}$.

(ii) *(relatively complete recourse) For any fixed $\xi_{[t]}$ and $x_{t-1} \in \mathcal{X}_{t-1}(\xi_{[t-1]})$, $Q_t(x_{t-1}, \xi_{[t]})$ is finite.*

(iii) *(strong duality) There are no duality gaps between problems (4) and (6) for $t = 2, \cdots, T-1$, and between problems (5) and (7).*

(iv) *(domination) There exists an integrable function $\kappa_t : \Xi_{[t]} \to \mathbb{R}_+$ with $\mathbb{E}_{\xi_{[t]}|\xi_{[t-1]}}[\kappa_t(\xi_{[t]})] < \infty$ for any fixed $\xi_{[t-1]} \in \Xi_{[t-1]}$, such that*

$$\left|Q_t(x_{t-1}, \xi_{[t]})\right| \leq \kappa_t(\xi_{[t]})$$

for a.e. $\xi_{[t]} \in \Xi_{[t]}$, $x_{t-1} \in \mathcal{X}_{t-1}(\xi_{[t-1]})$ and $t = 2, \cdots, T$.

Remark 2.1. *Assumption (i) is quite standard and easy to be verified. The relatively complete recourse condition in (ii) is commonly-used in two-stage or multistage stochastic programs. Some sufficient conditions of the strong duality (iii) can be found in [27, Proposition 2.21]. (iv) is important to make the expectation of cost-to-go functions to be well-defined.*

The following proposition directly follows from Lemma 2.2 (see [27, Proposition 3.5] for the linear constrained case).

Proposition 2.1. *Let Assumption 1 hold. For $t = 2, \cdots, T$ and given $\xi_{[t]}$, we have that $Q_t(\cdot, \xi_{[t]})$ is convex and $\partial Q_t(x_{t-1}, \xi_{[t]}) = \nabla H_t(x_{t-1}, \xi_t)^\top \Pi_t(x_{t-1}, \xi_{[t]})$.*

Now we are ready to give the main result of this subsection.

Theorem 2.1. *Let Assumption 1 hold. The MSP (2) is equivalent to the following MSVI:*

$$\begin{cases} 0 \in \nabla f_1(x_1, \xi_1) + \mathbb{E}_{\xi_{[2]}}[\nabla H_2(x_1, \xi_2)^\top \pi_2(\xi_{[2]})] + \mathcal{N}_{\mathcal{M}_1}(x_1), \\ \left.\begin{cases} 0 \in \nabla f_t(x_t(\xi_{[t]}), \xi_t) + \mathbb{E}_{\xi_{[t+1]}|\xi_{[t]}}[\nabla H_{t+1}(x_t(\xi_{[t]}), \xi_{t+1})^\top \pi_{t+1}(\xi_{[t+1]})] \\ \qquad\qquad + \nabla W_t(x_t(\xi_{[t]}), \xi_t)^\top \pi_t(\xi_{[t]}) + \mathcal{N}_{\mathcal{X}_t(\xi_{[t]})}(x_t(\xi_{[t]})), \\ 0 \in -W_t(x_t(\xi_{[t]}), \xi_t) - H_t(x_{t-1}, \xi_t) + \mathcal{N}_{\mathcal{D}_t^\circ(\xi_{[t]})}(\pi_t(\xi_{[t]})), \end{cases}\right\} \text{for } t = 2, \cdots, T-1, \\ \begin{cases} 0 \in \nabla f_T(x_T(\xi_{[T]}), \xi_T) + \nabla W_T(x_T(\xi_{[T]}), \xi_T)^\top \pi_T(\xi_{[T]}) + \mathcal{N}_{\mathcal{X}_T(\xi_{[T]})}(x_T(\xi_{[T]})), \\ 0 \in -W_T(x_T(\xi_{[T]}), \xi_T) - H_T(x_{T-1}, \xi_T) + \mathcal{N}_{\mathcal{D}_T^\circ(\xi_{[T]})}(\pi_T(\xi_{[T]})) \end{cases} \end{cases} \tag{11}$$

in the sense that: $(x_1, x_2(\cdot), \cdots, x_T(\cdot))$ is an optimal solution of (2), and $\pi_2(\cdot)$ is an optimal solution of (6) for $t = 2, \cdots, T-1$ and $\pi_T(\cdot)$ is an optimal solution of (7), if and only if $(x_1, x_2(\cdot), \pi_2(\cdot), \cdots, x_T(\cdot), \pi_T(\cdot))$ is a solution of (11).

Proof. First, we consider problem (5). It is known from Lemma 2.3 that: For given x_{T-1} and $\xi_{[T]}$, $x_T(\xi_{[T]})$ solves problem (5) if and only if there exists a $\pi_T(\xi_{[T]})$ such that the tuple $(x_T(\xi_{[T]}), \pi_T(\xi_{[T]}))$ satisfies

$$\begin{cases} 0 \in \nabla f_T(x_T(\xi_{[T]}), \xi_T) + \nabla W_T(x_T(\xi_{[T]}), \xi_T)^\top \pi_T(\xi_{[T]}) + \mathcal{N}_{\mathcal{X}_T(\xi_{[T]})}(x_T(\xi_{[T]})), \\ 0 \in -W_T(x_T(\xi_{[T]}), \xi_T) - H_T(x_{T-1}, \xi_T) + \mathcal{N}_{\mathcal{D}_T^\circ(\xi_{[T]})}(\pi_T(\xi_{[T]})). \end{cases} \tag{12}$$

For given x_{t-1} and $\xi_{[t]}$ with $t = 2, \cdots, T-1$, $x_t(\xi_{[t]})$ solves problem (4) if and only if there exists a $\pi_t(\xi_{[t]})$ such that the tuple $(x_t(\xi_{[t]}), \pi_t(\xi_{[t]}))$ satisfies

$$\begin{cases} 0 \in \nabla f_t(x_t(\xi_{[t]}), \xi_t) + \partial \mathbb{E}_{\xi_{[t+1]}|\xi_{[t]}}[Q_{t+1}(x_t(\xi_{[t]}), \xi_{[t+1]})] + \nabla W_t(x_t(\xi_{[t]}), \xi_t)^\top \pi_t(\xi_{[t]}) + \mathcal{N}_{\mathcal{X}_t(\xi_{[t]})}(x_t(\xi_{[t]})), \\ 0 \in -W_t(x_t(\xi_{[t]}), \xi_t) - H_t(x_{t-1}, \xi_t) + \mathcal{N}_{\mathcal{D}_t^\circ(\xi_{[t]})}(\pi_t(\xi_{[t]})). \end{cases}$$

By using Lemma 2.1 and Proposition 2.1, we know that

$$\begin{aligned}\partial\mathbb{E}_{\xi_{[t+1]}|\xi_{[t]}}[Q_{t+1}(x_t(\xi_{[t]}),\xi_{[t+1]})] &= \mathbb{E}_{\xi_{[t+1]}|\xi_{[t]}}[\partial Q_{t+1}(x_t(\xi_{[t]}),\xi_{[t+1]})]\\ &= \mathbb{E}_{\xi_{[t+1]}|\xi_{[t]}}[\nabla H_{t+1}(x_t(\xi_{[t]}),\xi_{t+1})^\top \Pi_{t+1}(x_t(\xi_{[t]}),\xi_{[t+1]})],\end{aligned}$$

where the last equality follows from the Moreau-Rockafellar theorem.

Then, we obtain, for $t=2,\cdots,T-1$, that

$$\begin{cases}0\in\nabla f_t(x_t(\xi_{[t]}),\xi_t)+\mathbb{E}_{\xi_{[t+1]}|\xi_{[t]}}[\nabla H_{t+1}(x_t(\xi_{[t]}),\xi_{t+1})^\top \pi_{t+1}(\xi_{[t+1]})]\\ \qquad\qquad +\nabla W_t(x_t(\xi_{[t]}),\xi_t)^\top \pi_t(\xi_{[t]})+\mathcal{N}_{\mathcal{X}_t(\xi_{[t]})}(x_t(\xi_{[t]})),\\ 0\in -W_t(x_t(\xi_{[t]}),\xi_t)-H_t(x_{t-1},\xi_t)+\mathcal{N}_{\mathcal{D}_t^\circ(\xi_{[t]})}(\pi_t(\xi_{[t]})),\end{cases} \tag{13}$$

where $\pi_{t+1}(\xi_{[t+1]})\in\Pi_{t+1}(x_t(\xi_{[t]}),\xi_{[t+1]})$.

For problem (3), the necessary and sufficient optimality condition can be written as

$$0\in\partial\left(f_1(x_1,\xi_1)+\mathbb{E}[Q_2(x_1,\xi_{[2]})]\right)+\mathcal{N}_{\mathcal{M}_1}(x_1).$$

Similarly, by the differentiability of $f_1(\cdot,\xi_1)$ and using Lemma 2.1 and Proposition 2.1, we obtain

$$0\in\nabla f_1(x_1,\xi_1)+\mathbb{E}_{\xi_{[2]}}[\nabla H_2(x_1,\xi_2)^\top \pi_2(\xi_{[2]})]+\mathcal{N}_{\mathcal{M}_1}(x_1). \tag{14}$$

Combining (12), (13) and (14), we complete the proof. ■

Remark 2.2. *Obviously, by regarding x_1 as the first stage variable and (x_t,π_t) as the t-th stage variable for $t=2,\cdots,T$, (11) is essentially a MSVI in accord with (1), which, therefore, illustrates the motivation as well as the reasonability of our MSVI model (1).*

The following corollary gives some equivalent forms of (2) by directly using Theorem 2.1 when the feasible sets of (2) are given by some specific polyhedral sets.

