Journal of Computational Mathematics, Vol. 24, No. 6, 2006, 733–748.

EXPONENTIAL CONVERGENCE OF SAMPLE AVERAGE APPROXIMATION METHODS FOR A CLASS OF STOCHASTIC MATHEMATICAL PROGRAMS WITH COMPLEMENTARITY CONSTRAINTS *1)

Fan-wen Meng
(*The Logistics Institute - Asia Pacific, National University of Singapore, 1 Law Link, Singapore*
School of Mathematics, University of Southampton, Highfield Southampton, SO17 1BJ, UK)
Hui-fu Xu
(*School of Mathematics, University of Southampton, Highfield Southampton, SO17 1BJ, UK*)

Abstract

In this paper, we propose a Sample Average Approximation (SAA) method for a class of Stochastic Mathematical Programs with Complementarity Constraints (SMPCC) recently considered by Birbil, Gürkan and Listes [3]. We study the statistical properties of obtained SAA estimators. In particular we show that under moderate conditions a sequence of weak stationary points of SAA programs converge to a weak stationary point of the true problem with probability approaching one at exponential rate as the sample size tends to infinity. To implement the SAA method more efficiently, we incorporate the method with some techniques such as Scholtes' regularization method and the well known smoothing NCP method. Some preliminary numerical results are reported.

Mathematics subject classification: 90C15, 90C30, 90C31, 90C33.
Key words: Stochastic mathematical programs with complementarity constraints, Sample average approximation, Weak stationary points, Exponential convergence.

1. Introduction

In this paper, we investigate the following Stochastic Mathematical Programs with Complementarity Constraints (SMPCC)

$$\begin{array}{ll} \min\limits_{x\in\mathcal{X}} & \mathbb{E}[f(x,\xi(\omega))] \\ \text{s.t.} & \mathbb{E}[F(x,\xi(\omega))]\geq 0,\ \mathbb{E}[G(x,\xi(\omega))]\geq 0, \\ & 0\leq \mathbb{E}[F(x,\xi(\omega))]\perp \mathbb{E}[G(x,\xi(\omega))]\geq 0, \end{array} \tag{1.1}$$

where $f:\mathbb{R}^n\times\mathbb{R}^k\to\mathbb{R}$, $F:\mathbb{R}^n\times\mathbb{R}^k\to\mathbb{R}^m$, $G:\mathbb{R}^n\times\mathbb{R}^k\to\mathbb{R}^m$ are twice continuously differentiable with respect to x for almost every ξ and continuous with respect to ξ, $\mathcal{X}\subset\mathbb{R}^n$ is a compact subset of $\mathbb{R}^n$, $\xi:\Omega\to\mathbb{R}^k$ is a vector of random variables defined on a probability space $(\Omega,\mathcal{F},P)$, $\mathbb{E}$ denotes the mathematical expectation.

The SMPCC model (1.1) was first considered by Birbil, Gürkan and Listes [3] and it is a natural extension of deterministic MPEC models [6]. The primary motivation for the model (1.1) is that the objective and constraint functions may involve some random data which reflect uncertainties in practical problems. For instance, in Stackelberg leader follower model [6], if

* Received November 19, 2005.
1) The work of the authors is supported by the United Kingdom Engineering and Physical Sciences Research Council grant GR/S90850/01.

players have to make a decision before the realization of uncertain demand, then each player has to consider the expected profit rather than the profit in a particular demand scenario. Subsequently the model can be reformulated as (1.1).

Birbil, Gürkan and Listes [3] applied the well known Sample Path (SP) method [9, 10] to solve (1.1). The basic idea of SP is to use computer simulation to approximate functions which are not observable. In this context, the expected value of functions in the objective and constraints of (1.1) are either not observable or very costly to be integrated out. Birbil, Gürkan and Listes [3] used simulation based average to construct successive approximate MPEC problems and showed that under some stability conditions the stationary points of approximate MPEC problems converge to their counterpart almost surely. More recently, Bastin, Cirillo and Toint [2] extended the discussion to investigate the convergence of stationary points for a broader class of stochastic optimization problems.

In this paper, we apply well known sample average approximation method to solve (1.1). Specifically, we consider an independent identically distributed (i.i.d) sample of $\xi(\omega)$, which is denoted by $\xi^1, ..., \xi^N$, and use the following *Sample Average Approximation* (SAA) problem to approximate the true problem (1.1):

$$\begin{array}{ll} \min\limits_{x\in\mathcal{X}} & \hat{f}_N(x) \\ \text{s.t.} & 0 \le \hat{F}_N(x) \perp \hat{G}_N(x) \ge 0, \end{array} \tag{1.2}$$

where $\hat{f}_N(x) = \frac{1}{N}\sum_{i=1}^N f(x,\xi^i)$, and $\hat{F}_N(x) = \frac{1}{N}\sum_{i=1}^N F(x,\xi^i)$, $\hat{G}_N(x) = \frac{1}{N}\sum_{i=1}^N G(x,\xi^i)$. SAA methods are essentially the same as SP methods and they have been extensively investigated in stochastic optimization. See recent work [1, 5, 16, 18]. More recently SAA methods have been applied to two stage stochastic mathematical programs with equilibrium constraints [17, 19, 21, 7] and various convergence results have been established.

In this paper, we analyze convergence of weak stationary points of SAA problem (1.2). This is motivated by the fact that due to the combinatorial nature, one may be more likely to obtain a stationary point than a local or global minimizer for MPEC problems. Consequently the notion of weak stationary point which was introduced by Scheel and Scholtes [12] is very relevant, indeed it has been well accepted and studied. The focus of this paper is on the convergence rate of weak stationary points of the SAA problem (1.2). Under some moderate conditions, we show the exponential convergence for weak stationary point of the SAA problems as sample size N tends to infinity. This result has significantly strengthened recent results of SAA methods for model (1.1).

The rest of this paper are organized as follows. In Section 2, we introduce some definitions about stationary points and present some preliminary results. In Section 3, we show that under some moderate stability conditions which are widely used in MPEC literature, the weak stationary points of SAA problem (1.2) converge to its counterpart of the true problem as the sample size tends to infinity. In Section 4, we discuss numerical implementation of the SAA method, and finally in Section 5 we report some preliminary test results.

