MATHEMATICS OF OPERATIONS RESEARCH
Vol. 32, No. 3, August 2007, pp. 648–668
ISSN 0364-765X | EISSN 1526-5471 | 07 | 3203 | 0648

DOI 10.1287/moor.1070.0260

Convergence Analysis of Sample Average Approximation Methods for a Class of Stochastic Mathematical Programs with Equality Constraints

Huifu Xu
School of Mathematics, University of Southampton, Highfield S017 1BJ, Southampton, United Kingdom,
h.xu@soton.ac.uk, http://www.maths.soton.ac.uk/staff/Xu/

Fanwen Meng
The Logistics Institute—Asia Pacific, National University of Singapore, 7 Engineering Drive 1, Singapore 117543,
tlimf@nus.edu.sg

In this paper we discuss the sample average approximation (SAA) method for a class of stochastic programs with nonsmooth equality constraints. We derive a uniform Strong Law of Large Numbers for random compact set-valued mappings and use it to investigate the convergence of Karush-Kuhn-Tucker points of SAA programs as the sample size increases. We also study the exponential convergence of global minimizers of the SAA problems to their counterparts of the true problem. The convergence analysis is extended to a smoothed SAA program. Finally, we apply the established results to a class of stochastic mathematical programs with complementarity constraints and report some preliminary numerical test results.

Key words: sample average approximations; strong law of large numbers; random set-valued mappings; stationary points
MSC2000 subject classification: Primary: 90C15, 90C31; secondary: 90C33, 90C30
OR/MS subject classification: Primary: probability, stochastic model; secondary: programming, nondifferentiable
History: Received April 11, 2005; revised December 11, 2005 and May 15, 2006.

1. Introduction. In this paper, we consider the following stochastic minimization program

$$\begin{aligned} &\min \ \mathbb{E}[f(x, y(x, \xi(\omega)), \xi(\omega))] \\ &\text{s.t.} \ \ x \in \mathscr{X}. \end{aligned} \tag{1}$$

Here $f: \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$ is a continuously differentiable real-valued function, $\xi: \Omega \to \Xi \subset \mathbb{R}^k$ is a random vector defined on probability space $(\Omega, \mathscr{F}, P)$, $\mathbb{E}$ denotes the mathematical expectation, $\mathscr{X}$ is the feasible set of decision variate x, which is a nonempty subset of $\mathbb{R}^m$, and $y(x, \xi(\omega))$ is a solution of the following system of equations

$$H(x, y, \xi(\omega)) = 0, \quad \text{for } x \in \mathscr{X}, \ \ \omega \in \Omega, \tag{2}$$

where $H: \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$ is a locally Lipschitz continuous vector-valued function. We restrict our discussion to the case when H is piecewise smooth because it represents a large class of locally Lipschitz continuous functions that cover most practical problems (Scholtes [29]). Throughout this paper, we assume that the probability measure P of our considered space $(\Omega, \mathscr{F}, P)$ is nonatomic and $\mathbb{E}[f(x, y(x, \xi(\omega)), \xi(\omega)]$ is well defined for every $x \in \mathscr{X}$. To ease the notation, we will write $\xi(\omega)$ as ξ, and this should be distinguished from ξ being a deterministic vector of Ξ in a context.

Problem (1) is essentially a *here and now* stochastic optimization model where a decision has to be made before the realization of uncertainty, and an optimal decision is chosen to minimize the objective on stochastic average. The key characteristic of model (1) is that $y(x, \xi)$ is an implicit function defined by a system of nonsmooth equations (2). If we can solve (2) and obtain an explicit expression for $y(x, \xi)$, then (1) reduces to an ordinary stochastic minimization problem that has been extensively investigated (Birge and Louveaux [5], Ruszczyński and Shapiro [27]).

In this paper, we assume that $y(x, \xi)$ may not necessarily be analytically obtainable by solving (2). We also assume that (2) has a unique solution for every $(x, \xi) \in \mathscr{X} \times \Xi$. In the case when the equation has multiple solutions, we may employ a regularization scheme to approximate (2) with a regularized equation that has a unique solution. See our follow-up work (Meng and Xu [17]) for details. An interesting application of this abstract formulation is that (2) may be reformulated from a variational inequality or a complementarity problem that describes a system equilibrium parameterized by x and ξ, and consequently $y(x, \xi)$ represents the equilibrium solution of the system. This type of problem is known as stochastic mathematical programs with equilibrium constraints (SMPEC). See Patriksson and Wynter [18], Lin et al. [15], Xu [36, 37], Shapiro [31], and Birbil et al. [4] for recent discussions.

This paper is concerned with numerical methods for solving (1). Throughout the paper, we assume that ξ has either a continuous or a discrete distribution. We also assume that the explicit expression of the distribution of ξ is not obtainable, but it can be estimated from past data. In other words, we assume that a sample of ξ can be obtained from time to time. We tackle (1) with the well-known Monte Carlo simulation-based method. The basic idea of such a method is to generate an independent identically distributed (i.i.d.) sample $\xi^1, \ldots, \xi^N$ of ξ and solve

$$\begin{aligned} \min_{\substack{x\in\mathscr{X}\\ y^1,\ldots,y^N\in\mathbb{R}^n}} \quad & \frac{1}{N}\sum_{i=1}^{N} f(x, y^i, \xi^i) \\ \text{s.t.} \quad & H(x, y^i, \xi^i) = 0, \quad i = 1, \ldots, N. \end{aligned} \tag{3}$$

We refer to (1) as the true problem and (3) as the sample average approximation (SAA) problem. SAA methods have been extensively investigated in stochastic optimization. This type of methods are also called sample path (SP) methods. There has been extensive literature on both SAA and SP. See the recent work of Artstein and Wets [2], King and Wets [14], Plambeck et al. [19], Robinson [22], Shapiro [30], and Shapiro and Homem-de-Mello [32].

The focus of this paper is on the convergence analysis of Karush-Kuhn-Tucker (KKT) points of the SAA problem (3). We use the Clarke generalized implicit function theorem (Clarke [8]) to eliminate variable y^i, $i = 1, \ldots, N$, in (3) and then investigate the convergence of the reduced SAA problem. Because the reduced problem is nonsmooth, we consider a kind of generalized KKT (GKKT) condition that is defined by the sample average of some random set-valued mappings, and we then investigate the convergence of the corresponding stationary points. We achieve this by establishing a uniform strong law of large numbers for random compact set-valued mappings.

We summarize the main results that are established in this paper. First, we obtain a uniform strong law of large numbers for a random compact set-valued mapping. Then we apply the result to convergence analysis of a sequence of GKKT points of SAA problems. We show that, under some moderate conditions, an accumulation point of a sequence of weak GKKT points of the reduced SAA problem is a weak GKKT point of (1) with probability 1 (w.p.1). We also show that with probability approaching one exponentially fast, a global minimizer of the reduced SAA problem becomes an ϵ-global minimizer of the true problem. We establish similar results when the SAA problem is smoothed. Finally, we apply the results to a class of stochastic mathematical programs with strong monotone complementarity constraints.

The rest of this paper is organized as follows. In §2, we investigate the existence and uniqueness of implicit solution $y(x, \xi)$ of system (2) and its smoothing. We then define the weak GKKT conditions for the true problem. In §3, we establish a uniform strong law of large numbers for a random compact set-valued mapping and use it to investigate the convergence of GKKT points of reduced SAA problems and exponential convergence of global minimizers of SAA problems. In §4, we present a similar analysis for the smoothed SAA problem. In §5, we apply the convergence results to stochastic mathematical programs with strong monotone complementarity constraints. Finally, in §6, we present some preliminary numerical test results on an SMPEC problem.

2. Preliminaries.

In this section, we present some preliminary discussions about the measurability of a random set-valued mapping, implicit function theorem based on Clarke generalized Jacobians and GKKT conditions.

Throughout this paper, we use the following notation. We use $\|\cdot\|$ to denote the Euclidean norm of a vector, a matrix, and a compact set of matrices. Specifically, if $\mathscr{M}$ is a compact set of matrices, then $\|\mathscr{M}\| := \max_{M\in\mathscr{M}} \|M\|$. We use $\text{dist}(x, \mathscr{D}) := \inf_{x'\in\mathscr{D}} \|x - x'\|$ to denote the distance from point x to set $\mathscr{D}$. Here, $\mathscr{D}$ may be a subset of $\mathbb{R}^k$ or a subset of $\mathbb{R}^{k\times k}$. Given two compact sets $\mathscr{C}$ and $\mathscr{D}$, we use $\mathbb{D}(\mathscr{C}, \mathscr{D}) := \sup_{x\in\mathscr{C}} \text{dist}(x, \mathscr{D})$ to denote the distance from set $\mathscr{C}$ to set $\mathscr{D}$, and use $\mathbb{H}(\mathscr{C}, \mathscr{D})$ to denote the Hausdorff distance between the two sets, that is, $\mathbb{H}(\mathscr{C}, \mathscr{D}) := \max(\mathbb{D}(\mathscr{C}, \mathscr{D}), \mathbb{D}(\mathscr{D}, \mathscr{C}))$. We use $S(x, \delta)$ to denote the closed ball in $\mathbb{R}^m$ with radius δ and center x; that is, $S(x, \delta) := \{x' \in \mathbb{R}^m\colon \|x' - x\| \le \delta\}$. For a vector-valued function $g\colon \mathbb{R}^m \to \mathbb{R}^l$, we use $\nabla g(x)$ to denote the classical Jacobian of $g(x)$ if $g(x)$ is Frechét differentiable at x. In the case when $l = 1$, that is, $g(x)$ is a real-valued function, $\nabla g(x)$ denotes the gradient of $g(x)$, which is a row vector.

For a set-valued mapping $\mathscr{A}\colon \mathbb{R}^n \times \mathbb{R}^m \to 2^{\mathbb{R}^{n\times(n+m)}}$, we use $\pi_y\mathscr{A}(y, z)$ to denote the set of all $n \times n$ matrices M such that, for some $n \times m$ matrix N, the $n \times (n + m)$ matrix $[M \;\; N]$ belongs to $\mathscr{A}(y, z)$. We use $\overline{\lim}$ to denote the outer limit of a sequence of vectors and set-valued mappings.

2.1. Clarke generalized Jacobians and random sets.

Let $P\colon \mathbb{R}^j \to \mathbb{R}^l$ be a locally Lipschitz continuous vector-valued function. Recall that the Clarke generalized Jacobian (Clarke [8]) of P at $x \in \mathbb{R}^j$ is defined as

$$\partial P(x) := \text{conv}\left\{ \lim_{\substack{y\in D_P\\ y\to x}} \nabla P(y) \right\},$$

where D_P denotes the set of points at which P is Frechét differentiable, $\nabla P(y)$ denotes the usual Jacobian of P, which is an $l \times j$ matrix, "conv" denotes the convex hull of a set. It is well known that the Clarke generalized Jacobian $\partial P(x)$ is a convex compact set (Clarke [8]). In the case that $j = l$, $\partial P(x)$ consists of square matrices. We say $\partial P(x)$ is nonsingular if every matrix in set $\partial P(x)$ is nonsingular, and in this case we use $\partial P(x)^{-1}$ to denote the set of inverse matrices of $\partial P(x)$.

In later discussions, we will be concerned with the Clarke generalized Jacobians of the random function $H(x, y, \xi)$, such as $\partial_x H(x, y, \xi)$ and $\partial_y H(x, y, \xi)$. For fixed x and y, these Jacobians are random sets. We need to deal with the expectation of them, which is related to the measurability of random sets. In what follows, we make some preparations for this.

Let $V \subset \mathbb{R}^n$ be a compact set of $\mathbb{R}^n$ and let ξ: $\Omega \to \Xi$ be a random vector (note that we use the same ξ and Ξ as in (1), although we do not have to in this general discussion). A random (matrix) set-valued mapping $\mathscr{A}(\cdot, \xi)$: $V \to 2^{\mathbb{R}^{n\times n}}$ is said to be *compact valued* if for every $v \in V$ and $\xi \in \Xi$ (a realization of $\xi(\omega)$), $\mathscr{A}(v, \xi)$ is a compact set. Let $\mathscr{B}$ denote the space of nonempty, compact subsets of $\mathbb{R}^{n\times n}$ equipped with the Hausdorff distance. Then, $\mathscr{A}(v, \xi(\cdot))$ can also be viewed as a single-valued mapping from Ω to $\mathscr{B}$. $\mathscr{A}(v, \xi(\cdot))$ is said to be measurable if and only if for every $B \in \mathscr{B}$, $\mathscr{A}(v, \xi(\cdot))^{-1}B$ is measurable.

We now define the expectation of $\mathscr{A}(x, \xi(\omega))$. A *selection* of the random set $\mathscr{A}(x, \xi(\omega))$ is a random matrix $A(x, \xi(\omega)) \in \mathscr{A}(x, \xi(\omega))$, which means $A(x, \xi(\omega))$ is measurable. Note that such selections exist; see Artstein and Vitale [1] and references therein. The *expectation of* $\mathscr{A}(x, \xi(\omega))$, denoted by $\mathbb{E}[\mathscr{A}(x, \xi(\omega))]$, is defined as the collection of $\mathbb{E}[A(x, \xi(\omega))]$ where $A(x, \xi(\omega))$ is a selection. We regard $\mathbb{E}[\mathscr{A}(x, \xi(\omega))]$ as well defined if $\mathbb{E}[\mathscr{A}(x, \xi(\omega))] \in \mathscr{B}$. A necessary and sufficient condition of the well definedness of the expectation is

$$\mathbb{E}[\|\mathscr{A}(x, \xi(\omega))\|] := \mathbb{E}[\mathbb{H}(0, \mathscr{A}(x, \xi(\omega)))] < \infty,$$

see Artstein and Vitale [1]. For simplicity of discussion, we make a blanket assumption throughout the rest of this paper that

$$\mathbb{E}[\|\pi_x \partial H(x, y, \xi)\|] < \infty, \qquad \mathbb{E}[\|\pi_y \partial H(x, y, \xi)\|] < \infty, \tag{4}$$

which implies that every selection from $\partial_x H(x, y, \xi)$ and $\partial_y H(x, y, \xi)$ has a finite expected value. Consequently, both $\mathbb{E}[\partial_x H(x, y, \xi)]$ and $\mathbb{E}[\partial_y H(x, y, \xi)]$ are well defined. It is expected that (4) is satisfied by many stochastic piecewise smooth Lipschitz functions in practice. Note that except for some pathological examples, most functions in practice are piecewise smooth. See Scholtes [29] for a comprehensive discussion of piecewise smooth functions.

For fixed $\xi \in \Xi$, $\mathscr{A}(\cdot, \xi)$ is said to be *upper semicontinuous* on V if for every $v \in V$, and a neighborhood $B(\mathscr{A}(v, \xi))$ of $\mathscr{A}(v, \xi)$, there exists a neighborhood $B(v)$ of v such that $\mathscr{A}(w, \xi) \subset B(\mathscr{A}(v, \xi))$ for all $w \in B(v)$. $\mathscr{A}(\cdot, \xi)$ is said to be *Hausdorff continuous* if $\lim_{w\to v} \mathscr{A}(w, \xi) = \mathscr{A}(v, \xi)$. Note that the Clarke generalized Jacobian of a locally Lipschitz-continuous function that involves random variables is a random compact set-valued mapping that is upper semicontinuous with respect to its deterministic variables.

2.2. An implicit function theorem. In this subsection, we investigate the parametric nonsmooth system of Equations (2). The existence and uniqueness of $y(x, \xi)$ for (2) is essential to model (1). For simplicity of discussion, we assume throughout that for every $(x, \xi) \in \mathscr{X} \times \Xi$, system (2) has a solution. The following results deal with the existence and uniqueness of implicit function $y(x, \xi)$, its Lipschitz continuity, and the estimate of its Clarke generalized Jacobian.