Corollary 2.1. *Let $f_t(x_t,\xi_t)$ be continuously differentiable and convex w.r.t. x_t for a.e. ξ_t for $t=1,2,\cdots,T$ and (ii)-(iv) in Assumption 1 hold.*

(a) If

$$\mathcal{M}_t(x_{t-1},\xi_t):=\left\{x_t\in\mathbb{R}^{n_t}: W_t(\xi_t)x_t+H_t(\xi_t)x_{t-1}\le h_t(\xi_t)\right\}, \tag{15}$$

where $W_t:\mathbb{R}^{s_t}\to\mathbb{R}^{r_t\times n_t}$, $H_t:\mathbb{R}^{s_t}\to\mathbb{R}^{r_t\times n_{t-1}}$ and $h_t:\mathbb{R}^{s_t}\to\mathbb{R}^{r_t}$ for $t=2,\cdots,T$, then (2) is equivalent to

$$\begin{cases}0\in\nabla f_1(x_1,\xi_1)+\mathbb{E}_{\xi_{[2]}}[H_2(\xi_2)^\top\pi_2(\xi_{[2]})]+\mathcal{N}_{\mathcal{M}_1}(x_1),\\ \left.\begin{cases}\nabla f_t(x_t(\xi_{[t]}),\xi_t)+\mathbb{E}_{\xi_{[t+1]}|\xi_{[t]}}[H_{t+1}(\xi_{t+1})^\top\pi_{t+1}(\xi_{[t+1]})]\\ \qquad +W_t(\xi_t)^\top\pi_t(\xi_{[t]})=0,\\ 0\le\pi_t(\xi_{[t]})\perp -W_t(\xi_t)x_t(\xi_{[t]})-H_t(\xi_t)x_{t-1}(\xi_{[t-1]})+h_t(\xi_t)\ge0,\end{cases}\right\}\text{for } t=2,\cdots,T-1,\\ \begin{cases}\nabla f_T(x_T(\xi_{[T]}),\xi_T)+W_T(\xi_T)^\top\pi_T(\xi_{[T]})=0,\\ 0\le\pi_T(\xi_{[T]})\perp -W_T(\xi_T)x_T(\xi_{[T]})-H_T(\xi_T)x_{T-1}(\xi_{[T-1]})+h_T(\xi_T)\ge0.\end{cases}\end{cases}$$

(b) If $\mathcal{M}_t(x_{t-1},\xi_t):=\left\{x_t\in\mathbb{R}^{n_t}_+: W_t(\xi_t)x_t+H_t(\xi_t)x_{t-1}=h_t(\xi_t)\right\}$ for $t=2,\cdots,T$, where all coefficients are defined in (15), then (2) is equivalent to

$$\begin{cases}0\in\nabla f_1(x_1,\xi_1)+\mathbb{E}_{\xi_{[2]}}[H_2(\xi_2)^\top\pi_2(\xi_{[2]})]+\mathcal{N}_{\mathcal{M}_1}(x_1),\\ \left.\begin{cases}0\le x_t(\xi_{[t]})\perp\nabla f_t(x_t(\xi_{[t]}),\xi_t)+\mathbb{E}_{\xi_{[t+1]}|\xi_{[t]}}[H_{t+1}(\xi_{t+1})^\top\pi_{t+1}(\xi_{[t+1]})]\\ \qquad +W_t(\xi_t)^\top\pi_t(\xi_{[t]})\ge0,\\ -W_t(\xi_t)x_t(\xi_{[t]})-H_t(\xi_t)x_{t-1}(\xi_{[t-1]})+h_t(\xi_t)=0,\end{cases}\right\}\text{for } t=2,\cdots,T-1,\\ \begin{cases}0\le x_T(\xi_{[T]})\perp\nabla f_T(x_T(\xi_{[T]}),\xi_T)+W_T(\xi_T)^\top\pi_T(\xi_{[T]})\ge0,\\ -W_T(\xi_T)x_T(\xi_{[T]})-H_T(\xi_T)x_{T-1}(\xi_{[T-1]})+h_T(\xi_T)=0.\end{cases}\end{cases}$$

(c) *If* $\mathcal{M}_t(x_{t-1},\xi_t) := \left\{x_t \in \mathbb{R}^{n_t}_+ : W_t(\xi_t)x_t + H_t(\xi_t)x_{t-1} \le h_t(\xi_t)\right\}$ *for* $t = 2,\cdots,T$, *where all coefficients are defined in* (15), *then* (2) *is equivalent to*

$$\begin{cases} 0 \in \nabla f_1(x_1,\xi_1) + \mathbb{E}_{\xi_{[2]}}[H_2(\xi_2)^\top \pi_2(\xi_{[2]})] + \mathcal{N}_{\mathcal{M}_1}(x_1), \\ \left.\begin{cases} 0 \le x_t(\xi_{[t]}) \perp \nabla f_t(x_t(\xi_{[t]}),\xi_t) + \mathbb{E}_{\xi_{[t+1]}|\xi_{[t]}}[H_{t+1}(\xi_{t+1})^\top \pi_{t+1}(\xi_{[t+1]})] \\ \qquad\qquad + W_t(\xi_t)^\top \pi_t(\xi_{[t]}) \ge 0, \\ 0 \le \pi_t(\xi_{[t]}) \perp -W_t(\xi_t)x_t(\xi_{[t]}) - H_t(\xi_t)x_{t-1}(\xi_{[t-1]}) + h_t(\xi_t) \ge 0, \end{cases}\right\} \textit{for } t = 2,\cdots,T-1, \\ \begin{cases} 0 \le x_T(\xi_{[T]}) \perp \nabla f_T(x_T(\xi_{[T]}),\xi_T) + W_T(\xi_T)^\top \pi_T(\xi_{[T]}) \ge 0, \\ 0 \le \pi_T(\xi_{[T]}) \perp -W_T(\xi_T)x_T(\xi_{[T]}) - H_T(\xi_T)x_{T-1}(\xi_{[T-1]}) + h_T(\xi_T) \ge 0. \end{cases} \end{cases}$$

2.3 Multistage N-player Noncooperative Game Problems

A two-stage N-player ($N \ge 2$) noncooperative quadratic game under uncertainty is studied in [17], which can be equivalently reformulated as two-stage SVIs. In this subsection, we consider a multistage N-player noncooperative game problem. The existing works (see e.g. [10, 11, 17]) have shown that the VI has its advantages in investigating game problems. Specifically, we equivalently transfer a multistage N-player noncooperative game problem into an MSVI under mild conditions. Therefore, MSVIs have provided an alternative and efficient avenue to deal with multistage N-player game problems.

To model the multistage N-player noncooperative game problem, we need some notations. Denote the i-th player's decision at the t-th stage by $x_t^i \in \mathbb{R}^{n_t^i}$ for $i = 1,\cdots,N$ and $t = 1,\cdots,T$. Denote $x_t = (x_t^1, x_t^2, \cdots, x_t^N) \in \mathbb{R}^{n_t}$ with $n_t = \sum_{j=1}^N n_t^j$ the whole decisions of the N players at the t-th stage. Let

$$x_t^{-i} := (x_t^1,\cdots,x_t^{i-1},x_t^{i+1},\cdots,x_t^N) \in \mathbb{R}^{n_t^{-i}},$$

where $n_t^{-i} = n_t - n_t^i$, be the rivals' decisions of the i-th player at the t-th stage. Then we can also write $x_t = (x_t^i, x_t^{-i})$. Similarly, let the discrete stochastic process $\xi_1,\xi_2,\cdots,\xi_T$ be defined in subsection 2.1.

For the given rivals' decisions $(x_1^{-i}, x_2^{-i}, \cdots, x_T^{-i})$, the i-th player's problem can be formulated as follows:

$$\min_{x_1^i \in X_1^i} \theta_1^i(x_1,\xi_1) + \mathbb{E}_{\xi_{[2]}}[\psi_2^i(x_1, x_2^{-i},\cdots,x_T^{-i},\xi_{[2]})], \tag{16}$$

where $X_1^i \subseteq \mathbb{R}^{n_1^i}$ is a nonempty closed convex subset, $\theta_1^i : \mathbb{R}^{n_1} \times \mathbb{R}^{s_1} \to \mathbb{R}$, $\psi_t^i : \mathbb{R}^{n_{t-1}+\sum_{j=t}^T n_j^{-i}} \times \mathbb{R}^{s^t} \to \mathbb{R}$ for $t = 2,\cdots,T$ are defined as

$$\begin{aligned} \psi_t^i(x_{t-1}, x_t^{-i},\cdots,x_T^{-i},\xi_{[t]}) := & \min_{x_t^i \in X_t^i(x_{t-1},x_t^{-i},\xi_t)} \theta_t^i(x_{t-1},x_t,\xi_t) \\ & + \mathbb{E}_{\xi_{[t+1]}|\xi_{[t]}}[\psi_{t+1}^i(x_t, x_{t+1}^{-i},\cdots,x_T^{-i},\xi_{[t+1]})] \end{aligned} \tag{17}$$

for $t = 2,\cdots,T-1$, and

$$\psi_T^i(x_{T-1}, x_T^{-i},\xi_{[T]}) := \min_{x_T^i \in X_T^i(x_{T-1},x_T^{-i},\xi_T)} \theta_T^i(x_{T-1},x_T,\xi_T), \tag{18}$$

where $\theta_t^i : \mathbb{R}^{n_{t-1}+n_t} \times \mathbb{R}^{s_t} \to \mathbb{R}$ and $X_t : \mathbb{R}^{n_{t-1}+n_t^{-i}} \times \mathbb{R}^{s_t} \rightrightarrows \mathbb{R}^{n_t^i}$ for $t = 2,\cdots,T$.

Recall that we say $(x_1, x_2(\cdot),\cdots,x_T(\cdot))$ is a Nash equilibrium for the multistage N-player game problem if for given $(x_1^{-i}, x_2^{-i}(\cdot),\cdots,x_T^{-i}(\cdot))$ with $i = 1,\cdots,N$, $(x_1^i, x_2^i(\cdot),\cdots,x_T^i(\cdot))$ is the optimal solution of problem (16)-(18).

To equivalently reformulate multistage N-player game problem (16)-(18) as a MSVI, we consider a quadratic game and make the following specific setting. For $i = 1,\cdots,N$ and

$t = 2, \cdots, T$, set

$$\theta_1^i(x_1, \xi_1) := \frac{1}{2}(x_1^i)^\top Q_1^i x_1^i + (c_1^i)^\top x_1^i + (x_1^i)^\top R_1^i x_1^{-i},$$

$$\theta_t^i(x_{t-1}, x_t, \xi_t) := \frac{1}{2}(x_t^i)^\top Q_t^i(\xi_t) x_t^i + c_t^i(\xi_t)^\top x_t^i + (x_t^i)^\top S_t^i(\xi_t) x_{t-1}^i + (x_t^i)^\top O_t^i(\xi_t) x_{t-1}^{-i} + (x_t^i)^\top R_t^i(\xi_t) x_t^{-i},$$

$$X_t^i(x_{t-1}, x_t^{-i}, \xi_t) := \left\{ x_t^i \in \mathbb{R}_+^{n_t^i} : A_t^i(\xi_t) x_{t-1}^i + A_t^{-i}(\xi_t) x_{t-1}^{-i} + B_t^i(\xi_t) x_t^i + B_t^{-i}(\xi_t) x_t^{-i} \le b_t^i(\xi_t) \right\},$$

where $Q_1^i \in \mathbb{R}^{n_1^i \times n_1^i}$, $c_1^i \in \mathbb{R}^{n_1^i}$, $R_1^i \in \mathbb{R}^{n_1^i \times n_1^{-i}}$ and $Q_t^i : \mathbb{R}^{s_t} \to \mathbb{R}^{n_t^i \times n_t^i}$, $c_t^i : \mathbb{R}^{s_t} \to \mathbb{R}^{n_t^i}$, $S_t^i : \mathbb{R}^{s_t} \to \mathbb{R}^{n_t^i \times n_{t-1}^i}$, $O_t^i : \mathbb{R}^{s_t} \to \mathbb{R}^{n_t^i \times n_{t-1}^{-i}}$, $R_t^i : \mathbb{R}^{s_t} \to \mathbb{R}^{n_t^i \times n_t^{-i}}$, $A_t^i : \mathbb{R}^{s_t} \to \mathbb{R}^{r_t^i \times n_{t_1}^i}$, $A_t^{-i} : \mathbb{R}^{s_t} \to \mathbb{R}^{r_t^i \times n_{t-1}^{-i}}$, $B_t^i : \mathbb{R}^{s_t} \to \mathbb{R}^{r_t^i \times n_t^i}$, $B_t^{-i} : \mathbb{R}^{s_t} \to \mathbb{R}^{r_t^i \times n_t^{-i}}$, $b_t^i : \mathbb{R}^{s_t} \to \mathbb{R}^{r_t^i}$ are vector- or matrix-valued mappings.