2. Preliminaries

In this section, we recall some basic notions and definitions in mathematical programs with complementarity constraints. For simplicity of notation, let $\bar{f}(x) = \mathbb{E}[f(x,\xi(\omega))]$, $\bar{F}(x) =$

$\mathbb{E}[F(x,\xi(\omega))]$ and $\bar{G}(x) = \mathbb{E}[G(x,\xi(\omega))]$. Then, (1.1) can be written componentwise as

$$\begin{array}{ll} \min\limits_{x\in\mathcal{X}} & \bar{f}(x) \\ \text{s.t.} & \bar{F}_j(x) \geq 0,\ \ \bar{G}_j(x) \geq 0,\ \ j = 1,\cdots,m, \\ & \bar{F}_j(x)\bar{G}_j(x) = 0,\ \ j = 1,\cdots,m. \end{array} \tag{2.1}$$

Under some appropriate conditions, we can show that the underlying functions in (2.1) are twice continuously differentiable.

Proposition 2.1. *Suppose that: (a) $f(x,\xi(\omega))$, $F(x,\xi(\omega))$, $G(x,\xi(\omega))$ are twice continuously differentiable in x w.p.1; (b) $\mathbb{E}[\nabla_x f(x,\xi(\omega))]$, $\mathbb{E}[\nabla_x F_j(x,\xi(\omega))]$ and $\mathbb{E}[\nabla_x G_j(x,\xi(\omega))]$ are well defined and $\nabla_x f(x,\xi(\omega))$, $\nabla_x F_j(x,\xi(\omega))$, $\nabla_x G_j(x,\xi(\omega))$ are dominated by an integrable function $\kappa_1(\xi)$, $j = 1,\cdots,m$. Then $\bar{f}(x)$, $\bar{F}(x)$ and $\bar{G}(x)$ are continuously differentiable and*

$$\nabla\bar{f}(x) = \mathbb{E}[\nabla_x f(x,\xi(\omega))]$$

and

$$\nabla\bar{F}_j(x) = \mathbb{E}[\nabla_x F_j(x,\xi(\omega))];\quad \nabla\bar{G}_j(x) = \mathbb{E}[\nabla_x G_j(x,\xi(\omega))],\ \ j = 1,\cdots,m.$$

Suppose in addition that (c) $\mathbb{E}[\nabla^2_{xx} f(x,\xi(\omega))]$, $\mathbb{E}[\nabla^2_{xx} F_j(x,\xi(\omega))]$ and $\mathbb{E}[\nabla^2_{xx} G_j(x,\xi(\omega))]$ are well defined; (d) $\nabla^2_{xx} f(x,\xi(\omega))$, $\nabla^2_{xx} F_j(x,\xi(\omega))$, $\nabla^2_{xx} G_j(x,\xi(\omega))$ are dominated by an integrable function $\kappa_2(\xi)$, $j = 1,\cdots,m$. Then $\bar{f}(x)$, $\bar{F}(x)$ and $\bar{G}(x)$ are twice continuously differentiable and

$$\nabla^2\bar{f}(x) = \mathbb{E}[\nabla^2_{xx} f(x,\xi(\omega))]$$

and

$$\nabla^2\bar{F}_j(x) = \mathbb{E}[\nabla^2_{xx} F_j(x,\xi(\omega))];\quad \nabla^2\bar{G}_j(x) = \mathbb{E}[\nabla^2_{xx} G_j(x,\xi(\omega))], j = 1,\cdots,m.$$

The results follow straightforwardly from [15, Proposition 3.2]. We omit the details of the proof.

Throughout the paper, we make a blanket assumption that conditions in Proposition 2.1 hold. Next, we define the following active index sets at a feasible point x of problem (2.1):

$$\begin{aligned} \mathcal{I}_{\bar{F}}(x) &:= \{i \in \{1,\cdots,m\} \mid \bar{F}_i(x) = 0\}, \\ \mathcal{I}_{\bar{G}}(x) &:= \{i \in \{1,\cdots,m\} \mid \bar{G}_i(x) = 0\}. \end{aligned}$$

Clearly $\mathcal{I}_{\bar{F}}(x) \cup \mathcal{I}_{\bar{G}}(x) = \{1,2,\cdots,m\}$. We call $\mathcal{I}_{\bar{F}}(x) \cap \mathcal{I}_{\bar{G}}(x)$ the *biactive set*.

Recall that the MPEC-LICQ is satisfied at a feasible point x^* if the gradients

$$\left\{\nabla\bar{F}_i(x^*), \nabla\bar{G}_j(x^*) : i \in \mathcal{I}_{\bar{F}}(x^*),\ j \in \mathcal{I}_{\bar{G}}(x^*)\right\}$$

are linearly independent. The notion of MPEC-LICQ was first considered by Scholtes [13] to deal with the problem in MPEC that the natural LICQ may not hold at any feasible point of an MPEC problem. The notion has been well accepted and widely used in the MPEC literature.

We consider the Lagrangian function associated with problem (1.1) (or (2.1))

$$L(x,\lambda,\mu) = \mathbb{E}[f(x,\xi(\omega))] - \lambda^T\mathbb{E}[F(x,\xi(\omega))] - \mu^T\mathbb{E}[G(x,\xi(\omega))].$$

A feasible point x^* of (2.1) is called a *weakly stationarity point* if there exist multipliers λ^* and μ^* such that

$$\left\{\begin{array}{l} \nabla_x L(x^*,\lambda^*,\mu^*) = 0, \\ \lambda_i^*\mathbb{E}[F_i(x^*,\xi)] = 0,\ i = 1,\cdots,m, \\ \mu_i^*\mathbb{E}[G_i(x^*,\xi)] = 0,\ i = 1,\cdots,m. \end{array}\right. \tag{2.2}$$

A feasible point x^* is called a *strongly stationarity point* if $\lambda_i^* \geq 0$, $\mu_i^* \geq 0$ for $i \in \mathcal{I}_{\bar{F}}(x^*) \cap \mathcal{I}_{\bar{G}}(x^*)$. Note that, in this case, if the MPEC-LICQ is satisfied at the point x^*, then the Lagrange multipliers λ^* and μ^* are unique. A solution (x^*, λ^*, μ^*) of (2.2) is said to satisfy *Upper-level Strict Complementarity* (USC) condition if $\lambda_i^* \mu_i^* \neq 0$, for $i \in \mathcal{I}_{\bar{F}}(x^*) \cap \mathcal{I}_{\bar{G}}(x^*)$, and it is said to satisfy *Lower-level Strict Complementarity* (LSC) condition if $\mathcal{I}_{\bar{F}}(x^*) \cap \mathcal{I}_{\bar{G}}(x^*) = \emptyset$.

