CONVERGENCE ANALYSIS OF SAMPLE AVERAGE APPROXIMATION OF TWO-STAGE STOCHASTIC GENERALIZED EQUATIONS*

XIAOJUN CHEN†, ALEXANDER SHAPIRO‡, AND HAILIN SUN§

Abstract. A solution of two-stage stochastic generalized equations is a pair: a first stage solution which is independent of realization of the random data and a second stage solution which is a function of random variables. This paper studies convergence of the sample average approximation of two-stage stochastic nonlinear generalized equations. In particular an exponential rate of the convergence is shown by using the perturbed partial linearization of functions. Moreover, sufficient conditions for the existence, uniqueness, continuity and regularity of solutions of two-stage stochastic generalized equations are presented under an assumption of monotonicity of the involved functions. These theoretical results are given without assuming relatively complete recourse, and are illustrated by two-stage stochastic non-cooperative games of two players.

Key words. Two-stage stochastic generalized equations, sample average approximation, convergence, exponential rate, monotone multifunctions

AMS subject classifications. 90C15, 90C33

1. Introduction.

Consider the following two-stage Stochastic Generalized Equations (SGE)

$$0 \in \mathbb{E}[\Phi(x, y(\xi), \xi)] + \Gamma_1(x), \ x \in X, \tag{1.1}$$

$$0 \in \Psi(x, y(\xi), \xi) + \Gamma_2(y(\xi), \xi), \quad \text{for a.e. } \xi \in \Xi. \tag{1.2}$$

Here $X \subseteq \mathbb{R}^n$ is a nonempty closed convex set, $\xi : \Omega \to \mathbb{R}^d$ is a random vector defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, whose probability distribution $P = \mathbb{P} \circ \xi^{-1}$ is supported on set $\Xi := \xi(\Omega) \subseteq \mathbb{R}^d$, $\Phi : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}^n$ and $\Psi : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}^m$, and $\Gamma_1 : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, $\Gamma_2 : \mathbb{R}^m \times \Xi \rightrightarrows \mathbb{R}^m$ are multifunctions (point-to-set mappings). We assume throughout the paper that $\Phi(\cdot, \cdot, \xi)$ and $\Psi(\cdot, \cdot, \xi)$ are *Lipschitz continuous* with Lipschitz modulus $\kappa_\Phi(\xi)$ and $\kappa_\Psi(\xi)$, respectively, and $y(\cdot) \in \mathcal{Y}$ with $\mathcal{Y}$ being the space of measurable functions from Ξ to $\mathbb{R}^m$ such that the expected value in (1.1) is well defined.

Solutions of (1.1)–(1.2) are searched over $x \in X$ and $y(\cdot) \in \mathcal{Y}$ to satisfy the corresponding inclusions, where the second stage inclusion (1.2) should hold for almost every (a.e.) realization of ξ. The first stage decision x is made before observing realization of the random data vector ξ and the second stage decision $y(\xi)$ is a function of ξ.

When the multifunctions Γ_1 and Γ_2 have the following form

$$\Gamma_1(x) := \mathcal{N}_C(x) \quad \text{and} \quad \Gamma_2(y, \xi) := \mathcal{N}_{K(\xi)}(y),$$

*Submitted to the editors DATE.

†Department of Applied Mathematics, The Hong Kong Polytechnic University, (xiaojun.chen@polyu.edu.hk). Research of this author was partly supported by Hong Kong Research Grant Council PolyU153016/16p.

‡School of Industrial and Systems Engineering, Georgia Institute of Technology, (alex.shapiro@isye.gatech.edu). Research of this author was partly supported by NSF grant 1633196 and DARPA EQUiPS program, grant SNL 014150709.

§School of Economics and Management, Nanjing University of Science and Technology; Department of Applied Mathematics, The Hong Kong Polytechnic University, (hlsun@njust.edu.cn). Research of this author was partly supported by National Natural Science Foundation of China 11401308, 11571178.

where $\mathcal{N}_C(x)$ is the normal cone to a nonempty closed convex set $C \subseteq \mathbb{R}^n$ at x, and similarly for $\mathcal{N}_{K(\xi)}(y)$, the SGE (1.1)–(1.2) reduce to the two-stage Stochastic Variational Inequalities (SVI) as in [2, 25]. The two-stage SVI represents first order optimality conditions for the two-stage stochastic optimization problem [1, 27] and models several equilibrium problems in stochastic environment [2, 4]. Moreover, if the sets C and $K(\xi)$, $\xi \in \Xi$, are closed convex *cones*, then

$$\mathcal{N}_C(x) = \{x^* \in C^* : x^\top x^* = 0\},\ x \in C,$$

where $C^* = \{x^* : x^\top x^* \le 0,\ \forall x \in C\}$ is the (negative) dual of cone C. In that case the SGE (1.1)–(1.2) reduce to the following two-stage stochastic cone VI

$$\begin{aligned} &C \ni x \perp \mathbb{E}[\Phi(x, y(\xi), \xi)] \in -C^*,\ x \in X, \\ &K(\xi) \ni y(\xi) \perp \Psi(x, y(\xi), \xi) \in -K^*(\xi), \quad \text{for a.e. } \xi \in \Xi. \end{aligned}$$

In particular when $C := \mathbb{R}^n_+$ with $C^* = -\mathbb{R}^n_+$, and $K(\xi) := \mathbb{R}^m_+$ with $K^*(\xi) = -\mathbb{R}^m_+$ for all $\xi \in \Xi$, the SGE (1.1)–(1.2) reduce to the two-stage Stochastic Nonlinear Complementarity Problem (SNCP):

$$\begin{aligned} &0 \le x \perp \mathbb{E}[\Phi(x, y(\xi), \xi)] \ge 0, \\ &0 \le y(\xi) \perp \Psi(x, y(\xi), \xi) \ge 0, \quad \text{for a.e. } \xi \in \Xi, \end{aligned}$$

which is a generalization of the two-stage Stochastic Linear Complementarity Problem (SLCP):

$$0 \le x \perp Ax + \mathbb{E}[B(\xi)y(\xi)] + q_1 \ge 0, \tag{1.3}$$

$$0 \le y(\xi) \perp L(\xi)x + M(\xi)y(\xi) + q_2(\xi) \ge 0, \quad \text{for a.e. } \xi \in \Xi, \tag{1.4}$$

where $A \in \mathbb{R}^{n\times n}$, $B : \Xi \to \mathbb{R}^{n\times m}$, $L : \Xi \to \mathbb{R}^{m\times n}$, $M : \Xi \to \mathbb{R}^{m\times m}$, $q_1 \in \mathbb{R}^n$, $q_2 : \Xi \to \mathbb{R}^m$. The two-stage SLCP arises from the KKT condition for the two-stage stochastic linear programming [2]. Existence of solutions of (1.3)-(1.4) has been studied in [3]. Moreover, the progressive hedging method has been applied to solve (1.3)-(1.4), with a finite number of realizations of ξ, in [23].

Most existing studies for two-stage stochastic problems assume *relatively complete recourse*, that is, for every $x \in X$ and a.e. $\xi \in \Xi$ the second stage problem has at least one solution. However, for the SGE (1.1)–(1.2), it could happen that for a certain first stage decision $x \in X$, the second stage generalized equation

$$0 \in \Psi(x, y, \xi) + \Gamma_2(y, \xi) \tag{1.5}$$

does not have a solution for some $\xi \in \Xi$. For such x and ξ the second stage decision $y(\xi)$ is not defined. If this happens for ξ with positive probability, then the expected value of the first stage problem is not defined and such x should be avoided. In practice, relatively complete recourse condition may not hold in many real world applications. For example, when considering to make a decision on building a power station for providing electrical power to satisfy the demand, it could be practically impossible to make sure that the uncertain demand will be satisfied under *any* possible circumstances.

In this paper, without assuming *relatively complete recourse*, we study convergence of the Sample Average Approximation (SAA)

$$0 \in N^{-1} \sum_{j=1}^{N} \Phi(x, y_j, \xi^j) + \Gamma_1(x),\ x \in X, \tag{1.6}$$

$$0 \in \Psi(x, y_j, \xi^j) + \Gamma_2(y_j, \xi^j), \quad j = 1, ..., N, \tag{1.7}$$

of the two-stage SGE (1.1)–(1.2) with y_j being a copy of the second stage vector for $\xi = \xi^j$, $j = 1, ..., N$, where $\xi^1, ..., \xi^N$ is an independent identically distributed (iid) sample of random vector ξ. Note that (1.1)-(1.2) is a two-stage extension of one-stage SGE. The convergence analysis and exponential rate of convergence of one-stage SGE has been investigated in a number of publications (e.g., [19, 27, 30] and references there in). We extend those convergence analysis results from one-stage SGE to two-stage SGE in a significant way. Our SAA method for the two-stage SGE (1.1)-(1.2) is different from the discretization scheme for the two-stage SLCP in [3]. The main difference is that the discretization scheme in [3] uses the partition of the support set Ξ and the conditional expectations of random functions, but our SAA method does not.

The paper is organized as follows. In section 2 we investigate almost sure and exponential rate of convergence of solutions of the SAA of the two-stage SGE. In section 3 convergence analysis of the mixed two-stage SVI-NCP is presented. In particular we give sufficient conditions for the existence, uniqueness, continuity and regularity of solutions by using the perturbed linearization of functions Φ and Ψ. Theoretical results, given in sections 2 and 3, are illustrated by numerical examples, using the Progressive Hedging Method (PHM), in section 4. It is worth noting that PHM is well-defined for two-stage monotone SVI without relatively complete recourse. Finally section 5 is devoted to conclusion remarks.

We use the following notation and terminology throughout the paper. Unless stated otherwise $\|x\|$ denotes the Euclidean norm of vector $x \in \mathbb{R}^n$. By $\mathcal{B} := \{x : \|x\| \le 1\}$ we denote unit ball in a considered vector space. For two sets $A, B \subset \mathbb{R}^m$ we denote by $d(x, B) := \inf_{y \in B} \|x - y\|$ distance from a point $x \in \mathbb{R}^m$ to the set B, $d(x, B) = +\infty$ if B is empty, by $\mathbb{D}(A, B) := \sup_{x \in A} d(x, B)$ the deviation of set A from the set B, and $\mathbb{H}(A, B) := \max\{\mathbb{D}(A, B), \mathbb{D}(B, A)\}$. The indicator function of a set A is defined as $I_A(x) = 0$ for $x \in A$ and $I_A(x) = +\infty$ for $x \notin A$. By $\text{bd}(A)$, int(A) and $\text{cl}(A)$ we denote the boundary, interior and topological closure of a set $A \subset \mathbb{R}^m$. By $\text{reint}(A)$ we denote the relative interior of a convex set $A \subset \mathbb{R}^m$. A multifunction (point-to-set mappings) $\Gamma : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ assigns to a point $x \in \mathbb{R}^n$ to a set $\Gamma(x) \subset \mathbb{R}^m$. A multifunction $\Gamma : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is said to be *closed* if $x_k \to x$, $x_k^* \in \Gamma(x_k)$ and $x_k^* \to x^*$, then $x^* \in \Gamma(x)$. It is said that a multifunction $\Gamma : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is *monotone*, if $(x - x')^\top (y - y') \ge 0$, for all $x, x' \in \mathbb{R}^n$, and $y \in \Gamma(x)$, $y' \in \Gamma(x')$. Note that for a nonempty closed convex set C, the normal cone multifunction $\Gamma(x) := \mathcal{N}_C(x)$ is closed and monotone. Note also that the normal cone $\mathcal{N}_C(x)$, at $x \in C$, is the (negative) dual of the tangent cone $\mathcal{T}_C(x)$. We use the same notation for ξ considered as a random vector and as a variable $\xi \in \mathbb{R}^d$. Which of these two meanings is used will be clear from the context. For vector $d \in \mathbb{R}^n$, d_J is a subvector of d whose entries are in the index $J \subseteq \{1, \cdots, n\}$. Similarly, for matrix $D \in \mathbb{R}^{n \times m}$, $D_{J_1 J_2}$ is a submatrix of D whose entries are in the index $J_1 \times J_2 \subseteq \{1, \cdots, n\} \times \{1, \cdots, m\}$.

2. Sample average approximation of the two-stage SGE.

In this section we discuss statistical properties of the first stage solution $\hat{x}_N$ of the SAA problem (1.6)–(1.7). In particular we investigate conditions ensuring convergence of $\hat{x}_N$, with probability one (w.p.1) and exponential, to its counterpart of the true problem (1.1)–(1.2).

Denote by $\mathcal{X}$ the set of $x \in X$ such that the second stage generalized equation (1.5) has a solution for a.e. $\xi \in \Xi$. The condition of relatively complete recourse means that $\mathcal{X} = X$. Note that $\mathcal{X}$ is a subset of X, and if $(\bar{x}, \bar{y}(\cdot))$ is a solution of (1.1)–(1.2), then $\bar{x} \in \mathcal{X}$. It is possible to formulate the two-stage SGE (1.1)–(1.2) in the following equivalent way. Let $\hat{y}(x,\xi)$ be a solution function of the second stage problem (1.5) for $x \in \mathcal{X}$ and $\xi \in \Xi$, i.e.,

$$0 \in \Psi(x, \hat{y}(x,\xi), \xi) + \Gamma_2(\hat{y}(x,\xi), \xi), \;\; x \in \mathcal{X}, \text{ a.e. } \xi \in \Xi.$$

Then the first stage problem becomes

$$0 \in \mathbb{E}[\Phi(x, \hat{y}(x,\xi), \xi)] + \Gamma_1(x), \; x \in \mathcal{X}. \tag{2.1}$$

If $\bar{x}$ is a solution of (2.1), then $(\bar{x}, \hat{y}(\bar{x}, \cdot))$ is a solution of (1.1)–(1.2). Conversely if $(\bar{x}, \bar{y}(\cdot))$ is a solution of (1.1)–(1.2), then $\bar{x}$ is a solution of (2.1). Note that problem (2.1) is a generalized equation which has been studied in the past decades, e.g. [19, 22, 24, 26].

It could happen that the second stage problem (1.5) has more than one solution for some $x \in \mathcal{X}$. In that case choice of $\hat{y}(x,\xi)$ is somewhat arbitrary and the corresponding SGE are not well posed. This motivates the following condition.

Assumption 2.1. *For a.e. $\xi \in \Xi$, problem (1.5) has a* unique *solution for all $x \in \mathcal{X}$.*

Under Assumption 2.1 the value $\hat{y}(x,\xi)$ is uniquely defined for all $x \in \mathcal{X}$ and a.e. $\xi \in \Xi$, and the first stage problem (2.1) can be written as the following generalized equation

$$0 \in \phi(x) + \Gamma_1(x), \; x \in \mathcal{X}, \tag{2.2}$$

where

$$\phi(x) := \mathbb{E}[\hat{\Phi}(x,\xi)] \text{ and } \hat{\Phi}(x,\xi) := \Phi(x, \hat{y}(x,\xi), \xi). \tag{2.3}$$

If the SGE have relatively complete recourse, then under Assumption 2.1 the SAA problem (1.6)–(1.7) can be written as

$$0 \in \hat{\phi}_N(x) + \Gamma_1(x), \; x \in X, \tag{2.4}$$

where $\hat{\phi}_N(x) := N^{-1} \sum_{j=1}^N \hat{\Phi}(x, \xi^j)$ with $\hat{\Phi}(x,\xi)$ defined in (2.3). Problem (2.4) can be viewed as the SAA of the first stage problem (2.2). If $(\hat{x}_N, \hat{y}_{jN})$ is a solution of the SAA problem (1.6)–(1.7), then $\hat{x}_N$ is a solution of (2.4) and $\hat{y}_{jN} = \hat{y}(\hat{x}_N, \xi^j)$, $j = 1, ..., N$. Note that the SAA problem (1.6)–(1.7) can be considered without assuming the relatively complete recourse, although in that case it could happen that $\hat{\phi}_N(x)$ is not defined for some $x \in X \setminus \mathcal{X}$ and solution $\hat{x}_N$ of (1.6) is not implementable at the second stage for some realizations of the random vector ξ. Our aim is the convergence analysis of the SAA problem (1.6)–(1.7) when sample size N increases.