Assumption 2. *For fixed* $(x_1^{-i}, x_2^{-i}(\cdot), \cdots, x_T^{-i}(\cdot))$ *with* $i = 1, \cdots, N$, (16)-(18), *which can be viewed as the MSP* (2), *satisfies assertions (i)-(iv) in Assumption 1.*

By directly using Corollary 2.1, we can obtain the following theorem, and thus we omit its proof here.

Theorem 2.2. *Let Assumption 2 hold. The multistage* N*-player game problem* (16)-(18) *is equivalent to the following MSVI: For* $i = 1, 2, \cdots, N$,

$$\begin{cases} 0 \in Q_1^i x_1^i + c_1^i + R_1^i x_1^{-i} + \mathbb{E}_{\xi_{[2]}}[A_2^i(\xi_2)^\top \pi_2^i(\xi_{[2]})] + \mathcal{N}_{X_1^i}(x_1^i), \\ \left.\begin{cases} 0 \le x_t^i(\xi_{[t]}) \perp Q_t^i(\xi_t) x_t^i + c_t^i(\xi_t) + S_t^i(\xi_t) x_{t-1}^i + O_t^i(\xi_t) x_{t-1}^{-i} + R_t^i(\xi_t) x_t^{-i} \\ \qquad + \mathbb{E}_{\xi_{[t+1]}|\xi_{[t]}}[A_{t+1}^i(\xi_{t+1})^\top \pi_{t+1}^i(\xi_{[t+1]})] + B_t^i(\xi_t)^\top \pi_t^i(\xi_{[t]}) \ge 0, \\ 0 \le \pi_t^i(\xi_{[t]}) \perp -A_t^i(\xi_t) x_{t-1}^i - A_t^{-i}(\xi_t) x_{t-1}^{-i} - B_t^i(\xi_t) x_t^i - B_t^{-i}(\xi_t) x_t^{-i} + b_t^i(\xi_t) \ge 0, \end{cases}\right\} \text{for } t = 2, \cdots, T-1, \\ \begin{cases} 0 \le x_T^i(\xi_{[T]}) \perp Q_T^i(\xi_T) x_T^i + c_T^i(\xi_T) + S_T^i(\xi_T) x_{T-1}^i + O_T^i(\xi_T) x_{T-1}^{-i} + R_T^i(\xi_T) x_T^{-i} + B_T^i(\xi_T)^\top \pi_T^i(\xi_{[T]}) \ge 0, \\ 0 \le \pi_T^i(\xi_{[T]}) \perp A_T^i(\xi_T) x_{T-1}^i + A_T^{-i}(\xi_T) x_{T-1}^{-i} + B_T^i(\xi_T) x_T^i + B_T^{-i}(\xi_T) x_T^{-i} + b_T^i(\xi_T) \ge 0. \end{cases} \end{cases} \tag{19}$$

in the sense that: $(\pi_2(\cdot), \cdots, \pi_T(\cdot))$ *with* $\pi_t(\cdot) = (\pi_t^1(\cdot), \cdots, \pi_t^N(\cdot))$, *with* $\pi_t^i(\cdot)$ *being the dual solution of* (17) *for* $t = 2, \cdots, T-1$, $i = 1, \cdots, N$ *and* $\pi_T^i(\cdot)$ *being the dual solution of* (18) *for* $i = 1, \cdots, N$ *and* $(x_1, x_2(\cdot), \cdots, x_T(\cdot))$ *is a Nash equilibrium of* (16)-(18), *if and only if* $(x_1, x_2(\cdot), \pi_2(\cdot), \cdots, x_T(\cdot), \pi_T(\cdot))$ *is a solution of* (19).

Analogously, we know from the formulation (19) that it applies to (1) by viewing $x_1 = (x_1^1, \cdots, x_1^N)$ as the first stage variable and (x_t, π_t) as the t-th stage variable for $t = 2, \cdots, T$, where $x_t = (x_t^1, \cdots, x_t^N)$ and $\pi_t = (\pi_t^1, \cdots, \pi_t^N)$.

3 Monotonicity

In this section, we give some monotonicity assertions under relatively general settings.

To proceed further, denote

$$\Lambda_T(x_1, x_2, \cdots, x_T, \xi_{[T]}) := \begin{pmatrix} \Phi_1(x_1, x_2, \xi_2) \\ \Phi_2(x_1, x_2, x_3, \xi_3) \\ \vdots \\ \Phi_{T-1}(x_{T-2}, x_{T-1}, x_T, \xi_T) \\ \Phi_T(x_{T-1}, x_T, \xi_T) \end{pmatrix}$$

and, for $t = 1, 2, \cdots, T-1$, denote

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

where $\hat{x}_{t+1}(x_t, \xi_{t+1})$ is an arbitrarily selected solution of

$$0 \in \mathbb{E}_{\xi_{t+2}|\xi_{[t+1]}}[\Phi_{t+1}(x_t, x_{t+1}, \hat{x}_{t+2}(x_{t+1}, \xi_{t+2}), \xi_{t+1})] + \mathcal{N}_{\mathcal{C}_{t+1}(\xi_{[t+1]})}(x_{t+1})$$

for $t = 1, \cdots, T-2$ and $0 \in \Phi_T(x_{T-1}, x_T, \xi_T) + \mathcal{N}_{\mathcal{C}_T(\xi_{[T]})}(x_T)$ for $t = T-1$. It is noteworthy that we always make a blanket assumption that such $\hat{x}_{t+1}(x_t, \xi_{t+1})$ exists for $t = 1, 2, \cdots, T-1$ in the following discussion. This blanket assumption can be satisfied trivially under monotonicity with some additional mild assumptions (see e.g. [1, Chapter 2]).

We make the following assumptions.

Assumption 3. *For any $(u_1, u_2, \cdots, u_T)$, $(z_1, z_2, \cdots, z_T) \in \mathbb{R}^n$, we have*

$$\left\langle \Lambda_T(u_1, u_2, \cdots, u_T, \xi_{[T]}) - \Lambda_T(z_1, z_2, \cdots, z_T, \xi_{[T]}), (u_1 - z_1, u_2 - z_2, \cdots, u_T - z_T) \right\rangle \geq 0$$

for a.e. $\xi_{[T]} \in \Xi_{[T]}$.

Assumption 4. *For any $(u_1, u_2, \cdots, u_T)$, $(z_1, z_2, \cdots, z_T) \in \mathbb{R}^n$, there exists $\eta : \Xi_{[T]} \to \mathbb{R}_{++}$ such that*

$$\begin{aligned} &\left\langle \Lambda_T(u_1, u_2, \cdots, u_T, \xi_{[T]}) - \Lambda_T(z_1, z_2, \cdots, z_T, \xi_{[T]}), (u_1 - z_1, u_2 - z_2, \cdots, u_T - z_T) \right\rangle \\ &\qquad \geq \eta(\xi_{[T]}) \left(\|u_1 - z_1\|^2 + \|u_2 - z_2\|^2 + \cdots + \|u_T - z_T\|^2 \right) \end{aligned}$$

for a.e. $\xi_{[T]} \in \Xi_{[T]}$.

Assumptions 3 and 4 refer to monotonicity and strong monotonicity of MSVIs respectively, which are widely adopted in SVIs (see e.g. [10, 11, 23, 24]).

We first consider the following auxiliary VIs:

$$\begin{cases} 0 \in f(x, y) + \mathcal{N}_X(x), \\ 0 \in g(x, y) + \mathcal{N}_Y(y), \end{cases} \tag{20}$$

where $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ are closed convex sets, $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ and $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$. For the brevity of description, denote

$$H(x, y) := \begin{pmatrix} f(x, y) \\ g(x, y) \end{pmatrix}.$$

We make blanket assumptions that problem (20) always has a solution and, for any fixed $x \in X$, $0 \in g(x, y) + \mathcal{N}_Y(y)$ has at least a solution.

Next, we have the following lemma w.r.t. (20).

Lemma 3.1. *If $H(x, y)$ is monotone (strongly monotone with modulus $\gamma > 0$) w.r.t. (x, y) on $X \times Y$, then $H(x, \hat{y}(x))$ is monotone (strongly monotone with modulus $\gamma > 0$) w.r.t. x on X, where $\hat{y}(x)$ is an arbitrarily selected solution of $0 \in g(x, y) + \mathcal{N}_Y(y)$ for fixed $x \in X$.*

Proof. We consider the monotone case firstly and verify that $H(x, \hat{y}(x))$ is monotone w.r.t. x on X. According to the definition of monotonicity, we have

$$\left\langle \begin{pmatrix} f(x_1, y_1) - f(x_2, y_2) \\ g(x_1, y_1) - g(x_2, y_2) \end{pmatrix}, \begin{pmatrix} x_1 - x_2 \\ y_1 - y_2 \end{pmatrix} \right\rangle \geq 0$$

for arbitrary $x_1, x_2 \in X$ and $y_1, y_2 \in Y$, which is equivalent to

$$\langle f(x_1, y_1) - f(x_2, y_2), x_1 - x_2 \rangle + \langle g(x_1, y_1) - g(x_2, y_2), y_1 - y_2 \rangle \geq 0.$$

Take y_1 and y_2 to be $\hat{y}_1 := \hat{y}(x_1)$ and $\hat{y}_2 := \hat{y}(x_2)$ respectively, where $\hat{y}(x)$ denotes an arbitrary solution of $0 \in g(x, y) + \mathcal{N}_Y(y)$ for given $x \in X$, and we obtain

$$\langle f(x_1, \hat{y}_1) - f(x_2, \hat{y}_2), x_1 - x_2\rangle + \langle g(x_1, \hat{y}_1) - g(x_2, \hat{y}_2), \hat{y}_1 - \hat{y}_2\rangle \geq 0. \tag{21}$$

Since $\hat{y}_1$ and $\hat{y}_2$ are solutions of VIs $0 \in g(x_1, y)+\mathcal{N}_Y(y)$ and $0 \in g(x_2, y)+\mathcal{N}_Y(y)$, respectively, we have

$$\begin{cases} \langle g(x_1, \hat{y}_1), y - \hat{y}_1\rangle \geq 0, \\ \langle g(x_2, \hat{y}_2), y - \hat{y}_2\rangle \geq 0 \end{cases}$$

for any $y \in Y$. Taking $y = \hat{y}_2$ and $y = \hat{y}_1$ in the above two inequalities respectively, add them together, and we obtain

$$\langle g(x_2, \hat{y}_2) - g(x_1, \hat{y}_1), \hat{y}_1 - \hat{y}_2\rangle \geq 0. \tag{22}$$

Then, (22) together with (21) yields $\langle f(x_1, \hat{y}_1) - f(x_2, \hat{y}_2), x_1 - x_2\rangle \geq 0$, which implies the monotonicity of $f(\cdot, \hat{y}(\cdot))$.