In this paper, we focus on the convergence of the weak stationary points of SAA program (1.2) when the sample size N tends to infinity although the results can be easily extended to the strong stationary points. For convenience, we make the following assumption.

Assumption 2.1. *The LSC condition holds at a weak stationary point of the true problem (1.1).*

Under Assumption 2.1, the optimality conditions (2.2) for the weak stationary point x^* can be written as

$$\begin{cases} \nabla_x L(x^*, \lambda^*, \mu^*) = 0, \\ \bar{F}_i(x^*) = 0, \ \ i \in \mathcal{I}_{\bar{F}}(x^*), \\ \bar{G}_i(x^*) = 0, \ \ i \in \mathcal{I}_{\bar{G}}(x^*), \\ \lambda_i^* = 0, i \in \mathcal{I}_{\bar{G}}(x^*), \\ \mu_i^* = 0, i \in \mathcal{I}_{\bar{F}}(x^*). \end{cases} \tag{2.3}$$

Define the mapping: $H : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ by

$$H(x, \lambda, \mu) = \begin{pmatrix} \nabla_x L(x, \lambda, \mu) \\ \bar{F}_{\mathcal{I}_{\bar{F}}(x)}(x) \\ \bar{G}_{\mathcal{I}_{\bar{G}}(x)}(x) \\ \lambda_{\mathcal{I}_{\bar{G}}(x)} \\ \mu_{\mathcal{I}_{\bar{F}}(x)} \end{pmatrix}, \tag{2.4}$$

where $\bar{F}_{\mathcal{I}_{\bar{F}}(x)}(x)$ denotes the column vector consisting of $\bar{F}_i(x)$ for each $i \in \mathcal{I}_{\bar{F}}(x)$, $\bar{G}_{\mathcal{I}_{\bar{G}}(x)}(x)$ denotes the column vector consisting of $\bar{G}_i(x)$ for each $i \in \mathcal{I}_{\bar{G}}(x)$, $\lambda_{\mathcal{I}_{\bar{G}}(x)}$ denotes the column vector consisting of λ_i for each $i \in \mathcal{I}_{\bar{G}}(x)$, $\mu_{\mathcal{I}_{\bar{F}}(x)}$ denotes the column vector consisting of μ_i for each $i \in \mathcal{I}_{\bar{F}}(x)$. Then (2.3) can be written as

$$H(x^*, \lambda^*, \mu^*) = 0.$$

Similarly, we can define the weak stationary point x^N of the SAA problem (1.2) associated with multipliers λ^N and μ^N:

$$\begin{cases} \nabla_x L_N(x^N, \lambda^N, \mu^N) = 0, \\ \lambda_i^N \hat{F}_i^N(x^N) = 0, \ i = 1, \cdots, m, \\ \mu_i^N \hat{G}_i^N(x^N) = 0, \ i = 1, \cdots, m, \end{cases} \tag{2.5}$$

where $\hat{F}_i^N$ denotes the i-th component of $\hat{F}_N$ and $\hat{G}_i^N$ denotes the i-th component of $\hat{G}_N$, and the Lagrangian function is defined as

$$L_N(x, \lambda, \mu) = \hat{f}_N(x) - \lambda^T \hat{F}_N(x) - \mu^T \hat{G}_N(x).$$

Let $\mathcal{I}_{\hat{F}_N}(x)$ and $\mathcal{I}_{\hat{G}_N}(x)$ be the active index sets at a feasible point x of (1.2). We make the following assumption for (2.5).

Assumption 2.2. *The LSC condition holds at a weak stationary point of the SAA problem (1.2).*

Under Assumption 2.2, condition (2.5) can be written as

$$\begin{pmatrix} \nabla_x L_N(x,\lambda,\mu) \\ \hat{F}^N_{\mathcal{I}_{\hat{F}_N}(x)}(x) \\ \hat{G}^N_{\mathcal{I}_{\hat{G}_N}(x)}(x) \\ \lambda_{\mathcal{I}_{\hat{G}_N}(x)} \\ \mu_{\mathcal{I}_{\hat{F}_N}(x)} \end{pmatrix} = 0.$$

Note that the index sets $\mathcal{I}_{\hat{F}_N}$ and $\mathcal{I}_{\hat{G}_N}$ depends on sample size N. This is not convenient for convergence analysis. In the following lemma, we show that they can be replaced by $\mathcal{I}_{\bar{F}}$ and $\mathcal{I}_{\bar{G}}$ respectively.

Lemma 2.1. *Let x^* be a weak stationary point of the true problem (1.1) and x^N be a weak stationary point of SAA problem (1.2). Suppose that Assumption 2.1 hold at x^* and Assumption 2.2 holds at x^N. Suppose also that $x^N \to x^*$ as $N \to \infty$. Then for N sufficiently large*

$$\mathcal{I}_{\bar{F}}(x^*) = \mathcal{I}_{\hat{F}_N}(x^N), \quad \mathcal{I}_{\bar{G}}(x^*) = \mathcal{I}_{\hat{G}_N}(x^N), \text{ w.p.1.}$$

Proof. Let $\mathcal{I} := \{1, 2, \cdots, m\}$. Under Assumptions 2.1–2.2, we have $\mathcal{I} \setminus \mathcal{I}_{\bar{F}}(x^*) = \mathcal{I}_{\bar{G}}(x^*)$ and $\mathcal{I} \setminus \mathcal{I}_{\hat{F}_N}(x^N) = \mathcal{I}_{\hat{G}_N}(x^N)$.