Denote by $\mathcal{S}^*$ the set of solutions of the first stage problem (2.2) and by $\hat{\mathcal{S}}_N$ the set of solutions of the SAA problem (1.6) (in case of relatively complete recourse, $\hat{\mathcal{S}}_N$ is the set of solutions of problem (2.4) as well). By $\bar{\mathcal{X}}(\xi)$ we denote the set of $x \in X$ such that problem (1.5) has a solution, and by $\bar{\mathcal{X}}_N := \cap_{j=1}^N \bar{\mathcal{X}}(\xi^j)$ the set of x such that problems (1.7) have a solution. Note that the set $\mathcal{X}$ is equal to the intersection of $\bar{\mathcal{X}}(\xi)$, a.e. $\xi \in \Xi$; i.e., $\mathcal{X} = \cap_{\xi \in \Xi \setminus \Upsilon} \bar{\mathcal{X}}(\xi)$ for some set $\Upsilon \subset \Xi$ such that $P(\Upsilon) = 0$. Note also that if the two-stage SGE have relatively complete recourse, then $\bar{\mathcal{X}}(\xi) = X$ for a.e. $\xi \in \Xi$.

2.1. Almost sure convergence. Consider the solution $\hat{y}(x,\xi)$ of the second stage problem (1.5). To ensure continuity of $\hat{y}(x,\xi)$ in $x \in \mathcal{X}$ for $\xi \in \Xi$, in addition to Assumption 2.1, we need the following boundedness condition.

ASSUMPTION 2.2. *For every ξ and every $x \in \bar{\mathcal{X}}(\xi)$ there is a neighborhood $\mathcal{V}$ of x and a measurable function $v(\xi)$ such that $\|\hat{y}(x',\xi)\| \le v(\xi)$ for all $x' \in \mathcal{V} \cap \bar{\mathcal{X}}(\xi)$.*

Note that function $v(\xi)$ depends on point x and its neighborhood $\mathcal{V}$. We suppress this in the notation of $v(\xi)$.

LEMMA 2.1. *Suppose that Assumptions* 2.1 and 2.2 *hold, and for a.e. $\xi \in \Xi$ the multifunction $\Gamma_2(\cdot,\xi)$ is closed. Then for a.e. $\xi \in \Xi$ the solution $\hat{y}(x,\xi)$ is a continuous function of $x \in \mathcal{X}$.*

Proof. The proof is quite standard. We argue by a contradiction. Suppose that for some $\bar{x} \in \mathcal{X}$ and $\xi \in \Xi$ the solution $\hat{y}(\cdot,\xi)$ is not continuous at $\bar{x}$. That is, there is a sequence $x_k \in \mathcal{X}$ converging to $\bar{x} \in \mathcal{X}$ such that $y_k := \hat{y}(x_k,\xi)$ does not converge to $\bar{y} := \hat{y}(\bar{x},\xi)$. Then by the boundedness assumption, by passing to a subsequence if necessary we can assume that y_k converges to a point y^* different from $\bar{y}$. Consequently $0 \in \Psi(x_k,y_k,\xi) + \Gamma_2(y_k,\xi)$ and $\Psi(x_k,y_k,\xi)$ converges to $\Psi(\bar{x},y^*,\xi)$. Since $\Gamma_2(\cdot,\xi)$ is closed, it follows that $0 \in \Psi(\bar{x},y^*,\xi) + \Gamma_2(y^*,\xi)$. Hence by the uniqueness assumption, $y^* = \bar{y}$ which gives the required contradiction. ∎

Suppose for the moment that in addition to the assumptions of Lemma 2.1, the SGE have relatively complete recourse. We can apply then general results to verify consistency of the SAA estimates. Consider function $\hat{\Phi}(x,\xi)$ defined in (2.3). By continuity of $\Phi(\cdot,\cdot,\xi)$ and $\hat{y}(\cdot,\xi)$, we have that $\hat{\Phi}(\cdot,\xi)$ is continuous on X. Assuming further that there is a compact set $X' \subseteq X$ such that $\mathcal{S}^* \subseteq X'$ and $\|\hat{\Phi}(x,\xi)\|_{x\in X'}$ is dominated by an integrable function, we have that the function $\phi(x) = \mathbb{E}[\hat{\Phi}(x,\xi)]$ is continuous on X' and $\hat{\phi}_N(x)$ converges w.p.1 to $\phi(x)$ uniformly on X'. Note that the boundedness condition of Lemma 2.1 and continuity of $\Phi(\cdot,\cdot,\xi)$ imply that $\hat{\Phi}(\cdot,\xi)$ is bounded on X'. Then consistency of SAA solutions follows by [27, Theorem 5.12]. We give below a more general result without the assumption of relatively complete recourse.

LEMMA 2.2. *Suppose that Assumptions* 2.1 and 2.2 *hold. Then for a.e. $\xi \in \Xi$ the set $\bar{\mathcal{X}}(\xi)$ is closed.*

Proof. For a given $\xi \in \Xi$ let $x_k \in \bar{\mathcal{X}}(\xi)$ be a sequence converging to a point $\bar{x}$. We need to show that $\bar{x} \in \bar{\mathcal{X}}(\xi)$. Let y_k be the solution of (1.5) for $x = x_k$ and ξ. Then by Assumption 2.2, there is a neighborhood $\mathcal{V}$ of $\bar{x}$ and a measurable function $v(\xi)$ such that $\|y_k\| \le v(\xi)$ when $x_k \in \mathcal{V}$. Hence by passing to a subsequence we can assume that y_k converges to a point $\bar{y} \in \mathbb{R}^m$. Since $\Psi(\cdot,\cdot,\xi)$ is continuous and $\Gamma_2(\cdot,\xi)$ is closed it follows that $\bar{y}$ is a solution of (1.5) for $x = \bar{x}$, and hence $\bar{x} \in \bar{\mathcal{X}}(\xi)$. ∎

By saying that a property holds w.p.1 for N large enough we mean that there is a set $\Sigma \subset \Omega$ of $\mathbb{P}$-measure zero such that for every $\omega \in \Omega \setminus \Sigma$ there exists a positive integer $N^* = N^*(\omega)$ such that the property holds for all $N \ge N^*(\omega)$ and $\omega \in \Omega \setminus \Sigma$.

For $\delta \in (0,1)$ consider a compact set $\bar{\Xi}_\delta \subset \Xi$ such that $\mathbb{P}(\bar{\Xi}_\delta) \ge 1-\delta$, and the multifunction $\Delta_\delta : X \rightrightarrows \bar{\Xi}_\delta$ defined as

$$\Delta_\delta(x) := \{\xi \in \bar{\Xi}_\delta : x \in \bar{\mathcal{X}}(\xi)\}. \tag{2.5}$$

ASSUMPTION 2.3. *For any $\delta \in (0,1)$ the multifunction $\Delta_\delta(\cdot)$ is outer semicontinuous.*

The following lemma shows that this assumption holds under mild conditions. Note that since the set $\bar{\Xi}_\delta$ is compact, the multifunction $\Delta_\delta(\cdot)$ is outer semicontinuous iff it is closed (cf., [24, Chapter 5(B)]).

LEMMA 2.3. *Suppose $\Psi(\cdot,\cdot,\cdot)$ is continuous, $\Gamma_2(\cdot,\cdot)$ is closed and Assumption 2.2 holds. Then the multifunction $\Delta_\delta(\cdot)$ is outer semicontinuous.*

Proof. Consider the second stage generalized equation (1.2) and any sequence $\{(x_k, y_k, \xi_k)\}$ such that $x_k \in X$, $\xi_k \in \Delta_\delta(x_k)$ with a corresponding second stage solution y_k and $(x_k, \xi_k) \to (x^*, \xi^*) \in X \times \Xi$. Since Ψ is continuous w.r.t. (x, y, ξ) and $\Gamma_2(\cdot,\cdot)$ is closed, we have that under Assumption 2.2, $\{y_k\}$ has accumulation points and any accumulation point y^* satisfies

$$0 \in \Psi(x^*, y^*, \xi^*) + \Gamma_2(y^*, \xi^*),$$

which implies $\xi^* \in \Delta_\delta(x^*)$. This shows that the multifunction $\Delta_\delta(\cdot)$ is closed. Since $\bar{\Xi}_\delta$ is compact, the closeness of $\Delta_\delta(\cdot)$ implies the outer semicontinuity of $\Delta_\delta(\cdot)$. ◻

Note that in the case when Ξ is compact, we can set $\delta = 0$ and replace $\bar{\Xi}_\delta$ by Ξ.

THEOREM 2.4. *Suppose that:* (i) *Assumptions 2.1-2.3 hold,* (ii) *the multifunctions $\Gamma_1(\cdot)$ and $\Gamma_2(\cdot,\xi)$, $\xi \in \Xi$, are closed,* (iii) *there is a compact subset X' of X such that $\mathcal{S}^* \subset X'$ and w.p.1 for all N large enough the set $\hat{\mathcal{S}}_N$ is nonempty and is contained in X',* (iv) *$\|\hat{\Phi}(x,\xi)\|_{x\in\mathcal{X}}$ is dominated by an integrable function,* (v) *the set $\mathcal{X}$ is nonempty. Let $\mathfrak{d}_N := \mathbb{D}\big(\bar{\mathcal{X}}_N \cap X', \mathcal{X} \cap X'\big)$. Then $\mathcal{S}^*$ is nonempty and the following statements hold.*

(a) *$\mathfrak{d}_N \to 0$ and $\mathbb{D}(\hat{\mathcal{S}}_N, \mathcal{S}^*) \to 0$ w.p.1 as $N \to \infty$.*

(b) *In addition assume that:* (vi) *for any $\delta > 0$, $\tau > 0$ and a.e. $\omega \in \Omega$, there exist $\gamma > 0$ and $N_0 = N_0(\omega)$ such that for any $x \in \mathcal{X} \cap X' + \gamma\,\mathcal{B}$ and $N \geq N_0$, there exists $z_x \in \mathcal{X} \cap X'$ such that*[1]

$$\|z_x - x\| \leq \tau, \;\; \Gamma_1(x) \subseteq \Gamma_1(z_x) + \delta\mathcal{B}, \;\text{ and }\; \|\hat{\phi}_N(z_x) - \hat{\phi}_N(x)\| \leq \delta. \tag{2.6}$$

Then w.p.1 for N large enough it follows that

$$\mathbb{D}(\hat{\mathcal{S}}_N, \mathcal{S}^*) \leq \tau + \mathcal{R}^{-1}\left(\sup_{x\in\mathcal{X}\cap X'} \|\phi(x) - \hat{\phi}_N(x)\|\right), \tag{2.7}$$

where for $\varepsilon \geq 0$ and $t \geq 0$,

$$\mathcal{R}(\varepsilon) := \inf_{x\in\mathcal{X}\cap X',\, d(x,\mathcal{S}^*)\geq\varepsilon} d\big(0, \phi(x) + \Gamma_1(x)\big),$$

$$\mathcal{R}^{-1}(t) := \inf\{\varepsilon \in \mathbb{R}_+ : \mathcal{R}(\varepsilon) \geq t\}.$$

Proof. Part (a). Let $\xi^j = \xi^j(\omega)$, $j = 1, \ldots$, be the iid sample, defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and $\bar{\mathcal{X}}_N = \bar{\mathcal{X}}_N(\omega)$ be the corresponding feasibility set of the SAA problem. Consider a point $\bar{x} \in X' \setminus \mathcal{X}$ and its neighborhood $\mathcal{V}_{\bar{x}} = \bar{x} + \gamma\mathcal{B}$ for some $\gamma > 0$. We have that probability $p := \mathbb{P}\{\xi \in \Xi : \bar{x} \notin \bar{\mathcal{X}}(\xi)\}$ is positive. Moreover it follows by Assumption 2.3 that we can choose $\gamma > 0$ such that probability $\mathbb{P}\left\{\mathcal{V}_{\bar{x}} \cap \bar{\mathcal{X}}(\xi) = \emptyset\right\}$ is positive. Indeed, for $\delta := p/4$ consider the multifunction Δ_δ

[1]Recall that $\hat{\phi}_N(x) = \hat{\phi}_N(x, \omega)$ are random functions defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

defined in (2.5). By outer semicontinuity of Δ_δ we have that for any $\varepsilon > 0$ there is $\gamma > 0$ such that for all $x \in \mathcal{V}_{\bar{x}}$ it follows that $\Delta_\delta(x) \subset \Delta_\delta(\bar{x}) + \varepsilon\mathcal{B}$. That is

$$\cup_{x\in\mathcal{V}_{\bar{x}}}\{\xi \in \bar{\Xi}_\delta : x \in \bar{\mathcal{X}}(\xi)\} \subset \{\xi \in \bar{\Xi}_\delta : \bar{x} \in \bar{\mathcal{X}}(\xi)\} + \varepsilon\mathcal{B} \subset \{\xi \in \Xi : \bar{x} \in \bar{\mathcal{X}}(\xi)\} + \varepsilon\mathcal{B}.$$

It follows that we can choose $\varepsilon > 0$ small enough such that

$$\mathbb{P}\big(\cup_{x\in\mathcal{V}_{\bar{x}}} \{\xi \in \bar{\Xi}_\delta : x \in \bar{\mathcal{X}}(\xi)\}\big) \le 1 - p/2.$$

Since $\delta = p/4$ we obtain

$$\mathbb{P}\big(\cup_{x\in\mathcal{V}_{\bar{x}}} \{\xi \in \Xi : x \in \bar{\mathcal{X}}(\xi)\}\big) \le 1 - p/4.$$

Noting that the event $\big\{\mathcal{V}_{\bar{x}} \cap \bar{\mathcal{X}}(\xi) = \emptyset\big\}$ is complement of the event $\big\{ \cup_{x\in\mathcal{V}_{\bar{x}}} \{\xi \in \Xi : x \in \bar{\mathcal{X}}(\xi)\}\big\}$, we obtain that $\mathbb{P}\big\{\mathcal{V}_{\bar{x}} \cap \bar{\mathcal{X}}(\xi) = \emptyset\big\} \ge p/4$.

Consider the event $E_N := \big\{\mathcal{V}_{\bar{x}} \cap \bar{\mathcal{X}}_N \ne \emptyset\big\}$. The complement of this event is $E_N^c = \big\{\mathcal{V}_{\bar{x}} \cap \bar{\mathcal{X}}_N = \emptyset\big\}$. Since the sample ξ^j, $j = 1, ...,$ is iid, we have

$$\begin{aligned} \mathbb{P}\big\{\mathcal{V}_{\bar{x}} \cap \bar{\mathcal{X}}_N \ne \emptyset\big\} &\le \textstyle\prod_{j=1}^N \mathbb{P}\big\{\mathcal{V}_{\bar{x}} \cap \bar{\mathcal{X}}(\xi^j) \ne \emptyset\big\} \\ &= \textstyle\prod_{j=1}^N \big(1 - \mathbb{P}\big\{\mathcal{V}_{\bar{x}} \cap \bar{\mathcal{X}}(\xi^j) = \emptyset\big\}\big) \le (1 - p/4)^N, \end{aligned}$$

and hence $\sum_{N=1}^\infty \mathbb{P}\big\{\mathcal{V}_{\bar{x}} \cap \bar{\mathcal{X}}_N \ne \emptyset\big\} < \infty$. It follows by Borel-Cantelli Lemma that $\mathbb{P}\left(\limsup_{N\to\infty} E_N\right) = 0$. That is for all N large enough the events E_N^c happen w.p.1. Now for a given $\varepsilon > 0$ consider the set $\mathcal{X}_\varepsilon := \{x \in X' : d(x, \mathcal{X}) < \varepsilon\}$. Since the set $X' \setminus \mathcal{X}_\varepsilon$ is compact we can choose a finite number of points $x_1, ..., x_K \in X' \setminus \mathcal{X}_\varepsilon$ and their respective neighborhoods $\mathcal{V}_1, ..., \mathcal{V}_K$ covering the set $X' \setminus \mathcal{X}_\varepsilon$ such that for all N large enough the events $\{\mathcal{V}_k \cap \bar{\mathcal{X}}_N = \emptyset\}$, $k = 1, ..., K$, happen w.p.1. It follows that w.p.1 for all N large enough $\bar{\mathcal{X}}_N$ is a subset of $\mathcal{X}_\varepsilon$. This shows that $\mathfrak{d}_N$ tends to zero w.p.1.