For the strong monotonicity case, we have

$$\langle f(x_1, y_1) - f(x_2, y_2), x_1 - x_2\rangle + \langle g(x_1, y_1) - g(x_2, y_2), y_1 - y_2\rangle \geq \gamma(\|x_1 - x_2\|^2 + \|y_1 - y_2\|^2)$$

for any $x_1, x_2 \in X$ and $y_1, y_2 \in Y$ and some strong monotonicity modulus $\gamma > 0$. Analogously, take y_1 and y_2 to be $\hat{y}_1 := \hat{y}(x_1)$ and $\hat{y}_2 := \hat{y}(x_2)$, where $\hat{y}(x)$ denotes the unique solution of $0 \in g(x, y) + \mathcal{N}_Y(y)$ for fixed $x \in X$, and we obtain

$$\langle f(x_1, \hat{y}_1) - f(x_2, \hat{y}_2), x_1 - x_2\rangle + \langle g(x_1, \hat{y}_1) - g(x_2, \hat{y}_2), \hat{y}_1 - \hat{y}_2\rangle \geq \gamma(\|x_1 - x_2\|^2 + \|\hat{y}_1 - \hat{y}_2\|^2).$$

Similarly, we can derive that $\langle g(x_1, \hat{y}_1) - g(x_2, \hat{y}_2), \hat{y}_1 - \hat{y}_2\rangle \leq 0$, which implies that

$$\begin{aligned} \langle f(x_1, \hat{y}(x_1)) - f(x_2, \hat{y}(x_2)), x_1 - x_2\rangle &\geq \gamma(\|x_1 - x_2\|^2 + \|\hat{y}_1 - \hat{y}_2\|^2) \\ &\geq \gamma \|x_1 - x_2\|^2 . \end{aligned}$$

Then we complete the proof. ■

Remark 3.1. *In the existing works [11, 12, 23], to derive the monotonicity property of $f(x, \hat{y}(x))$ w.r.t. x, they compute the subdifferential $\partial_x f(x, \hat{y}(x))$ and verify the positive semidefiniteness of every element contained in $\partial_x f(x, \hat{y}(x))$. However, in order to calculate $\partial_x f(x, \hat{y}(x))$, some strong conditions are required, for instance, the solution $\hat{y}(x)$ should be unique and the set Y should be polyhedral (see [25]). All these limit the application range of two-stage SVIs or MSVIs. In Lemma 3.1, we avoid these restrictive assumptions.*

Based on Lemma 3.1, we have the following monotonicity assertions.

Proposition 3.1. *Under Assumption 3, $\mathbb{E}_{\xi_{t+1}|\xi_{[t]}}\left[\Lambda_t(x_1, x_2, \cdots, x_t, \xi_{[t+1]})\right]$ is monotone w.r.t. $(x_1, \cdots, x_t)$ for $t = 1, \cdots, T-1$. If, further, Assumption 4 holds, $\mathbb{E}_{\xi_{t+1}|\xi_{[t]}}\left[\Lambda_t(x_1, x_2, \cdots, x_t, \xi_{[t+1]})\right]$ is strongly monotone w.r.t. $(x_1, \cdots, x_t)$ with modulus $\mathbb{E}_{\xi_{[T]}|\xi_{[t]}}[\eta(\xi_{[T]})]$ for $t = 1, \cdots, T-1$.*

Proof. We only verify the first assertion under Assumption 3. The strongly monotone case under Assumption 4 can be similarly derived as that in Lemma 3.1.

For $t = T-1$, by viewing $\Lambda_T(x_1, x_2, \cdots, x_T, \xi_{[T]})$ as

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

we know from Lemma 3.1 and Assumption 3 that

$$\Lambda_{T-1}(x_1, x_2, \cdots, x_{T-1}, \xi_{[T]}) = \begin{pmatrix} \Phi_1(x_1, x_2, \xi_2) \\ \Phi_2(x_1, x_2, x_3, \xi_3) \\ \vdots \\ \Phi_{T-1}(x_{T-2}, x_{T-1}, \hat{x}_T(x_{T-1}, \xi_T), \xi_T) \end{pmatrix}$$

is monotone w.r.t. $(x_1, \cdots, x_{T-1})$, which implies that

$$\mathbb{E}_{\xi_T|\xi_{[T-1]}}[\Lambda_{T-1}(x_1, x_2, \cdots, x_{T-1}, \xi_{[T]})] = \begin{pmatrix} \Phi_1(x_1, x_2, \xi_2) \\ \Phi_2(x_1, x_2, x_3, \xi_3) \\ \vdots \\ \mathbb{E}_{\xi_T|\xi_{[T-1]}}[\Phi_{T-1}(x_{T-2}, x_{T-1}, \hat{x}_T(x_{T-1}, \xi_T), \xi_T)] \end{pmatrix}$$

is monotone. Similarly, for $t = T-2$, consider

$$\begin{pmatrix} \begin{pmatrix} \Phi_1(x_1, x_2, \xi_2) \\ \Phi_2(x_1, x_2, x_3, \xi_3) \\ \vdots \\ \Phi_{T-2}(x_{T-3}, x_{T-2}, x_{T-1}, \xi_{T-1}), \xi_{T-1}) \end{pmatrix} \\ \mathbb{E}_{\xi_T|\xi_{[T-1]}}[\Phi_{T-1}(x_{T-2}, x_{T-1}, \hat{x}_T(x_{T-1}, \xi_T), \xi_T)] \end{pmatrix},$$

and we know from Lemma 3.1 that $\Lambda_{T-2}(x_1, x_2, \cdots, x_{T-2}, \xi_{[T-1]})$ is monotone w.r.t. $(x_1, \cdots, x_{T-2})$, and thus $\mathbb{E}_{\xi_{T-1}|\xi_{[T-2]}}[\Lambda_{T-2}(x_1, x_2, \cdots, x_{T-2}, \xi_{[T-1]})]$ is monotone. Repeat this procedure till $t = 1$, and we complete the proof. ∎

Theorem 3.1. *The following assertions hold.*

(i) Let Assumption 3 hold. For any given x_{t-1} and $\xi_{[t]}$ with $t = 1, \cdots, T-1$,

$$\mathbb{E}_{\xi_{t+1}|\xi_{[t]}}[\Phi_t(x_{t-1}, x_t, \hat{x}_{t+1}(x_t, \xi_{t+1}), \xi_{t+1})]$$

is monotone w.r.t. x_t.

(ii) If, further, Assumption 4 holds, for any given x_{t-1} and $\xi_{[t]}$ with $t = 1, \cdots, T$,

$$\mathbb{E}_{\xi_{t+1}|\xi_{[t]}}[\Phi_t(x_{t-1}, x_t, \hat{x}_{t+1}(x_t, \xi_{t+1}), \xi_{t+1})]$$

is strongly monotone w.r.t. x_t with modulus $\mathbb{E}_{\xi_{[T]}|\xi_{[t]}}[\eta(\xi_{[T]})]$.

Proof. Part (i): By Proposition 3.1, we know that $\mathbb{E}_{\xi_{t+2}|\xi_{[t+1]}}\left[\Lambda_{t+1}(x_1, x_2, \cdots, x_{t+1}, \xi_{[t+2]})\right]$ is monotone w.r.t. $(x_1, \cdots, x_{t+1})$. Thus we have

$$\begin{aligned} &\left\langle \mathbb{E}_{\xi_{t+2}|\xi_{[t+1]}}\left[\Lambda_{t+1}(x_1, x_2, \cdots, x_t', x_{t+1}', \xi_{[t+2]})\right] - \mathbb{E}_{\xi_{t+2}|\xi_{[t+1]}}\left[\Lambda_{t+1}(x_1, x_2, \cdots, x_t, x_{t+1}, \xi_{[t+2]})\right], \right\rangle \\ &= \left\langle \begin{pmatrix} \Phi_t(x_{t-1}, x_t', x_{t+1}', \xi_{t+1}) - \Phi_t(x_{t-1}, x_t, x_{t+1}, \xi_{t+1}) \\ \mathbb{E}_{\xi_{t+2}|\xi_{[t+1]}}[\Phi_{t+1}(x_t', x_{t+1}', \hat{x}_{t+2}(x_{t+1}', \xi_{t+2}), \xi_{t+2}) - \Phi_{t+1}(x_t, x_{t+1}, \hat{x}_{t+2}(x_{t+1}, \xi_{t+2}), \xi_{t+2})] \end{pmatrix}, \begin{pmatrix} x_t' - x_t \\ x_{t+1}' - x_{t+1} \end{pmatrix} \right\rangle \\ &\geq 0 \end{aligned}$$

for any $x_1, \cdots, x_{t-1}$ and $x_t', x_t, x_{t+1}', x_{t+1}$, which implies that: for given x_{t-1} and $\xi_{[t+1]}$,

$$\begin{pmatrix} \Phi_t(x_{t-1}, x_t, x_{t+1}, \xi_{t+1}) \\ \mathbb{E}_{\xi_{t+2}|\xi_{[t+1]}}[\Phi_{t+1}(x_t, x_{t+1}, \hat{x}_{t+2}(x_{t+1}, \xi_{t+2}), \xi_{t+2})] \end{pmatrix}$$

is monotone w.r.t. (x_t, x_{t+1}). By Lemma 3.1, we have

$$\Phi_t(x_{t-1}, x_t, \hat{x}_{t+1}(x_t, \xi_{t+1}), \xi_{t+1})$$

is monotone w.r.t. x_t, which implies that

$$\mathbb{E}_{\xi_{t+1}|\xi_{[t]}}[\Phi_t(x_{t-1}, x_t, \hat{x}_{t+1}(x_t, \xi_{t+1}), \xi_{t+1})]$$

is monotone w.r.t. x_t.