Let $i \in \mathcal{I}_{\bar{G}}(x^*)$. Then $\bar{F}_i(x^*) > 0$. Since $\bar{F}_i$ is continuous, there exists a closed neighborhood $\mathcal{B}(x^*)$ of x^* such that $\bar{F}_i(x) > \delta$ for any $x \in \mathcal{B}(x^*)$ where $\delta > 0$ is some constant. Since $x^N \to x^*$, we assume that N is sufficiently large such that $x^N \in \mathcal{B}(x^*)$. Hence $\bar{F}_i(x^N) > \delta$, therefore $\mathcal{I}_{\bar{G}}(x^*) \subset \mathcal{I}_{\hat{G}_N}(x^N)$.

Under the assumptions of Proposition 2.1 (a blanket assumption), we can apply the uniform strong law of large numbers [11, Lemma A1] to $\hat{F}^N_i$ on $\mathcal{B}(x^*)$ and conclude that $\hat{F}^N_i \to \bar{F}_i$ w.p.1 uniformly as $N \to \infty$. Therefore for N sufficiently large, we have that $\hat{F}^N_i(x^N) > \delta/2 > 0$ w.p.1. This shows $i \in \mathcal{I}_{\hat{G}_N}(x^N)$.

Similarly we can show that $\mathcal{I}_{\bar{F}}(x^*) \subset \mathcal{I}_{\hat{F}_N}(x^N)$. Since $\mathcal{I}_{\bar{G}}(x^*)$ and $\mathcal{I}_{\bar{F}}(x^*)$ are complementary to each other and $\mathcal{I}_{\hat{G}_N}(x^N)$ and $\mathcal{I}_{\hat{F}_N}(x^N)$ are complementary to each other, the conclusion follows.

In the following analysis, we are interested in the local convergence analysis of the weak stationary points of the SAA problem when N tends to infinity. Based on the observations obtained in Lemma 2.1, we only need to consider the following equation for SAA program (1.2) for N sufficiently large:

$$H_N(x,\lambda,\mu) := \begin{pmatrix} \nabla_x L_N(x,\lambda,\mu) \\ \hat{F}^N_{\mathcal{I}_{\bar{F}}(x)}(x) \\ \hat{G}^N_{\mathcal{I}_{\bar{G}}(x)}(x) \\ \lambda_{\mathcal{I}_{\bar{G}}(x)} \\ \mu_{\mathcal{I}_{\bar{F}}(x)} \end{pmatrix} = 0.$$

We are now ready to state our first convergence result for the weak stationary points of the SAA problem.

Proposition 2.2. *Suppose that Assumptions 2.1 and 2.2 hold. Let $\{x^N\}$ be a sequence of weak stationary points of the SAA problem (1.2) and $\{(\lambda^N, \mu^N)\}$ be the corresponding multipliers. Then w.p.1 an accumulation point of $\{x^N\}$ with finite corresponding accumulation point of $\{(\lambda^N, \mu^N)\}$ is a weak stationary point of (1.1).*

Proof. Assume without loss of generality that $\{(x^N, \lambda^N, \mu^N)\} \to (x^*, \lambda^*, \mu^*)$. Under Assumptions 2.1-2.2, we have from Lemma 2.1 that for N sufficiently large, $\mathcal{I}_{\bar{F}}(x^N) = \mathcal{I}_{\bar{F}}(x^*)$ and $\mathcal{I}_{\bar{G}}(x^N) = \mathcal{I}_{\bar{G}}(x^*)$. Then (x^N, λ^N, μ^N) is a solution of the following equation

$$H_N(x,\lambda,\mu) = \begin{pmatrix} \nabla_x L_N(x,\lambda,\mu) \\ \hat{F}^N_{\mathcal{I}_{\bar{F}}(x^*)}(x) \\ \hat{G}^N_{\mathcal{I}_{\bar{G}}(x^*)}(x) \\ \lambda_{\mathcal{I}_{\bar{G}}(x^*)} \\ \mu_{\mathcal{I}_{\bar{F}}(x^*)} \end{pmatrix} = 0.$$

Under condition (d) of Proposition 2.1 and the uniform law of large numbers [11, Lemma A1], H_N converges to H uniformly w.p.1. in any closed neighborhood of (x^*, λ^*, μ^*) w.p.1 as $N \to \infty$. Therefore

$$\lim_{N\to\infty} H_N(x^N, \lambda^N, \mu^N) = H(x^*, \lambda^*, \mu^*) = 0 \text{ w.p.1.}$$

which indicates that w.p.1 x^* is a weak stationary point of the true problem.

3. Exponential Convergence of Weak Stationary Points

In this section, we study the exponential convergence of the weak stationary points of the SAA program when the sample size N tends to infinity. For this purpose, we need to make the following assumption.

Assumption 3.1. *Let $\mathcal{X}$ be a compact set in $\mathbb{R}^n$ and $\upsilon(x, \xi(\omega))$ denote any element in the collection of functions $\{(\nabla_x f(x,\xi(\omega)))_i, F_j(x,\xi(\omega)), G_j(x,\xi(\omega)), (\nabla_x F_j(x,\xi(\omega)))_i, (\nabla_x G_j(x,\xi(\omega)))_i, i = 1,\cdots,n, j = 1,\cdots,m.\}$. Then $\upsilon(x,\xi(\omega))$ possesses the following properties:*

(A1) for every $x \in \mathcal{X}$, the moment generating function

$$M(t) := \mathbb{E}\left[e^{t(\upsilon(x,\xi(\omega)) - \mathbb{E}[\upsilon(x,\xi(\omega))])}\right]$$

of random variable $\upsilon(x,\xi(\omega)) - \mathbb{E}[\upsilon(x,\xi(\omega))]$ is finite valued for all t in a neighborhood of zero;

(A2) there exists a (measurable) function $\kappa_3 : \mathbb{R}^n \to \mathbb{R}_+$ and constant $\gamma > 0$ such that

$$|\upsilon(x',\xi(\omega)) - \upsilon(x,\xi(\omega))| \le \kappa_3(\xi(\omega))\|x' - x\|^\gamma \tag{3.1}$$

for all $\xi(\omega) \in \Xi$ and all $x', x \in \mathcal{X}$;

(A3) the moment generating function $M_{\kappa_3}(t)$ of $\kappa_3(\xi(\omega))$ is finite valued for all t in a neighborhood of zero.