To show that $\mathbb{D}(\hat{\mathcal{S}}_N, \mathcal{S}^*) \to 0$ w.p.1 the arguments now basically are deterministic, i.e., $\mathfrak{d}_N$ and $\hat{x}_N \in \hat{\mathcal{S}}_N$ are viewed as random variables, $\mathfrak{d}_N = \mathfrak{d}_N(\omega)$, $\hat{x}_N = \hat{x}_N(\omega)$, defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and we want to show that $d(\hat{x}_N(\omega), \mathcal{S}^*)$ tends to zero for all $\omega \in \Omega$ except on a set of $\mathbb{P}$-measure zero. Therefore we consider sequences $\mathfrak{d}_N$ and $\hat{x}_N$ as deterministic, for a particular (fixed) $\omega \in \Omega$, and drop mentioning "w.p.1". Since $\mathfrak{d}_N \to 0$, there is $\tilde{x}_N \in \mathcal{X}$ such that $\|\hat{x}_N - \tilde{x}_N\|$ tends to zero. Note that as an intersection of closed sets, the set $\mathcal{X}$ is closed. By the assumption (iv) and continuity of $\hat{\Phi}(\cdot, \xi)$ we have that $\hat{\phi}_N(\cdot)$ converges w.p.1 to $\phi(\cdot)$ uniformly on the compact set $\mathcal{X} \cap X'$ (this is the so-called uniform Law of Large Numbers, e.g., [27, Theorem 7.48]), i.e., for all $\omega \in \Omega$ except on a set of $\mathbb{P}$-measure zero

$$\sup_{x\in\mathcal{X}\cap X'} \|\hat{\phi}_N(x) - \phi(x)\| \to 0, \ \text{ as } N \to \infty.$$

By passing to a subsequence if necessary we can assume that $\hat{x}_N$ converges to a point x^*. It follows that $\tilde{x}_N \to x^*$ and hence $\hat{\phi}_N(\tilde{x}_N) \to \phi(x^*)$. Thus $\hat{\phi}_N(\hat{x}_N) \to \phi(x^*)$. Since Γ_1 is closed it follows that $0 \in \phi(x^*) + \Gamma_1(x^*)$, i.e., $x^* \in \mathcal{S}^*$. This completes the proof of part (a), and also implies that the set $\mathcal{S}^*$ is nonempty.

Proof of part (b).

By [19, Theorem 3.1 (ii)], $\mathcal{R}(0) = 0$, $\mathcal{R}(\varepsilon)$ is nondecreasing on $[0, \infty)$ and $\mathcal{R}(\varepsilon) > 0$ for all $\varepsilon > 0$. Note that it follows that $\mathcal{R}^{-1}(t)$ is nondecreasing on $[0, \infty)$ and tends to zero as $t \downarrow 0$.

Let $\delta = \mathcal{R}(\varepsilon)/4$. By part (a) and the uniform Law of Large Numbers, we have w.p.1 that for N large enough

$$\sup_{x \in \mathcal{X} \cap X'} \|\phi(x) - \hat{\phi}_N(x)\| \leq \delta.$$

Then w.p.1 for N large enough such that $\mathfrak{d}_N \leq \varepsilon$, for any point $x \in \bar{\mathcal{X}}_N \cap X'$ with $d(z_x, \mathcal{S}^*) \geq \varepsilon$ it follows that

$$\begin{aligned}
& d(0, \hat{\phi}_N(x) + \Gamma_1(x)) \\
\geq \quad & d(0, \hat{\phi}_N(z_x) + \Gamma_1(z_x) + \delta\mathcal{B}) - \mathbb{D}(\hat{\phi}_N(x) + \Gamma_1(x), \hat{\phi}_N(z_x) + \Gamma_1(z_x) + \delta\mathcal{B}) \\
\geq \quad & d(0, \phi(z_x) + \Gamma_1(z_x) + \delta\mathcal{B}) - \mathbb{D}(\hat{\phi}_N(z_x) + \Gamma_1(z_x) + \delta\mathcal{B}, \phi(z_x) + \Gamma_1(z_x) + \delta\mathcal{B}) \\
& - \mathbb{D}(\hat{\phi}_N(x) + \Gamma_1(x), \hat{\phi}_N(z_x) + \Gamma_1(z_x) + \delta\mathcal{B}) \\
\geq \quad & d(0, \phi(z_x) + \Gamma_1(z_x) + \delta\mathcal{B}) - \|\hat{\phi}_N(z_x), \phi(z_x)\| - \|\hat{\phi}_N(x), \hat{\phi}_N(z_x)\| \\
& - \mathbb{D}(\Gamma_1(x), \Gamma_1(z_x) + \delta\mathcal{B}) \\
\geq \quad & 3\delta - \delta - \delta - 0 = \delta.
\end{aligned}$$

which implies $x \notin \hat{\mathcal{S}}_N$. Then

$$d(x, \mathcal{S}^*) \leq \|x - z_x\| + d(z_x, \mathcal{S}^*) \leq \tau + \mathcal{R}^{-1}\left(\sup_{x \in \mathcal{X} \cap X'} \|\phi(x) - \hat{\phi}_N(x)\|\right).$$

This completes the proof. ∎

The assumption that the set $\hat{\mathcal{S}}_N$ is nonempty means existence of solutions of the SAA problem (1.6)-(1.7). Existence of the solutions of deterministic VI and infinite dimensional VI has been well investigated in [10] and [12], respectively. Existence of a solution to the perturbed generalized equations has been investigated in the literature of deterministic generalized equations. For instance, in [17] a number of sufficient conditions is derived which ensure solvability (existence of a solution) of perturbed generalized equations. Similar conditions were further investigated in [16] and their one-stage stochastic extension has been presented in [19]. Those results can be applied to one-stage version (2.2) of (1.1)-(1.2) and its SAA problem (2.4) directly. Moreover, in section 3, based on the results in [12] for infinite dimensional VI, we propose sufficient conditions of existence and uniqueness of the solutions of two-stage SVI-NCP, a special case of two-stage SGE (1.1)-(1.2).

In case of the relatively complete recourse there is no need for condition (vi), the estimate (2.7) holds with $\tau = 0$ and the derivations can follow the similar results in [19, 27, 30] directly. It is interesting to consider how strong condition (vi) is. In the following remark we show that condition (vi) can also hold without the assumption of relatively complete recourse under mild conditions.

Remark 2.1. In condition (vi), the third inequality of (2.6) can be easily verified when N sufficiently large and $\hat{\Phi}(\cdot, \xi)$ is Lipschitz continuous with Lipschitz module $\kappa_{\hat{\Phi}}(\xi)$ and $\mathbb{E}[\kappa_{\hat{\Phi}}(\xi)] < \infty$. In Lemma 2.7 and Theorem 3.7 below, we verify the third inequality of (2.6) under moderate conditions.

Moreover, in the case when $\Gamma_1(\cdot) := \mathcal{N}_C(\cdot)$ with a nonempty polyhedral convex set C, the first and second inequalities of (2.6) hold automatically. Let $\mathfrak{F} = \{F_1, \cdots, F_K\}$ be the family of all nonempty faces of C and

$$\mathcal{K} := \{k : \mathcal{X} \cap X' \cap F_k \neq \emptyset, k = 1, \cdots, K\}.$$

Then w.p.1 for N sufficiently large, $\bar{\mathcal{X}}_N \cap X' \cap F_k = \emptyset$ for all $k \notin \mathcal{K}$. Note that for all $k \in \mathcal{K}$, $\bar{\mathcal{X}}_N \cap X' \cap F_k \neq \emptyset$. Moreover, it is important to note that for all $x_1 \in \mathrm{reint}(F_k)$

and $x_2 \in F_k$, $k \in \{1, \cdots, K\}$, $\mathcal{N}_C(x_1) \subseteq \mathcal{N}_C(x_2)$. Then for any $x \in \bar{\mathcal{X}}_N \cap X' \setminus \mathcal{X}$, there exists $k \in \mathcal{K}$ such that $x \in \text{reint}(F_k)$. To see this, we assume for contradiction that $x \in F_k \setminus \text{reint}(F_k)$ for some $k \in \mathcal{K}$ and there is no $k \in \mathcal{K}$ such that $x \in \text{reint}(F_k)$. Then there exist some $\bar{k} \in \{1, \cdots, K\}$ such that $x \in \text{reint}(F_{\bar{k}})$ (if $F_{\bar{k}}$ is singleton, then $\text{reint}(F_{\bar{k}}) = F_{\bar{k}}$) and $\bar{k} \notin \mathcal{K}$. This contradicts that $\bar{\mathcal{X}}_N \cap X' \cap F_k = \emptyset$ for all $k \notin \mathcal{K}$.

Note that $\mathbb{H}\left(\bar{\mathcal{X}}_N \cap X', \mathcal{X} \cap X'\right) \leq \mathfrak{d}_N$ and $\mathfrak{d}_N \to 0$ as $N \to \infty$ w.p.1. Let $z_x = \arg\min_{z \in \mathcal{X} \cap X' \cap F_k} \|z - x\|$. Then $\mathcal{N}_C(x) \subseteq \mathcal{N}_C(z_x)$ and for

$$\tau_N := \max_{k \in \mathcal{K}} \max_{x \in \bar{\mathcal{X}}_N \cap X' \cap F_k} \min_{z \in \mathcal{X} \cap X' \cap F_k} \|z - x\|,$$

we have that $\tau_N \to 0$ as $\mathfrak{d}_N \to 0$. Hence (2.6) is verified.

From Figure 1, it is easy to observe the relationship between $x \in \bar{\mathcal{X}}_N \cap X'$ and $z_x \in \mathcal{X} \cap X'$: they are in the same face of the polyhedral convex set $C = \mathbb{R}^2_+$ and $\mathcal{N}_{\mathbb{R}^2_+}(x) \subseteq \mathcal{N}_{\mathbb{R}^2_+}(z_x)$, where $\mathcal{X}$, $\bar{\mathcal{X}}_N$ and X' are indicated in the figure. Moreover, $\tau \to 0$ with $\gamma \to 0$. In the general case when C is not polyhedral, let $\Gamma_1(x) = \mathcal{N}_C(x)$. Without complete recourse, even x and z_x are sufficiently close to each other, $\mathbb{D}(\mathcal{N}_C(x), \mathcal{N}_C(z_x))$ may still be the infinity. Then condition (2.6) fails.

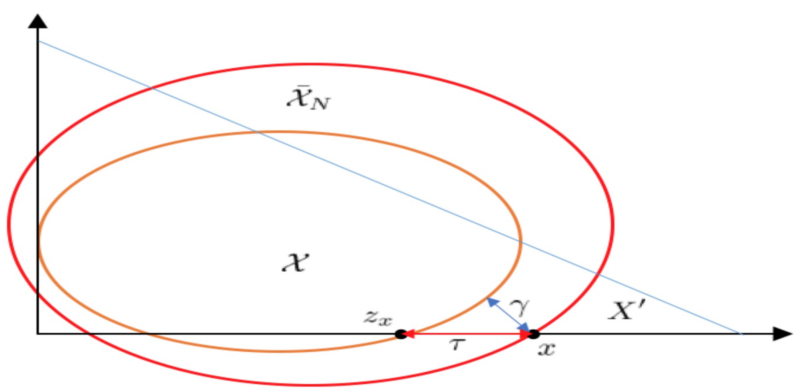

FIG. 1. *Relationship between x and z_x*

2.2. Exponential rate of convergence. We assume in this section that the set $\mathcal{S}^*$ of solutions of the first stage problem is nonempty, and the set X is *compact.* The last assumption of compactness of X can be relaxed to assuming that there is a compact subset X' of X such w.p.1 $\hat{\mathcal{S}}_N \subset X'$, and to deal with the set X' rather than X. For simplicity of notation we assume directly compactness of X.

Under Assumption 2.2 and by Lemma 2.1, we have that $\hat{\Phi}(x, \xi)$, defined in (2.3), is continuous in $x \in \mathcal{X}$. However to investigate the exponential rate of convergence, we need to verify Lipschitz continuity of $\hat{\Phi}(\cdot, \xi)$. To this end, we assume the *Clarke Differential* (CD) regularity property of the second stage generalized equation (1.2). By $\pi_y \partial_{(x,y)}(\Psi(\bar{x}, \bar{y}, \bar{\xi}))$, we denote the projection of the Clarke generalized Jacobian $\partial_{(x,y)}\Psi(\bar{x}, \bar{y}, \bar{\xi})$ in $\mathbb{R}^{m \times n} \times \mathbb{R}^{m \times m}$ onto $\mathbb{R}^{m \times m}$: the set $\pi_y \partial_{(x,y)}\Psi(\bar{x}, \bar{y}, \bar{\xi})$ consists of matrices $J \in \mathbb{R}^{m \times m}$ such that the matrix (S, J) belongs to $\partial_{(x,y)}\Psi(\bar{x}, \bar{y}, \bar{\xi})$ for some $S \in \mathbb{R}^{m \times n}$.

DEFINITION 2.5. *For $\bar{\xi} \in \Xi$ a solution $\bar{y}$ of the second stage generalized equation* (1.2) *is said to be* parametrically CD-regular*, at $x = \bar{x} \in \bar{\mathcal{X}}(\bar{\xi})$, if for each $J \in$*

$\pi_y \partial_{(x,y)} \Psi(\bar{x}, \bar{y}, \bar{\xi})$ *the solution* $\bar{y}$ *of the following SGE is strongly regular*

$$0 \in \Psi(\bar{x}, \bar{y}, \bar{\xi}) + J(y - \bar{y}) + \Gamma_2(y, \bar{\xi}). \tag{2.8}$$

That is, there exist neighborhoods $\mathcal{U}$ *of* $\bar{y}$ *and* $\mathcal{V}$ *of* 0 *such that for every* $\eta \in \mathcal{V}$ *the perturbed (partially) linearized SGE of* (2.8)

$$\eta \in \Psi(\bar{x}, \bar{y}, \bar{\xi}) + J(y - \bar{y}) + \Gamma_2(y, \bar{\xi})$$

has in $\mathcal{U}$ *a unique solution* $\hat{y}_{\bar{x}}(\eta)$, *and the mapping* $\eta \to \hat{y}_{\bar{x}}(\eta) : \mathcal{V} \to \mathcal{U}$ *is Lipschitz continuous.*

Assumption 2.4. *For a.e.* $\xi \in \Xi$, *there exists a unique, parametrically CD-regular solution* $\bar{y} = \hat{y}(\bar{x}, \xi)$ *of the second stage generalized equation* (1.2) *all* $\bar{x} \in \mathcal{X}$.

Proposition 2.6. *Suppose Assumption* 2.4 *holds. Then for a.e.* $\xi \in \Xi$, *the solution mapping* $\hat{y}(x, \xi)$ *of the second stage generalized equation* (1.2) *is a Lipschitz continuous function of* $x \in \mathcal{X}$, *with Lipschitz constant* $\kappa(\xi)$.

The result is implied directly by [14, Theorem 4] and the compactness of $\mathcal{X} \subseteq X$. Moreover, note that for any $\bar{x} \in \mathcal{X}$, if the generalized equation

$$0 \in G_{\bar{x}}(y) := \Psi(\bar{x}, \bar{y}, \bar{\xi}) + J(y - \bar{y}) + \Gamma_2(y, \bar{\xi}) \ \text{ for which } \ G_{\bar{x}}(\bar{y}) \ni 0,$$

has a locally Lipschitz continuous solution function at 0 for $\bar{y}$ with Lipschitz constant $\kappa_G(\bar{x}, \xi)$. Then by [9, Theorem 1.1], we have

$$\kappa_{\bar{x}}(\xi) = \kappa_G(\bar{x}, \xi)\kappa_\Psi(\xi) < \infty$$

is a Lipschitz constant of the second stage solution function $\hat{y}(x, \xi)$ at $\bar{x}$.

Assumption 2.5. *The set* $\mathcal{X}$ *is convex, its interior* $\text{int}(\mathcal{X}) \neq \emptyset$, *and for a.e.* $\xi \in \Xi$, *the generalized equation*

$$0 \in G_{\bar{x}}(y) = \Psi(\bar{x}, \bar{y}, \xi) + J(y - \bar{y}) + \Gamma_2(y, \xi), \ \text{ for which } \ G_{\bar{x}}(\bar{y}) \ni 0,$$

has a locally Lipschitz continuous solution function at 0 *for* $\bar{y}$ *with Lipschitz constant* $\kappa_G(\bar{x}, \xi)$ *for all* $\bar{x} \in \mathcal{X}$ *and there exists a measurable function* $\bar{\kappa}_G : \Xi \to \mathbb{R}_+$ *such that,* $\kappa_G(x, \xi) \le \bar{\kappa}_G(\xi)$ *and* $\mathbb{E}[\bar{\kappa}_G(\xi)\kappa_\Psi(\xi)] < \infty$.