Part (ii): By a similar procedure as Part (i), we have, for given x_{t-1} and $\xi_{[t+1]}$, that

$$\begin{pmatrix} \Phi_t(x_{t-1}, x_t, x_{t+1}, \xi_{t+1}) \\ \mathbb{E}_{\xi_{t+2}|\xi_{[t+1]}}[\Phi_{t+1}(x_t, x_{t+1}, \hat{x}_{t+2}(x_{t+1}, \xi_{t+2}), \xi_{t+2})] \end{pmatrix}$$

is strongly monotone w.r.t. (x_t, x_{t+1}) with modulus $\mathbb{E}_{\xi_{[T]}|\xi_{[t+1]}}[\eta(\xi_{[T]})]$. Thus, by using Lemma 3.1, we know that

$$\Phi_t(x_{t-1}, x_t, \hat{x}_{t+1}(x_t, \xi_{t+1}), \xi_{t+1})$$

is strongly monotone w.r.t. x_t with modulus $\mathbb{E}_{\xi_{[T]}|\xi_{[t+1]}}[\eta(\xi_{[T]})]$, which implies that

$$\mathbb{E}_{\xi_{t+1}|\xi_{[t]}}[\Phi_t(x_{t-1}, x_t, \hat{x}_{t+1}(x_t, \xi_{t+1}), \xi_{t+1})]$$

is strongly monotone w.r.t. x_t with modulus $\mathbb{E}_{\xi_{t+1}|\xi_{[t]}}[\mathbb{E}_{\xi_{[T]}|\xi_{[t+1]}}[\eta(\xi_{[T]})]] = \mathbb{E}_{\xi_{[T]}|\xi_{[t]}}[\eta(\xi_{[T]})]$. ■

Lemma 3.2. *Let Assumption 4 hold. Suppose that $\Phi_t(\cdot, \cdot, \xi_{t+1})$ for $t = 1, \cdots, T$ are locally Lipschitz continuous. Then $\hat{x}_{t+1}(x_t, \xi_{t+1})$ is unique and locally Lipschitz continuous w.r.t. x_t.*

Proof. Note that $\hat{x}_{t+1}(x_t, \xi_{t+1})$ is the solution of

$$0 \in \mathbb{E}_{\xi_{t+1}|\xi_{[t]}}[\Phi_t(x_{t-1}, x_t, \hat{x}_{t+1}(x_t, \xi_{t+1}), \xi_{t+1})] + \mathcal{N}_{\mathcal{C}_t(\xi_{[t]})}(x_t).$$

Since $\mathbb{E}_{\xi_{t+1}|\xi_{[t]}}[\Phi_t(x_{t-1}, x_t, \hat{x}_{t+1}(x_t, \xi_{t+1}), \xi_{t+1})]$ is strongly monotone with modulus $\mathbb{E}_{\xi_{[T]}|\xi_{[t]}}[\eta(\xi_{[T]})]$, we know from [23, Lemma 2] that the assertions hold. ■

Theorem 3.2. *Let Assumption 4 hold. Problem (1) is guaranteed to have a unique solution.*

Proof. By Theorem 3.1, we know that

$$\mathbb{E}_{\xi_{[2]}}[\Phi_1(x_1, \hat{x}_2(x_1, \xi_2), \xi_2)]$$

is strongly monotone w.r.t. x_1, which implies that

$$0 \in \mathbb{E}_{\xi_{[2]}}[\Phi_1(x_1, \hat{x}_2(x_1, \xi_2), \xi_2)] + \mathcal{N}_{\mathcal{C}_1}(x_1)$$

has a unique solution, denoted by x_1^*. Similarly, we know that

$$0 \in \mathbb{E}_{\xi_{[3]}|\xi_{[2]}}[\Phi_2(x_1, x_2, \hat{x}_3(x_2, \xi_3), \xi_3)] + \mathcal{N}_{\mathcal{C}_2(\xi_{[2]})}(x_2)$$

has a unique solution $x_2^*(\xi_{[2]})$. By this way, we can find a unique sequence $(x_1^*, x_2^*(\cdot), \cdots, x_T^*(\cdot))$ such that it uniquely solves problem (1). ■

Theorems 3.1 and 3.2 are important for driving the convergence rate of SAA with conditional sampling problem in the following section.

4 Complexity Analysis

In [23], the convergence analysis of SAA with conditional sampling for multistage stochastic nonlinear CPs has been investigated. However, it fails to address an important issue that is interesting from both the conceptual and the computational viewpoints. That is, how large the sample size should be to achieve a desired accuracy? In this section, we consider the complexity of SAA with conditional sampling of MSVIs under Assumption 4.

In what follows, for brevity, we focus only on the three-stage case because the number of stages does not affect the validity of our analysis for $T > 3$, but a large number of stages does increase lots of tedious notations. By letting $T = 3$ in (1), we can write the three-stage SVI as follows:

$$\begin{cases} 0 \in \mathbb{E}_{\xi_2}[\Phi_1(x_1, x_2(\xi_{[2]}), \xi_2)] + \mathcal{N}_{\mathcal{C}_1}(x_1), \\ 0 \in \mathbb{E}_{\xi_3|\xi_{[2]}}\left[\Phi_2(x_1, x_2(\xi_{[2]}), x_3(\xi_{[3]}), \xi_3)\right] + \mathcal{N}_{\mathcal{C}_2(\xi_{[2]})}(x_2(\xi_{[2]})), \text{ a.e. } \xi_{[2]} \in \Xi_{[2]}, \\ 0 \in \Phi_3(x_2(\xi_{[2]}), x_3(\xi_{[3]}), \xi_3) + \mathcal{N}_{\mathcal{C}_3(\xi_{[3]})}(x_3(\xi_{[3]})), \text{ a.e. } \xi_{[3]} \in \Xi_{[3]}. \end{cases} \tag{23}$$

Under Assumption 4, we know from Section 3 that for each x_2 and $\xi_3 \in \Xi_3$, the third stage problem has a unique solution, denoted by $\hat{x}_3(x_2, \xi_{[3]})$. Thus, we can equivalently rewrite (23) as

$$\begin{cases} 0 \in \mathbb{E}_{\xi_2}[\Phi_1(x_1, x_2(\xi_{[2]}), \xi_2)] + \mathcal{N}_{\mathcal{C}_1}(x_1), \\ 0 \in \mathbb{E}_{\xi_3|\xi_{[2]}}\left[\Phi_2(x_1, x_2(\xi_{[2]}), \hat{x}_3(x_2(\xi_{[2]}), \xi_{[3]}), \xi_3)\right] + \mathcal{N}_{\mathcal{C}_2(\xi_{[2]})}(x_2(\xi_{[2]})), \text{ a.e. } \xi_{[2]} \in \Xi_{[2]}. \end{cases} \tag{24}$$

Let $\xi_2^1, \xi_2^2, \cdots, \xi_2^{N_1}$ be i.i.d. samples of random vector ξ_2. Further, we have, according to ξ_2^i $(1 \le i \le N_1)$, let $\xi_3^{i1}, \xi_3^{i2}, \cdots, \xi_3^{iN_2}$ be i.i.d. conditional samples of ξ_3. Thus, we have total $N_1 N_2$ scenarios or sample paths, and the SAA problem of (24) with conditional sampling is as follows:

$$\begin{cases} 0 \in \dfrac{1}{N_1}\displaystyle\sum_{i=1}^{N_1} \Phi_1(x_1, x_2(\xi_2^i), \xi_2^i) + \mathcal{N}_{\mathcal{C}_1}(x_1), \\ 0 \in \dfrac{1}{N_2}\displaystyle\sum_{k=1}^{N_2} \Phi_2(x_1, x_2(\xi_2^j), \hat{x}_3(x_2(\xi_2^j), \xi_{[3]}^{jk}), \xi_3^{jk}) + \mathcal{N}_{\mathcal{C}_2(\xi_2^j)}(x_2(\xi_2^j)), \text{ for } j = 1, 2, \cdots, N_1, \end{cases} \tag{25}$$

where $\xi_{[3]}^{jk} := (\xi_1, \xi_2^j, \xi_3^{jk})$ for $j = 1, 2, \cdots, N_1$ and $k = 1, 2, \cdots, N_2$. Since ξ_1 is deterministic, in what follows, we do not distinguish $\xi_{[2]}^j = (\xi_1, \xi_2^j)$ from ξ_2^j. According to Theorem 3.2, we know that both problems (24) and (25) have unique solutions. Then we use $\hat{x}_2(x_1, \xi_2)$ to denote the unique solution of the second stage problem of (24) for given x_1 and ξ_2. We use $\tilde{x}_2^{N_2}(x_1, \xi_2^j)$ to denote the unique solution of the second stage problem of (25) for given x_1 and ξ_2^j.

Then we can further write problems (24) and (25) as

$$0 \in \mathbb{E}_{\xi_2}[\Phi_1(x_1, \hat{x}_2(x_1, \xi_{[2]}), \xi_2)] + \mathcal{N}_{\mathcal{C}_1}(x_1) \tag{26}$$

and

$$0 \in \frac{1}{N_1}\sum_{i=1}^{N_1} \Phi_1(x_1, \tilde{x}_2^{N_2}(x_1, \xi_2^i), \xi_2^i) + \mathcal{N}_{\mathcal{C}_1}(x_1). \tag{27}$$

Let x_1^* and $\tilde{x}_1^{N_1,N_2}$ denote the unique solutions of problems (26) and (27), respectively. In the reminder of this section, we investigate the convergence rate between x_1^* and $\tilde{x}_1^{N_1,N_2}$ w.r.t. N_1, N_2.

We make the following assumptions.

Assumption 5. *Let the following assertions hold.*

(i) There exist convex and compact sets $\mathcal{X}_1 \subseteq \mathbb{R}^{n_1}$ and $\mathcal{X}_2 \subseteq \mathbb{R}^{n_2}$ such that $x_1^*, \tilde{x}_1^{N_1,N_2} \in \mathcal{X}_1$ for any $N_1, N_2 \in \mathbb{N}$ and $\hat{x}_2(x_1,\xi_2), \tilde{x}_2^{N_2}(x_1,\xi_2) \in \mathcal{X}_2$ for any $N_2 \in \mathbb{N}$, $x_1 \in \mathcal{X}_1$ and $\xi_2 \in \Xi_2$.

(ii) For each $\xi_2 \in \Xi_2$ and $\xi_3 \in \Xi_3$, $\Phi_1(\cdot,\cdot,\xi_2)$, $\Phi_2(\cdot,\cdot,\cdot,\xi_3)$ and $\Phi_3(\cdot,\cdot,\xi_3)$ are Lipschitz continuous with moduli $L_1(\xi_2)$, $L_2(\xi_3)$, $L_3(\xi_3)$, respectively.

Assumption 5 is standard. Specifically, (i) in Assumption 5 holds when $\mathcal{C}_1 \subseteq \mathcal{X}_1$ and $\mathcal{C}_2(\xi_{[2]}) \subseteq \mathcal{X}_2$. (ii) in Assumption 5 is commonly-used in multistage stochastic equilibrium [23].