Assumption 3.1 (A1) means that the random variable $\upsilon(x,\xi(\omega))$ does not have a heavy tail distribution. In particular, it holds if this random variable has a distribution supported on a bounded subset of $\mathbb{R}$. Assumption 3.1 (A2) requires $\upsilon(x,\xi(\omega))$ to be globally Holder continuous with respect to x. Assumption 3.1 (A3) is satisfied if $\mathbb{E}[\kappa_3(\xi(\omega))]$ is finite.

Lemma 3.1. *Let $\upsilon(x,\xi(\omega)) = (\upsilon_1(x,\xi(\omega)),\cdots,\upsilon_p(x,\xi(\omega)))^T$ be such that each component $\upsilon_i(x,\xi(\omega))$ is given as in Assumption 3.1 and Assumption 3.1 hold. Then for any $\epsilon > 0$, there exist positive constants $C(\epsilon)$ and $\beta(\epsilon)$, independent of N, such that*

$$\text{Prob}\left\{\sup_{x\in\mathcal{X}} \|\upsilon_N(x) - \mathbb{E}[\upsilon(x,\xi(\omega))]\| \ge \epsilon\right\} \le pC(\epsilon)e^{-N\beta(\epsilon)}, \tag{3.2}$$

where $\upsilon_N(x) = N^{-1}\sum_{i=1}^N \upsilon(x,\xi^i)$ and $\xi^1,\cdots,\xi^N$ is an i.i.d sample of $\xi(\omega)$.

Proof. By [19, Theorem 5.1], for each $i \in \{1, \cdots, p\}$, there exist positive constants $C_i(\epsilon)$ and $\beta_i(\epsilon)$, independent of N such that

$$\text{Prob}\left\{\sup_{x\in\mathcal{X}} |\upsilon_i^N(x) - \mathbb{E}[\upsilon_i(x,\xi(\omega))]| \geq \epsilon\right\} \leq C_i(\epsilon)e^{-N\beta_i(\epsilon)}, \tag{3.3}$$

where $\upsilon_i^N(x)$ denotes the i-th component of $\upsilon_N(x)$. Consequently, we have

$$\begin{aligned}\text{Prob}\left\{\sup_{x\in\mathcal{X}} \|\upsilon_N(x) - \mathbb{E}[\upsilon(x,\xi(\omega))]\| \geq \epsilon\right\} &\leq \sum_{i=1}^{p} \text{Prob}\left\{\sup_{x\in\mathcal{X}} |\upsilon_i^N(x) - \mathbb{E}[\upsilon_i(x,\xi(\omega))]| \geq \epsilon\right\} \\ &\leq \sum_{i=1}^{p} C_i(\epsilon)e^{-N\beta_i(\epsilon)} \\ &\leq pC(\epsilon)e^{-N\beta(\epsilon)}\end{aligned}$$

where $C(\epsilon) := \max\{C_1(\epsilon), \cdots, C_p(\epsilon)\}$ and $\beta(\epsilon) := \min\{\beta_1(\epsilon), \cdots, \beta_p(\epsilon)\}$.

We are now ready to state our main convergence result, that is, if $\{x^N\}$ converges to x^* as in Proposition 2.2, then it does so at exponential rate under some mild conditions. The proof relies on a stability theorem [4, Theorem 5.2.4].

Let $\Phi : \mathcal{D} \to \mathbb{R}^n$ be a continuous function, where $\mathcal{D}$ is a subset of $\mathbb{R}^n$. Let $\mathcal{N}$ be an open subset of $\mathcal{D}$. We use $\mathcal{B}(\Phi; \epsilon, cl\mathcal{N})$, where cl denotes the closure of a set, to denote the set of continuous functions Ψ such that

$$\sup_{y\in cl\mathcal{N}} \|\Psi(y) - \Phi(y)\| < \epsilon.$$

An isolated zero $x \in \mathcal{D}$ of Φ is said to be *stable* if for every open neighborhood $\mathcal{N}$ of x which satisfies

$$cl\mathcal{N} \subset \mathcal{D} \text{ and } \Phi^{-1}(0) \bigcap cl\mathcal{N} = \{x\},$$

there exist positive scalars ϵ and c such that, for all $\Psi \in \mathcal{B}(\Phi; \epsilon, cl\mathcal{N})$, $\Psi^{-1}(0) \cap \mathcal{N}$ is nonempty and

$$\|x' - x\| \leq c\|\Phi(x')\|, \ \ \forall x' \in \Psi^{-1}(0) \cap \mathcal{N}.$$

Theorem 3.1. *Suppose that all conditions in Proposition 2.2 are satisfied. Suppose also that Assumptions 2.1, 2.2 and 3.1 hold. Suppose further that a sequence of weak stationary points of (1.2), $\{x^N\}$, converges to x^* and (λ^*, μ^*) is the corresponding Lagrange multipliers as Proposition 2.2 and*

$$W := \begin{pmatrix} \nabla^2_{xx}L(x^*,\lambda^*,\mu^*) & -\nabla_x \bar{F}_{\mathcal{I}_{\bar{F}}(x^*)}(x^*) & -\nabla_x \bar{G}_{\mathcal{I}_{\bar{G}}(x^*)}(x^*) \\ \nabla_x \bar{F}_{\mathcal{I}_{\bar{F}}(x^*)}(x^*)^T & 0 & 0 \\ \nabla_x \bar{G}_{\mathcal{I}_{\bar{G}}(x^*)}(x^*)^T & 0 & 0 \end{pmatrix} \tag{3.4}$$

is nonsingular. Then $\{x^N\}$ converges to x^* with probability approaching 1 exponentially fast with the increase of sample size N and x^* is the weak stationary points of the SAA program (1.2).