Under Assumption 2.5, it can be seen that $\mathbb{E}[\hat{y}(x, \xi)]$ is Lipschitz continuous over $x \in \mathcal{X}$ with Lipschitz constant $\mathbb{E}[\bar{\kappa}_G(\xi)\kappa_\Psi(\xi)]$. We consider then the first stage (1.1) of the SGE as the generalized equation (2.2) with the respective second stage solution $\hat{y}(x, \xi)$ (recall definition (2.3) of $\hat{\Phi}(x, \xi)$ and $\phi(x)$).

Lemma 2.7. *Suppose that Assumptions* 2.4–2.5 *hold,* $\mathbb{E}[\kappa_\Phi(\xi)] < \infty$ *and*

$$\mathbb{E}\left[\kappa_\Phi(\xi)\bar{\kappa}_G(\xi)\kappa_\Psi(\xi)\right] < \infty.$$

Then for a.e. $\xi \in \Xi$, $\hat{\Phi}(x, \xi)$ *and* $\phi(x)$ *are Lipschitz continuous over* $x \in \mathcal{X}$ *with respective Lipschitz modulus*

$$\kappa_\Phi(\xi) + \kappa_\Phi(\xi)\bar{\kappa}_G(\xi)\kappa_\Psi(\xi) \text{ and } \mathbb{E}[\kappa_\Phi(\xi)] + \mathbb{E}[\kappa_\Phi(\xi)\bar{\kappa}_G(\xi)\kappa_\Psi(\xi)].$$

Remark 2.2. Specifically we study Assumptions 2.2–2.5 in the framework of the following SGE:

$$0 \in \mathbb{E}[\Phi(x, y(\xi), \xi)] + \Gamma_1(x), \ x \in X, \tag{2.9}$$

$$0 \in \Psi(x, y(\xi), \xi) + \mathcal{N}_{\mathbb{R}^m_+}(H(x, y(\xi), \xi)), \quad \text{for a.e. } \xi \in \Xi, \tag{2.10}$$

where $H(x,y,\xi) : \mathbb{R}^n \times \mathbb{R}^m \times \Xi \to \mathbb{R}^m$. Let $h(x,y,\xi) := \min\{\Psi(x,y,\xi), H(x,y,\xi)\}$. Then the second stage VI (2.10) is equivalent to

$$h(x,y,\xi) = 0, \quad \text{for a.e. } \xi \in \Xi. \tag{2.11}$$

For $x = \bar{x}$ and $\xi \in \Xi$ let $\bar{y}$ be a solution of (2.11), and suppose that each matrix $J \in \pi_y \partial h(\bar{x},\bar{y},\xi)$ is nonsingular for a.e. ξ. Then by Clarke's Inverse Function Theorem, there exists a Lipschitz continuous solution function $\hat{y}(x,\xi)$ such that $\hat{y}(\bar{x},\xi) = \bar{y}$ and the Lipschitz constant is bounded by $\|J^{-1}(x,y,\xi)S(x,y,\xi)\|$ for all

$$(S(x,y,\xi), J(x,y,\xi))^\top \in \pi_{x,y}\partial h(x,y,\xi).$$

Then Assumption 2.4 holds. Moreover, if we assume

$$\mathbb{E}\left[\|J^{-1}(x,\hat{y}(x,\xi),\xi)S(x,\hat{y}(x,\xi),\xi)\|\right] < \infty$$

for all $x \in \mathcal{X}$, then Assumption 2.5 holds.

Now we investigate exponential rate of convergence of the two-stage SAA problem (1.6)–(1.7) by using a uniform Large Deviations Theorem (cf., [27, 28, 30]). Let

$$M_x^i(t) := \mathbb{E}\left\{\exp\big(t[\hat{\Phi}_i(x,\xi) - \phi_i(x)]\big)\right\}$$

be the moment generating function of the random variable $\hat{\Phi}_i(x,\xi) - \phi_i(x)$, $i = 1,\ldots,n$, and

$$M_\kappa(t) := \mathbb{E}\left\{\exp\left(t\big[\kappa_\Phi(\xi) + \kappa_\Phi(\xi)\kappa(\xi) - \mathbb{E}[\kappa_\Phi(\xi) + \kappa_\Phi(\xi)\kappa(\xi)]\big]\right)\right\}.$$

ASSUMPTION 2.6. *For every $x \in \mathcal{X}$ and $i = 1,\ldots,n$, the moment generating functions $M_x^i(t)$ and $M_\kappa(t)$ have finite values for all t in a neighborhood of zero.*

THEOREM 2.8. *Suppose:* (i) *assumptions 2.1, 2.3–2.6 hold,* (ii) $\mathcal{S}^*$ *is nonempty and w.p.1 for N large enough, $\hat{\mathcal{S}}_N$ are nonempty,* (iii) *the multifunctions $\Gamma_1(\cdot)$ and $\Gamma_2(\cdot,\xi)$, $\xi \in \Xi$, are closed and monotone. Then the following statements hold.*

(a) *For sufficiently small $\varepsilon > 0$ there exist positive constants $\varrho = \varrho(\varepsilon)$ and $\varsigma = \varsigma(\varepsilon)$, independent of N, such that*

$$\mathbb{P}\left\{\sup_{x\in\mathcal{X}} \big\|\hat{\phi}_N(x) - \phi(x)\big\| \geq \varepsilon\right\} \leq \varrho(\varepsilon)e^{-N\varsigma(\varepsilon)}. \tag{2.12}$$

(b) *Assume in addition:* (iv) *The condition of part* (b) *in Theorem 2.4 holds and w.p.1 for N sufficiently large,*

$$\mathcal{S}^* \cap \mathrm{cl}\big(\mathrm{bd}(\mathcal{X}) \cap \mathrm{int}(\bar{\mathcal{X}}_N)\big) = \emptyset. \tag{2.13}$$

(v) *$\phi(\cdot)$ has the following strong monotonicity property for every $x^* \in \mathcal{S}^*$:*

$$(x - x^*)^\top(\phi(x) - \phi(x^*)) \geq g(\|x - x^*\|), \ \forall x \in \mathcal{X}, \tag{2.14}$$

where $g : \mathbb{R}_+ \to \mathbb{R}_+$ is such a function that function $\mathfrak{r}(\tau) := g(\tau)/\tau$ is monotonically increasing for $\tau > 0$.
Then $\mathcal{S}^* = \{x^*\}$ is a singleton and for any sufficiently small $\varepsilon > 0$, there exists N sufficiently large such that

$$\mathbb{P}\left\{\mathbb{D}(\hat{\mathcal{S}}_N, \mathcal{S}^*) \geq \varepsilon\right\} \leq \varrho\left(\mathfrak{r}^{-1}(\varepsilon)\right)\exp\left(-N\varsigma\big(\mathfrak{r}^{-1}(\varepsilon)\big)\right), \tag{2.15}$$

where $\varrho(\cdot)$ and $\varsigma(\cdot)$ are defined in (2.12), and $\mathfrak{r}^{-1}(\varepsilon) := \inf\{\tau > 0 : \mathfrak{r}(\tau) \geq \varepsilon\}$ is the inverse of $\mathfrak{r}(\tau)$.

Proof. Part (a). By Lemma 2.7, because of conditions (i) and (ii) and compactness of X, we have by [27, Theorem 7.67] that for every $i \in \{1, \ldots, n\}$ and $\varepsilon > 0$ small enough, there exist positive constants $\varrho_i = \varrho_i(\varepsilon)$ and $\varsigma_i = \varsigma_i(\varepsilon)$, independent of N, such that

$$\mathbb{P}\left\{\sup_{x\in\mathcal{X}} \left|(\hat{\phi}_N)_i(x) - \phi_i(x)\right| \geq \varepsilon\right\} \leq \varrho_i(\varepsilon)e^{-N\varsigma_i(\varepsilon)},$$

and hence (2.12) follows.

Part (b). By condition (iv) we have that $\mathbb{D}(\mathcal{S}^*, \bar{\mathcal{X}}_N \setminus \mathcal{X}) > 0$. Let ε be sufficiently small such that w.p.1 for N sufficiently large,

$$\mathbb{D}(\mathcal{S}^*, \bar{\mathcal{X}}_N \setminus \mathcal{X}) \geq 3\varepsilon.$$

Note that since $\mathcal{X} \subseteq \bar{\mathcal{X}}_{N+1} \subseteq \bar{\mathcal{X}}_N$, $\mathbb{D}(\mathcal{S}^*, \bar{\mathcal{X}}_N \setminus \mathcal{X})$ is nondecreasing with $N \to \infty$.

By Theorem 2.4, part (b), w.p.1 for N sufficiently large such that $\tau \leq \varepsilon$, we have

$$\mathcal{R}^{-1}\left(\sup_{x\in\mathcal{X}} \|\hat{\phi}_N(x) - \phi(x)\|\right) \leq \varepsilon$$

and

$$\mathbb{D}(\hat{\mathcal{S}}_N, \mathcal{S}^*) \leq \tau + \mathcal{R}^{-1}\left(\sup_{x\in\mathcal{X}} \|\hat{\phi}_N(x) - \phi(x)\|\right) \leq 2\varepsilon.$$

Since by condition (iv), when N sufficiently large w.p.1, for any point $\tilde{x} \in \bar{\mathcal{X}}_N \setminus \mathcal{X}$, $\mathbb{D}(\tilde{x}, \mathcal{S}^*) \geq 3\varepsilon$, which implies $\hat{\mathcal{S}}_N \subset \mathcal{X}$ and then

$$\mathbb{D}(\hat{\mathcal{S}}_N, \mathcal{S}^*) \leq \mathcal{R}^{-1}\left(\sup_{x\in\mathcal{X}} \|\hat{\phi}_N(x) - \phi(x)\|\right). \tag{2.16}$$

In order to use (2.16) to derive an exponential rate of convergence of the SAA estimators we need an upper bound for $\mathcal{R}^{-1}(t)$, or equivalently a lower bound for $\mathcal{R}(\varepsilon)$. Note that because of the monotonicity assumptions we have that $\mathcal{S}^* = \{x^*\}$.

For $x \in \mathcal{X}$ and $z \in \Gamma_1(x)$ we have

$$(x - x^*)^\top(\phi(x) - \phi(x^*)) = (x - x^*)^\top(\phi(x) + z - \phi(x^*) - z) \leq (x - x^*)^\top(\phi(x) + z),$$

where the last inequality holds since $-\phi(x^*) \in \Gamma_1(x^*)$ and because of monotonicity of Γ_1. It follows that

$$(x - x^*)^\top(\phi(x) - \phi(x^*)) \leq \|x - x^*\|\,\|\phi(x) + z\|,$$

and since $z \in \Gamma_1(x)$ was arbitrary that

$$(x - x^*)^\top(\phi(x) - \phi(x^*)) \leq \|x - x^*\|\, d\big(0, \phi(x) + \Gamma_1(x)\big).$$

Together with (2.14) this implies

$$d\big(0, \phi(x) + \Gamma_1(x)\big) \geq \mathfrak{r}(\|x - x^*\|).$$

It follows that $\mathcal{R}(\varepsilon) \geq \mathfrak{r}(\varepsilon)$, $\varepsilon \geq 0$, and hence

$$\mathcal{R}^{-1}(t) \leq \mathfrak{r}^{-1}(t),$$

where $\mathfrak{r}^{-1}(\cdot)$ is the inverse of function $\mathfrak{r}(\cdot)$. Then by (2.12), (2.15) holds. ∎

Note that if $g(\tau) := c\,\tau^{\alpha}$ for some constants $c > 0$ and $\alpha > 1$, then $\mathfrak{r}^{-1}(t) = (t/c)^{1/(\alpha-1)}$. In particular for $\alpha = 2$, condition (2.14) assumes strong monotonicity of $\phi(\cdot)$. Note also that condition (iv) is not needed if the relatively complete recourse condition holds.

It is also interesting to consider how strong condition (2.13) is. Note that when $\mathcal{S}^* \subset \mathrm{int}(\mathcal{X})$, condition (2.13) holds. Moreover, we can also see from the following simple example that even when $\mathcal{S}^* \cap \mathrm{bd}(\mathcal{X}) \neq \emptyset$, condition (2.13) may still hold.

EXAMPLE 2.1. *Consider a two-stage SLCP*

$$0 \le \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \perp \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} \mathbb{E}[y_1(\xi)] \\ \mathbb{E}[y_2(\xi)] \end{pmatrix} \ge 0,$$

$$0 \le \begin{pmatrix} y_1(\xi) \\ y_2(\xi) \end{pmatrix} \perp \begin{pmatrix} \alpha(x_1,\xi) & 0 \\ 0 & \alpha(x_2,\xi) \end{pmatrix} \begin{pmatrix} y_1(\xi) \\ y_2(\xi) \end{pmatrix} - \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \ge 0, \ \textit{a.e. } \xi \in \Xi,$$

where

$$\alpha(t,\xi) = \begin{cases} \frac{1}{t+\xi+51}, & \textit{if } \ t+\xi \le 100, \\ 0, & \textit{otherwise}, \end{cases}$$

and ξ follows uniform distribution in $[-50, 50]$.

By simple calculation, we have that $\mathcal{S}^* = \{(0,0)\}$ and $\mathcal{X} = [0,50] \times [0,50]$. Moreover, consider an iid samples $\{\xi^j\}_{j=1}^N$ with $\max_j \xi^j = 49$, $\bar{\mathcal{X}}_N = [0,51] \times [0,51]$. Let $X = \{x : 0 \le x_1, x_2 \le 100\}$. It is easy to observe that although $\mathcal{S}^* = \{(0,0)\}$ is at the boundary of $\mathcal{X} \cap X$, condition (2.13) still holds.

REMARK 2.3. It is also interesting to estimate the required sample size of the SAA problem for the two-stage SGE. Similar to a discussion in [28, p.410], if there exists a positive constant $\sigma > 0$ such that

$$M_x^i(t) \le \exp\{\sigma^2 t^2/2\}, \ \ \forall t \in \mathbb{R}, \ i = 1, ..., n, \tag{2.17}$$

then it can be verified that $I_x^i(z) \ge \frac{z^2}{2\sigma^2}$ for all $z \in \mathbb{R}$, where $I_x^i(z) := \sup_{t\in\mathbb{R}}\{zt - \log M_x^i(t)\}$ is the large deviations rate function of random variable $\hat{\Phi}_i(x,\xi) - \phi_i(x)$, $i = 1, \cdots, n$. Note that if $\hat{\Phi}_i(x,\xi) - \phi_i(x)$ is subgaussian random variable, (2.17) holds, $i = 1, ..., n$. Then it can be verified that if

$$N \ge \frac{32 n \sigma}{\varepsilon^2} \left[\ln(n(2\Pi + 1)) + \ln\left(\frac{1}{\alpha}\right) \right],$$

then

$$\mathbb{P}\left\{ \sup_{x\in\mathcal{X}} \left\| \hat{\phi}_N(x) - \phi(x) \right\| \ge \varepsilon \right\} \le \alpha,$$

where $\Pi := (O(1) D \mathbb{E}[\kappa_\Phi(\xi) + \kappa_\Phi(\xi)\kappa(\xi)]/\varepsilon)^n$ and D is the diameter of X. Consequently it follows by (2.16) that if

$$N \ge \frac{32 n \sigma}{(\mathfrak{r}^{-1}(\varepsilon))^2} \left[\ln(n(2\hat{\Pi} + 1)) + \ln\left(\frac{1}{\alpha}\right) \right],$$

with $\hat{\Pi} := \left(O(1) D \mathbb{E}[\kappa_\Phi(\xi) + \kappa_\Phi(\xi)\kappa(\xi)]/\mathfrak{r}^{-1}(\varepsilon)\right)^n$, then we have

$$\mathbb{P}\left\{ \mathbb{D}(\hat{\mathcal{S}}_N, \mathcal{S}^*) \ge \varepsilon \right\} \le \alpha.$$

Confidence intervals based on the sample average approximations were studied in [18] for one-stage SVI problems. It could be possible to extend those results to two-stage SGE under mild conditions. This could be a topic for a future research.