LEMMA 2.1. *Suppose that* (a) H: $\mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$ *is piecewise smooth*; (b) *for every* $(x, \xi) \in \mathscr{X} \times \Xi$, *there exists* y *such that* $H(x, y, \xi) = 0$; (c) $\pi_y \partial H(x, y, \xi)$ *is uniformly nonsingular on* $\mathscr{X} \times \mathbb{R}^n \times \Xi$; *that is, there exists a constant* $\delta > 0$ *such that* $\|[\pi_y \partial H(x, y, \xi)]^{-1}\| \le \delta$ *for all* $(x, y, \xi) \in \mathscr{X} \times \mathbb{R}^n \times \Xi$. *Then,*

(i) *there exists a unique globally Lipschitz continuous and piecewise smooth function* y: $\mathscr{X} \times \Xi \to \mathbb{R}^n$ *such that* $y(x, \xi)$ *solves* (2);

(ii) *for* $(x, \xi) \in \mathscr{X} \times \Xi$,

$$\partial_x y(x, \xi) \subset \operatorname{conv}\{-R^{-1}U\colon (U, R, V) \in \partial H(x, y(x, \xi), \xi),\, U \in \mathbb{R}^{n\times m},\, R \in \mathbb{R}^{n\times n},\, V \in \mathbb{R}^{n\times k}\}; \tag{5}$$

(iii) $y(x, \xi(\cdot))$: $\Omega \to \mathbb{R}^n$ *is measurable.*

The results follow from the Clarke generalized implicit function theorem (Clarke [8]) and the calculus of implicit generalized Jacobians. See a proof in the appendix.

Recall that a function $g(x, \cdot)$: $\mathscr{X} \times \Omega \to \mathbb{R}^n$ is called a *Carathéodory function* if $g(x, \cdot)$ is measurable, and for every $\omega \in \Omega$, $g(\cdot, \omega)$ is continuous. Obviously, $y(x, \xi(\omega))$ is a Carathéodory function, and by Aubin and Frankowska [3, Lemma 8.2.3], $f(x, y(x, \xi(\omega)), \xi(\omega))$ is also a Carathéodory function.

2.3. GKKT conditions for the true problem. Because $y(x,\xi)$ is nonsmooth, we consider GKKT conditions for (1) as follows:

$$\begin{aligned} 0 &\in \mathbb{E}[\partial_x(f(x,y(x,\xi),\xi))]+\mathcal{N}_{\mathscr{X}}(x)\\ &=\mathbb{E}[\nabla_x f(x,y(x,\xi),\xi)+\nabla_y f(x,y(x,\xi),\xi)\partial_x y(x,\xi)]+\mathcal{N}_{\mathscr{X}}(x) \end{aligned}\tag{6}$$

where $\partial_x y(x,\xi)$ denotes the Clarke generalized Jacobian of $y(x,\xi)$ with respect to x, and $\mathcal{N}_{\mathscr{X}}(x)$ denotes the normal cone of $\mathscr{X}$ at $x\in\mathscr{X}$; that is,

$$\mathcal{N}_{\mathscr{X}}(x):=[T_{\mathscr{X}}(x)]^-=\{\varsigma\in\mathbb{R}^m \mid \langle\varsigma,d\rangle\le 0 \text{ for all } d\in T_{\mathscr{X}}(x)\},$$

where $T_{\mathscr{X}}(x):=\limsup_{t\downarrow 0}(\mathscr{X}-x)/t$.

We say that $x\in\mathscr{X}$ is a *GKKT point* of (1) if it satisfies (6). Note that we implicitly assume that $\mathbb{E}[\nabla_x f(x,y(x,\xi),\xi)+\nabla_y f(x,y(x,\xi),\xi)\partial_x y(x,\xi)]$ is well defined, a sufficient condition for which is that the gradients and subgradients under integration are bounded by an integrable function. Note also that using the expected value of subdifferentials to characterize the optimality conditions for stochastic programs is well documented in the literature and can be traced back to the earlier work by Rockafellar and Wets [24], where "basic Kuhn-Tucker conditions" in terms of Rockafellar's convex subdifferential (Rockafellar [23]) are derived for a class of stochastic programs with convex objective and convex constraints. See also Wets [35]. This type of optimality condition is weaker than optimality condition $0\in\partial\mathbb{E}[f(x,y(x,\xi))]+\mathcal{N}_{\mathscr{X}}(x)$, in that in general $\mathbb{E}[\partial_x f(x,y(x,\xi),\xi)]$ is larger than $\partial\mathbb{E}[f(x,y(x,\xi),\xi)]$. However, the latter is disadvantaged by requiring the derivative information of the expected value of $f(x,y(x,\xi),\xi)$, which may be difficult to obtain. Moreover, under some regularity conditions (Clarke [8, Definition 2.3.4]) such as convexity of $f(x,y(x,\xi),\xi)$ in x or continuous differentiability of the function in x, these two sets of subdifferentials may coincide. See similar discussions in Wets [35, Proposition 2.10] and Homen-De-Mello [11, Proposition 5.1].

To ease the notation, let $\Lambda(x,\xi):=-\nabla_y f(x,y(x,\xi),\xi)\,\mathrm{conv}([\pi_y\partial H(x,y(x,\xi),\xi)]^{-1})$. Using the estimate of $\partial_x y(x,\xi)$ in part (ii) of Lemma 2.1, we may consider a *weak* GKKT condition for (1) as follows:

$$0\in\mathbb{E}[\nabla_x f(x,y(x,\xi),\xi)+\Lambda(x,\xi)\pi_x\partial H(x,y(x,\xi),\xi)]+\mathcal{N}_{\mathscr{X}}(x).\tag{7}$$

A point $x\in\mathscr{X}$ is called a *weak GKKT point* of (1) if x satisfies (7). Because

$$\partial_x y(x,\xi)\subset\mathrm{conv}([\pi_y\partial H(x,y(x,\xi),\xi)]^{-1})\pi_x\partial H(x,y(x,\xi),\xi),$$

we have

$$\nabla_y f(x,y(x,\xi),\xi)\partial_x y(x,\xi)\subset\Lambda(x,\xi)\pi_x\partial H(x,y(x,\xi),\xi),\tag{8}$$

which implies that a GKKT point must be a weak GKKT point, but not vice versa. This means that the set of weak GKKT points gives a bound or an estimate of the set of GKKT points. To see how precise the estimate is, we need to look at the inclusion in (8). Observe that if H is strictly differentiable, then ∂H reduces to a singleton (the classical Jacobian), and consequently the two sets are equal and coincide with the usual KKT points. This is a very important case because in many practical instances, function H is piecewise smooth, which means H is continuously differentiable everywhere except for some kinks. In SMPECs, when strict complementarity holds, H becomes continuously differentiable. In general, the relationship between the two sets depends on the index consistency (Ralph and Xu [21]) of the piecewise smooth function H in y at $y(x,\xi)$ and the structure of $\pi_x\partial H$. See a discussion on a similar relation for the deterministic case in Ralph and Xu [21, §5]. Here we list two main reasons for us to consider the weak GKKT condition (7): (a) This condition only utilizes the derivative information of $H(x,y,\xi)$ instead of that of the implicit function $y(x,\xi)$, which is often difficult to obtain; (b) under some bounded derivative conditions, GKKT points (KKT points) of SAA problems converge to the set of weak GKKT points as sample size tends to infinity; see Theorem 3.1 and Theorem 4.1. Note that in this paper the weak GKKT condition is used as a convenient and unified approach to effectively address the optimality conditions of our problem at both smooth and nonsmoothness points.

2.4. Implicit smoothing and KKT conditions. The nonsmoothness of H is a hurdle of a direct application of existing smooth optimization methods. One of the popular ways to deal with the nonsmoothness is to consider a smoothing of H so that we obtain a smoothed system to approximate (2).

Let $\epsilon\in\mathbb{R}$ be a parameter. We say $\hat{H}(x,y,\xi,\epsilon)$: $\mathbb{R}^m\times\mathbb{R}^n\times\mathbb{R}^k\times\mathbb{R}\to\mathbb{R}^n$ is a *smoothing* of H if it satisfies the following: (a) for every $(x,y,\xi)\in\mathbb{R}^m\times\mathbb{R}^n\times\mathbb{R}^k$, $\hat{H}(x,y,\xi,0)=H(x,y,\xi)$; (b) for every $x\in\mathbb{R}^m$, $\hat{H}$ is locally Lipschitz at $(x,y,\xi,0)$; and (c) $\hat{H}$ is continuously differentiable on $\mathbb{R}^m\times\mathbb{R}^n\times\mathbb{R}^k\times\mathbb{R}\backslash\{0\}$.

This type of general smoothing was considered by Ralph and Xu [21], and it covers many practically useful smoothing functions in complementarity problems and deterministic mathematical programs with equilibrium constraints (MPECs). Our purpose here is to replace (2) with a smoothed system, and then we consider a smooth approximation of (1) by

$$\begin{aligned} \min\ & \mathbb{E}[f(x, \hat{y}(x, \xi, \epsilon), \xi)] \\ \text{s.t.}\ & x \in \mathscr{X}, \end{aligned} \tag{9}$$

where $\hat{y}(x, \xi, \epsilon)$ is a solution of the following smooth system of equations

$$\widehat{H}(x, y, \xi, \epsilon) = 0, \quad \forall\, x \in \mathscr{X},\ \ \xi \in \Xi,\ \ \epsilon \in \mathbb{R}. \tag{10}$$

To ensure that the smoothed program (9) is well defined, the smoothed system (10) must have a unique solution. This requires some extra assumptions on the smoothing. The following lemma addresses this issue.

Lemma 2.2. *Let $\widehat{H}$ be a smoothing of H. Suppose that the conditions in Lemma 2.1 are satisfied. Then, for any $(x, \xi) \in \mathscr{X} \times \Xi$ and any $|\epsilon| \in [0, \epsilon_0)$ where ϵ_0 is a small positive scalar, there exists $y \in \mathbb{R}^n$ such that $\widehat{H}(x, y, \xi, \epsilon) = 0$.*

The result is expected. We attach a standard proof in the appendix. Using Lemma 2.2, we can derive the following result.

Lemma 2.3. *Let the assumptions of Lemma 2.1 hold. Suppose that there exists $\epsilon_0 > 0$ such that for $|\epsilon| \in [0, \epsilon_0]$, $\pi_y \partial \widehat{H}(x, y, \xi, \epsilon)$ is uniformly nonsingular for any $(x, y, \xi) \in \mathscr{X} \times \mathbb{R}^n \times \Xi$. Then, Equation (10) defines a unique implicit function $\hat{y}$: $\mathscr{X} \times \mathbb{R}^n \times \Xi \times [-\epsilon_0, \epsilon_0] \to \mathbb{R}^m$. Moreover, $\hat{y}$ is locally Lipschitz continuous on $\mathscr{X} \times \Xi \times [-\epsilon_0, \epsilon_0]$ and continuously differentiable on $\mathscr{X} \times \Xi \times [-\epsilon_0, \epsilon_0] \backslash \{0\}$.*

Proof. Observe first that by Lemma 2.2, for any $(x, \xi) \in \mathscr{X} \times \Xi$ and small $|\epsilon| > 0$, there exists y such that $\widehat{H}(x, y, \xi, \epsilon) = 0$. Because by assumption $\pi_y \partial \widehat{H}(x, y, \xi, \epsilon)$ is uniformly nonsingular on $\mathscr{X} \times \mathbb{R}^n \times \Xi \times [-\epsilon_0, \epsilon_0]$, then with a similar argument as that in the proof of Lemma 2.1 and by virtue of the continuous differentiability of $\widehat{H}$ on $\mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^k \times \mathbb{R} \backslash \{0\}$, we can prove that the claimed results hold. We omit the details. □

From Lemma 2.3 and Aubin and Frankowska [3, Theorem 8.2.9], we know that $\hat{y}(x, \xi(\omega), \epsilon)$ is a Carathéodory function, and so is $f(x, \hat{y}(x, \xi(\omega), \epsilon), \epsilon)$. By Lemma 2.3, the first-order necessary condition for (9) can be written as follows

$$0 \in \mathbb{E}[\nabla_x f(x, \hat{y}(x, \xi, \epsilon), \xi) + \nabla_y f(x, \hat{y}(x, \xi, \epsilon), \xi) \nabla_x \hat{y}(x, \xi, \epsilon)] + \mathcal{N}_{\mathscr{X}}(x), \quad \text{for } \epsilon \neq 0, \tag{11}$$

where $\nabla_x \hat{y}(x, \xi, \epsilon) = -\nabla_y \widehat{H}(x, \hat{y}(x, \xi, \epsilon), \xi, \epsilon)^{-1} \nabla_x \widehat{H}(x, \hat{y}(x, \xi, \epsilon), \xi, \epsilon)$. We would like to investigate the limiting behaviour of the KKT points based on (11) as $\epsilon \to 0$ and compare them with the GKKT points based on (6) in subsequent analysis.

3. Sample average approximations.

In this section, we discuss SAA problem (3). In particular, we investigate the convergence of GKKT points of (3) as the sample size N tends to infinity.

3.1. GKKT conditions.

Observe that (3) is a deterministic optimization problem with $N \times n$ nonsmooth equality constraints. In convergence analysis, we will consider, at least theoretically, the case in which N tends to ∞. Obviously, (3) is not well defined for N being ∞ because the problem has an infinite number of equality constraints. To tackle this problem, we use Lemma 2.1 to reformulate (3) as an *implicit* problem

$$\begin{aligned} \min\ & f_N(x) := \frac{1}{N} \sum_{i=1}^{N} f(x, y(x, \xi^i), \xi^i) \\ \text{s.t.}\ & x \in \mathscr{X}. \end{aligned} \tag{12}$$

For a finite N, (3) is equivalent to (12) in the sense that if $(x, y^1, \ldots, y^N)$ is a local minimizer of (3), then x is a local minimizer of (12). Conversely, if x is a local minimizer of (12), then there exist unique $y^1, \ldots, y^N$ such that $H(x, y^i, \xi^i) = 0$ and $(x, y^1, \ldots, y^N)$ is a local minimizer of (3).

The relation is, however, not as clear about GKKT points. Ralph and Xu [21] considered a general minimization problem with nonsmooth equality constraints and investigated the relationship between the GKKT conditions of the original problem and their counterpart of equivalent implicit reformulated problems (where some variables are eliminated). It is concluded that the two GKKT conditions are "roughly equivalent" if the underlying functions are piecewise smooth, but neither implies the other. See Ralph and Xu [21, §5] for details. Here we consider the weak GKKT conditions of the implicitly reformulated program and show that they are equivalent to the GKKT conditions of SAA problem (3).