To derive the rate of convergence between $\tilde{x}_1^{N_1,N_2}$ and x_1^*, we first investigate the uniform rate of convergence between

$$\frac{1}{N_1}\sum_{i=1}^{N_1} \Phi_1(x_1, \tilde{x}_2^{N_2}(x_1,\xi_2^i), \xi_2^i) \text{ and } \mathbb{E}_{\xi_2}[\Phi_1(x_1, \hat{x}_2(x_1,\xi_{[2]}), \xi_2)]$$

over $\mathcal{X}_1$. For this purpose, we introduce the following intermediate term:

$$\frac{1}{N_1}\sum_{i=1}^{N_1} \Phi_1(x_1, \hat{x}_2(x_1,\xi_2^i), \xi_2^i).$$

Then we have

$$\begin{aligned} &\text{Prob}\left\{\sup_{x_1\in\mathcal{X}_1}\left\|\frac{1}{N_1}\sum_{i=1}^{N_1} \Phi_1(x_1, \tilde{x}_2^{N_2}(x_1,\xi_2^i), \xi_2^i) - \mathbb{E}_{\xi_2}[\Phi_1(x_1, \hat{x}_2(x_1,\xi_{[2]}), \xi_2)]\right\| \geq \epsilon\right\} \\ &\leq \text{Prob}\left\{\sup_{x_1\in\mathcal{X}_1}\left\|\frac{1}{N_1}\sum_{i=1}^{N_1} \Phi_1(x_1, \tilde{x}_2^{N_2}(x_1,\xi_2^i), \xi_2^i) - \frac{1}{N_1}\sum_{i=1}^{N_1} \Phi_1(x_1, \hat{x}_2(x_1,\xi_2^i), \xi_2^i)\right\| \geq \frac{\epsilon}{2}\right\} \quad (28) \\ &+ \text{Prob}\left\{\sup_{x_1\in\mathcal{X}_1}\left\|\frac{1}{N_1}\sum_{i=1}^{N_1} \Phi_1(x_1, \hat{x}_2(x_1,\xi_2^i), \xi_2^i) - \mathbb{E}_{\xi_2}[\Phi_1(x_1, \hat{x}_2(x_1,\xi_{[2]}), \xi_2)]\right\| \geq \frac{\epsilon}{2}\right\}. \end{aligned}$$

Thus, in the sequel, we estimate

$$\text{Prob}\left\{\sup_{x_1\in\mathcal{X}_1}\left\|\frac{1}{N_1}\sum_{i=1}^{N_1} \Phi_1(x_1, \tilde{x}_2^{N_2}(x_1,\xi_2^i), \xi_2^i) - \frac{1}{N_1}\sum_{i=1}^{N_1} \Phi_1(x_1, \hat{x}_2(x_1,\xi_2^i), \xi_2^i)\right\| \geq \frac{\epsilon}{2}\right\} \quad (29)$$

and

$$\text{Prob}\left\{\sup_{x_1\in\mathcal{X}_1}\left\|\frac{1}{N_1}\sum_{i=1}^{N_1} \Phi_1(x_1, \hat{x}_2(x_1,\xi_2^i), \xi_2^i) - \mathbb{E}_{\xi_2}[\Phi_1(x_1, \hat{x}_2(x_1,\xi_{[2]}), \xi_2)]\right\| \geq \frac{\epsilon}{2}\right\}, \quad (30)$$

respectively.

By directly invoking Lemma 3.2, due to the convexity and compactness of $\mathcal{X}_1$ and $\mathcal{X}_2$, we have the following lemma.

Lemma 4.1. *Let Assumptions 4 and 5 hold. Then, for any given $\xi_{[2]}$ and $\xi_{[3]}$, $\hat{x}_2(\cdot,\xi_{[2]})$ and $\hat{x}_3(\cdot,\xi_{[3]})$ are Lipschitz continuous over $\mathcal{X}_1$ and $\mathcal{X}_2$.*

Denote the Lipschitz moduli of $\hat{x}_2(\cdot,\xi_{[2]})$ over $\mathcal{X}_1$ and $\hat{x}_3(\cdot,\xi_{[3]})$ over $\mathcal{X}_2$ by $\hat{L}_2(\xi_{[2]})$ and $\hat{L}_3(\xi_{[3]})$, respectively.

Next, we consider the following auxiliary variational inequalities:

$$0 \in F(x) + \mathcal{N}_X(x), \quad (31)$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ and $X \subseteq \mathbb{R}^n$ is a closed convex set, and its perturbation:

$$0 \in \tilde{F}(x) + \mathcal{N}_X(x), \tag{32}$$

where $\tilde{F} : \mathbb{R}^n \to \mathbb{R}^n$.

Then we have the following proposition.

Proposition 4.1. *Let F be strongly monotone and Lipschitz continuous and $\tilde{F}$ be strongly monotone and continuous. Then both problems (31) and (32) have unique solutions, denoted by x^* and $\tilde{x}^*$, respectively. Moreover,*

$$\|x^* - \tilde{x}^*\| \leq L_F \sup_{x \in X} \left\| F(x) - \tilde{F}(x) \right\|,$$

where L_F is a positive constant depends only on the strong monotonicity modulus and Lipschitz modulus of F.

Proof. It is known from [1, Theorem 2.3.3] that both VIs (31) and (32) have unique solutions. Based on [29, Theorem 3.1], we have that, for any $x \in X$,

$$\|x - x^*\| \leq L_F \|x - \text{Proj}_X(x - F(x))\|, \tag{33}$$

where $L_F > 0$ depends only on the strong monotonicity modulus and Lipschitz constant of F.

Let $x = \tilde{x}^* = \text{Proj}_X(\tilde{x}^* - \tilde{F}(\tilde{x}^*))$. By substituting it into (33), we obtain

$$\begin{aligned}
\|\tilde{x}^* - x^*\| &\leq L_F \left\| \text{Proj}_X(\tilde{x}^* - \tilde{F}(\tilde{x}^*)) - \text{Proj}_X(\tilde{x}^* - F(\tilde{x}^*)) \right\| \\
&\leq L_F \left\| \tilde{x}^* - \tilde{F}(\tilde{x}^*) - \tilde{x}^* + F(\tilde{x}^*) \right\| \\
&= L_F \left\| \tilde{F}(\tilde{x}^*) - F(\tilde{x}^*) \right\| \\
&\leq L_F \sup_{x \in X} \left\| \tilde{F}(x) - F(x) \right\|.
\end{aligned}$$

The proof is complete. ■

Immediately, by applying Proposition 4.1 to the second stage problems of (24) and (25), we have the following proposition.

Proposition 4.2. *Let Assumptions 4 and 5 hold. For given $x_1 \in \mathcal{X}_1$ and $\xi_2^j \in \Xi_2$, we have*

$$\left\| \tilde{x}_2^{N_2}(x_1, \xi_2^j) - \hat{x}_2(x_1, \xi_2^j) \right\| \leq$$

$$L_{\Phi_2}(\xi_{[2]}^j) \sup_{x_2 \in \mathcal{X}_2} \left\| \frac{1}{N_2} \sum_{k=1}^{N_2} \Phi_2(x_1, x_2, \hat{x}_3(x_2, \xi_{[3]}^{jk}), \xi_3^{jk}) - \mathbb{E}_{\xi_3 | \xi_{[2]}^j} \left[\Phi_2(x_1, x_2, \hat{x}_3(x_2, \xi_{[3]}), \xi_3) \right] \right\|,$$

where $L_{\Phi_2}(\xi_{[2]}^j)$ is a positive constant depending on $\xi_{[2]}^j$.

Proof. By Theorem 3.1, we known that $\mathbb{E}_{\xi_3|\xi_{[2]}^j} \left[\Phi_2(x_1, \cdot, \hat{x}_3(\cdot, \xi_{[3]}), \xi_3) \right]$ is strongly monotone with modulus $\mathbb{E}_{\xi_3|\xi_{[2]}^j}[\eta(\xi_{[3]})]$. Moreover, we have from Lemma 4.1 and Assumption 5 that $\mathbb{E}_{\xi_3|\xi_{[2]}^j} \left[\Phi_2(x_1, \cdot, \hat{x}_3(\cdot, \xi_{[3]}), \xi_3) \right]$ is also Lipschitz continuous, where the Lipschitz modulus is independent of x_1. Consider the following two problems:

$$0 \in \mathbb{E}_{\xi_3|\xi_{[2]}^j} \left[\Phi_2(x_1, x_2, \hat{x}_3(x_2(\xi_{[2]}), \xi_{[3]}), \xi_3) \right] + \mathcal{N}_{\mathcal{C}_2(\xi_{[2]}^j)}(x_2),$$

$$0 \in \frac{1}{N_2} \sum_{k=1}^{N_2} \Phi_2(x_1, x_2, \hat{x}_3(x_2(\xi_2^j), \xi_{[3]}^{jk}), \xi_3^{jk}) + \mathcal{N}_{\mathcal{C}_2(\xi_{[2]}^j)}(x_2).$$

Then, by directly invoking Proposition 4.1, we complete the proof. ■

We proceed to have from (29) that

$$\begin{aligned}
&\sup_{x_1\in\mathcal{X}_1}\left\|\frac{1}{N_1}\sum_{i=1}^{N_1}\Phi_1(x_1,\tilde{x}_2^{N_2}(x_1,\xi_2^i),\xi_2^i)-\frac{1}{N_1}\sum_{i=1}^{N_1}\Phi_1(x_1,\hat{x}_2(x_1,\xi_2^i),\xi_2^i)\right\|\\
&\le \frac{1}{N_1}\sum_{i=1}^{N_1}\sup_{x_1\in\mathcal{X}_1}\left\|\Phi_1(x_1,\tilde{x}_2^{N_2}(x_1,\xi_2^i),\xi_2^i)-\Phi_1(x_1,\hat{x}_2(x_1,\xi_2^i),\xi_2^i)\right\|\\
&\le \frac{1}{N_1}\sum_{i=1}^{N_1}L_1(\xi_2^i)\sup_{x_1\in\mathcal{X}_1}\left\|\tilde{x}_2^{N_2}(x_1,\xi_2^i)-\hat{x}_2(x_1,\xi_2^i)\right\|\\
&\le \frac{1}{N_1}\sum_{i=1}^{N_1}\left(\kappa(\xi_{[2]}^i)\sup_{\substack{x_1\in\mathcal{X}_1\\x_2\in\mathcal{X}_2}}\left\|\frac{1}{N_2}\sum_{k=1}^{N_2}\Phi_2(x_1,x_2,\hat{x}_3(x_2,\xi_{[3]}^{ik}),\xi_3^{ik})-\mathbb{E}_{\xi_3|\xi_{[2]}^i}\left[\Phi_2(x_1,x_2,\hat{x}_3(x_2,\xi_{[3]}),\xi_3)\right]\right\|\right),
\end{aligned}\tag{34}$$

where $\kappa(\xi_{[2]}^i):=L_1(\xi_2^i)L_{\Phi_2}(\xi_{[2]}^i)$ and the last inequality follows from Proposition 4.2.

To analyse the rate of convergence, we need the following useful lemma.