In the next section, we will verify the conditions of Theorems 2.4 and 2.8 for the two-stage SVI-NCP under moderate assumptions.

3. Two-stage SVI-NCP and its SAA problem. In this section, we investigate convergence properties of the two-stage SGE (1.1)–(1.2) when $\Phi(x, y, \xi)$ and $\Psi(x, y, \xi)$ are continuously differentiable w.r.t. (x, y) for a.e. $\xi \in \Xi$ and $\Gamma_1(x) := \mathcal{N}_C(x)$ and $\Gamma_2(y) := \mathcal{N}_{\mathbb{R}^m_+}(y)$ with $C \subseteq \mathbb{R}^n$ being a nonempty, polyhedral, convex set. That is, we consider the mixed two-stage SVI-NCP

$$0 \in \mathbb{E}[\Phi(x, y(\xi), \xi)] + \mathcal{N}_C(x), \tag{3.1}$$

$$0 \le y(\xi) \perp \Psi(x, y(\xi), \xi) \ge 0, \quad \text{for a.e. } \xi \in \Xi, \tag{3.2}$$

and study convergence analysis of its SAA problem

$$0 \in N^{-1} \sum_{j=1}^{N} \Phi(x, y(\xi^j), \xi^j) + \mathcal{N}_C(x), \tag{3.3}$$

$$0 \le y(\xi^j) \perp \Psi(x, y(\xi^j), \xi^j) \ge 0, \;\; j = 1, ..., N. \tag{3.4}$$

We first give some required definitions. Let $\mathcal{Y}$ be the space of measurable functions $u : \Xi \to \mathbb{R}^m$ with finite value of $\int \|u(\xi)\|^2 P(d\xi)$ and $\langle \cdot, \cdot \rangle$ denotes the scalar product in the Hilbert space $\mathbb{R}^n \times \mathcal{Y}$ equipped with $\mathcal{L}_2$-norm, that is, for $x, z \in \mathbb{R}^n$ and $y, u \in \mathcal{Y}$,

$$\langle (x, y), (z, u) \rangle := x^\top z + \int_\Xi y(\xi)^\top u(\xi) P(d\xi).$$

Consider mapping $\mathcal{G} : \mathbb{R}^n \times \mathcal{Y} \to \mathbb{R}^n \times \mathcal{Y}$ defined as

$$\mathcal{G}(x, y(\cdot)) := \big(\mathbb{E}[\Phi(x, y(\xi), \xi)], \Psi(x, y(\cdot), \cdot)\big).$$

Monotonicity properties of this mapping are defined in the usual way. In particular the mapping $\mathcal{G}$ is said to be strongly monotone if there exists a positive number $\bar\kappa$ such that for any $(x, y(\cdot)), (z, u(\cdot)) \in \mathbb{R}^n \times \mathcal{Y}$, we have

$$\left\langle \mathcal{G}(x, y(\cdot)) - \mathcal{G}(z, u(\cdot)), \begin{pmatrix} x - z \\ y(\cdot) - u(\cdot) \end{pmatrix} \right\rangle \ge \bar\kappa(\|x - z\|^2 + \mathbb{E}[\|y(\xi) - u(\xi)\|^2]).$$

DEFINITION 3.1. ([12, Definition 12.1]) *The mapping $\mathcal{G} : \mathbb{R}^n \times \mathcal{Y} \to \mathbb{R}^n \times \mathcal{Y}$ is hemicontinuous on $\mathbb{R}^n \times \mathcal{Y}$ if $\mathcal{G}$ is continuous on line segments in $\mathbb{R}^n \times \mathcal{Y}$, i.e., for every pair of points $(x, y(\cdot)), (z, u(\cdot)) \in \mathbb{R}^n \times \mathcal{Y}$, the following function is continuous*

$$t \mapsto \left\langle \mathcal{G}(tx + (1-t)z, ty(\cdot) + (1-t)u(\cdot)), \begin{pmatrix} x - z \\ y(\cdot) - u(\cdot) \end{pmatrix} \right\rangle.$$

DEFINITION 3.2. ([12, Definition 12.3 (i)]) *The mapping $\mathcal{G} : \mathbb{R}^n \times \mathcal{Y} \to \mathbb{R}^n \times \mathcal{Y}$ is coercive if there exists $(x_0, y_0(\cdot)) \in \mathbb{R}^n \times \mathcal{Y}$ such that*

$$\frac{\left\langle \mathcal{G}(x, y(\cdot)), \begin{pmatrix} x - x_0 \\ y(\cdot) - y_0(\cdot) \end{pmatrix} \right\rangle}{\|x - x_0\| + \mathbb{E}[\|y(\xi) - y_0(\xi)\|]} \to \infty \;\text{ as }\; \|x\| + \mathbb{E}[\|y(\xi)\|] \to \infty \;\text{ and } (x, y(\cdot)) \in \mathbb{R}^n \times \mathcal{Y}.$$

Note that the strong monotonicity of $\mathcal{G}$ implies the coerciveness of $\mathcal{G}$, see [12, Chapter 12]. In section 3.1, we consider the properties in the second stage SNCP.

3.1. Lipschitz properties of the second stage solution mapping. Strong regularity of VI was investigated in Dontchev and Rockafellar [8]. We apply their results to the second stage SNCP. Consider a linear VI

$$0 \in Hz + q + \mathcal{N}_U(z), \tag{3.5}$$

where U is a closed nonempty, polyhedral, convex subset of $\mathbb{R}^l$.

DEFINITION 3.3. [8, Definition 2] *The critical face condition is said to hold at (q_0, z_0) if for any choice of faces F_1 and F_2 of the critical cone $\mathcal{C}_0$ with $F_2 \subset F_1$,*

$$u \in F_1 - F_2, \;\; H^\top u \in (F_1 - F_2)^* \implies u = 0,$$

where critical cone $\mathcal{C}_0 = \mathcal{C}(z_0, v_0) := \{z' \in \mathcal{T}_U(z_0) : z' \perp v_0\}$ with $v_0 = Hz_0 + q_0$.

THEOREM 3.4. [8, Theorem 2] *The linear variational inequality (3.5) is strongly regular at (q_0, z_0) if and only if the critical face condition holds at (q_0, z_0), where z_0 is the solution of the linear VI: $0 \in Hz + q_0 + \mathcal{N}_U(z)$.*

COROLLARY 3.1. [8, Corollary 1] *A sufficient condition for strong regularity of the linear variational inequality (3.5) at (q_0, z_0) is that $u^\top Hu > 0$ for all vectors $u \neq 0$ in the subspace spanned by the critical cone $\mathcal{C}_0$.*

Note that when H is a positive definite matrix, the condition in Corollary 3.1 holds and we do not need to assume the critical face condition in Definition 3.3. Then we apply Corollary 3.1 to the two-stage SVI-NCP and consider the Clarke generalized Jacobian of $\hat{y}(x,\xi)$. To this end, we introduce some notations: let

$$\begin{aligned}
\alpha(\hat{y}(x,\xi)) &= \{i : (\hat{y}(x,\xi))_i > (\Psi(x, \hat{y}(x,\xi), \xi))_i\}\\
\beta(\hat{y}(x,\xi)) &= \{i : (\hat{y}(x,\xi))_i = (\Psi(x, \hat{y}(x,\xi), \xi))_i\}\\
\gamma(\hat{y}(x,\xi)) &= \{i : (\hat{y}(x,\xi))_i < (\Psi(x, \hat{y}(x,\xi), \xi))_i\}.
\end{aligned}$$

Note that for any $x \in \mathcal{X}$ and a.e. $\xi \in \Xi$, $\hat{y}(x,\xi)$, $\alpha(\hat{y}(x,\xi))$, $\beta(\hat{y}(x,\xi))$ and $\gamma(\hat{y}(x,\xi)$ are uniquely defined. For simplicity, we use $\alpha = \alpha(\hat{y}(x,\xi))$, $\beta = \beta(\hat{y}(x,\xi))$ and $\gamma = \gamma(\hat{y}(x,\xi))$. Let $\nabla_x \Psi(x,y,\xi)$ and $\nabla_y \Psi(x,y,\xi)$ be the Jacobian of $\Psi(x,y,\xi)$ w.r.t. x and y respectively.

ASSUMPTION 3.1. *For a.e. $\xi \in \Xi$ and all $x \in \mathcal{X} \cap C$, $\Psi(x,\cdot,\xi)$ is strongly monotone, that is there exists a positive valued measurable $\kappa_y(\xi)$ such that for all $y, u \in \mathbb{R}^m$,*

$$\langle \Psi(x,y,\xi) - \Psi(x,u,\xi), y - u\rangle \geq \kappa_y(\xi)\|y-u\|^2$$

with $\mathbb{E}[\kappa_y(\xi)] < +\infty$.

Applying Corollary 2.1 in [15] to the second stage of the SVI-NCP, we have the following lemma.

LEMMA 3.5. *Suppose Assumption 3.1 holds and for a fixed $\bar{\xi} \in \Xi$, $\Psi(x,y,\xi)$ is continuously differentiable w.r.t. (x,y). Then for the fixed $\bar{\xi} \in \Xi$, (a) $\hat{y}(x,\bar{\xi})$ is an unique solution of the second stage NCP (3.2), (b) $\hat{y}(x,\bar{\xi})$ is F-differentiable at $\bar{x} \in \mathcal{X} \cap C$ if and only if $\beta(\hat{y}(\bar{x},\bar{\xi}))$ is empty and*

$$(\nabla_x \hat{y}(\bar{x},\xi))_\alpha = -(\nabla_y \Psi_{\alpha\alpha}(\bar{x}, \hat{y}(\bar{x},\xi),\xi))^{-1}\nabla_x \Psi_\alpha(\bar{x}, \hat{y}(\bar{x},\xi),\xi), \;\; (\nabla_x \hat{y}(\bar{x},\xi))_\gamma = 0$$

or

$$\nabla_x \Psi_\beta(\bar{x}, \hat{y}(\bar{x},\xi),\xi) = \nabla_y \Psi_{\beta\alpha}(\bar{x}, \hat{y}(\bar{x},\xi),\xi)(\nabla_y \Psi_{\alpha\alpha}(\bar{x}, \hat{y}(\bar{x},\xi),\xi))^{-1}\nabla_x \Psi_\alpha(\bar{x}, \hat{y}(\bar{x},\xi),\xi)$$

in this case, the F-derivative of $\hat{y}(\cdot,\xi)$ at $\bar{x}$ is given by

$$(\nabla_x \hat{y}(\bar{x},\xi))_\alpha = -(\nabla_y \Psi_{\alpha\alpha}(\bar{x},\hat{y}(\bar{x},\xi),\xi))^{-1}\nabla_x \Psi_\alpha(\bar{x},\hat{y}(\bar{x},\xi),\xi),$$
$$(\nabla_x \hat{y}(\bar{x},\xi))_\beta = 0, \;\; (\nabla_x \hat{y}(\bar{x},\xi))_\gamma = 0.$$

THEOREM 3.6. *Let $\Psi : \mathbb{R}^n \times \mathbb{R}^m \times \Xi \to \mathbb{R}^m$ be Lipschitz continuous and continuously differentiable over $\mathbb{R}^n \times \mathbb{R}^m$ for a.e. $\xi \in \Xi$. Suppose Assumption 3.1 holds and $\Phi(x,y,\xi)$ is continuously differentiable w.r.t. (x,y) for a.e. $\xi \in \Xi$. Then for a.e. $\xi \in \Xi$ and $x \in \mathcal{X}$, the following holds.*

(a) *The second stage SNCP (3.2) has a unique solution $\hat{y}(x,\xi)$ which is parametrically CD-regular and the mapping $x \mapsto \hat{y}(x,\xi)$ is Lipschitz continuous over $\mathcal{X} \cap X'$, where X' is a compact subset of $\mathbb{R}^n$.*

(b) *The Clarke Jacobian of $\hat{y}(x,\xi)$ w.r.t. x is as follows*

$$\partial \hat{y}(x,\xi) = \mathrm{conv}\Big\{ \lim_{z\to x} \nabla_z \hat{y}(z,\xi) : \nabla_z \hat{y}(z,\xi) = -[I - D_\alpha(I - M(z,\hat{y}(z,\xi),\xi))]^{-1} D_\alpha L(z,\hat{y}(z,\xi),\xi)\Big\},$$

where $M(x,y,\xi) = \nabla_y \Psi(x,y,\xi)$, $L(x,\hat{y}(x,\xi),\xi) = \nabla_x \Psi(x,\hat{y}(x,\xi),\xi)$.

Proof. Part (a). Note that by Lemma 3.5 (a), for almost all $\bar{\xi} \in \Xi$ and every $\bar{x} \in \mathcal{X} \cap X'$, there exists a unique solution $\hat{y}(\bar{x},\bar{\xi})$ of the second stage SNCP (3.2). Moreover, consider the LCP

$$0 \le y \perp \Psi(\bar{x},\bar{y},\bar{\xi}) + \nabla_y \Psi(\bar{x},\bar{y},\bar{\xi})(\bar{y} - y) \ge 0, \tag{3.6}$$

where $\bar{y} = \hat{y}(\bar{x},\bar{\xi})$. By the strong monotonicity of $\Psi(\bar{x},\cdot,\bar{\xi})$, $\nabla_y \Psi(\bar{x},\bar{y},\bar{\xi})$ is positive definite. Then by Corollary 3.1, the LCP (3.6) is strongly regular at $\bar{y}$. This implies the parametrically CD-regular of the second stage SNCP (3.2) with $\bar{x}$ at solution $\bar{y}$. Then the Lipschitz property follows from [14, Theorem 4] and the compactness of X'.

Part (b). For any fixed $\bar{\xi}$, by Part (a), there exists a unique Lipschitz function $\hat{y}(\cdot,\bar{\xi})$ such that $\hat{y}(x,\bar{\xi})$ over $\mathcal{X}$ which solves

$$0 \le y \perp \Psi(x,y,\bar{\xi}) \ge 0.$$

Note that $\hat{y}(\cdot,\bar{\xi})$ is Lipschitz continuous and hence F-differentiable almost everywhere over $\mathcal{B}_\delta(\bar{x})$. Then for any $x' \in \mathcal{B}_\delta(\bar{x})$ such that $\hat{y}(x',\bar{\xi})$ is F-differentiable, by Lemma 3.5 (b), we have $\beta(\hat{y}(x',\xi))$ is empty and

$$(\nabla_x \hat{y}(x',\xi))_\alpha = -(\nabla_y \Psi(x',\hat{y}(x',\xi),\xi))^{-1}_{\alpha\alpha}(\nabla_x \Psi(x',\hat{y}(x',\xi),\xi))_\alpha, \;\; (\nabla_x \hat{y}(x',\xi))_\gamma = 0 \tag{3.7}$$

or $\beta(\hat{y}(x',\xi))$ is not empty and

$$\begin{aligned}(\nabla_x \hat{y}(x',\xi))_\alpha &= -(\nabla_y \Psi(x',\hat{y}(x',\xi),\xi))^{-1}_{\alpha\alpha}(\nabla_x \Psi(x',\hat{y}(x',\xi),\xi))_\alpha,\\ (\nabla_x \hat{y}(x',\xi))_\beta &= 0, \;\; (\nabla_x \hat{y}(x',\xi))_\gamma = 0.\end{aligned} \tag{3.8}$$

Let $D_J \in \mathcal{D}$ be an m-dimensional diagonal matrix with $J \in \mathcal{J}$ and

$$(D_J)_{jj} := \begin{cases} 1, & \text{if } j \in J,\\ 0, & \text{otherwise,}\end{cases} \tag{3.9}$$

$M(x,y,\xi) = \nabla_y \Psi(x,y,\xi)$ and $W(x,\xi) = [I - D_\alpha(I - M(x,y,\xi))]^{-1}D_\alpha$. Then by (3.7) and (3.8), similar as in [5, Theorem 2.1],

$$\nabla_x \hat{y}(x',\xi) = -[I - D_\alpha(I - M(x',\hat{y}(x',\bar{\xi}),\xi))]^{-1} D_\alpha L(x',\hat{y}(x',\bar{\xi}),\bar{\xi}),$$

where $L(x,\hat{y}(x,\xi),\xi) = \nabla_x \Psi(x,\hat{y}(x,\xi),\xi)$. Let

$$U_J(M) = (I - D_J(I-M))^{-1} D_J, \ \ \forall J \in \mathcal{J}. \tag{3.10}$$

By the definition and outer semicontinuity of Clarke generalized Jacobian, we have

$$\begin{aligned} \partial \hat{y}(x,\xi) &= \operatorname{conv}\Big\{\lim_{z\to x} \nabla_z \hat{y}(z,\xi) : \nabla_z \hat{y}(z,\xi) = \\ &\quad -[I - D_\alpha(I - M(z,\hat{y}(z,\xi),\xi))]^{-1} D_\alpha L(z,\hat{y}(z,\xi),\xi)\Big\} \\ &\subseteq \operatorname{conv}\{-U_J(M(x,\hat{y}(x,\xi),\xi))L(x,\hat{y}(x,\xi),\xi) : J \in \mathcal{J}\}. \end{aligned}$$

We complete the proof. ∎

It is easy to observe that

$$\begin{aligned} &\partial \hat{y}(x,\xi) = \operatorname{conv}\Big\{\lim_{z\to x} \nabla_z \hat{y}(z,\xi) : \nabla_z \hat{y}(z,\xi) \\ &\qquad = -[I - D_\alpha(I - M(z,\hat{y}(z,\xi),\xi))]^{-1} D_\alpha L(z,\hat{y}(z,\xi),\xi)\Big\} \\ &\subseteq \operatorname{conv}\{-U_J(M(x,\hat{y}(x,\xi),\xi))L(x,\hat{y}(x,\xi),\xi) : J \in \mathcal{J}\}, \end{aligned} \tag{3.11}$$

where $\mathcal{J} := 2^{\{1,\dots,m\}}$, D_J and U_J are defined in (3.9) and (3.10) respectively.