We start our discussion with the GKKT condition for (12), which can be written as

$$0 \in \partial f_N(x) + \mathcal{N}_{\mathcal{X}}(x).$$

Suppose that $\pi_y \partial H(x, y(x, \xi^i), \xi^i)$, $i = 1, \ldots, N$, is nonsingular. Note that

$$\partial f_N(x) \subset \frac{1}{N}\sum_{i=1}^{N}[\nabla_x f(x, y(x, \xi^i), \xi^i) + \Lambda(x, \xi^i)\pi_x \partial H(x, y(x, \xi^i), \xi^i)].$$

We may consider the *weak* GKKT condition of (12) as follows

$$0 \in \frac{1}{N}\sum_{i=1}^{N}[\nabla_x f(x, y(x, \xi^i), \xi^i) + \Lambda(x, \xi^i)\pi_x \partial H(x, y(x, \xi^i), \xi^i)] + \mathcal{N}_{\mathcal{X}}(x). \tag{13}$$

We next consider the GKKT conditions of (3). Following Hiriart-Urruty [12, Corollary 2.5], we have

$$0 \in \frac{1}{N}\sum_{i=1}^{N}\nabla_x f(x, y^i, \xi^i) + \sum_{i=1}^{N}\lambda^i \partial_x H(x, y^i, \xi^i) + \mathcal{N}_{\mathcal{X}}(x), \tag{14}$$

$$0 \in \frac{1}{N}\begin{pmatrix} \nabla_y f(x, y^1, \xi^1) \\ \vdots \\ \nabla_y f(x, y^N, \xi^N) \end{pmatrix} + \begin{pmatrix} \lambda^1 \partial_y H(x, y^1, \xi^1) \\ \vdots \\ \lambda^N \partial_y H(x, y^N, \xi^N) \end{pmatrix}, \tag{15}$$

$$0 = H(x, y^i, \xi^i), \quad i = 1, \ldots, N. \tag{16}$$

Because $\partial_y H(x, y^i, \xi^i) \subset \pi_y \partial H(x, y^i, \xi^i)$, $i = 1, \ldots, N$, is nonsingular, we can find λ^i such that

$$\lambda^i \in -\frac{1}{N}\nabla_y f(x, y^i, \xi^i)[\partial_y H(x, y^i, \xi^i)]^{-1} \subset -\frac{1}{N}\nabla_y f(x, y^i, \xi^i)\,\mathrm{conv}([\pi_y \partial H(x, y^i, \xi^i)]^{-1}), \quad i = 1, \ldots, N.$$

Replacing λ^i with $-(1/N)\nabla_y f(x, y^i, \xi^i)\,\mathrm{conv}([\pi_y \partial H(x, y^i, \xi^i)]^{-1})$ in (14) and y^i with $y(x, \xi^i)$, we can obtain (13). In what follows, we will investigate the convergence of *weak GKKT points* that satisfy (13).

3.2. A uniform law of large numbers for a random set-valued mapping. Before proceeding to the convergence analysis of the GKKT sequence based on (13), we need to establish some kind of uniform (strong) law of large numbers for a random set-valued mapping in Lemma 3.2.

Let $\mathcal{A}(\cdot, \xi)$: $V \to 2^{\mathbb{R}^{n\times n}}$ be a random compact set-valued mapping, where $V \subset \mathbb{R}^n$ is a compact set of $\mathbb{R}^n$ and ξ: $\Omega \to \Xi$ is a random vector (note that we use the same ξ and Ξ as in (1)). Let $\xi^1, \ldots, \xi^N$ be an i.i.d. sample of ξ. Obviously, for every $v \in V$, $\mathcal{A}(v, \xi^i)$, $i = 1, \ldots, N$, are independent, identically distributed random sets of matrices. Let

$$\mathcal{A}_N(v) := \frac{1}{N}\sum_{i=1}^{N}\mathcal{A}(v, \xi^i). \tag{17}$$

Here the addition of compact sets is in the sense of Minkowski; that is, for two compact subsets K and L, $K + L = \{x + y\colon x \in K, y \in L\}$. Obviously, $\mathcal{A}_N(v)$ is a compact subset of $\mathbb{R}^{n\times n}$.

In what follows, we establish some kind of uniform strong law of large numbers for the convergence of $\mathcal{A}_N(v)$ as N tends to ∞ in Lemma 3.2. For this purpose, we need a strong law of large numbers for random compact sets, which is essentially due to Artstein and Vitale [1].

Lemma 3.1. *Let $C_i \subset \mathbb{R}^{n\times n}$, $i = 1, \ldots, N$ be i.i.d. random sets such that $\mathbb{E}[\|C_i\|] < \infty$. Then*

$$\lim_{N\to\infty}\frac{C_1 + \cdots + C_N}{N} = \mathbb{E}[\mathrm{conv}\, C] := \{\mathbb{E}[c]\colon c \in \mathrm{conv}\, C\}, \quad \text{w.p. } 1$$

where C is a random set that has the same distribution as C_i, $i = 1, \ldots, N$. Moreover, if the probability measure P is nonatomic, then $\mathbb{E}[\mathrm{conv}\, C] = \mathbb{E}[C]$.

By Lemma 3.1, we can establish the following result, which is one of the essential results of this paper.

Lemma 3.2. *Let $V \subset \mathbb{R}^n$ be a compact set, and $\mathcal{A}(\cdot, \xi)$: $V \to 2^{\mathbb{R}^{n\times n}}$ be a compact, upper-semicontinuous set-valued mapping for every $\xi \in \Xi$. Let $\mathcal{A}_N(v)$ be defined as in (17) for every $v \in V$. Suppose that there exists*

$\sigma(\xi)$ such that

$$\|\mathcal{A}(v,\xi)\| := \sup_{A\in\mathcal{A}(v,\xi)} \|A\| \le \sigma(\xi), \tag{18}$$

where $\mathbb{E}[\sigma(\xi)] < \infty$. Then for any $\gamma > 0$, there exists $\delta > 0$ such that

$$\overline{\lim_{N\to\infty}} \mathcal{A}_N(v) \subset \mathbb{E}[\mathcal{A}_\delta(v,\xi)] + \gamma \boldsymbol{B}, \quad w.p.\ 1 \tag{19}$$

uniformly for $v \in V$, where $\mathcal{A}_\delta(v,\xi) := \bigcup_{w\in S(v,\delta)} \mathcal{A}(w,\xi)$, $\mathcal{B}$ denotes the unit ball in $\mathbb{R}^{n\times n}$, and the mathematical expectation of $A(v,\xi)$ is taken entrywise on matrix $A(v,\xi)$.

Proof. Let $v \in V$. Let W_k be a series of neighborhoods of point v shrinking to $\{v\}$ as $k\to\infty$. Let $b_k(\xi) := \sup_{w\in W_k} \mathbb{D}(\mathcal{A}(w,\xi), \mathcal{A}(v,\xi))$ (see the definition of $\mathbb{D}$ in §2.1). Note that if $\mathbb{D}(\mathcal{C},\mathcal{D}) = 0$, then $\mathcal{C}\subset\mathcal{D}$, but the equality may not hold. Under the condition (18), $b_k(\xi)$ is bounded by $2\sigma(\xi)$, where by assumption $\mathbb{E}[\sigma(\xi)]<\infty$. By the Lebesgue dominated convergence theorem and the upper semicontinuity of $\mathcal{A}(\cdot,\xi)$

$$\lim_{k\to\infty} \mathbb{E}[b_k(\xi)] = \mathbb{E}\Big[\lim_{k\to\infty} b_k(\xi)\Big] = 0. \tag{20}$$

For any $w\in W_k$, it is not difficult to verify that

$$\begin{aligned}\mathbb{D}(\mathcal{A}_N(w), \mathcal{A}_N(v)) &= \mathbb{D}\left(\frac{1}{N}\sum_{i=1}^N \mathcal{A}(w,\xi^i), \frac{1}{N}\sum_{i=1}^N \mathcal{A}(v,\xi^i)\right)\\ &\le \frac{1}{N}\sum_{i=1}^N \mathbb{D}(\mathcal{A}(w,\xi^i), \mathcal{A}(v,\xi^i)).\end{aligned}$$

Thus,

$$\sup_{w\in W_k}\{\mathbb{D}(\mathcal{A}_N(w),\mathcal{A}_N(v))\} \le \frac{1}{N}\sum_{i=1}^N \sup_{w\in W_k}\{\mathbb{D}(\mathcal{A}(w,\xi^i),\mathcal{A}(v,\xi^i))\} = \frac{1}{N}\sum_{i=1}^N b_k(\xi^i). \tag{21}$$

By the strong law of large numbers, it follows that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N b_k(\xi^i) = \mathbb{E}[b_k(\xi)], \quad \text{w.p. } 1. \tag{22}$$

It then follows from (20)–(22) that for any $\gamma>0$, there exists a closed δ-neighborhood $S(v,\delta)$ of v such that

$$\sup_{w\in S(v,\delta)} \mathbb{D}(\mathcal{A}_N(w),\mathcal{A}_N(v)) \le \gamma, \quad \text{w.p. } 1$$

for sufficiently large N.

Because V is compact, we may choose $v_1,\dots,v_J\in V$ and their corresponding closed neighborhoods $S(v_1,\delta),\dots,S(v_J,\delta)$ such that $V\subset\bigcup_{j=1}^J S(v_j,\delta)$, and for N sufficiently large,

$$\sup_{w\in S(v_j,\delta)} \mathbb{D}(\mathcal{A}_N(w),\mathcal{A}_N(v_j)) \le \gamma, \quad \text{for } j=1,\dots,J, \ \text{w.p. } 1.$$

This implies that

$$\overline{\lim_{N\to\infty}} \mathcal{A}_N(w) \subset \overline{\lim_{N\to\infty}} \mathcal{A}_N(v_j) + \gamma\boldsymbol{B}, \quad \forall\, w\in S(v_j,\delta), \ j=1,\dots,J. \tag{23}$$

By applying Lemma 3.1 to $\mathcal{A}_N(v_j)$, we have

$$\overline{\lim_{N\to\infty}} \mathcal{A}_N(v_j) \subset \mathbb{E}[\mathcal{A}(v_j,\xi)], \quad j=1,\dots,J, \ \text{w.p. } 1. \tag{24}$$

Moreover, for each j,

$$\mathbb{E}[\mathcal{A}(v_j,\xi)] \subset \mathbb{E}[\mathcal{A}_\delta(w,\xi)], \quad \forall\, w\in S(v_j,\delta). \tag{25}$$

Combining (23)–(25), we have

$$\overline{\lim_{N\to\infty}} \mathcal{A}_N(w) \subset \mathbb{E}[\mathcal{A}_\delta(w,\xi)] + \gamma\boldsymbol{B}, \quad \text{w.p. } 1 \tag{26}$$

uniformly for $w\in S(v_j,\delta)$. Because V is covered by the union of a finite number of neighborhoods, (26) holds for all $w\in V$. This completes the proof. □

Note that the existence of $\mathbb{E}[A(v,\xi)]$ for every matrix $A(v,\xi)\in\mathcal{A}(v,\xi)$ is guaranteed by our assumption (18). Of course, not every stochastic compact set-valued mapping has this property. We are interested in those that enjoy this property. Note also that set V, the domain of $\mathcal{A}(\cdot,\xi)$, is assumed to be compact. This is purely for the convenience of presentation. It is obvious that the lemma holds true on any compact set of a noncompact domain of the mapping. The same comment applies to Corollary 3.1, Example 3.1, and Lemma 3.3.

Corollary 3.1. *If the set-valued mapping $\mathcal{A}(\cdot,\xi)\colon V \to 2^{\mathbb{R}^{n\times n}}$ in Lemma 3.2 is Hausdorff continuous, then the following relation*

$$\lim_{N\to\infty} \mathcal{A}_N(v) = \mathbb{E}[\mathcal{A}(v,\xi)], \quad \textit{w.p. } 1$$

holds uniformly for all $v \in V$.

Proof. The results follow straightforwardly from Lemma 3.2 by replacing distance $\mathbb{D}$ with the Hausdorff distance $\mathbb{H}(\mathcal{A}(w,\xi),\mathcal{A}(v,\xi))$. We omit the details. □

To better explain both Lemma 3.2 and Corollary 3.1, we provide an example below.

Example 3.1. Consider a vector-valued Lipschitz-continuous function $h(\cdot,\xi)\colon \mathcal{X}\to\mathbb{R}^n$, where $\xi\colon \Omega\to\Xi\subset\mathbb{R}^l$ is a random vector and $\mathcal{X}$ is a compact subset of $\mathbb{R}^n$. For every $\xi\in\Xi$, it is well known (Clarke [8]) that the Clarke generalized Jacobian $\partial_x h(\cdot,\xi)$ is an upper-semicontinuous, convex, and compact set-valued mapping. Let $\xi^1,\ldots,\xi^N$ be an i.i.d. sample of ξ. Then $\partial_x h(x,\xi^i)$, $i=1,\ldots,N$, are convex, compact i.i.d. random sets. Let

$$\mathcal{A}_N(x) := \frac{1}{N}\sum_{i=1}^{N} \partial_x h(x,\xi^i).$$

Assume that there exists a function $\sigma(\xi)$ such that $\|\partial_x h(x,\xi(\omega))\| < \sigma(\xi)$, and $\mathbb{E}[\sigma(\xi)]<\infty$. Then, by using Lemma 3.2, for any $\gamma>0$, there exists $\delta>0$ such that

$$\overline{\lim_{N\to\infty}}\, \mathcal{A}_N(x) \subset \mathbb{E}[(\partial_x)_\delta h(x,\xi)] + \gamma \boldsymbol{B}, \quad \text{w.p. } 1$$

uniformly for $x\in\mathcal{X}$.

We next consider the Clarke ϵ-generalized Jacobian of $h(x,\xi)$ with respect to x for a given $\xi\in\Xi$ below

$$\partial_x^\epsilon h(x,\xi) := \bigcup_{x'\in\mathcal{S}(x,\epsilon)} \partial_x h(x',\xi),$$

where ϵ is a small positive number. The notion of the ϵ-generalized Jacobian was that it is a generalization of the Clarke ϵ-subdifferential (Polak et al. [20]) of a real-valued Lipschitz function. For fixed ϵ and $\xi\in\Xi$, it is not difficult to verify that $\partial_x^\epsilon h(\cdot,\xi)$ is a compact and upper-semicontinuous set-valued mapping. However, $\partial_x^\epsilon h(\cdot,\xi)$ may not be Hausdorff continuous in general; see a counterexample in Xu et al. [38].

Let

$$\mathcal{A}_N^\epsilon(x) := \frac{1}{N}\sum_{i=1}^{N} \partial_x^\epsilon h(x,\xi^i).$$

Then, by Lemma 3.2, we have that, for any $\gamma>0$, there exists $\delta\ge\epsilon$ such that

$$\overline{\lim_{N\to\infty}}\, \mathcal{A}_N^\epsilon(x) \subset \mathbb{E}[\partial_x^\delta h(x,\xi)] + \gamma \boldsymbol{B}, \quad \text{w.p. } 1$$

uniformly for $x\in\mathcal{X}$.

In the case that $h(x,\xi)$ is a real-valued convex function of x, there is an alternative way to define the ϵ-subdifferential using Rockafellar's subdifferential in convex analysis (Rockafellar [23]):

$$\partial_x^\epsilon h(x,\xi) := \{\eta\colon\, h(x',\xi) \ge h(x,\xi) + \eta^T(x'-x) - \epsilon,\ \forall\, x'\in\mathbb{R}^n\}.$$

This type of ϵ-subdifferential was first considered by Brønsted and Rockafellar [6], and it is Hausdorff continuous in x. Using Corollary 3.1, we have

$$\lim_{N\to\infty} \mathcal{A}_N^\epsilon(x) = \mathbb{E}[\partial^\epsilon h(x,\xi)], \quad \text{w.p. } 1$$

uniformly with respect to x in $\mathcal{X}$. This result can be found in Shapiro and Wardi [33, Proposition 3.4].

3.3. Convergence of the GKKT point of the SAA problem. With Lemma 3.2, we are able to establish convergence results of the weak GKKT points for SAA problem (12).

Lemma 3.3. *Let V be a compact subset of $\mathbb{R}^n$, and $\mathcal{A}(\cdot,\xi)\colon V\to 2^{\mathbb{R}^{n\times n}}$ be a compact, upper-semicontinuous set-valued mapping for every $\xi\in\Xi$. Suppose that there exists $\sigma(\xi)$ such that $\sup_{x\in V}\|\mathcal{A}(x,\xi)\|\le\sigma(\xi)$ and*

$\mathbb{E}[\sigma(\xi)] < \infty$. *If* $x^* \in V$ *is such a point that for any* $\gamma > 0$ *there exists* $\delta > 0$ *such that* $0 \in \mathbb{E}[\mathcal{A}_\delta(x^*, \xi)] + \gamma \boldsymbol{B}$, *then* $0 \in \mathbb{E}[\mathcal{A}(x^*, \xi)]$, *where* $\mathcal{A}_\delta(x^*, \xi) = \bigcup_{x \in S(x^*, \delta)} \mathcal{A}(x, \xi)$.