Lemma 4.2 ([24, Lemma 3.1]). *Let $p\ge 2$, ω be a random vector supported over Ω, $F:\mathbb{R}^n\times\Omega\to\mathbb{R}$ and $\tau:\Omega\to\mathbb{R}_+$ be a measurable function with $\mathbb{E}[\tau(\omega)^p]<+\infty$. Suppose that: (i) $\mathcal{X}\subseteq\mathbb{R}^n$ is compact; (ii) $\mathbb{E}[|F(x,\omega)|^p]<+\infty$ for each $x\in\mathcal{X}$; (iii) $\omega^1,\cdots,\omega^N$ are iid samples of ω; (iv) $F(\cdot,\omega)$ is Lipschitz continuous over $\mathcal{X}$ with Lipschitz modulus $\tau(\omega)$. Then, for arbitrary $\epsilon>0$, there exists a positive scalar θ independent of sample size N, such that*

$$\mathrm{Prob}\left\{\sup_{x\in\mathcal{X}}\left|\hat{f}_N(x)-f(x)\right|\ge\epsilon\right\}\le\frac{\theta}{N^{\frac{p}{2}}\epsilon^p}$$

for sufficiently large N, where $\hat{f}_N(x):=\frac{1}{N}\sum_{i=1}^N F(x,\omega^i)$ and $f(x):=\mathbb{E}[F(x,\omega)]$.

Denote the Lipschitz moduli of $\Phi_1(x_1,\hat{x}_2(x_1,\xi_{[2]}),\xi_2)$ w.r.t. x_1 by $\kappa_1(\xi_{[2]}):=L_1(\xi_2)\cdot\hat{L}_2(\xi_{[2]})$ and $\Phi_2(x_1,x_2,\hat{x}_3(x_2,\xi_{[3]}),\xi_3)$ w.r.t. x_2 by $\kappa_2(\xi_{[3]}):=L_2(\xi_3)\cdot\hat{L}_3(\xi_{[3]})$.

Then we make the following standard assumptions.

Assumption 6. *For $p\ge 2$, the following assertions hold:*

(i) For given $x_1\in\mathcal{X}_1$, $\mathbb{E}_{\xi_{[2]}}\left[\left\|\Phi_1(x_1,\hat{x}_2(x_1,\xi_{[2]}),\xi_2)\right\|^p\right]$ is finite valued.

(ii) $\mathbb{E}_{\xi_{[2]}}\left[\kappa_1(\xi_{[2]})^p\right]$ is finite valued.

(iii) For given $\xi_{[2]}\in\Xi_{[2]}$, $x_1\in\mathcal{X}_1$ and $x_2\in\mathcal{X}_2$, $\mathbb{E}_{\xi_{[3]}|\xi_{[2]}}\left[\left\|\Phi_2(x_1,x_2,\hat{x}_3(x_2,\xi_{[3]}),\xi_3)\right\|^p\right]$ is finite valued.

(iv) For given $\xi_{[2]}\in\Xi_{[2]}$, $\mathbb{E}_{\xi_{[3]}|\xi_{[2]}}\left[\kappa_2(\xi_{[3]})^p\right]$ is finite valued.

Based on Assumption 6, we know from Lemma 4.2 that: there exists positive constants θ_1 and $\theta_2(\xi_{[2]}^i)$ such that

$$\mathrm{Prob}\left\{\sup_{x_1\in\mathcal{X}_1}\left\|\frac{1}{N_1}\sum_{i=1}^{N_1}\Phi_1(x_1,\hat{x}_2(x_1,\xi_2^i),\xi_2^i)-\mathbb{E}_{\xi_2}[\Phi_1(x_1,\hat{x}_2(x_1,\xi_{[2]}),\xi_2)]\right\|\ge\frac{\epsilon}{2}\right\}\le\frac{\theta_1}{N_1^{\frac{p}{2}}\epsilon^p}\tag{35}$$

and

$$\begin{aligned}
&\mathrm{Prob}\left\{\sup_{\substack{x_1\in\mathcal{X}_1\\x_2\in\mathcal{X}_2}}\left\|\frac{1}{N_2}\sum_{k=1}^{N_2}\Phi_2(x_1,x_2,\hat{x}_3(x_2,\xi_{[3]}^{ik}),\xi_3^{ik})-\mathbb{E}_{\xi_3|\xi_{[2]}^i}\left[\Phi_2(x_1,x_2,\hat{x}_3(x_2,\xi_{[3]}),\xi_3)\right]\right\|\ge\frac{\epsilon}{2}\right\}\\
&\le\frac{\theta_2(\xi_{[2]}^i)}{N_2^{\frac{p}{2}}\epsilon^p}.
\end{aligned}\tag{36}$$

We further make the following assumption.

Assumption 7. *There exists a scalar $\theta_2 > 0$ such that $\theta_2(\xi_{[2]}^i) \leq \theta_2$ for all $\xi_{[2]}^i \in \Xi_{[2]}$.*

Remark 4.1. *A similar assumption can be found in [24, Assumption 2]. [24, Remark 3.3] comments it detailedly. Generally speaking, Assumption 7 imposes restrictions on how the conditional distribution of ξ_3 can be influenced by $\xi_{[2]}^i$. For example, if $\theta_2(\xi_{[2]}^i)$ depends on $\xi_{[2]}^i$ continuously and Ξ_2 is bounded, then Assumption 7 holds. Specially, when ξ_3 is independent of ξ_2 (see e.g. [30]), Assumption 7 holds automatically.*

Now we are ready to give the main result of this section.

Theorem 4.1. *Let $p \geq 2$ and $\epsilon > 0$. Suppose that Assumptions 4-7 hold, and $\mathbb{E}_{\xi_{[2]}}[\kappa(\xi_{[2]})^p] < \infty$, where $\kappa(\xi_{[2]})$ is defined in (34). Then the following assertions hold.*

(i) There exists a constant $\theta_\kappa > 0$ such that

$$\begin{aligned}&\text{Prob}\left\{\sup_{x_1\in\mathcal{X}_1}\left\|\frac{1}{N_1}\sum_{i=1}^{N_1}\Phi_1(x_1,\tilde{x}_2^{N_2}(x_1,\xi_2^i),\xi_2^i)-\mathbb{E}_{\xi_2}[\Phi_1(x_1,\hat{x}_2(x_1,\xi_{[2]}),\xi_2)]\right\|\geq\epsilon\right\}\\&\leq\frac{\theta_1}{N_1^{\frac{p}{2}}\epsilon^p}+\frac{\theta_\kappa}{N_1^{\frac{p}{2}}}+\frac{\theta_2(\bar{\kappa}+1)^pN_1}{N_2^{\frac{p}{2}}\epsilon^p},\end{aligned}$$

where 'Prob' denotes the joint probability measure induced by $\xi_{[3]}^{jk}$ for $j = 1, \cdots, N_1$ and $k = 1, \cdots, N_2$, θ_1 is defined in (35), θ_2 is defined in Assumption 7 and $\bar{\kappa} := \mathbb{E}_{\xi_{[2]}}[\kappa(\xi_{[2]})]$.

(ii) There exists a constant $L > 0$, independent of N_1, N_2, such that

$$\text{Prob}\left\{\left\|\tilde{x}_1^{N_1,N_2}-x_1^*\right\|\geq\epsilon\right\}\leq L\left(\frac{\theta_1}{N_1^{\frac{p}{2}}\epsilon^p}+\frac{\theta_\kappa}{N_1^{\frac{p}{2}}}+\frac{\theta_2(\bar{\kappa}+1)^pN_1}{N_2^{\frac{p}{2}}\epsilon^p}\right).$$

Proof. Part (i): We know from (28) and (34) that

$$\begin{aligned}&\text{Prob}\left\{\sup_{x_1\in\mathcal{X}_1}\left\|\frac{1}{N_1}\sum_{i=1}^{N_1}\Phi_1(x_1,\tilde{x}_2^{N_2}(x_1,\xi_2^i),\xi_2^i)-\mathbb{E}_{\xi_2}[\Phi_1(x_1,\hat{x}_2(x_1,\xi_{[2]}),\xi_2)]\right\|\geq\epsilon\right\}\\&\leq\text{Prob}\left\{\sup_{x_1\in\mathcal{X}_1}\left\|\frac{1}{N_1}\sum_{i=1}^{N_1}\Phi_1(x_1,\hat{x}_2(x_1,\xi_2^i),\xi_2^i)-\mathbb{E}_{\xi_2}[\Phi_1(x_1,\hat{x}_2(x_1,\xi_{[2]}),\xi_2)]\right\|\geq\frac{\epsilon}{2}\right\}\\&\quad+\text{Prob}\left\{\frac{1}{N_1}\sum_{i=1}^{N_1}\kappa(\xi_{[2]}^i)\Psi(\xi_2^i,N_2)\geq\frac{\epsilon}{2}\right\}\\&\leq\frac{\theta_1}{N_1^{\frac{p}{2}}\epsilon^p}+\text{Prob}\left\{\left(\frac{1}{N_1}\sum_{i=1}^{N_1}\kappa(\xi_{[2]}^i)\right)\cdot\max_{1\leq i\leq N_1}\Psi(\xi_2^i,N_2)\geq\frac{\epsilon}{2}\right\},\end{aligned}$$

where

$$\Psi(\xi_2^i,N_2):=\sup_{\substack{x_1\in\mathcal{X}_1\\x_2\in\mathcal{X}_2}}\left\|\frac{1}{N_2}\sum_{k=1}^{N_2}\Phi_2(x_1,x_2,\hat{x}_3(x_2,\xi_{[3]}^{ik}),\xi_3^{ik})-\mathbb{E}_{\xi_3|\xi_{[2]}^i}\left[\Phi_2(x_1,x_2,\hat{x}_3(x_2,\xi_{[3]}),\xi_3)\right]\right\|$$

and the last inequality follows from (35).

Note that

$$\begin{aligned}\text{Prob}\left\{\frac{1}{N_1}\sum_{i=1}^{N_1}\kappa(\xi_{[2]}^i)\geq\bar{\kappa}+1\right\}&=\text{Prob}\left\{\frac{1}{N_1}\sum_{i=1}^{N_1}\kappa(\xi_{[2]}^i)-\bar{\kappa}\geq 1\right\}\\&\leq\text{Prob}\left\{\left|\frac{1}{N_1}\sum_{i=1}^{N_1}\kappa(\xi_{[2]}^i)-\bar{\kappa}\right|\geq 1\right\}\\&\leq\frac{\theta_\kappa}{N_1^{\frac{p}{2}}}\end{aligned}\tag{37}$$

for some $\theta_\kappa>0$, where $\bar{\kappa}:=\mathbb{E}_{\xi_{[2]}}[\kappa(\xi_{[2]})]$ and the last inequality directly follows from Lemma 4.2.