Under Assumption 3.1, the two-stage SVI-NCP can be reformulated as a single stage SVI with $\hat{\Phi}(x,\xi) = \Phi(x,\hat{y}(x,\xi),\xi)$ and $\phi(x) = \mathbb{E}[\hat{\Phi}(x,\xi)]$ as follows

$$0 \in \phi(x) + \mathcal{N}_C(x). \tag{3.12}$$

With the results in Theorem 3.6, SVI (3.12) has the following properties. Let

$$\Theta(x,y(\xi),\xi) = \begin{pmatrix} \Phi(x,y(\xi),\xi) \\ \Psi(x,y(\xi),\xi) \end{pmatrix}$$

and $\nabla\Theta(x,y,\xi)$ be the Jacobian of Θ. Then

$$\nabla\Theta(x,y,\xi) = \begin{pmatrix} A(x,y,\xi) & B(x,y,\xi) \\ L(x,y,\xi) & M(x,y,\xi) \end{pmatrix},$$

where $A(x,y,\xi) = \nabla_x\Phi(x,y,\xi)$, $B(x,y,\xi) = \nabla_y\Phi(x,y,\xi)$, $L(x,y,\xi) = \nabla_x\Psi(x,y,\xi)$ and $M(x,y,\xi) = \nabla_y\Psi(x,y,\xi)$.

THEOREM 3.7. *Suppose the conditions of Theorem 3.6 hold. Let $X' \subseteq C$ be a compact set, for any $\xi \in \Xi$, $Y(\xi) = \{\hat{y}(x,\xi) : x \in X'\}$ and $\nabla\Theta(x,y,\xi)$ be the Jacobian of Θ. Assume*

$$\mathbb{E}[\|A(x,\hat{y}(x,\xi),\xi) - B(x,\hat{y}(x,\xi),\xi)M(x,\hat{y}(x,\xi),\xi)^{-1}L(x,\hat{y}(x,\xi),\xi)\|] < +\infty \tag{3.13}$$

over $\mathcal{X} \cap X'$. Then

(a) *$\hat{\Phi}(x,\xi)$ is Lipschitz continuous w.r.t. x over $\mathcal{X} \cap X'$ for all $\xi \in \Xi$.*

(b) *$\mathbb{E}[\hat{\Phi}(x,\xi)]$ is Lipschitz continuous w.r.t. x over $\mathcal{X} \cap X'$.*

Proof. Part (a). By the compactness of X' and Theorem 3.6 (a), $Y(\xi)$ is compact for almost all $\xi \in \Xi$. By the continuity of $\nabla\Theta(x,\hat{y}(x,\xi),\xi)$, we have

$$A(x,\hat{y}(x,\xi),\xi) - B(x,\hat{y}(x,\xi),\xi)M(x,\hat{y}(x,\xi),\xi)^{-1}L(x,\hat{y}(x,\xi),\xi)$$

is continuous over X'. Then we have

$$\sup_{x\in X'} \|A(x,\hat{y}(x,\xi),\xi) - B(x,\hat{y}(x,\xi),\xi)M(x,\hat{y}(x,\xi),\xi)^{-1}L(x,\hat{y}(x,\xi),\xi)\| < +\infty.$$

Moreover, by Theorem 3.6 (b) and (3.11), the Lipschitz module of $\hat{\Phi}(x,\xi)$, denote by $\mathrm{lip}_\Phi(\xi)$ satisfies

$$\begin{aligned} & \mathrm{lip}_\Phi(\xi) \\ \leq\ & \sup_{x\in X'} \|A(x,\hat{y}(x,\xi),\xi) - B(x,\hat{y}(x,\xi),\xi)M(x,\hat{y}(x,\xi),\xi)^{-1}L(x,\hat{y}(x,\xi),\xi)\| \\ <\ & +\infty. \end{aligned}$$

Part (b). it comes from Part (a) and (3.13) directly. ∎

3.2. Existence, uniqueness and CD-regularity of the solutions. Consider the mixed SVI-NCP (3.1)-(3.2) and its one stage reformulation (3.12). If we replace Assumption 3.1 by the following assumption, we can have stronger results.

ASSUMPTION 3.2. *For a.e. $\xi \in \Xi$, $\Theta(x, y(\xi), \xi)$ is strongly monotone with parameter $\kappa(\xi)$ at $(x, y(\cdot)) \in C \times \mathcal{Y}$, where $\mathbb{E}[\kappa(\xi)] < +\infty$.*

Note that Assumption 3.1 can be implied by Assumption 3.2 over $C \times \mathcal{Y}$.

THEOREM 3.8. *Suppose Assumption 3.2 holds over $C \times \mathcal{Y}$ and $\Phi(x, y, \xi)$ and $\Psi(x, y, \xi)$ are continuously differentiable w.r.t. (x, y) for a.e. $\xi \in \Xi$. Then*

(a) *$\mathcal{G} : C \times \mathcal{Y} \to C \times \mathcal{Y}$ is strongly monotone and hemicontinuous.*
(b) *For all x and almost all $\xi \in \Xi$, $\Psi(x, y(\xi), \xi)$ is strongly monotone and continuous w.r.t. $y(\xi) \in \mathbb{R}^m$.*
(c) *The two-stage SVI-NCP (3.1)-(3.2) has a unique solution.*
(d) *The two-stage SVI-NCP (3.1)-(3.2) has relatively complete recourse, that is for all x and almost all $\xi \in \Xi$, the NCP (3.2) has a unique solution.*

Proof. Parts (a) and (b) come from Assumption 3.2 over $C \times \mathcal{Y}$ directly. Since the strong monotonicity of $\mathcal{G}$ and Ψ implies the coerciveness of $\mathcal{G}$ and Ψ, see [12, Chapter 12], by [12, Theorem 12.2 and Lemma 12.2], we have Part (c) and Part (d). ∎

With the results in sections 3.1 and above, we have the following theorem by only assume that Assumption 3.2 holds in a neighborhood of $\mathrm{Sol}^* \cap X' \times \mathcal{Y}$. Our result extends [3, Proposition 2.1] for two-stage SLCP .

THEOREM 3.9. *Let Sol^* be the solution set of the mixed SVI-NCP (3.1)-(3.2). Suppose (i) there exists a compact set X' such that $\mathrm{Sol}^* \cap X' \times \mathcal{Y}$ is nonempty, (ii) Assumption 3.2 holds over $\mathrm{Sol}^* \cap X' \times \mathcal{Y}$ and (iii) the conditions of Theorem 3.7 hold. Then*

(a) *For any $(x, y(\cdot)) \in \mathrm{Sol}^*$, every matrix in $\partial\hat{\Phi}(x)$ is positive definite and $\hat{\Phi}$ and ϕ are strongly monotone at x.*
(b) *Any solution $x^* \in \mathcal{S}^* \cap X'$ of SVI (3.12) is CD-regular and an isolate solution.*
(c) *Moreover, if replacing conditions (i) and (ii) by supposing (iv) Assumption 3.2 holds over $\mathbb{R}^n \times \mathcal{Y}$, then SVI (3.12) has a unique solution x^* and the solution is CD-regular.*

Proof. Part (a). Note that under Assumption 3.2, for any $(x, y(\cdot)) \in \mathrm{Sol}^*$, the matrix

$$\begin{pmatrix} A(x,y(\xi),\xi) & B(x,y(\xi),\xi) \\ L(x,y(\xi),\xi) & M(x,y(\xi),\xi) \end{pmatrix} \succ 0.$$

From (ii) of Lemma 2.1 in [3], we have

$$A(x,y(\xi),\xi) - B(x,y(\xi),\xi)U_J(M(x,y(\xi),\xi))L(x,y(\xi),\xi) \succ 0, \ \ \forall J \in \mathcal{J}.$$

For any $\bar{x}$ such that $(\bar{x}, \bar{y}(\cdot)) \in \mathrm{Sol}^*$, let $\mathcal{B}_\delta(\bar{x})$ be a small neighborhood of $\bar{x}$,

$$\mathcal{D}_{\hat{y}}(\bar{x}) := \{x' : x' \in \mathcal{B}_\delta(\bar{x}),\ \hat{y}(x',\xi) \text{ is F-differentiable w.r.t. } x \text{ at } x'\}$$

and

$$\mathcal{D}_{\hat{\Phi}}(\bar{x}) := \{x' : x' \in \mathcal{B}_\delta(\bar{x}),\ \hat{\Phi}(x', \xi) \text{ is F-differentiable w.r.t. } x \text{ at } x'\}.$$

Since $\Phi(x, y, \xi)$ is continuously differentiable w.r.t. (x, y), $\hat{y}(\cdot, \xi)$ is F-differentiable w.r.t. x, which implies $\hat{\Phi}(\cdot, \xi)$ is F-differentiable w.r.t. x. Then $\mathcal{D}_{\hat{y}}(\bar{x}) \subseteq \mathcal{D}_{\hat{\Phi}}(\bar{x})$. Moreover, since $\hat{y}(x, \xi)$ and $\hat{\Phi}(x, \xi)$ are Lipschitz continuous w.r.t. x over $\mathcal{B}_\delta(\bar{x})$, they are F-differentiable almost everywhere over $\mathcal{B}_\delta(\bar{x})$. Then the measure of $\mathcal{D}_{\hat{\Phi}}(\bar{x}) \backslash \mathcal{D}_{\hat{y}}(\bar{x})$ is zero. By Theorem 3.6 (b), (3.11) and the definition of Clarke generalized Jacobian, we have

(3.14)

$$\begin{aligned}
& \partial_x \hat{\Phi}(\bar{x}, \xi) \\
= \ & \operatorname{conv}\left\{ \lim_{x' \to \bar{x}} \nabla_x \hat{\Phi}(x', \xi) : x' \in \mathcal{D}_{\hat{\Phi}}(\bar{x}) \right\} \\
= \ & \operatorname{conv}\left\{ \lim_{x' \to \bar{x}} \nabla_x \Phi(x', \hat{y}(x', \xi), \xi) + \nabla_y \Phi(x', \hat{y}(x', \xi), \xi) \nabla_x \hat{y}(x', \xi) : x' \in \mathcal{D}_{\hat{y}}(\bar{x}) \right\} \\
= \ & \operatorname{conv}\left\{ \lim_{x' \to \bar{x}} A(x', \hat{y}(x', \xi), \xi) \right. \\
& \left. - B(x', \hat{y}(x', \xi), \xi) U_{\alpha(\hat{y}(x', \xi))}(M(x', \hat{y}(x', \xi), \xi)) L(x', \hat{y}(x', \xi), \xi) : x' \in \mathcal{D}_{\hat{y}}(\bar{x}) \right\} \\
\subset \ & \operatorname{conv}\{A(x, \hat{y}(x, \xi), \xi) \\
& - B(x, \hat{y}(x, \xi), \xi) U_J(M(x, \hat{y}(x, \xi), \xi)) L(x, \hat{y}(x, \xi), \xi) : J \in \mathcal{J}\},
\end{aligned}$$

where the second equation is from [29, Theorem 4] and the fact that the measure of $\mathcal{D}_{\hat{\Phi}}(\bar{x}) \backslash \mathcal{D}_{\hat{y}}(\bar{x})$ is 0. By (3.14), every matrix in $\partial_x \hat{\Phi}(\bar{x}, \xi)$ is positive definite. And then $\hat{\Phi}$ is strongly monotone which implies ϕ is strongly monotone at $\bar{x}$.

Part (b). By Corollary 3.1, the linearized SVI

$$0 \in V_{x^*}(x - x^*) + \mathbb{E}[\hat{\Phi}(x^*, \xi)] + \mathcal{N}_C(x),$$

is strongly regular for all $V_{x^*} \in \partial \phi(x^*) \subseteq \mathbb{E}[\partial_x \hat{\Phi}(x^*, \xi)]$. Then the NCP (3.12) at x^* is CD-regular. Moreover, by the definition of CD regular, x^* is a unique solution of the NCP (3.12) over a neighborhood of x^*.

Part (c). By Part (a) and Theorem 3.8, NCP (3.12) has a unique solution x^*. The CD regular of NCP (3.12) at x^* follows from Part (b). $\square$

3.3. Convergence analysis of the SAA two-stage SVI-NCP.

Consider the two-stage SVI-NCP (3.1)-(3.2) and its SAA problem (3.3)-(3.4).

We discuss the existence and uniqueness of the solutions of SAA two-stage SVI (3.3)-(3.4) under Assumption 3.2 over $C \times \mathcal{Y}$ firstly. Define

$$\mathcal{G}_N(x, y(\cdot)) := \begin{pmatrix} N^{-1} \sum_{j=1}^N \Phi(x, y(\xi^j), \xi^j) \\ \Psi(x, y(\xi^1), \xi^1) \\ \vdots \\ \Psi(x, y(\xi^N), \xi^N) \end{pmatrix}.$$

THEOREM 3.10. *Suppose Assumption 3.2 holds over $C \times \mathcal{Y}$ and $\Phi(x, y, \xi)$ and $\Psi(x, y, \xi)$ are continuously differentiable w.r.t. (x, y) for a.e. $\xi \in \Xi$. Then*

(a) *$\mathcal{G}_N : C \times \mathcal{Y} \to C \times \mathcal{Y}$ is strongly monotone with $N^{-1} \sum_{j=1}^N \kappa(\xi^j)$ and hemi-continuous.*

(b) *The SAA two-stage SVI (3.3)-(3.4) has a unique solution.*

Proof. By Assumption 3.2, we have Parts (a) and (b). $\square$

Then we investigate the almost sure convergence and convergence rate of the first stage solution $\bar{x}_N$ of (3.3)-(3.4) to optimal solutions of the true problem by only supposing Assumption 3.2 holds at a neighborhood of $\mathrm{Sol}^* \cap X' \times \mathcal{Y}$.

Note that the normal cone multifunction $x \mapsto \mathcal{N}_C(x)$ is closed. Note also that function $\hat{\Phi}(x,\xi) = \Phi(x, \hat{y}(x,\xi), \xi)$, where $\hat{y}(x,\xi)$ is a solution of the second stage problem (3.2). Then the first stage of SAA problem with second stage solution can be written as

$$0 \in N^{-1} \sum_{j=1}^{N} \hat{\Phi}(x, \xi^j) + \mathcal{N}_C(x). \tag{3.15}$$

Under the conditions (i)-(iii) of Theorem 3.9, the two-stage SVI-NCP (3.1)-(3.2) and its SAA problem (3.3)-(3.4) satisfy conditions of Theorem 2.4 and with $\mathcal{R}^{-1}(t) \le \frac{t}{c}$ for some positive number c (by Remark 2.1, the strongly monotone of ϕ and the argument in the proof of Part (b), Theorem 2.8). Then Theorem 2.4 can be applied directly.