Proof. By assumption, $\mathcal{A}_\delta(x^*, \xi)$ is bounded by $\sigma(\xi)$. Moreover, because $\mathcal{A}_\delta(x^*, \xi)$ is monotonic in δ and $\mathcal{A}(\cdot, \xi)$ is upper semicontinuous, then $\lim_{\delta \to 0} \mathcal{A}_\delta(x^*, \xi) = \mathcal{A}(x^*, \xi)$. By the Lebesgue dominated convergence theorem,

$$\lim_{\delta \to 0} \mathbb{E}[\mathcal{A}_\delta(x^*, \xi)] = \mathbb{E}\Big[\lim_{\delta \to 0} \mathcal{A}_\delta(x^*, \xi)\Big] = \mathbb{E}[\mathcal{A}(x^*, \xi)].$$

Note also that $\delta \to 0$ as $\gamma \to 0$. The conclusion follows. □

Assumption 3.1. *Let* $\mathcal{C}$ *be a compact subset of* $\mathcal{X}$. *There exists a function* $\kappa_0(\xi)$ *such that for all* $(x, \xi) \in \mathcal{C} \times \Xi$,

$$\max\{\|\nabla_x f(x, y(x, \xi), \xi)\|, \|\Lambda(x, \xi)\|, \|\pi_x \partial H(x, y(x, \xi), \xi)\|\} \le \kappa_0(\xi), \tag{27}$$

where $\mathbb{E}[\kappa_0(\xi) + \kappa_0(\xi)^2] < \infty$.

Note that set $\mathcal{C}$ in the assumption is deliberately left unspecified, and it will be specified in the context of various convergence theorems in this and the next subsections. Note also that $\|\Lambda(x, \xi)\| \le \|\nabla_y f(x, y(x, \xi), \xi)\| \|[\pi_y \partial H(x, y(x, \xi), \xi)]^{-1}\|$; therefore, (27) holds if $\pi_y \partial H$ is uniformly nonsingular and $\nabla_{x,y} f$ and $\pi_x \partial H$ are bounded by a positive integrable function. We are now ready to state one of our main convergence results of this paper.

Theorem 3.1. *Let the assumptions of Lemma 2.1 hold. Let* $\{x_N\}$ *be a sequence of weak GKKT points that satisfy* (13) *and let* x^* *be an accumulation point. Let Assumption 3.1 hold w.p. 1 on a compact set* $\mathcal{C}$, *which contains a neighborhood of* x^*, $S(x^*, \eta) \cap \mathcal{X}$, *where* $\eta > 0$ *is a small number. Then, w.p. 1,* x^* *satisfies*

$$0 \in \mathbb{E}[\mathcal{G}(x^*, \xi)] + \mathcal{N}_{\mathcal{X}}(x^*),$$

where $\mathcal{G}(x, \xi) := \nabla_x f(x, y(x, \xi), \xi) + \Lambda(x, \xi) \pi_x \partial H(x, y(x, \xi), \xi)$; *that is,* x^* *is a weak GKKT point of* (1).

Proof. The main idea of the proof is to apply Lemma 3.2 to set-valued mapping $\mathcal{G}(x, \xi)$ on the compact set $\mathcal{C}$. First, observe that, under condition (27), it is easy to prove that $\|\mathcal{G}(x, \xi)\|$ is bounded by $\kappa_0(\xi) + \kappa_0(\xi)^2$. Because the latter is integrable by assumption, then $\mathbb{E}[\mathcal{G}(x, \xi)]$ is well defined for $x \in \mathcal{C}$.

By Lemma 3.3, we only need to prove that w.p. 1 point x^* satisfies that for any (small) $\gamma > 0$, there exists a $\delta > 0$ such that

$$0 \in \mathbb{E}[\mathcal{G}_{2\delta}(x^*, \xi)] + \mathcal{N}_{\mathcal{X}}(x^*) + 2\gamma \boldsymbol{B}, \tag{28}$$

where $\mathcal{G}_\delta(x^*, \xi) := \bigcup_{x \in S(x^*, \delta)} \mathcal{G}(x, \xi)$. Under the assumptions of Lemma 2.1, there exists a unique $y(x, \xi)$ such that $H(x, y(x, \xi), \xi) = 0$ for every $(x, \xi) \in \mathcal{X} \times \Xi$. Because $\partial H(\cdot, y(\cdot, \xi), \xi)$ is an upper-semicontinuous, compact set-valued mapping and by assumption $\pi_y \partial H(x, y, \xi)$ is uniformly nonsingular, then $\mathcal{G}(\cdot, \cdot)$ is also an upper-semicontinuous and compact set-valued mapping on $\mathcal{X}$ for every $\xi \in \Xi$. Moreover, for $x \in \mathcal{C}$, we have from (27) that $\mathcal{G}(x, \xi)$ is uniformly dominated by an integrable function $\kappa_0(\xi)(1 + \kappa_0(\xi))$. Because $\mathcal{C}$ is compact, it follows from Lemma 3.2 that for any $\gamma > 0$, there exists $\delta > 0$ such that

$$\overline{\lim_{N \to \infty}} \frac{1}{N} \sum_{i=1}^N \mathcal{G}(x, \xi^i) \subset \mathbb{E}[\mathcal{G}_\delta(x, \xi)] + \gamma \boldsymbol{B}, \quad \text{w.p. } 1 \tag{29}$$

uniformly for $x \in \mathcal{C}$. Let x^* be an accumulation point of $\{x_N\}$. Assume without loss of generality that $\{x_N\} \to \{x^*\}$ and $\{x_N\} \subset \mathcal{C}$. Because x_N is a GKKT point of problem (12), by definition,

$$0 \in \frac{1}{N} \sum_{i=1}^N \mathcal{G}(x_N, \xi^i) + \mathcal{N}_{\mathcal{X}}(x_N). \tag{30}$$

In what follows, we estimate term $\mathbb{D}((1/N) \sum_{i=1}^N \mathcal{G}(x_N, \xi^i), \mathbb{E}[\mathcal{G}_{2\delta}(x^*, \xi)] + 2\gamma \boldsymbol{B})$. Note that

$$\mathbb{D}\bigg(\frac{1}{N} \sum_{i=1}^N \mathcal{G}(x_N, \xi^i), \mathbb{E}[\mathcal{G}_{2\delta}(x^*, \xi)] + 2\gamma \boldsymbol{B}\bigg) \le \mathbb{D}\bigg(\frac{1}{N} \sum_{i=1}^N \mathcal{G}(x_N, \xi^i), \frac{1}{N} \sum_{i=1}^N \mathcal{G}_\delta(x^*, \xi^i) + \gamma \boldsymbol{B}\bigg) + \mathbb{D}\bigg(\frac{1}{N} \sum_{i=1}^N \mathcal{G}_\delta(x^*, \xi^i) + \gamma \boldsymbol{B}, \mathbb{E}[\mathcal{G}_{2\delta}(x^*, \xi)] + 2\gamma \boldsymbol{B}\bigg).$$

By (29), the second term at the right-hand side of the equation tends to zero w.p. 1 as $N \to \infty$. On the other hand, let N be sufficiently large that $x_N \in S(x^*, \delta)$. Then, by definition, $\mathscr{G}(x_N, \xi^i) \subset \mathscr{G}_\delta(x^*, \xi^i)$; hence,

$$\mathbb{D}\left(\frac{1}{N}\sum_{i=1}^{N}\mathscr{G}(x_N, \xi^i), \frac{1}{N}\sum_{i=1}^{N}\mathscr{G}_\delta(x^*, \xi^i) + \gamma\boldsymbol{B}\right) = 0 \tag{31}$$

for every N sufficiently large. This shows

$$\lim_{N\to\infty}\mathbb{D}\left(\frac{1}{N}\sum_{i=1}^{N}\mathscr{G}(x_N, \xi^i), \mathbb{E}[\mathscr{G}_{2\delta}(x^*, \xi)] + 2\gamma\boldsymbol{B}\right) = 0, \quad \text{w.p. 1.}$$

Combining this with (30), we obtain (28). □

Note that if the condition (27) holds in a compact subset of $\mathscr{X}$ that contains all accumulation points of $\{x_N\}$, then the result of Theorem 3.1 can be strengthened to that w.p. 1 any accumulation point of $\{x_N\}$ is a weak GKKT point of (1).

From a computational perspective, it might be more preferable to consider an ϵ-generalized Jacobian rather than a Clarke generalized Jacobian because the latter is difficult to compute for general locally Lipschitz functions, whereas the former may be computed by a smoothing method or an approximation method. See Ralph and Xu [21] and references therein for a detailed discussion on this regard. Consequently, we may consider the following weak ϵ-GKKT condition

$$\begin{aligned}0 \in \frac{1}{N}\sum_{i=1}^{N}[&\nabla_x f(x, y(x, \xi^i), \xi^i) - \nabla_y f(x, y(x, \xi^i), \xi^i)\,\mathrm{conv}([\pi_y\partial^\epsilon H(x, y(x, \xi^i), \xi^i)]^{-1})\\ &\cdot \pi_x\partial^\epsilon H(x, y(x, \xi^i), \xi^i)] + \mathscr{N}_{\mathscr{X}}(x)\end{aligned} \tag{32}$$

where $\epsilon > 0$, $\partial^\epsilon H$ denotes the ϵ-generalized Jacobian of H. Because

$$[\pi_y\partial^\epsilon H(x, y(x, \xi^i), \xi^i)]^{-1}\pi_x\partial^\epsilon H(x, y(x, \xi^i), \xi^i) \to [\pi_y\partial H(x, y(x, \xi^i), \xi^i)]^{-1}\pi_x\partial H(x, y(x, \xi^i), \xi^i)$$

as $\epsilon \to 0$, the weak GKKT condition (32) reduces to (13).

We say that x^ϵ is an ϵ-*weak GKKT point* if it satisfies (32). We have the following result for the ϵ-weak GKKT sequence.

COROLLARY 3.2. *Let the assumptions of Lemma 2.1 hold. Let $\{x_N^\epsilon\}$ be a sequence of ϵ-weak GKKT points that satisfies (32), and let x^* be an accumulation point. Let Assumption 3.1 hold with $\mathscr{C} := S(x^*, \eta) \cap \mathscr{X}$ where $\eta > 0$ is a small number. Then, w.p. 1, x^* satisfies*

$$0 \in \mathbb{E}[\mathscr{R}^\epsilon(x^*, \xi)] + \mathscr{N}_{\mathscr{X}}(x^*),$$

where

$$\mathscr{R}^\epsilon(x, \xi) := \nabla_x f(x, y(x, \xi), \xi) - \nabla_y f(x, y(x, \xi), \xi)\,\mathrm{conv}([\pi_y\partial^\epsilon H(x, y(x, \xi), \xi)]^{-1})\pi_x\partial^\epsilon H(x, y(x, \xi), \xi).$$

3.4. Convergence of optimal solutions. Another useful way to investigate the convergence of the SAA method is to look into the rate at which a local minimizer or global minimizer of the true problem (1) becomes that of the SAA problem as the sample size N increases. Here we restrict our discussion to global minimizers.

Let $v(x, \xi) := f(x, y(x, \xi), \xi)$ and $\vartheta(x) := \mathbb{E}[v(x, \xi)]$ where $y(x, \xi)$ is the unique solution of (2). Obviously, $\vartheta(x)$ is locally Lipschitz and directionally differentiable.

Let $\delta > 0$ be a constant. We say that $x^\delta \in \mathscr{X}$ is a δ-*global minimizer* of (1), if $\vartheta(x^\delta) \le \vartheta^* + \delta$, where ϑ^* denotes the global minimum of (1). For the convergence analysis, we need the following result, which is a special case of the well-known Berge's stability theorem.

LEMMA 3.4. *Consider a general constrained minimization problem*

$$\begin{aligned}&\min \quad p(x)\\ &\ s.t. \quad x \in X,\end{aligned}$$

where p: $\mathbb{R}^n \to \mathbb{R}$ is continuous and X is a subset of $\mathbb{R}^n$, and a perturbed program

$$\begin{aligned}&\min \quad \tilde{p}(x)\\ &\ s.t. \quad x \in X,\end{aligned}$$

where $\tilde{p}$: $\mathbb{R}^n \to \mathbb{R}$ is continuous and $|\tilde{p}(x) - p(x)| \le \delta, \forall\, x \in X$. Suppose that x^* is a global minimizer of $p(x)$ over X, and $\tilde{x}^*$ is a global minimizer of $\tilde{p}(x)$ over X. Then $|p(x^*) - \tilde{p}(\tilde{x}^*)| \le \delta$.

This lemma indicates that a global minimizer of p is a δ-global minimizer of $\tilde{p}$ in that

$$|\tilde{p}(x^*) - \tilde{p}(\tilde{x}^*)| \le |\tilde{p}(x^*) - p(x^*)| + |p(x^*) - \tilde{p}(\tilde{x}^*)| \le 2\delta.$$

Note that, strictly speaking, x^* is a 2δ-global minimizer of $\tilde{p}$. However, by convention, we call it a δ-global minimizer.

Theorem 3.2. *Suppose that the set of global minimizers of (1) is bounded and contained in a compact set $\mathscr{C}$ on which the boundedness condition (27) of Assumption 3.1 holds. Suppose also that the assumptions of Lemma 2.1 are satisfied. Then,*

(i) *$v(\cdot, \xi)$ is Lipschitz continuous on $\mathscr{C}$; that is, there exists $L_0(\xi) > 0$ such that*

$$|v(x', \xi) - v(x'', \xi)| \le L_0(\xi)\|x' - x''\|, \quad \forall\, x', x'' \in \mathscr{C}.$$

In addition, the moment-generating function $M_{L_0}(t) := \mathbb{E}[e^{tL_0(\xi)}]$ of $L_0(\xi)$ is finite valued for all t in a neighborhood of zero.

(ii) *With probability approaching one exponentially fast with increase of the sample size N, a global minimizer of (1) becomes a δ-global minimizer of (12).*

Proof. Part (i). Because $\mathscr{C}$ is compact, it follows from part (i) of Lemma 2.1 that the locally Lipschitz-continuous implicit function $y(\cdot, \xi)$ is globally Lipschitz on $\mathscr{C}$. By part (ii) of Lemma 2.1 and (8),

$$\partial_x v(x, \xi) \subset \nabla_x f(x, y(x, \xi), \xi) + \Lambda(x, \xi)\pi_x \partial H(x, y(x, \xi), \xi).$$

Using condition (27), we get an estimate on the upper bound for $\|\partial_x v(x, \xi)\|$, which is $L_0(\xi) := \kappa_0(\xi)(1 + \kappa_0(\xi))$ where $\kappa_0(\xi)$ is defined as in Theorem 3.1. By assumption, $\mathbb{E}[L_0(\xi)] < \infty$. Then, it follows from the Lebesgue dominated convergence theorem that the moment-generating function $M_{L_0}(t)$ of $L_0(\xi)$ is finite valued for all t in a neighborhood of zero, and it is infinitely differentiable at $t = 0$ (Shapiro [30, §8]).