Proof. Observe first that H defined in (2.4) is continuously differentiable at (x^*, λ^*, μ^*). The Jacobian of H with respect to (x, λ, μ) at (x^*, λ^*, μ^*) is

$$\nabla H(x^*,\lambda^*,\mu^*) = \begin{bmatrix} \nabla^2_{xx}L(x^*,\lambda^*,\mu^*) & -\nabla_x \bar{F}_{\mathcal{I}_{\bar{F}}(x^*)}(x^*) & -\nabla_x \bar{F}_{\mathcal{I}_{\bar{G}}(x^*)}(x^*) & -\nabla_x \bar{G}_{\mathcal{I}_{\bar{G}}(x^*)}(x^*) & -\nabla_x \bar{G}_{\mathcal{I}_{\bar{F}}(x^*)}(x^*) \\ \nabla_x \bar{F}_{\mathcal{I}_{\bar{F}}(x^*)}(x^*)^T & 0 & 0 & 0 & 0 \\ \nabla_x \bar{G}_{\mathcal{I}_{\bar{G}}(x^*)}(x^*)^T & 0 & 0 & 0 & 0 \\ 0 & 0 & I_{\mathcal{I}_{\bar{G}}(x^*)} & 0 & 0 \\ 0 & 0 & 0 & 0 & I_{\mathcal{I}_{\bar{F}}(x^*)} \end{bmatrix}.$$

By swapping some columns in matrix $\nabla H(x^*, \lambda^*, \mu^*)$ which does not change the nonsingularity/singularity of the matrix, we obtain the following matrix

$$V = \begin{bmatrix} W & \alpha & \beta \\ 0 & I_{\mathcal{I}_{\bar{G}}(x^*)} & 0 \\ 0 & 0 & I_{\mathcal{I}_{\bar{F}}(x^*)} \end{bmatrix},$$

where

$$\alpha = \begin{bmatrix} -\nabla_x \bar{F}_{\mathcal{I}_{\bar{G}}(x^*)}(x^*) \\ 0 \\ 0 \end{bmatrix} \text{ and } \beta = \begin{bmatrix} -\nabla_x \bar{G}_{\mathcal{I}_{\bar{F}}(x^*)}(x^*) \\ 0 \\ 0 \end{bmatrix}.$$

It is easy to see that $\nabla H(x^*, \lambda^*, \mu^*)$ is nonsingular if and only if matrix W is nonsingular. This shows nonsingularity of $\nabla H(x^*, \lambda^*, \mu^*)$ under the assumption of the nonsingularity of W. Thereby we have

$$H'(x^*, \lambda^*, \mu^*; dx, d\lambda, d\mu) = 0 \implies dx = 0,\ d\lambda = 0,\ d\mu = 0,$$

where $H'(\cdot; h)$ denotes the directional derivative of H in direction h. To ease the notation, let w denote (x, λ, μ). By [4, Lemma 5.2.1], there exist an open neighbourhood $\mathcal{B}$ of w^* and a constant $c > 0$ such that

$$\|w - w^*\| \leq c\|H(w)\|$$

for any $w \in \mathcal{B}$. This indicates that w^* is an isolated solution of H. Moreover, since $\nabla H(w^*)$ is nonsingular, by the inverse function theorem, H is a local homeomorphism at w^*, that is, there exists an open neighborhood $\mathcal{N}$ of w^* such that the restriction $H|_{\mathcal{N}} : \mathcal{N} \to H(\mathcal{N})$ is a homeomorphism [4, Proposition 2.1.14]. Furthermore, following a discussion of [4, Section 2.1.1], we have that for any open neighborhood $\mathcal{N}'$ contained in $\mathcal{N}$, the degree $deg(H, \mathcal{N}', H(w^*))$ coincides with the index set of H at w^*, $ind(H, w^*)$, and the sign of the determinant of $\nabla H(w^*)$. Thus by [4, Theorem 5.2.4], w^* is a stable zero of H, that is, for any open neighborhood $\mathcal{N}$ of w^* satisfying

$$H^{-1}(0) \cap \mathrm{cl}(\mathcal{N}) = \{w^*\},$$

and there exist positive scalars ϵ and η such that for any $\Psi \in \mathcal{B}(H; \epsilon, \mathrm{cl}\mathcal{N})$ and any $w \in \Psi^{-1}(0) \cap \mathcal{N}$,

$$\|w - w^*\| \leq \eta\|H(w)\|.$$

Next we define the set

$$\mathcal{W} := \{w \mid \|w - w^*\| \leq \delta\}$$

where δ is a small positive scalar. Since $\mathcal{W}$ is compact, by assumption, we have

$$\mathcal{W} \cap H^{-1}(0) = \{w^*\}.$$

By the uniform strong law of large numbers [11, Lemma A1], H_N converges to H uniformly w.p.1. on set $\mathcal{W}$, that is, there exist $\epsilon > 0$ such that

$$\sup_{w \in \mathcal{W}} \|H_N(w) - H(w)\| < \epsilon$$

for N large enough, which leads to

$$H_N^{-1}(0) \cap \mathrm{int}(\mathcal{W}) \neq \emptyset,$$

where *int* denotes the interior of a set. Let $w^N := (x^N, \lambda^N, \mu^N)$. Then $w^N \in H_N^{-1}(0) \cap \mathrm{int}(\mathcal{W})$, that is,

$$H_N(x^N, \lambda^N, \mu^N) = 0, \quad (x^N, \lambda^N, \mu^N) \in \mathcal{W}.$$

Furthermore, we have

$$\|w^N - w^*\| \quad \leq \quad c\|H(w^N)\| = c\|H_N(w^N) - H(w^N)\| \leq c \sup_{w\in\mathcal{W}} \|H_N(w) - H(w)\| < c\epsilon.$$