Definition 3.11. [10, 20] *A solution x^* of the SVI (3.12) is said to be strongly stable if for every open neighborhood $\mathcal{V}$ of x^* such that $\mathrm{SOL}(C,\phi) \cap \mathrm{cl}\mathcal{V} = \{x^*\}$, there exist two positive scalars δ and ϵ such that for every continuous function $\tilde{\phi}$ satisfying*

$$\sup_{x \in C \cap \mathrm{cl}\mathcal{V}} \|\tilde{\phi}(x) - \phi(x)\| \le \epsilon,$$

the set $\mathrm{SOL}(C,\tilde{\phi}) \cap \mathcal{V}$ is a singleton; moreover, for another continuous function $\bar{\phi}$ satisfying the same condition as $\tilde{\phi}$, it holds that

$$\|x - x'\| \le \delta \|[\phi(x) - \tilde{\phi}(x)] - [\phi(x') - \bar{\phi}(x')]\|,$$

where x and x' are elements in the sets $\mathrm{SOL}(C,\tilde{\phi}) \cap \mathcal{V}$ and $\mathrm{SOL}(C,\bar{\phi}) \cap \mathcal{V}$, respectively.

Theorem 3.12. *Suppose conditions (i)-(iii) of Theorem 3.9 hold. Let x^* be a solution of the SVI (3.12) and X' be a compact set such that $x^* \in \mathrm{int}(X')$. Assume there exists $\varepsilon > 0$ such that for N sufficiently large,*

$$x^* \notin \mathrm{cl}(\mathrm{bd}(\mathcal{X}) \cap \mathrm{int}(\bar{\mathcal{X}}_N \cap X')). \tag{3.16}$$

Then there exist a solution $\hat{x}_N$ of the SAA problem (3.15) and a positive scalar δ such that $\|\hat{x}_N - x^*\| \to 0$ as $N \to \infty$ w.p.1 and for N sufficiently large w.p.1

$$\|\hat{x}_N - x^*\| \le \delta \sup_{x \in \mathcal{X} \cap X'} \|\hat{\phi}_N(x) - \phi(x)\|. \tag{3.17}$$

Proof. By Theorem 3.9 (b), the SVI (3.12) at x^* is CD-regular. By [20, Theorem 3] and [10], x^* is a strong stable solution of the SVI (3.12). Note that by Theorem 3.9 (a) and [27, Theorem 7.48], we have

$$\sup_{x \in \mathcal{X} \cap X'} \|\hat{\phi}_N(x) - \phi(x)\|$$

converges to 0 uniformly. Then by Definition 3.11 and (3.16), there exist two positive scalars δ, ϵ such that for N sufficiently large, w.p.1

$$\sup_{x \in \mathcal{X} \cap X'} \|\hat{\phi}_N(x) - \phi(x)\| \le \min\{\epsilon, \varepsilon/\delta\}$$

and

$$\|\hat{x}_N - x^*\| \le \delta \sup_{x\in\mathcal{X}\cap X'} \|\hat{\phi}_N(x) - \phi(x)\|,$$

which implies $\hat{x}_N \in \mathcal{X}$. ∎

Note that Theorem 3.12 guarantees that $\mathcal{R}^{-1}(t) \le \delta t$ and condition (3.16) is discussed after Theorem 2.8. Note also that replacing conditions (i) - (ii) and condition (3.16) by supposing condition (iv) of Theorem 3.9, conclusion (3.17) also holds. Moreover, in this case, by Theorem 3.9 (c) and Theorem 3.10, x^* and $\hat{x}_N$ are the unique solutions of the SVI (3.12) and its SAA problem (3.15) respectively.

Then we consider the exponential rate of convergence. Note that under Assumption 3.1, for SAA problem of the mixed two-stage SVI-NCP (3.3)-(3.4), Assumptions 2.1, 2.4, 2.5 and condition (iii) in Theorem 2.8 hold. If we replace Assumption 3.1 by Assumption 3.2 over $\mathrm{Sol}^* \cap X' \times \mathcal{Y}$, we have the following theorem.

THEOREM 3.13. *Let $X' \subset C$ be a convex compact subset such that $\mathcal{B}_\delta(x^*) \subset X'$. Suppose the conditions in Theorem 3.12 and Assumption 2.6 hold. Then for any $\varepsilon > 0$ there exist positive constants $\delta > 0$ (independent of ε), $\varrho = \varrho(\varepsilon)$ and $\varsigma = \varsigma(\varepsilon)$, independent of N, such that*

$$\mathsf{Pr}\left\{\sup_{x\in\mathcal{X}} \left\|\hat{\phi}_N(x) - \phi(x)\right\| \ge \varepsilon\right\} \le \varrho(\varepsilon) e^{-N\varsigma(\varepsilon)}, \tag{3.18}$$

and

$$\mathsf{Pr}\left\{\|x_N - x^*\| \ge \varepsilon\right\} \le \varrho(\varepsilon/\delta) e^{-N\varsigma(\varepsilon/\delta)}. \tag{3.19}$$

Proof. By Theorem 3.9 (a), Assumption 2.6 and [27, Theorem 7.67], the conditions of Theorem 2.8 (a) hold and then (3.18) holds. Under condition (3.16) in Theorem 3.12, (3.19) follows from (3.17) and (3.18). ∎

The two-stage SVI-NCP is a class of important two-stage SGE and can cover a wide class of real world applications. Moreover, the structure of the second stage NCP has been well investigated in the literature (e.g., [5, 15]). By combining those results in our case we can formulate the Clarke generalized Jacobian of the solution function of the second stage NCP and derive stability analysis of the first stage SVI. We will consider the two-stage SVI in further research.

4. Examples. In this section, we illustrate our theoretical results in the last sections by a two-stage stochastic non-cooperative game of two players [3, 21]. Let $\xi : \Omega \to \Xi \subseteq \mathbb{R}^d$ be a random vector, $x_i \in \mathbb{R}^{n_i}$ and $y_i(\cdot) \in \mathcal{Y}_i$ be the strategy vectors and policies of the ith player at the first stage and second stage, respectively, where $\mathcal{Y}_i$ is a measurable function space from Ξ to $\mathbb{R}^{m_i}$, $i = 1, 2$, $n = n_1 + n_2$, $m = m_1 + m_2$. In this two-stage stochastic game, the ith player solves the following optimization problem:

$$\min_{x_i\in[a_i,b_i]} \ \theta_i(x_i, x_{-i}) + \mathbb{E}[\psi_i(x_i, x_{-i}, y_{-i}(\xi), \xi)], \tag{4.1}$$

where $\theta_i(x_i, x_{-i}) := \frac{1}{2}x_i^T H_i x_i + q_i^T x_i + x_i^T P_i x_{-i}$,

$$\psi_i(x_i, x_{-i}, y_{-i}(\xi), \xi) := \min_{y_i\in[l_i(\xi),u_i(\xi)]} \ \phi_i(y_i, x_i, x_{-i}, y_{-i}(\xi), \xi) \tag{4.2}$$

is the optimal value function of the recourse action y_i at the second stage with

$$\phi_i(y_i, x_i, x_{-i}, y_{-i}(\xi), \xi) = \frac{1}{2} y_i^\top Q_i(\xi) y_i + c_i(\xi)^\top y_i + \sum_{j=1}^{2} y_i^\top S_{ij}(\xi) x_j + y_i^\top O_i(\xi) y_{-i}(\xi),$$

$a_i, b_i \in \mathbb{R}^{n_i}$, $l_i, u_i : \Xi \to \mathbb{R}^{m_i}$ are vector valued measurable functions, $l_i(\xi) < u_i(\xi)$ for all $\xi \in \Xi$, H_i and $Q_i(\xi)$ are symmetric positive definite matrices for a.e $\xi \in \Xi$, $x = (x_1, x_2)$, $y(\cdot) = (y_1(\cdot), y_2(\cdot))$, $x_{-i} = x_{i'}$ and $y_{-i} = y_{i'}$, for $i' \neq i$. We use $y_i(\xi)$ to denote the unique solution of (4.2).

By [11, Theorem 5.3 and Corollary 5.4], $\psi_i(x_i, x_{-i}, y_{-i}(\xi), \xi)$ is continuously differentiable w.r.t. x_i and

$$\nabla_{x_i} \psi_i(x_i, x_{-i}, y_{-i}(\xi), \xi) = S_{ii}^T(\xi) y_i(\xi).$$

Hence the two-stage stochastic game can be formulated as a two-stage stochastic linear VI

$$\begin{array}{rcl} -\nabla_{x_i}\theta_i(x_i, x_{-i}) - \mathbb{E}[\nabla_{x_i}\psi_i(x_i, x_{-i}, y_{-i}(\xi), \xi)] & \in & \mathcal{N}_{[a_i, b_i]}(x), \\ -\nabla_{y_i(\xi)}\phi_i(y_i(\xi), x_i, x_{-i}, y_{-i}(\xi), \xi) & \in & \mathcal{N}_{[l_i(\xi), u_i(\xi)]}(y_i(\xi)), \\ & & \text{for a.e. } \xi \in \Xi, \end{array}$$

for $i = 1, 2$, with the following matrix-vector form

$$\begin{array}{rcll} -Ax - \mathbb{E}[B(\xi)y(\xi)] - h_1 & \in & \mathcal{N}_{[a,b]}(x) & \\ -M(\xi)y(\xi) - L(\xi)x - h_2(\xi) & \in & \mathcal{N}_{[l(\xi), u(\xi)]}(y(\xi)), & \text{for a.e. } \xi \in \Xi, \end{array} \tag{4.3}$$

where

$$A = \begin{pmatrix} H_1 & P_1 \\ P_2 & H_2 \end{pmatrix}, \quad B(\xi) = \begin{pmatrix} S_{11}^T(\xi) & 0 \\ 0 & S_{22}^T(\xi) \end{pmatrix},$$

$$L(\xi) = \begin{pmatrix} S_{11}(\xi) & S_{12}(\xi) \\ S_{21}(\xi) & S_{22}(\xi) \end{pmatrix}, \quad M(\xi) = \begin{pmatrix} Q_1(\xi) & O_1(\xi) \\ O_2(\xi) & Q_2(\xi) \end{pmatrix},$$

$h_1 = (q_1, q_2)$ and $h_2(\xi) = (c_1(\xi), c_2(\xi))$. Moreover, if there exists a positive continuous function $\kappa(\xi)$ such that $\mathbb{E}[\kappa(\xi)] < +\infty$ and for a.e. $\xi \in \Xi$,

$$\left(z^\top, u^\top\right) \begin{pmatrix} A & B(\xi) \\ L(\xi) & M(\xi) \end{pmatrix} \begin{pmatrix} z \\ u \end{pmatrix} \geq \kappa(\xi)(\|z\|^2 + \|u\|^2), \quad \forall z \in \mathbb{R}^n, \ u \in \mathbb{R}^m, \tag{4.4}$$

the two-stage box constrained SVI (4.3) satisfy Assumption 3.2. By the Schur complement condition for positive definiteness [13], a sufficient condition for (4.4) is

$$4H_2 - (P_1 + P_2^\top) H_1^{-1} (P_1 + P_2^\top) \quad \text{is positive definite}$$

and for some $k_1 > 0$ and a.e. $\xi \in \Xi$,

$$\lambda_{\min}(M(\xi) + M(\xi)^\top - (B(\xi) + L(\xi)^\top)(A + A^\top)^{-1}(B(\xi) + L(\xi)^\top)) \geq k_1 > 0,$$

where $\lambda_{\min}(V)$ is the smallest eigenvalue of $V \in \mathbb{R}^{m \times m}$.

Under condition (4.4), by Corollary 3.1 and Theorem 3.8, the conditions in Theorem 2.8 hold for (4.3). To see this, we only need to show condition (vi) of Theorem 2.8 holds for (4.3). Consider the second stage VI of (4.3) for fixed ξ and x, by the proof of [7, Lemma 2.1], we have

$$\hat{y}(x, \xi) - \hat{y}(x', \xi) = -(I - D(x, x', \xi) + D(x, x', \xi)M(\xi))^{-1} D(x, x', \xi) L(\xi)(x - x'),$$

which implies

$$\partial_x \hat{y}(x,\xi) \subseteq \{-(I - D + DM(\xi))^{-1}DL(\xi) : D \in \mathcal{D}_0\}, \tag{4.5}$$

where $D(x, x', \xi)$ is a diagonal matrix with diagonal elements

$$d_i = \begin{cases} 0, & \text{if } (\hat{y}_i(x,\xi))_i - z_i(x,\xi), (\hat{y}(x',\xi))_i - z_i(x',\xi) \in [u_i(\xi), \infty), \\ 0, & \text{if } (\hat{y}(x,\xi))_i - z_i(x,\xi), (\hat{y}(x',\xi))_i - z_i(x',\xi) \in (-\infty, l_i(\xi)], \\ 1, & \text{if } (\hat{y}(x,\xi))_i - z_i(x,\xi), (\hat{y}(x',\xi))_i - z_i(x',\xi) \in (l_i(\xi), u_i(\xi)), \\ \frac{(\hat{y}(x,\xi))_i - (\hat{y}(x',\xi))_i}{(\hat{y}(x,\xi))_i - z_i(x,\xi) - ((\hat{y}(x',\xi))_i - z_i(x',\xi)}, & \text{otherwise}, \end{cases}$$

$z_i(x,\xi) = (M(\xi)\hat{y}(x,\xi) + L(\xi)x + h_2(\xi))_i$, $d_i \in [0,1]$, $i = 1, \cdots, m$, $\mathcal{D}_0$ is a set of diagonal matrices in $\mathbb{R}^{m\times m}$ with the diagonal elements in $[0,1]$. Then we consider the one stage SVI with $\hat{y}(x,\xi)$ as follows

$$- Ax - \mathbb{E}[B(\xi)\hat{y}(x,\xi)] - h_1 \in \mathcal{N}_{[a,b]}(x). \tag{4.6}$$

By using the similar arguments as in the proof of Theorem 3.9 and (4.5), every element of the Clarke Jacobian of $Ax + \mathbb{E}[B(\xi)\hat{y}(x,\xi)] + h_1$ is a positive definite matrix. Then (4.6) is strong monotone and hence condition (vi) of Theorem 2.8 holds. In what follows, we verify the convergence results in Theorem 2.8 numerically.

Let $\{\xi^j\}_{i=1}^N$ be an iid sample of random variable ξ. Then the SAA problem of (4.3) is

$$\begin{array}{rcl} -Ax - \frac{1}{N}\sum_{j=1}^N B(\xi^j)y(\xi^j) - h_1 & \in & \mathcal{N}_{[a,b]}(x) \\ -M(\xi^j)y(\xi^j) - L(\xi^j)x - h_2(\xi^j) & \in & \mathcal{N}_{[l(\xi^j),u(\xi^j)]}(y(\xi^j)), \quad j = 1, \ldots, N. \end{array} \tag{4.7}$$

PHM converges to a solution of (4.7) if condition (4.4) holds.

ALGORITHM 4.1 (PHM). *Choose $r > 0$ and initial points $x^0 \in \mathbb{R}^n$, $x_j^0 = x^0 \in \mathbb{R}^n$, $y_j^0 \in \mathbb{R}^m$ and $w_j^0 \in \mathbb{R}^n$, $j = 1, \cdots, N$ such that $\frac{1}{N}\sum_{j=1}^N w_j^0 = 0$. Let $\nu = 0$.*

Step 1. *For $j = 1, \cdots, N$, solve the box constrained VI*

$$\begin{array}{rcl} -Ax_j - B(\xi^j)y_j - h_1 - w_j^\nu - r(x_j - x_j^\nu) & \in & \mathcal{N}_{[a,b]}(x_j), \\ -M(\xi^j)y_j - L(\xi^j)x_j - h_2(\xi^j) - r(y_j - y_j^\nu) & \in & \mathcal{N}_{[l(\xi^j),u(\xi^j)]}(y_j), \end{array} \tag{4.8}$$

and obtain a solution $(\hat{x}_j^\nu, \hat{y}_j^\nu)$, $j = 1, \cdots, N$.