Part (ii). By part (i), it follows from Shapiro and Xu [34, Theorem 5.1] that for any $\tau > 0$, there exist positive constant $C = C(\tau)$ and $\beta(\tau)$, independent of N, such that

$$\text{Prob}\left\{ \sup_{x \in \mathscr{C}} |f_N(x) - \vartheta(x)| > \tau \right\} \le C(\tau)e^{-N\beta(\tau)}. \tag{33}$$

Let x^* denote a global minimizer of (1), and let x_N denote a global minimizer of (12). From (33), it follows by Lemma 3.4 that

$$\text{Prob}\{|f_N(x^N) - \vartheta(x^*)| > \tau\} \le C(\tau)e^{-N\beta(\tau)}.$$

This implies that with a probability of at least $1 - C(\tau)e^{-N\beta(\tau)}$, a global minimizer of (1) becomes a τ-global minimizer of (12). ☐

Note that we may make a statement analogous to part (ii) of Theorem 3.2 for a local minimizer; that is, if x^* is a local minimizer of (1), then with probability approaching one exponentially fast with increase of the sample size N, x^* becomes a δ-local minimizer of (12). This follows straightforwardly from part (ii) considering a closed neighborhood of a local minimizer so that it becomes a global minimizer within the neighborhood.

4. SAA of a smoothed problem. In this section, we investigate the convergence analysis of smoothed SAA problems. Consider the following smoothed SAA problem

$$\begin{aligned} \min_{\substack{x \in \mathscr{X} \\ y^1, \ldots, y^N \in \mathbb{R}^n}} \quad & \frac{1}{N}\sum_{i=1}^{N} f(x, y^i, \xi^i) \\ \text{s.t.} \quad & \widehat{H}(x, y^i, \xi^i, \epsilon) = 0, \quad i = 1, \ldots, N, \end{aligned} \tag{34}$$

where $\epsilon \neq 0$. The implicit version of (34) can be written as

$$\min_{x \in \mathscr{X}} \hat{f}_N(x, \epsilon) := \frac{1}{N}\sum_{i=1}^{N} f(x, \hat{y}(x, \xi^i, \epsilon), \xi^i), \tag{35}$$

where $\hat{y}(x, \xi^i, \epsilon)$ denotes the unique solution of $\widehat{H}(x, y, \xi^i, \epsilon) = 0$ for $i = 1, \ldots, N$.

4.1. Convergence of KKT points. The KKT conditions of (34) can be written as follows

$$0 \in \frac{1}{N}\sum_{i=1}^{N} \nabla_x f(x, y^i, \xi^i) + \sum_{i=1}^{N} \lambda^i \nabla_x \widehat{H}(x, y^i, \xi^i, \epsilon) + \mathcal{N}_{\mathcal{X}}(x), \tag{36}$$

$$0 = \frac{1}{N}\begin{pmatrix} \nabla_y f(x, y^1, \xi^1) \\ \vdots \\ \nabla_y f(x, y^N, \xi^N) \end{pmatrix} + \begin{pmatrix} \lambda^1 \nabla_y \widehat{H}(x, y^1, \xi^1, \epsilon) \\ \vdots \\ \lambda^N \nabla_y \widehat{H}(x, y^N, \xi^N, \epsilon) \end{pmatrix}, \tag{37}$$

$$0 = \widehat{H}(x, y^i, \xi^i, \epsilon), \quad i = 1, \ldots, N. \tag{38}$$

Because $\nabla_y \widehat{H}(x, y^i, \xi^i, \epsilon)$, $i = 1, \ldots, N$, is nonsingular, we can solve λ^i, $i = 1, \ldots, N$, from (37)

$$\lambda^i = -\frac{1}{N}\nabla_y f(x, y^i, \xi^i) \nabla_y \widehat{H}(x, y^i, \xi^i, \epsilon)^{-1}, \quad i = 1, \ldots, N.$$

To ease the notation, let

$$\hat{\Lambda}(x, \xi, \epsilon) := -\nabla_y f(x, \hat{y}(x, \xi, \epsilon), \xi)\,\mathrm{conv}[\pi_y \partial \widehat{H}(x, \hat{y}(x, \xi, \epsilon), \xi, \epsilon)]^{-1}.$$

Obviously, when $\epsilon \neq 0$, $\mathrm{conv}[\pi_y \partial \widehat{H}(x, \hat{y}(x, \xi^i, \epsilon), \xi, \epsilon)]^{-1}$ reduces to $[\nabla_y \widehat{H}(x, \hat{y}(x, \xi^i, \epsilon), \xi, \epsilon)]^{-1}$, and hence $\hat{\Lambda}(x, \xi, \epsilon)$ becomes a singleton! Observe also that $\nabla_y \widehat{H}(x, y^i, \xi^i, \epsilon)$ can be written as $\nabla_y \widehat{H}(x, y^i, \xi^i, \epsilon)$. Replacing λ^i in (36) with $(1/N)\hat{\Lambda}(x, \xi, \epsilon)$ and writing y^i as $\hat{y}(x, \xi^i, \epsilon)$, we obtain

$$0 \in \frac{1}{N}\sum_{i=1}^{N}[\nabla_x f(x, \hat{y}(x, \xi^i, \epsilon), \xi^i) + \hat{\Lambda}(x, \xi^i, \epsilon)\nabla_x \widehat{H}(x, \hat{y}(x, \xi^i, \epsilon), \xi^i, \epsilon)] + \mathcal{N}_{\mathcal{X}}(x). \tag{39}$$

Note that in numerical implementation, there are two ways to select ϵ in (34). One is to fix ϵ. The other is to choose $\epsilon = \epsilon_N$ and let ϵ_N tend to zero as $N \to \infty$. In Theorem 4.1, we establish the convergence results of KKT points in both cases. We need the following assumption.

ASSUMPTION 4.1. *Let $\mathcal{C}$ be a compact subset of $\mathcal{X}$. There exists a function $\kappa_1(\xi)$ and a constant $\epsilon_0 > 0$ such that for all $(x, \xi, \epsilon) \in \mathcal{C} \times \mathbb{R}^n \times \Xi \times [-\epsilon_0, \epsilon_0]$,*

$$\max\{\|\nabla_x f(x, \hat{y}(x, \xi, \epsilon), \xi)\|, \|\hat{\Lambda}(x, \xi, \epsilon)\|, \|\pi_x \partial \widehat{H}(x, \hat{y}(x, \xi, \epsilon), \xi, \epsilon)\|\} \leq \kappa_1(\xi), \tag{40}$$

where $\mathbb{E}[\kappa_1(\xi) + \kappa_1(\xi)^2] < \infty$.

Note that set $\mathcal{C}$ in the assumption is deliberately left unspecified, and it will be specified in the context of various convergence theorems in this and following subsections. Note also that $\|\hat{\Lambda}(x, \xi, \epsilon)\| \leq \|\nabla_y f(x, \hat{y}(x, \xi), \xi, \epsilon)\| \|[\pi_y \partial \widehat{H}(x, \hat{y}(x, \xi, \epsilon), \xi)]^{-1}\|$, therefore, (40) holds if $\pi_y \partial \widehat{H}$ is uniformly nonsingular and $\nabla_{x,y} f$ and $\pi_x \partial H$ are bounded by a positive integrable function. We are now ready to state one of the main convergence results of our paper.

THEOREM 4.1. *Let the conditions of Lemma 2.3 hold.*

(i) *Let $\epsilon \neq 0$ be fixed, let $\{x_N(\epsilon)\}$ be a sequence of KKT points that satisfies (39) and let $x^*(\epsilon)$ be an accumulation point $x^*(\epsilon)$. If there exists a compact set $\mathcal{C}$ that contains a neighborhood of $x^*(\epsilon)$ such that Assumption 4.1 holds, then, w.p. 1, $x^*(\epsilon)$ satisfies*

$$0 \in \mathbb{E}[\hat{a}(x^*(\epsilon), \xi, \epsilon)] + \mathcal{N}_{\mathcal{X}}(x^*(\epsilon)), \tag{41}$$

where $\hat{a}(x, \xi, \epsilon) := \nabla_x f(x, \hat{y}(x, \xi, \epsilon), \xi) + \hat{\Lambda}(x, \xi, \epsilon)\nabla_x \widehat{H}(x, \hat{y}(x, \xi, \epsilon), \xi, \epsilon)$.

(ii) *Let $\epsilon = \epsilon_N$ where $\epsilon_N \to 0$ as $N \to \infty$; let x^* be an accumulation point of $\{x(\epsilon_N)\}$. If Assumption 4.1 holds on a compact set $\mathcal{C}$ that contains a neighborhood of x^* w.p. 1, then, w.p. 1, x^* satisfies*

$$0 \in \mathbb{E}[\mathcal{L}(x^*, \xi)] + \mathcal{N}_{\mathcal{X}}(x^*), \tag{42}$$

where $\mathcal{L}(x, \xi) := \nabla_x f(x, y(x, \xi), \xi) + \hat{\Lambda}(x, \xi, 0)\pi_x \partial \widehat{H}(x, y(x, \xi), \xi, 0)$.

Before providing a proof, we note that from (42) we can easily see that if $\widehat{H}$ satisfies Jacobian consistency (Chen et al. [7], Ralph and Xu [21]) conditions, that is, $\pi_x\partial\widehat{H}(x, y(x,\xi),\xi,0) \subset \pi_x\partial H(x, y(x,\xi),\xi)$ and $\pi_y\partial\widehat{H}(x, y(x,\xi),\xi,0) \subset \pi_y\partial H(x, y(x,\xi),\xi)$, then $\hat{\Lambda}(x,\xi,0)\subset\Lambda(x,\xi)$. Consequently,

$$\mathscr{L}(x^*,\xi)\subset\nabla_x f(x^*, y(x^*,\xi),\xi)+\Lambda(x^*,\xi)\pi_x\partial H(x^*, y(x^*,\xi),\xi),$$

which implies (7); that is, x^* is a weak GKKT point of (1).

Proof of Theorem 4.1. Observe first that both (41) and (42) are well defined under condition (40).

Part (i). By assumption, $\hat{a}(\cdot,\xi,\epsilon)$ is continuous on $\mathscr{X}$ for every $\xi\in\Xi$. Moreover, from (40) it follows that $\hat{a}(x,\xi,\epsilon)$ is bounded by $\kappa_1(\xi)(1+\kappa_1(\xi))$, which has a finite expected value. Because $\mathscr{X}$ is compact and ϵ is a constant, by applying Rubinstein and Shapiro [26, Lemma A1] to $\hat{a}(x,\xi,\epsilon)$ componentwise, we have

$$\lim_{N\to\infty}\max_{x\in\mathscr{X}}\left\|\frac{1}{N}\sum_{i=1}^{N}\hat{a}(x,\xi^i,\epsilon)-b(x,\epsilon)\right\|_\infty=0,\quad \text{w.p. 1} \tag{43}$$

where $\|\cdot\|_\infty$ denotes the infinity norm of a vector and $b(x,\epsilon)=\mathbb{E}[\hat{a}(x,\xi,\epsilon)]$. Assume without loss of generality that $x_N(\epsilon)\to x^*(\epsilon)$ as $N\to\infty$. Because $x_N(\epsilon)$ is a KKT point that satisfies (39), then

$$0\in\frac{1}{N}\sum_{i=1}^{N}\hat{a}(x_N(\epsilon),\xi^i,\epsilon)+\mathscr{N}_{\mathscr{X}}(x_N(\epsilon)),$$

which can be rewritten as

$$\begin{aligned}0\in\ &\frac{1}{N}\sum_{i=1}^{N}\hat{a}(x^*(\epsilon),\xi^i,\epsilon)+\left[\frac{1}{N}\sum_{i=1}^{N}\hat{a}(x_N(\epsilon),\xi^i,\epsilon)-b(x_N(\epsilon),\epsilon)\right]\\&+\left[b(x_N(\epsilon),\epsilon)-\frac{1}{N}\sum_{i=1}^{N}\hat{a}(x^*(\epsilon),\xi^i,\epsilon)\right]+\mathscr{N}_{\mathscr{X}}(x_N(\epsilon)).\end{aligned}$$

By the strong law of large numbers, the first term on the right-hand side of the equation above tends to $\mathbb{E}[\hat{a}(x^*(\epsilon),\xi,\epsilon)]$ w.p. 1, the second term tends to zero w.p. 1 due to (43), and the third term tends to zero w.p. 1 as N tends to infinity. This shows part (i).

Part (ii). To show that x^* satisfies (42), it suffices, by Lemma 3.3, to prove that for any $\gamma>0$, there exists $\delta>0$, such that x^* satisfying

$$0\in\mathbb{E}[\mathscr{L}_{2\delta}(x,\xi)]+\mathscr{N}_{\mathscr{X}}(x)+2\gamma\boldsymbol{B},$$

where

$$\mathscr{L}_\delta(x,\xi):=\bigcup_{x'\in S(x,\delta)}\mathscr{L}(x',\xi).$$

We treat ϵ in $\hat{a}(x,\xi,\epsilon)$ as a variable. Let

$$\hat{\mathscr{A}}(x,\xi,\epsilon):=\begin{cases}\hat{a}(x,\xi,\epsilon), & \epsilon\neq 0,\\ \mathscr{L}(x,\xi), & \epsilon=0.\end{cases}$$

By assumption, $\hat{\mathscr{A}}(\cdot,\xi,\cdot)$: $\mathscr{X}\times[-\epsilon_0,\epsilon_0]\to 2^{\mathbb{R}^n}$ is an upper-semicontinuous and compact set-valued mapping for every $\xi\in\Xi$. From (40), it follows that $\hat{\mathscr{A}}(x,\xi,\epsilon)$ is bounded by $\kappa_1(\xi)(1+\kappa_1(\xi))$ for $x\in\mathscr{C}$ and $\mathbb{E}[\kappa_1(\xi)(1+\kappa_1(\xi))]<\infty$. By Lemma 3.2, for any $\gamma>0$, there exists $\delta>0$ such that

$$\overline{\lim_{N\to\infty}}\frac{1}{N}\sum_{i=1}^{N}\hat{\mathscr{A}}(x,\xi^i,\epsilon)\subset\mathbb{E}[\hat{\mathscr{A}}_\delta(x,\xi,\epsilon)]+\gamma\boldsymbol{B}\quad\text{w.p. 1} \tag{44}$$

uniformly with respect to $(x,\epsilon)\in\mathscr{C}\times[-\epsilon_0,\epsilon_0]$, where

$$\hat{\mathscr{A}}_\delta(x,\xi,\epsilon)=\bigcup_{(x',\epsilon')\in S((x,\epsilon),\delta)}\mathscr{A}(x',\xi,\epsilon').$$

On the other hand, because $x_N(\epsilon_N)$ is a KKT point of problem (34) with $\epsilon=\epsilon_N$,

$$0\in\frac{1}{N}\sum_{i=1}^{N}\hat{\mathscr{A}}(x_N(\epsilon_N),\xi^i,\epsilon_N)+\mathscr{N}_{\mathscr{X}}(x_N(\epsilon_N)). \tag{45}$$

Assume without loss of generality that $x_N(\epsilon_N)\to x^*$ as $N\to\infty$ and $x_N(\epsilon_N)\subset\mathscr{C}$. It suffices to prove that

$$\mathbb{D}\left(\frac{1}{N}\sum_{i=1}^{N}\hat{\mathscr{A}}(x_N(\epsilon_N),\xi^i,\epsilon_N),\mathbb{E}[\mathscr{L}_{2\delta}(x^*,\xi)]+2\gamma\boldsymbol{B}\right)\to 0,\quad\text{as } N\to\infty.$$

Note that

$$\mathbb{D}\left(\frac{1}{N}\sum_{i=1}^{N}\hat{\mathcal{A}}(x_N(\epsilon_N),\xi^i,\epsilon_N),\mathbb{E}[\mathcal{L}_{2\delta}(x^*,\xi)]+2\gamma\boldsymbol{B}\right)$$
$$\leq \mathbb{D}\left(\frac{1}{N}\sum_{i=1}^{N}\hat{\mathcal{A}}(x_N(\epsilon_N),\xi^i,\epsilon_N),\frac{1}{N}\sum_{i=1}^{N}\hat{\mathcal{A}}_\delta(x^*,\xi^i,0)+\gamma\boldsymbol{B}\right)$$
$$+\mathbb{D}\left(\frac{1}{N}\sum_{i=1}^{N}\hat{\mathcal{A}}_\delta(x^*,\xi^i,0)+\gamma\boldsymbol{B},\mathbb{E}[\mathcal{L}_{2\delta}(x^*,\xi)]+2\gamma\boldsymbol{B}\right).$$

It follows from (44) that the second term on the right-hand side of the equation above tends to zero w.p. 1 as $N\to\infty$. On the other hand, because

$$\hat{\mathcal{A}}(x_N(\epsilon_N),\xi^i,\epsilon_N)\subset\hat{\mathcal{A}}_\delta(x^*,\xi^i,0)$$

for $(x_N(\epsilon_N),\epsilon_N)\in S((x^*,0),\delta)$, the first term on the right is zero for N sufficiently large. The conclusion follows by the relation above with (45). The proof is complete. □

4.2. Convergence of optimal solutions. Similar to the analysis in §3, we can present the convergence of optimal solutions of (34) to those of the true problem (1). Let $\hat{v}(x,\xi,\epsilon):=f(x,\hat{y}(x,\xi,\epsilon),\xi)$.