On the other hand, we apply Lemma 3.1 to H_N and H over set $\mathcal{W}$, that is, for the above ϵ, under Assumption 3.1, there exist $\tilde{\kappa}_j(\epsilon) > 0$ and $\nu_j(\epsilon) > 0$, independent of N, such that

$$\text{Prob}\left\{\sup_{w\in\mathcal{W}} |H_{Nj}(w) - H_j(w)| \geq \epsilon\right\} \leq \tilde{\kappa}_j(\epsilon)e^{-N\nu_j(\epsilon)},$$

where $H_{Nj}(w)$ and $H_j(w)$ denote the j-th component of $H_N(w)$ and $H_j(w)$ respectively. Therefore

$$\begin{aligned}\text{Prob}\left\{\|w^N - w^*\| \geq c\epsilon\right\} &\leq \text{Prob}\left\{\sup_{w\in\mathcal{W}} \|H_N(w) - H(w)\| \geq \epsilon\right\} \\ &\leq (n+4m)\kappa(\epsilon)e^{-N\nu(\epsilon)},\end{aligned}$$

where $\kappa(\epsilon) = \max_j\{\tilde{\kappa}_j(\epsilon)\}$, $\nu(\epsilon) = \min_j\{\nu_j(\epsilon)\}$. In particular,

$$\text{Prob}\left\{\|x^N - x^*| \geq c\epsilon\right\} \leq (n+4m)\kappa(\epsilon)e^{-N\nu(\epsilon)}.$$

The proof is complete.

Note that Scheel and Scholtes [12] used a similar condition to (3.4) in their study of stability of deterministic MPEC. Note also that (3.4) implies that the MPEC-LICQ holds at x^*. Assumptions 2.1 and 2.2 are essential in establishing the exponential convergence rate in Theorem 3.1. It is an open and challenging question whether similar results can be obtained under some weaker conditions.

4. Numerical Implementation

In this section, we propose two approaches for practical implementation of the SAA program (1.2): regularization approach and smoothing NCP approach. The first approach is well known and has been widely used to deal with deterministic MPECs [13, 12]. Recently Meng, Ralph and Xu [8] used the approach for solving two stage SMPECs. Likewise, the NCP approach has been used a popular approach for solving deterministic MPECs. It has been recently used to solve two stage SMPECs by Xu [20], Xu and Meng [21] and Meng and Xu [7].

4.1. Regularization and smoothing NCP schemes for SAA program

The regularization approach for SAA programs is defined as follows:

$$\begin{array}{ll}\min\limits_{x\in\mathcal{X}} & \frac{1}{N}\sum_{i=1}^N f(x,\xi^i) \\ \text{s.t.} & N^{-1}\sum_{i=1}^N F_j(x,\xi^i) \geq 0,\ N^{-1}\sum_{i=1}^N G_j(x,\xi^i) \geq 0,\ j=1,\cdots,m, \\ & \left(N^{-1}\sum_{i=1}^N F_j(x,\xi^i)\right)\left(N^{-1}\sum_{i=1}^N G_j(x,\xi^i)\right) \leq \epsilon,\ j=1,\cdots,m,\end{array} \tag{4.1}$$

where $\epsilon > 0$ is called *regularization* parameter. The problem (4.1) is an sample average approximation of the following problem

$$\begin{array}{ll}\min\limits_{x\in\mathcal{X}} & \mathbb{E}[f(x,\xi(\omega))] \\ \text{s.t.} & \mathbb{E}[F_j(x,\xi(\omega))] \geq 0,\ \mathbb{E}[G_j(x,\xi(\omega))] \geq 0,\ j=1,\cdots,m, \\ & \mathbb{E}[F_j(x,\xi(\omega))]\mathbb{E}[G_j(x,\xi(\omega))] \leq \epsilon,\ j=1,\cdots,m,\end{array} \tag{4.2}$$

It is well known that under some mild conditions a weak stationary point of (4.2) approximates its counterpart of the true problem (1.1). See for instances [12]. Note that in numerical

implementation of (4.1), parameter ϵ_N may be related to sample size N, indeed, it is driven to zero as N tends to infinity. Consequently we expect that a weak stationary point of (4.1) converges to a weak stationary point of the true problem (1.1).

The smoothing NCP approach is defined as

$$\begin{array}{ll} \min\limits_{x\in\mathcal{X}} & \frac{1}{N}\sum_{i=1}^{N} f(x,\xi^i) \\ \text{s.t.} & \Psi\left(N^{-1}\sum_{i=1}^{N} F(x,\xi^i), N^{-1}\sum_{i=1}^{N} G(x,\xi^i),\epsilon\right)=0, \end{array} \tag{4.3}$$

where

$$\Psi\left(N^{-1}\sum_{i=1}^{N} F(x,\xi^i), N^{-1}\sum_{i=1}^{N} G(x,\xi^i),\epsilon\right)$$

$$:=\begin{pmatrix} \psi\left(N^{-1}\sum_{i=1}^{N} F_1(x,\xi^i), N^{-1}\sum_{i=1}^{N} G_1(x,\xi^i),\epsilon\right) \\ \vdots \\ \psi\left(N^{-1}\sum_{i=1}^{N} F_m(x,\xi^i), N^{-1}\sum_{i=1}^{N} G_m(x,\xi^i),\epsilon\right) \end{pmatrix},$$

and ψ is a smoothing function of an NCP function. Here, we use a smoothing function of the min-function $min(a,b)$

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

This problem is a sample average approximation of the following problem

$$\begin{array}{ll} \min\limits_{x\in\mathcal{X}} & \mathbb{E}[f(x,\xi(\omega))] \\ \text{s.t.} & \Psi\left(\mathbb{E}[F(x,\xi(\omega))],\mathbb{E}[G(x,\xi(\omega)],\epsilon\right)=0. \end{array} \tag{4.4}$$

It is not difficult to show that under some mild conditions, a weak stationary point of (4.4) is also a weak stationary point of the true problem (1.1). Note that in implementation the smoothing parameter $\epsilon>0$ is driven to zero as N increases. Consequently we expect that a weak stationary point of (4.3) converges to a weak stationary point of the true problem (1.1).

4.2. Estimating the optimal value

In practice, the SAA scheme involves repeated solutions of the SAA problem (4.1) and (4.3) with independent samples. We now discuss consistency of the SAA estimators, including the lower and upper bounds estimations, and the gap of optimal values estimation.