Thus, we have

$$\begin{aligned}&\text{Prob}\left\{\left(\frac{1}{N_1}\sum_{i=1}^{N_1}\kappa(\xi_{[2]}^i)\right)\cdot\max_{1\leq i\leq N_1}\Psi(\xi_2^i,N_2)\geq\frac{\epsilon}{2}\right\}\\&=\text{Prob}\left\{\left(\frac{1}{N_1}\sum_{i=1}^{N_1}\kappa(\xi_{[2]}^i)\right)\cdot\max_{1\leq i\leq N_1}\Psi(\xi_2^i,N_2)\geq\frac{\epsilon}{2}\right\}\cdot\text{Prob}\left\{\frac{1}{N_1}\sum_{i=1}^{N_1}\kappa(\xi_{[2]}^i)\geq\bar{\kappa}+1\right\}\\&+\text{Prob}\left\{\left(\frac{1}{N_1}\sum_{i=1}^{N_1}\kappa(\xi_{[2]}^i)\right)\cdot\max_{1\leq i\leq N_1}\Psi(\xi_2^i,N_2)\geq\frac{\epsilon}{2}\right\}\cdot\text{Prob}\left\{\frac{1}{N_1}\sum_{i=1}^{N_1}\kappa(\xi_{[2]}^i)\leq\bar{\kappa}+1\right\}\\&\leq\text{Prob}\left\{\frac{1}{N_1}\sum_{i=1}^{N_1}\kappa(\xi_{[2]}^i)\geq\bar{\kappa}+1\right\}+\text{Prob}\left\{(\bar{\kappa}+1)\cdot\max_{1\leq i\leq N_1}\Psi(\xi_2^i,N_2)\geq\frac{\epsilon}{2}\right\}\\&\overset{(a)}{\leq}\frac{\theta_\kappa}{N_1^{\frac{p}{2}}}+\text{Prob}\left\{\max_{1\leq i\leq N_1}\Psi(\xi_2^i,N_2)\geq\frac{\epsilon}{2(\bar{\kappa}+1)}\right\}\\&\leq\frac{\theta_\kappa}{N_1^{\frac{p}{2}}}+\sum_{i=1}^{N_1}\text{Prob}\left\{\Psi(\xi_2^i,N_2)\geq\frac{\epsilon}{2(\bar{\kappa}+1)}\right\}\\&\overset{(b)}{\leq}\frac{\theta_\kappa}{N_1^{\frac{p}{2}}}+\frac{\theta_2(\bar{\kappa}+1)^pN_1}{N_2^{\frac{p}{2}}\epsilon^p},\end{aligned}$$

where (a) follows from (37) and (b) follows from (36) and Assumption 7.

To sum up, we obtain

$$\begin{aligned}&\text{Prob}\left\{\sup_{x_1\in\mathcal{X}_1}\left\|\frac{1}{N_1}\sum_{i=1}^{N_1}\Phi_1(x_1,\tilde{x}_2^{N_2}(x_1,\xi_2^i),\xi_2^i)-\mathbb{E}_{\xi_2}[\Phi_1(x_1,\hat{x}_2(x_1,\xi_{[2]}),\xi_2)]\right\|\geq\epsilon\right\}\\&\leq\frac{\theta_1}{N_1^{\frac{p}{2}}\epsilon^p}+\frac{\theta_\kappa}{N_1^{\frac{p}{2}}}+\frac{\theta_2(\bar{\kappa}+1)^pN_1}{N_2^{\frac{p}{2}}\epsilon^p}.\end{aligned}$$

Part (ii): Consider problems (26) and (27). By Proposition 4.1, there exists a positive constant depending only on the strong monotonicity and Lipschitz moduli of $\mathbb{E}_{\xi_2}[\Phi_1(x_1,\hat{x}_2(x_1,\xi_{[2]}),\xi_2)]$ w.r.t. x_1, denoted by $\bar{L}$, such that

$$\left\|\tilde{x}_1^{N_1,N_2}-x_1^*\right\|\leq\bar{L}\left(\sup_{x_1\in\mathcal{X}_1}\left\|\frac{1}{N_1}\sum_{i=1}^{N_1}\Phi_1(x_1,\tilde{x}_2^{N_2}(x_1,\xi_2^i),\xi_2^i)-\mathbb{E}_{\xi_2}[\Phi_1(x_1,\hat{x}_2(x_1,\xi_{[2]}),\xi_2)]\right\|\right),$$

which, in turn, implies that

$$\text{Prob}\left\{\left\|\tilde{x}_1^{N_1,N_2}-x_1^*\right\|\geq\epsilon\right\}\leq\bar{L}^p\left(\frac{\theta_1}{N_1^{\frac{p}{2}}\epsilon^p}+\frac{\theta_\kappa}{N_1^{\frac{p}{2}}}+\frac{\theta_2(\bar{\kappa}+1)^pN_1}{N_2^{\frac{p}{2}}\epsilon^p}\right).$$

By letting $L := \bar{L}^p$, we complete the proof. ■

Remark 4.2. *In Theorem 4.1, we give the polynomial convergence rate under relatively mild conditions (see Assumption 6). By a similar procedure, we can also provide the exponential convergence rate as that in [27, Theorem 7.73], where some so-called light-tailed assumptions are required. Moreover, for sufficiently small* $\epsilon > 0$*, we have*

$$\frac{\theta_1}{N_1^{\frac{p}{2}}\epsilon^p} + \frac{\theta_\kappa}{N_1^{\frac{p}{2}}} + \frac{\theta_2(\bar{\kappa}+1)^p N_1}{N_2^{\frac{p}{2}}\epsilon^p} \approx \frac{\theta_1}{N_1^{\frac{p}{2}}\epsilon^p} + \frac{\theta_2(\bar{\kappa}+1)^p N_1}{N_2^{\frac{p}{2}}\epsilon^p}.$$

5 Conclusions

In this paper, we have studied MSVIs. We first give two source problems arising from MSPs and multistage multi-player noncooperative game problems, respectively. Next, the (strong) monotonicity properties have been investigated. Compared with the existing results, we give more general results. Finally, we have analysed the complexity of MSVIs. The polynomial rate of convergence has been derived.

Acknowledgements This work is supported by China Postdoctoral Science Foundation (Grant No. 2020M673117) and National Natural Science Foundation of China (Grant Nos. 11871276 and 12122108).

References

[1] Francisco Facchinei and Jong-Shi Pang. *Finite-dimensional Variational Inequalities and Complementarity Problems*. Springer, New York, 2007.

[2] Richard W. Cottle, Jong-Shi Pang, and Richard E. Stone. *The Linear Complementarity Problem*. SIAM, Philadelphia, 2009.

[3] Gül Gürkan, A Yonca Özge, and Stephen M Robinson. Sample-path solution of stochastic variational inequalities. *Mathematical Programming*, 84(2):313–333, 1999.

[4] Huifu Xu. Sample average approximation methods for a class of stochastic variational inequality problems. *Asia-Pacific Journal of Operational Research*, 27(01):103–119, 2010.

[5] Xiaojun Chen and Masao Fukushima. Expected residual minimization method for stochastic linear complementarity problems. *Mathematics of Operations Research*, 30(4):1022–1038, 2005.

[6] Xiaojun Chen, Roger J-B Wets, and Yanfang Zhang. Stochastic variational inequalities: residual minimization smoothing sample average approximations. *SIAM Journal on Optimization*, 22(2):649–673, 2012.

[7] Uday V Shanbhag. Stochastic variational inequality problems: Applications, analysis, and algorithms. In *INFORMS TutORials in Operations Research*, pages 71–107. INFORMS, 2013.

[8] R Tyrrell Rockafellar and Roger JB Wets. Stochastic variational inequalities: single-stage to multistage. *Mathematical Programming*, 165(1):331–360, 2017.

[9] Xiaojun Chen, Ting Kei Pong, and Roger JB Wets. Two-stage stochastic variational inequalities: an ERM-solution procedure. *Mathematical Programming*, 165(1):71–111, 2017.

[10] Xiaojun Chen, Hailin Sun, and Huifu Xu. Discrete approximation of two-stage stochastic and distributionally robust linear complementarity problems. *Mathematical Programming*, 177(1-2):255–289, 2019.

[11] Xiaojun Chen, Alexander Shapiro, and Hailin Sun. Convergence analysis of sample average approximation of two-stage stochastic generalized equations. *SIAM Journal on Optimization*, 29(1):135–161, 2019.

[12] Jie Jiang and Shengjie Li. Regularized sample average approximation approach for two-stage stochastic variational inequalities. *Journal of Optimization Theory and Applications*, 190(2):650–671, 2021.

[13] Jie Jiang, Xiaojun Chen, and Zhiping Chen. Quantitative analysis for a class of two-stage stochastic linear variational inequality problems. *Computational Optimization and Applications*, 76(2):431–460, 2020.

[14] Jianxun Liu, Shengjie Li, and Jie Jiang. Quantitative stability of two-stage stochastic linear variational inequality problems with fixed recourse. *Applicable Analysis*, pages 1–17, 2020.

[15] Xiaozhou Wang and Xiaojun Chen. Solving two-stage stochastic variational inequalities by a projection semismooth Newton algorithm. *Working paper*, 2022.

[16] Jie Jiang, Yun Shi, Xiaozhou Wang, and Xiaojun Chen. Regularized two-stage stochastic variational inequalities for Cournot-Nash equilibrium under uncertainty. *Journal of Computational Mathematics*, 37(6):813–842, 2019.

[17] Min Zhang, Jie Sun, and Honglei Xu. Two-stage quadratic games under uncertainty and their solution by progressive hedging algorithms. *SIAM Journal on Optimization*, 29(3):1799–1818, 2019.

[18] Min Li and Chao Zhang. Two-stage stochastic variational inequality arising from stochastic programming. *Journal of Optimization Theory and Applications*, 186(1):324–343, 2020.

[19] Min Li, Chao Zhang, Mingxv Ding, and Ruipu Lv. A two-stage stochastic variational inequality model for storage and dynamic distribution of medical supplies in epidemic management. *Applied Mathematical Modelling*, 102:35–61, 2022.

[20] Hailin Sun and Xiaojun Chen. Two-stage stochastic variational inequalities: theory, algorithms and applications. *Journal of the Operations Research Society of China*, 9(1):1–32, 2021.

[21] R Tyrrell Rockafellar and Jie Sun. Solving monotone stochastic variational inequalities and complementarity problems by progressive hedging. *Mathematical Programming*, 174(1-2):453–471, 2019.

[22] Xingbang Cui, Jie Sun, and Liping Zhang. Solvability of multistage pseudomonotone stochastic variational inequalities. *Available on arXiv:2201.01454*, 2022.

[23] Jie Jiang, Hailin Sun, and Bin Zhou. Convergence analysis of sample average approximation for a class of stochastic nonlinear complementarity problems: from two-stage to multistage. *Numerical Algorithms*, 89(1):167–194, 2022.

[24] Jie Jiang and Shengjie Li. On complexity of multistage stochastic programs under heavy tailed distributions. *Operations Research Letters*, 49(2):265–269, 2021.

[25] Jerzy Kyparisis. Solution differentiability for variational inequalities. *Mathematical Programming*, 48(1-3):285–301, 1990.

[26] R Tyrrell Rockafellar and Roger J-B Wets. *Variational Analysis*. Springer, New York, 2009.

[27] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, Philadelphia, 2ed edition, 2014.

[28] J Frédéric Bonnans and Alexander Shapiro. *Perturbation Analysis of Optimization Problems*. Springer, New York, 2013.

[29] Khanh Duy Pham and Nhut Minh Bui. Error bounds for strongly monotone and lipschitz continuous variational inequalities. *Optimization Letters*, 12(5):971–984, 2018.

[30] Alexander Shapiro. On complexity of multistage stochastic programs. *Operations Research Letters*, 34(1):1–8, 2006.