Theorem 4.2. *Suppose that the set of global minimizers of* (1) *is bounded and contained in a compact set* $\mathcal{C}$ *on which the boundedness condition* (40) *of Assumption* 4.1 *holds. Suppose also that the assumptions of Lemma* 2.1 *are satisfied. Then*

(i) $\hat{v}(x,\xi,\epsilon)$ *is uniformly calm* (*Rockafellar and Wets* [25]) *at zero; that is, there exists a constant* $C_1>0$ *such that*

$$|\hat{v}(x,\xi,\epsilon)-v(x,\xi)|\leq C_1|\epsilon|,\quad \forall x\in\mathcal{C} \tag{46}$$

for $\epsilon>0$ *sufficiently small.*

(ii) *With probability approaching one exponentially fast with increase of the sample size* N, *a global minimizer of* (1) *becomes a* δ*-global minimizer of* (35).

Proof. Part (i). Under (40) and assumptions of Lemma 2.3, there exists a unique implicit function $\hat{y}(x,\xi,\epsilon)$ for $(x,\xi,\epsilon)\in\mathcal{C}\times\Xi\times(0,\epsilon_0]$ such that $\widehat{H}(x,\hat{y}(x,\xi,\epsilon),\xi,\epsilon)=0$. Because for $\epsilon>0$,

$$\nabla_\epsilon\hat{v}(x,\xi,\epsilon)=-\nabla_y f(x,\hat{y}(x,\xi,\epsilon),\xi)\nabla_y\widehat{H}(x,\hat{y}(x,\xi,\epsilon),\xi,\epsilon)^{-1}\nabla_\epsilon\widehat{H}(x,\hat{y}(x,\xi,\epsilon),\xi,\epsilon).$$

Under condition (40), $\|-\nabla_y f(x,\hat{y}(x,\xi,\epsilon),\xi)\nabla_y\widehat{H}(x,\hat{y}(x,\xi,\epsilon),\xi,\epsilon)^{-1}\|$ is bounded by $\kappa_1(\xi)^2$. Note that by assumption, $\nabla_\epsilon\widehat{H}(x,\hat{y}(x,\xi,\epsilon),\xi,\epsilon)$ is bounded, and the bound is denoted by L. Note also that

$$\|\hat{v}(x,\xi,\epsilon)-v(x,\xi)\|\leq\int_0^1\|\nabla_\epsilon\hat{v}(x,\xi,t\epsilon)\|\epsilon\,dt.$$

Obviously, the right-hand side is bounded by $L\epsilon$. The conclusion follows.

The proof of part (ii) is similar to those of Theorem 3.2. We omit the details. □

As in Theorem 3.2, we can make a statement for a local minimizer, that is, with probability approaching one exponentially fast with increase of the sample size N, a local minimizer of (1) becomes a local minimizer (35).

5. Application in SMPECs. In this section, we apply the theory developed in the preceding sections to the following stochastic mathematical programs with equilibrium constraints (SMPEC):

$$\begin{aligned}&\min\ \mathbb{E}[f(x,y(x,\xi),\xi)]\\&\text{s.t.}\ \ x\in\mathcal{X},\end{aligned}\tag{47}$$

where $\mathcal{X}$ is a nonempty subset of $\mathbb{R}^m$, $f\colon\mathbb{R}^m\times\mathbb{R}^n\times\mathbb{R}^k\to\mathbb{R}$, $F\colon\mathbb{R}^m\times\mathbb{R}^n\times\mathbb{R}^k\to\mathbb{R}^n$ are continuously differentiable, $\xi\colon\Omega\to\mathbb{R}^k$ is a vector of random variables defined on sample space $(\Omega,\mathcal{F},P)$, $\mathbb{E}$ denotes the mathematical expectation; $y(x,\xi)$ solves the following complementarity problem

$$0\leq y\perp F(x,y,\xi(\omega))\geq 0,\quad \text{for }\omega\in\Omega.$$

It is known that the complementarity constraint in (47) can be reformulated as a system of equations using an elementary function ϕ: $\mathbb{R}\times\mathbb{R}\to\mathbb{R}$,

$$\Phi(x,y,\xi):=\begin{bmatrix}\phi(y_1,F_1(x,y,\xi))\\ \vdots \\ \phi(y_n,F_n(x,y,\xi))\end{bmatrix}=0, \tag{48}$$

where $\phi(a,b)=0$ if and only if $a,b\geq 0$ and $ab=0$. Such a function is known as an NCP function (where NCP stands for nonlinear complementarity problems). There are many NCP functions available in the literature. See, for instance, Kanzow [13], Luo et al. [16], and Facchinei and Pang [9]. Here we only consider the two most popular NCP functions. One is the min-function, which is defined as $\phi(a,b)=\min(a,b)$. The function is globally Lipschitz continuous, and it is continuously differentiable everywhere except at the line $a=b$. The other is the Fischer-Burmeister function (Fischer [10]), which is defined as $\phi(a,b)=\sqrt{a^2+b^2}-(a+b)$. This function is globally Lipschitz continuous, and it is continuously differentiable everywhere except at $(0,0)$. Using (48), we can reformulate (47) as a stochastic minimization problem with nonsmooth equality constraints:

$$\min_{x\in\mathscr{X}}\mathbb{E}[f(x,y(x,\xi),\xi)]\quad\text{where } y(x,\xi) \text{ solves } \Phi(x,y,\xi)=0. \tag{49}$$

In this section, we consider the case where F is *uniformly strongly monotone* with respect to y; that is, there exists a constant $\alpha>0$ such that

$$(F(x,y',\xi)-F(x,y'',\xi))^T(y'-y'')\geq\alpha\|y'-y''\|^2$$

for any $y',y''\in\mathbb{R}^n$, $x\in\mathscr{X}$, and $\xi\in\Xi$. Under the monotone condition, it is known that $\nabla_yF(x,y,\xi)$ is uniformly positive definite for all $y\in\mathbb{R}^n$, $x\in\mathscr{X}$, and $\xi\in\Xi$. Hence, we can easily see that the conditions of Lemma 2.1 are satisfied. Therefore, there exists a unique locally Lipschitz function y: $\mathscr{X}\times\Xi\to\mathbb{R}^n$ such that $\Phi(x,y(x,\xi),\xi)=0$. Let $\xi^1,\ldots,\xi^N$ be an i.i.d. sample of ξ. Consider the following SAA program for (49)

$$\begin{aligned}\min_{\substack{x\in\mathscr{X}\\ y^1,\ldots,y^N\in\mathbb{R}^n}}\ &\frac{1}{N}\sum_{i=1}^N f(x,y^i,\xi^i)\\ \text{s.t.}\ &\Phi(x,y^i,\xi^i)=0,\quad i=1,\ldots,N,\end{aligned} \tag{50}$$

and its reduced implicit form

$$\begin{aligned}\min\ & f_N(x):=\frac{1}{N}\sum_{i=1}^N f(x,y(x,\xi^i),\xi^i)\\ \text{s.t.}\ & x\in\mathscr{X},\end{aligned} \tag{51}$$

where $y(x,\xi^i)$ solves $\Phi(x,y^i,\xi^i)=0$. We have the convergence results for (51).

Proposition 5.1. *Let $F(x,y,\xi)$ be uniformly strongly monotone in y. Let $\nabla_xF(x,y(x,\xi),\xi)$ and $\nabla_{x,y}f(x,y(x,\xi),\xi)$ be uniformly bounded by $\kappa_2(\xi)$ for $x\in\mathscr{X}$, where $\mathbb{E}[\kappa_2(\xi)^2]<\infty$. Then*

(i) *if $\{x_N\}$ is a sequence of weak GKKT points of (51), then any accumulation point x^* of sequence $\{x_N\}$ satisfies*

$$0\in\mathbb{E}[\mathcal{J}(x^*,\xi)]+\mathcal{N}_{\mathscr{X}}(x^*),$$

where

$$\mathcal{J}(x,\xi):=\nabla_xf(x,y(x,\xi),\xi)+\Lambda(x,\xi)\pi_x\partial\Phi(x,y(x,\xi),\xi)$$

and $\Lambda(x,\xi):=-\nabla_yf(x,y(x,\xi),\xi)\,\mathrm{conv}[\pi_y\partial\Phi(x,y(x,\xi),\xi)]^{-1}$;

(ii) *if the set of global minimizers is bounded, then with probability approaching one exponentially fast with increase of the sample size N, a global minimizer of (47) becomes a δ-global minimizer of (51).*

Proof. We use Theorems 3.1 and 3.2 to prove the results. Observe that both the conditions and conclusion of part (i) of this proposition are stronger than their counterparts in Theorem 3.1 in that we can prove that

Assumption 3.1 holds for $H(x, y, \xi) := \Phi(x, y, \xi)$, with $\mathscr{C}$ being any compact subset of $\mathscr{X}$ that contains *all* accumulation points of $\{x_N\}$. We verify Assumption 3.1. By a simple calculation,

$$\pi_x \partial\Phi(x, y, \xi) \subset \operatorname{diag}(\partial_b \phi(y_i, F_i(x, y, \xi)))\nabla_x F(x, y, \xi)$$

and

$$\pi_y \partial\Phi(x, y, \xi) \subset \operatorname{diag}(\partial_a \phi(y_i, F_i(x, y, \xi))) + \operatorname{diag}(\partial_b \phi(y_i, F_i(x, y, \xi)))\nabla_y F(x, y, \xi),$$

where "diag" denotes a diagonal matrix. It is not difficult to observe that $\|\pi_x \partial\Phi(x, y, \xi)\| \leq \|\nabla_x F(x, y, \xi)\|$ Moreover, by Xu [37, Proposition 2.1], there exists a constant $C > 0$ such that $\|\pi_y \partial\Phi(x, y, \xi)^{-1}\| \leq C$. Thus,

$$\|\mathscr{J}(x, \xi)\| \leq \kappa_2(\xi) + C\kappa_2(\xi)^2, \quad \forall\, x \in \mathscr{X}, \quad \xi \in \Xi.$$

Because the right-hand side of the above equation is integrable by assumption, we have verified Assumption 3.1 for any $x \in \mathscr{X}$, and hence $\mathscr{C}$ can be taken as any compact subset of $\mathscr{X}$ that contains all accumulation points of $\{x_N\}$. The rest follows straightforwardly from Theorems 3.1 and 3.2. $\square$

Notice that Φ defined in (48) is a nonsmooth function. We investigate the smoothing NCP approach to solving (47). First, recall that a smoothing function ψ: $\mathbb{R} \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ of an NCP function $\phi(a, b)$ satisfies: (A1) $\psi(a, b, 0) = \phi(a, b)$; (A2) $\psi(a, b, c)$ is Lipschitz continuous and it is continuously differentiable everywhere except at $c = 0$; (A3) (strong Jacobian consistency, Chen et al. [7]) for $(a, b) \in \mathbb{R}^2$, $\pi_{a,b}\partial\psi(a, b, 0) = \partial\phi(a, b)$. In particular, it is known that a smoothing function of $\min(a, b)$ is $\psi(a, b, c) = -\frac{1}{2}(\sqrt{(a-b)^2 + c^2} - (a + b))$, and a smoothing function of the Fischer-Burmeister function is $\psi(a, b, c) = \sqrt{a^2 + b^2 + c^2} - (a + b)$. See, for instance, Kanzow [13].

Lemma 5.1. *Let $\psi(a, b, c)$ be the smoothing function of the Fischer-Burmeister function or min-function as defined above. Then, there exists $l > 0$ such that $\|\nabla_{a,b}\psi(a, b, \epsilon)\| \geq l$ for any $a, b \in \mathbb{R}$ and any $\epsilon \in (0, \epsilon_0]$ with $\epsilon_0 > 0$.*

The proof is elementary, and hence is moved to the appendix. By virtue of the smoothing function ψ, we can define a smoothing mapping Ψ: $\mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^k \times \mathbb{R}_+ \to \mathbb{R}^n$ of $\Phi(x, y, \xi)$ as $\Psi(x, y, \xi, \epsilon) := (\psi(y_1, F_1(x, y, \xi), \epsilon), \ldots, \psi(y_n, F_n(x, y, \xi), \epsilon))^T$. It is clear that $\Psi(x, y, \xi, 0) = \Phi(x, y, \xi)$. Consequently, we can consider a smoothing approximation of (49) as follows:

$$\begin{aligned} \min\ & \mathbb{E}[f(x, \hat{y}(x, \xi, \epsilon), \xi)] \\ \text{s.t.}\ & \hat{y}(x, \xi, \epsilon) \quad \text{solves } \Psi(x, y, \xi, \epsilon) = 0, \\ & x \in \mathscr{X}. \end{aligned} \tag{52}$$

Note that for $\epsilon \neq 0$, $\nabla_y \Psi(x, y, \xi, \epsilon) = D_a + D_b \nabla_y F(x, y, \xi)$, where $D_a = \operatorname{diag}(D_1^a, \ldots, D_n^a)$ and $D_b = \operatorname{diag}(D_1^b, \ldots, D_n^b)$ denote the diagonal $n \times n$ matrices with the following (i, i)th entries, respectively, $D_i^a = \nabla_a \psi(y_i, F_i(x, y, \xi), \epsilon)$; $D_i^b = \nabla_b \psi(y_i, F_i(x, y, \xi), \epsilon)$. By Xu [37, Proposition 4.2], Lemma 5.1, and the uniform strong monotonicity of $F(x, \cdot, \xi)$, we can see that conditions of Lemma 2.3 hold in this case. Hence, there exists a locally Lipschitz-continuous function on $\mathscr{X} \times \Xi \times [-\epsilon_0, \epsilon_0]$ such that $\Psi(x, y(x, \xi, \epsilon), \xi, \epsilon) = 0$, where ϵ_0 is a positive scalar. We consider the following smoothed SAA program of (52):

$$\begin{aligned} \min_{\substack{x \in \mathscr{X} \\ y^1, \ldots, y^N \in \mathbb{R}^n}}\ & \frac{1}{N}\sum_{i=1}^{N} f(x, y^i, \xi^i) \\ \text{s.t.}\ & \Psi(x, y^i, \xi^i, \epsilon) = 0, \quad i = 1, \ldots, N, \end{aligned} \tag{53}$$

and its implicit form

$$\begin{aligned} \min\ & \frac{1}{N}\sum_{i=1}^{N} f(x, \hat{y}(x, \xi^i, \epsilon), \xi^i) \\ \text{s.t.}\ & x \in \mathscr{X}, \end{aligned} \tag{54}$$

where $\hat{y}(x, \xi^i, \epsilon)$ solves $\Psi(x, y^i, \xi^i, \epsilon) = 0$. We apply the results derived in §4 to (53). The following result is derived by virtue of Theorem 4.1.