Step 2. *Let $\bar{x}^{\nu+1} = \frac{1}{N}\sum_{j=1}^N \hat{x}_j^\nu$. For $j = 1, \cdots, N$, set*

$$x_j^{\nu+1} = \bar{x}^{\nu+1}, \quad y_j^{\nu+1} = \hat{y}_j^\nu, \quad w_j^{\nu+1} = w_j^\nu + r(\hat{x}_j^\nu - x_j^{\nu+1}).$$

Note that PHM is well-defined if $\begin{pmatrix} A & B(\xi^j) \\ L(\xi^j) & M(\xi^j) \end{pmatrix}$, $j = 1, \cdots, N$ are positive semi-definite, that is, (4.8) has a unique solution for each j, even for some x and ξ^j the second stage problem

$$-M(\xi^j)y - L(\xi^j)x - h_2(\xi^j) \in \mathcal{N}_{[l(\xi^j),u(\xi^j)]}(y)$$

has no solution.

4.1. Generation of matrices satisfying condition (4.4). We generate matrices A, $B(\xi), L(\xi), M(\xi)$ by the following procedure. Randomly generate a symmetric positive definite matrix $H_1 \in \mathbb{R}^{n_1\times n_1}$, matrices $P_1 \in \mathbb{R}^{n_1\times n_2}, P_2 \in \mathbb{R}^{n_2\times n_1}$. Set $H_2 = \frac{1}{4}(P_1^\top + P_2)H_1^{-1}(P_1 + P_2^\top) + \alpha I_{n_2}$, where α is a positive number. Randomly generate matrices with entries within $[-1, 1]$:

$$\bar S_{11} \in \mathbb{R}^{m_1\times n_1}, \quad \bar S_{12} \in \mathbb{R}^{m_1\times n_2}, \quad \bar S_{21} \in \mathbb{R}^{m_2\times n_1},$$

$$\bar S_{22} \in \mathbb{R}^{m_2\times n_2}, \quad \bar O_1 \in \mathbb{R}^{m_1\times m_2}, \quad \bar O_2 \in \mathbb{R}^{m_2\times m_1}.$$

Randomly generate two symmetric matrices $\bar Q_1 \in \mathbb{R}^{m_1\times m_1}$ and $\bar Q_2 \in \mathbb{R}^{m_2\times m_2}$ whose diagonal entries are greater than $m - 1 + \alpha$, off-diagonal entries are in $[-1, 1]$, respectively.

Generate an iid sample $\{\xi^j\}_{j=1}^N \subset [0, 1]^{10} \times [-1, 1]^{10}$ of random variable $\xi \in \mathbb{R}^{20}$ following uniformly distribution over $\Xi = [0, 1]^{10} \times [-1, 1]^{10}$. Set

$$S_{11}(\xi) = \xi_1^j \bar S_{11},\ S_{12}(\xi) = \xi_2^j \bar S_{12},\ S_{21}(\xi) = \xi_3^j \bar S_{21},$$

$$S_{22}(\xi) = \xi_4^j \bar S_{22},\ O_1(\xi) = \xi_5^j \bar O_1,\ O_2(\xi) = \xi_6^j \bar O_2,$$

$$Q_1(\xi) = \bar Q_1 + (\xi_7^j + \frac{(n+m)^2}{\lambda_{\min}(A + A^T)})I_{m_1} \quad Q_2(\xi) = \bar Q_2 + (\xi_8^j + \frac{(n+m)^2}{\lambda_{\min}(A + A^T)})I_{m_2}.$$

Set $B(\xi^j), L(\xi^j), M(\xi^j)$ as in (4.3).

The matrices generated by this procedure satisfy condition (4.4). Indeed, since H_1 and $4H_2 - (P_1 + P_2^T)H_1^{-1}(P_1 + P_2^T)$ are positive definite, by the Schur complement condition for positive definiteness [13], $A + A^T$ is symmetric positive definite, and thus A is positive definite. Moreover, since the matrix $\bar M := \begin{pmatrix} \bar Q_1 & \bar O_1 \\ \bar O_2 & \bar Q_2 \end{pmatrix}$ is diagonal dominance with positive diagonal entries $\bar M_{ii} \geq m - 1 + \alpha$, it is positive definite and the eigenvalues $M + M^T$ are greater than 2α. Hence, for any $y \in \mathbb{R}^m$, we have

$$\begin{aligned} & y^T(M(\xi) + M(\xi)^T - (B(\xi)^T + L(\xi))(A + A^T)^{-1}(B(\xi) + L(\xi)^T))y \\ \geq{} & (2\alpha + \frac{(n+m)^2}{\lambda_{\min}(A + A^T)})\|y\|^2 - \frac{1}{\lambda_{\min}(A + A^T)}\|(B(\xi)^T + L(\xi))\|^2\|y\|^2 \geq 2\alpha\|y\|^2, \end{aligned}$$

where we use $\|B(\xi)^T + L(\xi)\|^2 \leq \|B(\xi)^T + L(\xi)\|_1^2 \leq (m + n)^2$. Using the Schur complement condition for positive definiteness [13] again, we obtain condition (4.4).

Finally, we generate the box constraints, h_1 and $h_2(\cdot)$. For the first stage, the lower bound is set as $a = 0\mathbf{1}_n$, and the upper bound of the box constraints b is randomly generated from $[1, 50]^6$. For the second stage, we set $l(\xi) = (1 + \xi_9)\bar l$ and $u(\xi) = (1 + \xi_{10})\bar u$, where $\mathbf{1}_n \in \mathbb{R}^n$ is a vector with all elements 1, $\bar l$ is randomly generated from $[0, 1]^{10}$ and $\bar u$ is randomly generated from $[3, 50]^{10}$. Moreover, the vector h_1 is randomly generated from $[-5, 5]^6$ and $h_2(\xi) = (\xi_{11}, \cdots, \xi_{20})$ is a random vector following uniform distribution over $[-1, 1]^{10}$.

4.2. Numerical results. For each sample size of $N = 10, 50, 250, 1250, 2250$, we randomly generate 20 test problems and solve the box-constrained VI in Step 1 of PHM by the homotopy-smoothing method [6]. We stop the iteration when

$$\mathbf{res} := \|x - \mathrm{mid}(x - Ax - \frac{1}{N}\sum_{j=1}^N B(\xi^j)\hat y(x, \xi^j) - h_1, a, b)\| \leq 10^{-5}, \tag{4.9}$$

or the iterations reach 5000, where mid$(\cdot)$ denotes the componentwise median operator, $\hat{y}(x, \xi^j)$ is the solution of the second stage box constrained VI with x and ξ^j.

Parameters for the numerical tests are chosen as follows: $n_1 = n_2 = 3, m_1 = m_2 = 5, \alpha = 1$ and maximize iteration number is 5000.

Figures 1 shows the convergence tendency of x_1, x_2, x_3, x_4, x_5 and x_6 respectively. Note that since we use the homotopy-smoothing method to solve the box-constrained VI in Step 1 of PHM and the stop criterion is 10^{-5}, x_2 is not always feasible. However, $[a_i - x_i]_+ + [x_i - b_i]_+ \leq 10^{-5}$, $i = 1, \ldots, 6$, which is related to the stopping criterion of the homotopy-smoothing method.

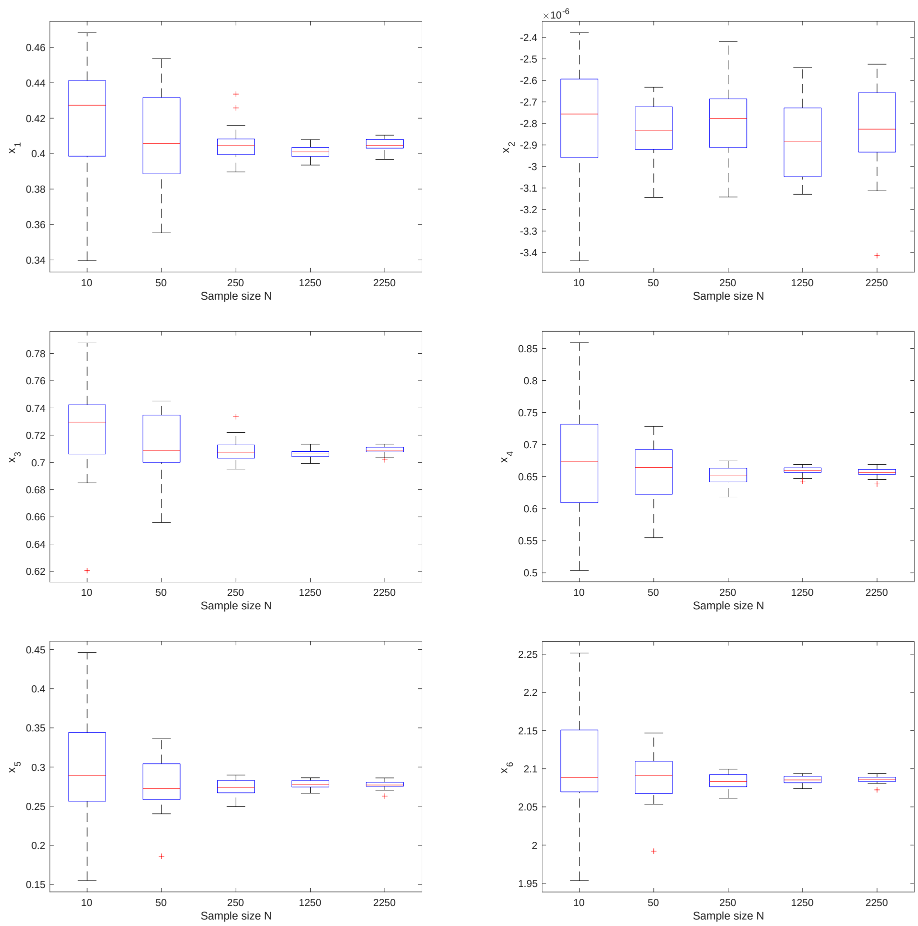

FIG. 2. *Convergence of* x_1 - x_6

We use $x^{N_t,j}$ $j = 1, \ldots, 3000, t = 1, \ldots, 5$ to denote the computed solutions with sample size N_t for the j-th test problem shown in Figure 1. Then we compute the mean, variance and 95% confidence interval (CI) of the corresponding **res** defined in (4.9) with $x = x^{N_t,j}$ by using a new set of 20 randomly generated test problems with sample size $N = 3000$ for computing $\hat{y}(x^{N_t,j}, \xi^j), j = 1, \ldots, 3000, t = 1, \ldots, 5$. We can see that the average of the mean, variance and width of 95% CI of **res** in Table 1 decrease as the sample size increases.

	$N_1 = 10$	$N_2 = 50$	$N_3 = 250$	$N_4 = 1250$	$N_5 = 2250$
mean	0.22449	0.13753	0.04820	0.02885	0.02500
variance	0.01984	0.00605	0.00118	0.00023	0.00016
95% CI	[0.2158, 0.2332]	[0.1349, 0.1402]	[0.0477, 0.0487]	[0.0287, 0.0290]	[0.0249, 0.0251]

TABLE 1
Mean, variance and 95% confidence interval (CI) of **res**

5. Conclusion remarks. Without assuming *relatively complete recourse*, we prove the convergence of the SAA problem (1.6)-(1.7) of the two-stage SGE (1.1)–(1.2) in Theorem 2.4, and show the exponential rate of the convergence in Theorem 2.9. When the two-stage SGE (1.1)–(1.2) has relatively complete recourse, Assumption 2.3, conditions (v)-(vi) in Theorem 2.4 and condition (iv) in Theorem 2.8 hold.

In section 3, we present sufficient conditions for the existence, uniqueness, continuity and regularity of solutions of the two-stage SVI-NCP (3.1)–(3.2) by using the perturbed linearization of functions Φ and Ψ and then show the almost sure convergence and exponential convergence of its SAA problem (3.3)-(3.4). Numerical examples in section 4 satisfy all conditions of Theorem 2.8 and we show the convergence of SAA method numerically.

Acknowledgments. We are grateful to the anonymous referees for their constructive comments which helped us to improve presentation of the paper.

REFERENCES

[1] J. Birge and F. Louveaux, *Introduction to Stochastic Programming*, Springer, 1997.
[2] X. Chen, T.K. Pong and R. B-J. Wets, Two-stage stochastic variational inequalities: an ERM-solution procedure, *Math. Program.,* 165 (2017), pp. 71-111.
[3] X. Chen, H. Sun and H. Xu, Discrete approximation of two-stage stochastic and distributionally robust linear complementarity problems, *Math. Program.,* (2018), https://doi.org/10.1007/s10107-018-1266-4.
[4] X. Chen, R. B-J. Wets and Y. Zhang, Stochastic variational inequalities: residual minimization smoothing/sample average approximations, *SIAM J. Optim.,* 22 (2012), pp. 649-673.
[5] X. Chen and S. Xiang, Newton iterations in implicit time-stepping scheme for differential linear complementarity systems, *Math. Program.,* 138 (2013), pp. 579-606.
[6] X. Chen and Y. Ye, On homotopy-smoothing methods for box-constrained variational inequalities, *SIAM J. Control Optim.*, 37 (1999), pp. 589-616.
[7] X. Chen and Z. Wang, Computational error bounds for a differential linear variational inequality, *IMA J. Numer. Anal.*, 32 (2012), pp. 957-982.
[8] A.L. Dontchev and R.T. Rockafellar, Characterizations of strong regularity for variational inequalities over polyhedral convex sets, *SIAM J. Optim.*, 6 (1996), pp. 1087-1105.
[9] A.L. Dontchev and R.T. Rockafellar, Newton's method for generalized equations: a sequential implicit function theorem, *Math. Program.,* 123 (2010), pp. 139-159.
[10] F. Facchinei and J-S Pang, Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer-Verlag, New York, 2003.
[11] J. Gauvin and F. Dubeau, Differential properties of the marginal function in mathematical programming, *Math. Program. Stud.,* 19 (1982), pp. 101-119.
[12] N. Hadjisavvas, S. Komlósi and S. Schaible, *Handbook of Generalized Convexity and Generalized Monotonicity*, Springer, New York, 2005.
[13] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, 21st printing, 2007.
[14] A.F. Izmailov, Strongly regular nonsmooth generalized equations, *Math. Program.,* 147 (2014), pp. 581-590.
[15] J. Kyparisis, Solution differentiability for variational inequalities, *Math. Program.,* 48 (1990), pp. 285-301.
[16] A. J. King and R. T. Rockafellar, Sensitivity analysis for nonsmooth generalized equations, *Math. Program.*, 55 (1992), pp. 191-212.
[17] B. Kummer, Generalized equations: solvability and regularity, *Math. Program. Stud.*, 21 (1984),

pp. 199-212.

[18] M. Lamm, S. Lu and A. Budhiraja, Individual confidence intervals for solutions to expected value formulations of stochastic variational inequalities, *Math. Program.,* 165 (2016), pp. 1-46.

[19] Y. Liu, W. Röemish and H. Xu, Quantitative stability analysis of stochastic generalized equations, *SIAM J. Optim.*, 24 (2014), pp. 467-497.

[20] J-S Pang, D. Sun and J. Sun, Semismooth homeomorphisms and strong stability of semidefinite and Lorentz cone complementarity problems, *Math. Oper. Res.*, 28 (2003), pp. 39-63.

[21] J.-S. Pang, S. Sen and U. Shanbhag, Two-stage non-cooperative games with risk-averse players, *Math. Program.*, 165 (2017), pp. 235-290.

[22] S. M., Robinson, Strongly regular generalized equations, *Math. Oper. Res.*, 5 (1980), pp. 43-62.

[23] R.T. Rockafellar and J. Sun, Solving monotone stochastic variational inequalities and complementarity problems by progressive hedging, *Math. Program.*, (2018), https://doi.org/10.1007/s10107-018-1251-y.

[24] R.T. Rockafellar and R. B-J. Wets, *Variational Analysis*, Springer-Verlag, Berlin Heidelberg, 1998.

[25] R.T. Rockafellar and R. B-J. Wets, Stochastic variational inequalities: single-stage to multistage, *Math. Program.*, 165 (2017), pp. 331-360.

[26] A. Shapiro, Sensitivity analysis of generalized equations, *J. Math. Sci.*, 115 (2003), pp. 2554-2565.

[27] A. Shapiro, D. Dentcheva, and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*, SIAM, Philadelphia, 2009.

[28] A. Shapiro and H. Xu, Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation, *Optim.*, 57 (2008), pp. 395-418.

[29] J. Warga, Fat homeomorphisms and unbounded derivate containers, *J. Math. Anal. Appl.* 81 (1981), pp. 545-560.

[30] H. Xu, Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming, *J. Math. Anal. Appl.*, 368 (2010), pp. 692-710.