Proposition 5.2. *Let $F(x, y, \xi)$ be uniformly strongly monotone in y. Let $\nabla_x F(x, \hat{y}(x, \xi, \epsilon), \xi)$ and $\nabla f(x, \hat{y}(x, \xi, \epsilon), \xi)$ be uniformly bounded by $\kappa_2(\xi)$ for $x \in \mathscr{X}$, where $\mathbb{E}[\kappa_2(\xi)^2] < \infty$. Then*

(i) *for fixed $\epsilon > 0$, any accumulation point $x(\epsilon)$ of the sequence of KKT points $\{x_N(\epsilon)\}$ of (54) satisfies the following*

$$0 \in \mathbb{E}[\hbar(x(\epsilon), \xi, \epsilon)] + \mathcal{N}_{\mathscr{X}}(x(\epsilon)),$$

where

$$\hbar(x,\xi,\epsilon) := \nabla_x f(x,\hat{y}(x,\xi,\epsilon),\xi) + \hat{\Lambda}(x,\xi,\epsilon)\nabla_x \Psi(x,\hat{y}(x,\xi,\epsilon),\xi,\epsilon),$$

and

$$\hat{\Lambda}(x,\xi,\epsilon) := -\nabla_y f(x,\hat{y}(x,\xi,\epsilon),\xi)\nabla_y \Psi(x,\hat{y}(x,\xi,\epsilon),\xi,\epsilon)^{-1};$$

(ii) *for $\epsilon = \epsilon_N$ where $\epsilon_N \to 0$ as $N \to \infty$, any accumulation point x^* of the sequence of KKT points $\{x(\epsilon_N)\}$ of* (54) *satisfies*

$$0 \in \mathbb{E}[\mathcal{M}(x^*,\xi)] + \mathcal{N}_{\mathcal{X}}(x^*),$$

where $\mathcal{M}(x,\xi) := \nabla_x f(x,y(x,\xi),\xi) + \Lambda(x,\xi)\pi_x \partial\Phi(x,y(x,\xi),\xi)$;

(iii) *with probability approaching one exponentially fast with increase of the sample size N, a global minimizer of* (47) *becomes a δ-global minimizer of* (54).

Proof. It is easy to verify that (40) is satisfied by any $x \in \mathcal{X}$, and hence Assumption 4.1 holds for any compact subset $\mathcal{C}$ of $\mathcal{X}$, which contains all accumulation points of $\{x_N\}$. The results follow from Theorems 4.1 and 4.2. □

6. An example. In this section, we use a simple example in SMPEC to illustrate how our proposed smoothing SAA method works. Consider

$$\begin{aligned} \min\ & \mathbb{E}[f(x,y(x,\xi),\xi)] \\ \text{s.t.}\ & x_1 \geq 0, \quad x_2 \geq 0, \end{aligned} \tag{55}$$

where

$$f(x,y,\xi) := \tfrac{1}{6}x_1^4 - \tfrac{1}{12}(1+x_1)^4 - \tfrac{1}{3}x_2^3 + (x_1-1)^2 + (x_2-1)^2 + y_1^2 + y_2^2,$$

and $y(x,\xi)$ solves the following complementarity problem

$$0 \leq y \perp F(x,y,\xi) \geq 0,$$

where $F(x,y,\xi) := (y_1 - x_1 + \xi_1 - \xi_2, y_2 - x_2 + \xi_2)$ and ξ_1, ξ_2 are independent random variables, both having uniform distribution on $[0,1]$. The example is varied from an example in Shapiro and Xu [34]. Observe first that $\nabla_y F(x,y,\xi) = \left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$. Therefore, F is uniformly strongly monotone in y. Using the min-function, the complementary constraints can rewritten as:

$$H(x,y,\xi) := (\min(y_1, y_1 - x_1 + \xi_1 - \xi_2)\min(y_2, y_2 - x_2 + \xi_2)) = 0.$$

The equation has a unique solution $(y_1(x,\xi), y_2(x,\xi))$ where

$$y_1(x,\xi) = \begin{cases} x_1 - \xi_1 + \xi_2 & \text{if } x_1 \geq \xi_1 - \xi_2, \\ 0 & \text{otherwise;} \end{cases}$$

$$y_2(x,\xi) = \begin{cases} x_2 - \xi_2 & \text{if } x_2 \geq \xi_2, \\ 0 & \text{otherwise.} \end{cases}$$

Substituting them into the objective function, we have

$$\mathbb{E}[f(x,y(x,\xi),\xi)] = \begin{cases} (x_1-1)^2 + (x_2-1)^2, & \text{if } 1 \geq x_1 \geq 0,\ \ x_2 \geq 0, \\ (x_1-1)^4/12 + (x_1-1)^2 + (x_2-1)^2, & \text{if } x_1 \geq 1,\ \ x_2 \geq 0. \end{cases}$$

Obviously, the optimal solution is $(1,1)$ and the optimal value is 0. We now consider the smoothing method. The smoothed constraints are

$$\widehat{H}(x,y,\xi,\epsilon) := -\tfrac{1}{2}\begin{pmatrix} \sqrt{(x_1-\xi_1+\xi_2)^2+\epsilon^2} - (2y_1 - x_1 + \xi_1 - \xi_2) \\ \sqrt{(x_2-\xi_2)^2+\epsilon^2} - (2y_2 - x_2 + \xi_2) \end{pmatrix} = 0.$$

The equation has a unique solution $\hat{y}(x, \xi, \epsilon) = (\hat{y}_1(x, \xi, \epsilon), \hat{y}_2(x, \xi, \epsilon))$ where

$$\hat{y}_1(x, \xi, \epsilon) = \tfrac{1}{2}\big(x_1 - \xi_1 + \xi_2 + \sqrt{(x_1 - \xi_1 + \xi_2)^2 + \epsilon^2}\big),$$
$$\hat{y}_2(x, \xi, \epsilon) = \tfrac{1}{2}\big(x_2 - \xi_2 + \sqrt{(x_2 - \xi_2)^2 + \epsilon^2}\big).$$

We next check the boundedness assumption of Proposition 5.2. Because $\nabla_x f(x, y, \xi) = (\frac{2}{3}x_1^3 - \frac{1}{3}(1 + x_1)^3 + 2(x_1 - 1), -x_2^2 + 2(x_2 - 1))$, $\nabla_x F(x, y, \xi) = -\left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$, both of which are independent of y and ξ, it suffices to verify the boundedness condition for $\nabla_y f(x, \hat{y}(x, \xi, \epsilon), \xi)$. Because $\nabla_y f(x, y, \xi) = (2y_1, 2y_2)$, it is easy to verify that $\nabla_y f(x, \hat{y}(x, \xi, \epsilon), \xi)$ is bounded by an integrable function $2(|x_1 - \xi_1 + \xi_2| + |x_2 - \xi_2| + \epsilon)$. Alternatively, we can verify the boundedness condition (40) in Assumption 4.1 directly. Note that because

$$\nabla_x \widehat{H}(x, y, \xi, \epsilon) = -\frac{1}{2}\begin{pmatrix} \dfrac{x_1 - \xi_1 + \xi_2}{\sqrt{(x_1 - \xi_1 + \xi_2)^2 + \epsilon^2}} + 1 & 0 \\ 0 & \dfrac{x_2 - \xi_2}{\sqrt{(x_2 - \xi_2)^2 + \epsilon^2}} + 1 \end{pmatrix},$$

and $\nabla_y \widehat{H}(x, y, \xi, \epsilon) = \left(\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}\right)$, we can easily obtain that

$$\hat{\Lambda}(x, \xi, \epsilon) = -\nabla_y f(x, \hat{y}(x, \xi, \epsilon), \xi)\nabla_y \widehat{H}(x, \hat{y}(x, \xi, \epsilon), \xi, \epsilon)^{-1} = -2\hat{y}(x, \xi, \epsilon).$$

Obviously, $\|\hat{\Lambda}(x, \xi, \epsilon)\|$ and $\|\nabla_x \widehat{H}(x, y, \xi, \epsilon)\|$ are bounded by some positive measurable function.

We carry out some numerical experiments for the problem (55) with the smoothing SAA method (53). We solve the discretised problem by implementing mathematical programming codes in Matlab 6.5 installed in a PC with Windows XP operating system. We use the Matlab built-in solver *fmincon* for solving the smoothing SAA problems.

Note that, in general, optimal solutions and optimal values are unknown. Therefore, we need to discuss how to estimate the lower and upper bounds of the optimal objective value of the smoothing problem. The following method of constructing statistical lower and upper bounds was suggested in Santoso et al. [28].

Given $\epsilon > 0$, let $\widehat{w}(\epsilon)$ denote the optimal value of the smoothing problem and $\widehat{w}_N(\epsilon)$ the optimal value of the smoothing SAA problem. Let $\hat{\vartheta}(x, \epsilon) = \mathbb{E}[\hat{v}(x, \xi, \epsilon)]$ where $\hat{v}(x, \xi, \mu) = f(x, \hat{y}(x, \xi, \epsilon), \xi)$. It is known (Santoso et al. [28]) that $\mathbb{E}[\widehat{w}_N(\epsilon)] \le \widehat{w}(\epsilon)$. To estimate $\mathbb{E}[\widehat{w}_N(\epsilon)]$, we generate M independent samples of ξ, $\{\xi_j^1, \ldots, \xi_j^N\}$, $j = 1, \ldots, M$, each of size N. Let $\widehat{w}_N^j(\epsilon)$, $j = 1, \ldots, M$, denote the corresponding optimal value of the problem. Compute

$$L_{N,M}(\epsilon) := \frac{1}{M}\sum_{j=1}^{M} \widehat{w}_N^j(\epsilon),$$

which is an unbiased estimate of $\mathbb{E}[\widehat{w}_N(\epsilon)]$. Then, $L_{N,M}(\epsilon)$ provides a statistical lower bound for $\widehat{w}(\epsilon)$. An estimate of the variance of the estimator $L_{N,M}(\epsilon)$ can be computed by

$$s_L^2(N, M; \epsilon) := \frac{1}{M(M-1)}\sum_{j=1}^{M}(\widehat{w}_N^j(\epsilon) - L_{N,M}(\epsilon))^2.$$

An upper bound for the optimal value $\widehat{w}(\epsilon)$ can be obtained by virtue of the fact that $\hat{\vartheta}(\bar{x}, \epsilon) \ge \widehat{w}(\epsilon)$ for any $\bar{x} \in \mathscr{X}$. Hence, by choosing $\bar{x}$ to be a near-optimal solution, for example, by solving one SAA problem, and using an unbiased estimator of $\hat{\vartheta}(\bar{x}, \epsilon)$, we can obtain an estimate of an upper bound for $\widehat{w}(\epsilon)$. To do so, generate an i.i.d sample of $\xi^1, \ldots, \xi^{N'}$; let $\hat{\nu}_{N'}(\bar{x}, \epsilon) := (1/N')\sum_{i=1}^{N'} \hat{v}(\bar{x}, \xi^i, \epsilon)$. Then we have $\mathbb{E}[\hat{\nu}_{N'}(\bar{x}, \epsilon)] = \hat{\vartheta}(\bar{x}, \epsilon)$, thereby, $\hat{\nu}_{N'}(\bar{x}, \epsilon)$ is an estimate of an upper bound on $\widehat{w}(\epsilon)$. Note that the generated sample is i.i.d, and then the variance of this estimate can be estimated as

$$s_U^2(\bar{x}, N'; \epsilon) := \frac{1}{N'(N'-1)}\sum_{j=1}^{N'}(\hat{v}(\bar{x}, \xi^i, \epsilon) - \hat{\nu}_{N'}(\bar{x}, \epsilon))^2.$$

Note also that in this part, we need to solve the following N' independent repeated subproblems

$$\min \; f(\bar{x}, y, \xi^i)$$
$$\text{s.t.} \; \widehat{H}(\bar{x}, y, \xi^i, \epsilon) = 0.$$

Table 1. Numerical results of the smoothing SAA method for problem (55).

ϵ	N	M	N'	$L_{N,M}$	s_L	$\bar{x}_N^j$	$\hat{v}_{N'}$	s_U	Gap	S_{Gap}
10^{-3}	200	10	400	−0.0272	0.0044	(1.0011, 0.9957)	0.0015	0.0332	0.0287	0.0335
10^{-4}	200	10	500	0.0021	0.0037	(0.9992, 0.9989)	0.0246	0.0261	0.0225	0.0264
10^{-5}	200	10	400	0.0035	0.0223	(1.0007, 1.0008)	0.0050	0.0343	0.0015	0.0409
10^{-3}	300	10	500	−0.0032	0.0035	(0.9991, 0.9997)	0.0275	0.0315	0.0307	0.0316
10^{-4}	300	10	700	−0.0029	0.0026	(1.0001, 0.9996)	0.0259	0.0266	0.0288	0.0268
10^{-5}	300	10	800	0.0076	0.0427	(1.0002, 1.0001)	0.0088	0.0216	0.0012	0.0478

In practice, we may choose $\bar{x}$ to be any of the solutions of the M regularized SAA problems, by generating independent samples $\{\xi_j^1, \ldots, \xi_j^N\}$, $j = 1, \ldots, M$. In fact, we will use $\bar{x}_N^j$, the best optimal solution that estimates the smallest optimal value $\widehat{w}(\epsilon)$, to compute the upper bound estimate and the optimality gap.

Using the lower bound estimate and the objective function value estimate as discussed above, we compute an estimate of the optimality gap of the solution $\bar{x}$ and the corresponding estimated variance as follows:

$$\text{Gap}_{N,M,N'}(\bar{x}) := \hat{v}_{N'}(\bar{x}, \epsilon) - L_{N,M}(\epsilon), \qquad S^2_{\text{Gap}} := s_L^2(N, M; \epsilon) + s_U^2(\bar{x}, N'; \epsilon).$$

We conduct our test with different values of the smoothing parameter ϵ and sample sizes N, M, and N'. We report the lower and upper bounds, $L_{N,M}$ and $\hat{v}_{N'}$, of $\widehat{w}(\epsilon)$; the sample variances, s_L, s_U; and the estimate of the optimality gap, Gap; of the solution $\bar{x}_N^j$, the variance of the gap estimator S_{Gap}. The test results are displayed in Table 1.

The numerical results show that both optimal solutions and values of the smoothing SAA problems approximate those of the true problem very well as the sample size increases and the smoothing parameter tends to zero. More extensive tests are required to evaluate the performance of the proposed method, but this is beyond the scope of this paper.

Appendix.

Lemma A.1 (Scholtes [29, Theorem 3.2.5]). *Let G: $\mathbb{R}^n \to \mathbb{R}^n$ be a B-differentiable locally Lipschitz homeomorphism at every point $x \in \mathbb{R}^n$. If there exists $l > 0$ such that*

$$\|G'(x; d)\| \geq l\|d\|, \quad \forall\, x, d \in \mathbb{R}^n,$$

then G maps $\mathbb{R}^n$ homeomorphically onto $\mathbb{R}^n$.

Proof of Lemma 2.1. Part (i). Let $\bar{x} \in \mathbb{R}^m$, $\bar{\xi} \in \Xi$ be fixed. Let $\bar{y}$ be such that $H(\bar{x}, \bar{y}, \bar{\xi}) = 0$. Because $\partial_y H(x, y, \xi)$ is nonsingular, by the Clarke implicit function theorem, which is a combination of Theorem 7.1.1, and following the corollary in Clarke [8], there exists a neighborhood U for $\bar{x}$ and V for $\bar{\xi}$ such that a locally Lipschitz function y: $U \times V \to \mathbb{R}^n$ such that $H(x, y(x, \xi), \xi) = 0$ for all $(x, \xi) \in U \times V$. In what follows, we show that the implicit function $y(x, \xi)$ is globally unique.

Let $G(x, y, \xi) := (x^T, H(x, y, \xi)^T, \xi^T)^T$. Because $\partial_y H(x, y, \xi)$ is uniformly nonsingular for $(x, y, \xi) \in \mathscr{X} \times \mathbb{R}^n \times \Xi$, then $\partial G(x, y, \xi)$ is uniformly nonsingular. Moreover, because H is Lipschitz and directionally differentiable, then it is B-differentiable. Hence, G is also B-differentiable. Therefore, there exists $l > 0$ such that the directional derivative of G with respect to y satisfies

$$\|G'(x, y, \xi; d)\| \geq l\|d\|, \quad \forall\, (x, y, \xi, d) \in \mathbb{R}^m \times \mathbb{R}^n \times \Xi \times \mathbb{R}^{m+n+k}.$$

By Lemma A.1, G maps $\mathbb{R}^m \times \mathbb{R}^n \times \Xi$ homeomorphically onto $\mathbb{R}^{m+n+k}$. This shows that the implicit function $y(x, \xi)$ is globally unique.

Part (ii) follows from Ralph and Xu [21, Theorem 4.8].

Part (iii) follows from the inverse image theorem (Aubin and Frankowska [3, Theorem 8.2.9]). □

Proof of Lemma 2.2. Let ϵ_0 be a small positive scalar. By assumption, for any $(\bar{x}, \bar{\xi}) \in \mathscr{X} \times \Xi$, there exist open neighborhoods $B(\bar{x}, \delta_1)$ and $B(\bar{\xi}, \delta_2)$, where δ_1, δ_2 denote radii of these two balls satisfying $0 < \max\{\delta_1, \delta_2\} < \epsilon_0$, and $y(x, \xi)$ defined on $B(\bar{x}, \delta_1) \times B(\bar{\xi}, \delta_2)$ such that $H(x, y(x, \xi), \xi) = 0$. Let $\widehat{G}(x, y, \xi, \epsilon) = (x^T, \xi^T, \epsilon, \widehat{H}(x, y, \xi, \epsilon)^T)^T$. Let $p = (\bar{x}, \bar{\xi}, 0, 0) \in \mathbb{R}^m \times \mathbb{R}^k \times \mathbb{R} \times \mathbb{R}^n$ and $\Im := B(\bar{x}, \delta_1) \times B(\bar{\xi}, \delta_2) \times \mathscr{U}(\bar{y}) \times (-\epsilon_0, \epsilon_0)$ where $\mathscr{U}(\bar{y}) := \{y$: $y = y(x, \xi), (x, \xi) \in B(\bar{x}, \delta_1) \times B(\bar{\xi}, \delta_2)\}$. Evidently, $\Im$ is a bounded open set, $\widehat{G}$ is a continuous mapping, and equation $\widehat{G}(x, y, \xi, \epsilon) = p$ has a solution on $\Im$ with $\deg(\widehat{G}, \Im, p) \neq 0$.

Now, for any $(x', \xi') \in B(\bar{x}, \delta_1) \times B(\bar{\xi}, \delta_2)$ and $\epsilon' \in (0, \epsilon_0)$, set $q = (x', \xi', \epsilon', 0) \in \mathbb{R}^m \times \mathbb{R}^k \times \mathbb{R} \times \mathbb{R}^n$. Then, it follows that $\|p-q\|_\infty = \|(\bar{x}-x', \bar{\xi}-\xi', -\epsilon', 0)\|_\infty = \max\{\|\bar{x}-x'\|_\infty, \|\bar{\xi}-\xi'\|_\infty, \epsilon'\}$. On the other hand, we have

$$\begin{aligned}\operatorname{dist}_\infty(p, \widehat{G}(\operatorname{bd}\Im)) &= \inf\{\|p-\tilde{p}\|_\infty \colon \tilde{p} \in \widehat{G}(\operatorname{bd}\Im)\}\\ &= \inf\{\max\{\|\bar{x}-\tilde{x}\|_\infty, \|\bar{\xi}-\tilde{\xi}\|_\infty, \epsilon_0, \|\widehat{H}(\tilde{x}, \tilde{y}, \tilde{\xi}, \pm\epsilon_0)\|_\infty\} \colon (\tilde{x}, \tilde{\xi}, \tilde{y}, \pm\epsilon_0) \in \operatorname{bd}\Im\}\\ &= \inf\{\max\{\delta_1, \delta_2, \epsilon_0, \|\widehat{H}(\tilde{x}, \tilde{y}, \tilde{\xi}, \pm\epsilon_0)\|_\infty\} \colon (\tilde{x}, \tilde{\xi}, \tilde{y}, \pm\epsilon_0) \in \operatorname{bd}\Im\},\end{aligned}$$

where "bd" denotes the boundary of a set. Because $\epsilon' < \epsilon_0$, $\delta_1 < \epsilon_0$, $\delta_2 < \epsilon_0$, it then follows that

$$\|p-q\|_\infty < \operatorname{dist}_\infty(p, \widehat{G}(\operatorname{bd}\Im)),$$

for any q as defined above. Thus, by Facchinei and Pang [9, Proposition 2.1.3], we have $\deg(\widehat{G}, \Im, p) = \deg(\widehat{G}, \Im, q) \neq 0$, which implies that equation $\widehat{G}(x, y, \xi, \epsilon) = q$ has a solution on $\Im$. So, for any $\epsilon' \in (0, \epsilon_0)$, $\widehat{H}(x, y, \xi, \epsilon') = 0$ has a solution y at any given point $(x', \xi') \in B(\bar{x}, \delta_1) \times B(\bar{\xi}, \delta_2)$. In addition, due to the arbitrariness of $\bar{x}$ and $\bar{\xi}$, it is not hard to see that given any $(x, \xi) \in \mathscr{X} \times \Xi$ and small positive ϵ, equation $\widehat{H}(x, y, \xi, \epsilon) = 0$ always has a solution y. □

Proof of Lemma 5.1. First, we consider the smoothing function of the Fischer-Burmeister function, $\psi(a, b, \epsilon) = \sqrt{a^2+b^2+\epsilon^2} - (a+b)$. It is evident that for any $\epsilon > 0$ and any $a, b \in \mathbb{R}$, we have

$$\nabla_{a,b}\psi(a, b, \epsilon) = \begin{pmatrix} \dfrac{a}{\sqrt{a^2+b^2+\epsilon^2}} - 1 \\ \dfrac{b}{\sqrt{a^2+b^2+\epsilon^2}} - 1 \end{pmatrix}.$$

Let $r := (a/\sqrt{a^2+b^2+\epsilon^2}, b/\sqrt{a^2+b^2+\epsilon^2})^T$ and $s := (1, 1)^T$. Then, $\nabla_{a,b}\psi(a, b, \epsilon) = r - s$. It is obvious that $\|r\| \le 1$, $\forall a, b, \epsilon$, and $\|s\| = \sqrt{2}$. Thus $\|r-s\|^2 = 2 + \|r\|^2 - 2r_1 - 2r_2$.

We now consider the following minimization problem:

$$\begin{aligned} \min \quad & \|r\|^2 - 2r_1 - 2r_2 + 2 \\ \text{s.t.} \quad & \|r\| \le 1, \quad r \in \mathbb{R}^2. \end{aligned}$$

Notice that the above problem can be written as $\min\{(r_1-1)^2 + (r_2-1)^2 \mid \|r\| \le 1,\ r \in \mathbb{R}^2\}$, the minimum of which is obviously equal to $3 - 2\sqrt{2}$. Consequently, $\|r-s\|^2 \ge 3 - 2\sqrt{2}$. Let $l = \sqrt{3-2\sqrt{2}}$. We have $\|\nabla_{a,b}\psi(a, b, \epsilon)\| = \|r-s\| \ge l$ for any $a, b \in \mathbb{R}$ and $\epsilon \in (0, \epsilon_0]$, $\epsilon_0 > 0$. This shows the conclusion for the Fischer and Burmeister function.

Next, we consider the smoothing function of min-function

$$\psi(a, b, \epsilon) = -\frac{1}{2}\Big(\sqrt{(a-b)^2+\epsilon^2} - (a+b)\Big).$$

Evidently, for any $\epsilon > 0$,

$$\nabla_{a,b}\psi(a, b, \epsilon) = -\frac{1}{2}\begin{pmatrix} \dfrac{a-b}{\sqrt{(a-b)^2+\epsilon^2}} - 1 \\ \dfrac{b-a}{\sqrt{(a-b)^2+\epsilon^2}} - 1 \end{pmatrix}.$$

Let $r = ((a-b)/\sqrt{(a-b)^2+\epsilon^2}, (b-a)/\sqrt{(a-b)^2+\epsilon^2})^T$, and $s = (1, 1)^T$. Noticing that $\|r\| \le 1$, hence with similar arguments as above, we can conclude that the proposed result is valid in this case. □

Acknowledgments. This research is supported by the United Kingdom Engineering and Physical Sciences Research Council Grant GR/S90850/01, and the work of the second author was carried out while he was working with the first author in the School of Mathematics, University of Southampton. The authors would like to thank Daniel Ralph for pointing out an error in Theorem 3.1 on an earlier version of this paper. They would also like to thank Alexander Shapiro for bringing to their attention a number of references on random sets, particularly the reference Artstein and Vitale [1], and are also grateful for many discussions which led to a significant improvement on the presentation of Lemma 3.2. Finally, they thank three anonymous referees for helpful comments and the area editor Jong-Shi Pang for organizing a quick review.

References

[1] Artstein, Z., R. A. Vitale. 1975. A strong law of large numbers for random compact sets. *Ann. Probab.* **3** 879–882.
[2] Artstein, Z., R. J-B. Wets. 1995. Consistency of minimizers and the SLLN for stochastic programs. *J. Convex Anal.* **2** 1–17.
[3] Aubin, J.-P., H. Frankowska. 1990. *Set-Valued Analysis*. Birkhauser, Boston, MA.
[4] Birbil, S. I., G. Gurkan, O. L. Listes. 2004. Simulation-based solution of stochastic mathematical programs with complementarity constraints: Sample-path analysis. Working paper, Center for Economic Research, Tilburg University, Tilburg, The Netherlands.
[5] Birge, J., F. Louveaux. 1997. *Introduction to Stochastic Programming*. Springer Series in Operations Research, Springer-Verlag, New York.
[6] Brønsted, A., R. T. Rockafellar. 1965. On the subdifferentiability of convex functions. *Proc. Amer. Math. Soc.* **16** 605–611.
[7] Chen, X., L. Qi, D. Sun. 1998. Global and superlinear convergence of the smoothing Newton's method and its application to general box constrained variational inequalities. *Math. Comput.* **67** 519–540.
[8] Clarke, F. H. 1983. *Optimization and Nonsmooth Analysis*. Wiley, New York.
[9] Facchinei, F., J.-S. Pang. 2003. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer, New York.
[10] Fischer, A. 1992. A special Newton-type optimization method. *Optimization* **24** 269–284.
[11] Homen-De-Mello, T. 2001. Estimation of derivatives of nonsmooth performance measures in regenerative systems. *Math. Oper. Res.* **26** 741–768.
[12] Hiriart-Urruty, J.-B. 1979. Refinements of necessary optimality conditions in nondifferentiable programming I. *Appl. Math. Optim.* **5** 63–82.
[13] Kanzow, C. 1996. Some noninterior continuation methods for linear complementarity problems. *SIAM J. Matrix Anal. Appl.* **17** 851–868.
[14] King, A. J., R. J-B. Wets. 1991. Epi-consistency of convex stochastic programs. *Stochastics Stochastics Rep.* **34** 83–92.
[15] Lin, G.-H., X. Chen, M. Fukushima. 2003. Smoothing implicit programming approaches for stochastic mathematical programs with linear complementarity constraints. http://www.amp.i.kyoto-u.ac.jp/tecrep/index-e.html.
[16] Luo, Z.-Q., J.-S. Pang, D. Ralph. 1996. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, Cambridge, UK.
[17] Meng, F., H. Xu. 2006. A regularized sample average approximation method for stochastic mathematical programs with nonsmooth equality constraints. *SIAM J. Optim.* **17** 891–919.
[18] Patriksson, M., L. Wynter. 1999. Stochastic mathematical programs with equilibrium constraints. *Oper. Res. Lett.* **25** 159–167.
[19] Plambeck, E. L., B. R. Fu, S. M. Robinson, R. Suri. 1996. Sample-path optimization of convex stochastic performance functions. *Math. Programming* **75** 137–176.
[20] Polak, E., D. Q. Mayne, Y. Wardi. 1983. On the extension of constrained optimization algorithms from differentiable to nondifferentiable problems. *SIAM J. Control Optim.* **21** 179–203.
[21] Ralph, D., H. Xu. 2005. Implicit smoothing and its application to optimization with piecewise smooth equality constraints. *Optim. Theory Appl.* **124** 673–699.
[22] Robinson, S. M. 1996. Analysis of sample-path optimization. *Math. Oper. Res.* **21** 513–528.
[23] Rockafellar, R. T. 1970. *Convex Analysis*. Princeton University Press, Princeton, NJ.
[24] Rockafellar, R. T., R. J-B. Wets. 1975. Stochastic convex programming: Kuhn-Tucker conditions. *J. Math. Econom.* **2** 349–370.
[25] Rockafellar, R. T., R. J-B. Wets. 1998. *Variational Analysis*. Springer-Verlag, Berlin, Germany.
[26] Rubinstein, R. Y., A. Shapiro. 1993. *Discrete Events Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Methods*. John Wiley and Sons, New York.
[27] Ruszczyński, A., A. Shapiro. 2003. Stochastic programming. A. Ruszczyński, A. Shapiro, eds. *Handbooks in OR & MS. Stochastic Programming*, Vol. 10. North-Holland Publishing Company, Amsterdam, The Netherlands.
[28] Santoso, T., S. Ahmed, M. Goetschalckx, A. Shapiro. 2005. A stochastic programming approach for supply chain network design under uncertainty. *Eur. J. Oper. Res.* **167** 96–115.
[29] Scholtes, S. 1994. Introduction to piecewise smooth equations. Habilitation thesis, University of Karlsruhe, Karlsruhe, Germany.
[30] Shapiro, A. 2003. Monte Carlo sampling methods. A. Ruszczyński, A. Shapiro, eds. *Handbooks in OR & MS. Stochastic Programming*, Vol. 10. North-Holland Publishing Company, Amsterdam, The Netherlands.
[31] Shapiro, A. 2006. Stochastic mathematical programs with equilibrium constraints. *J. Optim. Theor. Appl.* **128** 223–243.
[32] Shapiro, A., T. Homem-de-Mello. 2000. On rate of convergence of Monte Carlo approximations of stochastic programs. *SIAM J. Optim.* **11** 70–86.
[33] Shapiro, A., Y. Wardi. 1996. Convergence analysis of stochastic algorithms. *Math. Oper. Res.* **21** 615–628.
[34] Shapiro, A., H. Xu. 2005. Stochastic mathematical programs with equilibrium constraints, modeling and sample average approximation. http://www.optimization-online.org/DB_HTML/2005/01/1046.html.
[35] Wets, R. J.-B. 1989. Stochastic programming. G. L. Nemhauser, A. H. G. Rinnooy Kan, M. J. Todd, eds. *Handbooks in OR & MS*, Vol. 1. North-Holland Publishing Company, Amsterdam, The Netherlands, 573–629.
[36] Xu, H. 2005. An MPCC approach for stochastic Stackelberg-Nash-Cournot equilibrium. *Optimization* **54** 27–57.
[37] Xu, H. 2006. An implicit programming approach for a class of stochastic mathematical programs with complementarity constraints. *SIAM J. Optim.* **16** 670–696.
[38] Xu, H., A. M. Rubinov, B. M. Glover. 1999. Continuous approximations to generalized Jacobians. *Optimization* **46** 221–246.