Lower Bound Estimation. For small regularized parameter or smoothing parameter ϵ, let $v^*(\epsilon)$ and $\tilde{v}_N(\epsilon)$ denote the optimal values of the regularized problem (4.2) or the smoothing NCP problem (4.4) and the corresponding SAA problem. It is well known that

$$\mathbb{E}[\tilde{v}_N(\epsilon)]\le v^*(\epsilon).$$

The expected value $\mathbb{E}[\tilde{v}_N(\epsilon)]$ can be estimated as follows. Generate M independent samples, $\{\xi_j^1,\cdots,\xi_j^N\}$, $j=1,\cdots,M$, each of size N. For each sample j, solve the regularized SAA problem

$$\begin{array}{ll} \min\limits_{x\in\mathcal{X}} & \frac{1}{N}\sum_{i=1}^{N} f(x,\xi_j^i) \\ \text{s.t.} & N^{-1}\sum_{i=1}^{N} F(x,\xi_j^i)\ge 0,\ \ N^{-1}\sum_{i=1}^{N} G(x,\xi_j^i)\ge 0, \\ & \left(N^{-1}\sum_{i=1}^{N} F(x,\xi_j^i)\right)\circ\left(N^{-1}\sum_{i=1}^{N} G(x,\xi_j^i)\right)\le\epsilon, \end{array}$$

where $\circ$ denotes Hadmard product, or solve the smoothing NCP SAA problem

$$\begin{array}{ll}\min\limits_{x\in\mathcal{X}} & \frac{1}{N}\sum_{i=1}^{N} f(x,\xi_j^i)\\ \text{s.t.} & \Psi\left(N^{-1}\sum_{i=1}^{N} F(x,\xi_j^i), N^{-1}\sum_{i=1}^{N} G(x,\xi_j^i),\epsilon\right)=0.\end{array}$$

Let $\tilde{v}_N^j(\epsilon)$ denote the corresponding optimal value of the above problem, $j=1,\cdots,M$. We compute

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

which is an unbiased estimate of $\mathbb{E}[\tilde{v}_N(\epsilon)]$. It provides a statistical lower bound for the true optimal value $v^*(\epsilon)$. Let z_α be such that $P\{z\le z_\alpha\}=1-\alpha$ where z satisfies the standard normal distribution, that is, $z\backsim N(0,1)$. By the Central Limit Theorem, we can obtain an approximate $(1-\alpha)$-confidence interval for $\mathbb{E}[\tilde{v}_N(\epsilon)]$ as follows:

$$\left[L_{N,M}(\epsilon)-\frac{z_\alpha s_L(M;\epsilon)}{\sqrt{M}}, L_{N,M}(\epsilon)+\frac{z_\alpha s_L(M;\epsilon)}{\sqrt{M}}\right]$$

where the sample variance estimator of $Var[\tilde{v}_N(\epsilon)]$ is

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

Upper Bound Estimation. An upper bound can be obtained by observing that, for a feasible point $\bar{x}=\bar{x}(\epsilon)$ of the regularized SAA program or the smoothing NCP SAA program with a small regularization parameter or a small smoothing parameter $\epsilon>0$,

$$\tilde{\vartheta}(\bar{x},\epsilon) := \tilde{\vartheta}(\bar{x}(\epsilon)) \ge v^*(\epsilon).$$

Hence, by choosing $\bar{x}(\epsilon)$ to be an ϵ-optimal solution, for example, by solving an SAA problem, and using an unbiased estimator for $\tilde{\vartheta}(\bar{x}(\epsilon))$, we can obtain an estimate of the upper bound of $v^*(\epsilon)$. To do so, we may generate T independent batches of samples: $\{\xi_j^1,\cdots,\xi_j^{N'}\}$, $j=1,\cdots,T$, each of size N'. For any feasible $x(\epsilon)$, let

$$\tilde{v}_{N'}^j(x,\epsilon) := \frac{1}{N'}\sum_{i=1}^{N'} f(x(\epsilon),\xi_j^i).$$

Then $\mathbb{E}[\tilde{v}_{N'}^j(x,\epsilon)]=\tilde{\vartheta}(x,\epsilon)$. We calculate

$$U_{N',T}(\bar{x};\epsilon) := \frac{1}{T}\sum_{j=1}^{T}\tilde{v}_{N'}^j(\bar{x},\epsilon),$$

which is an estimate of $\tilde{\vartheta}(\bar{x},\epsilon)$. By the Central Limit Theorem again, we can obtain a $(1-\alpha)$-confidence interval for $\tilde{\vartheta}(\bar{x},\epsilon)$

$$\left[U_{N',T}(\bar{x};\epsilon)-\frac{z_\alpha s_U(\bar{x},T;\epsilon)}{\sqrt{T}}, U_{N',T}(\bar{x};\epsilon)+\frac{z_\alpha s_U(\bar{x},T;\epsilon)}{\sqrt{T}}\right],$$

where the sample variance estimator of $Var[\tilde{v}_{N\prime}(\bar{x},\epsilon)]$ is

$$s_U^2(\bar{x},T;\epsilon) := \frac{1}{T-1}\sum_{j=1}^{T}\left(\tilde{v}_{N\prime}^j(\bar{x},\epsilon)-U_{N',T}(\bar{x},\epsilon)\right)^2.$$

Note that, in practice, we may take $\bar{x}$ from any of the solutions of the M regularized SAA problems or the smoothing NCP SAA problems as discussed in the previous subsection by

generating independent samples $\{\xi_j^1, \cdots, \xi_j^N\}$, $j = 1, \cdots, M$. Indeed, we will use $\bar{x}_N^j$, the best optimal solution which estimates the smallest optimal value $v^*(\epsilon)$, to compute the upper bound estimates and the optimality gap in Section 5.

Estimating the Gap. We compute an estimate of the optimality gap of the solution $\bar{x}$ using the lower bound estimate and the objective function value estimate as follows:

$$Gap_{N,M,N',T}(\bar{x}) := U_{N',T}(\bar{x};\epsilon) - L_{N,M}(\epsilon),$$

where $U_{N',T}(\bar{x};\epsilon)$ and $L_{N,M}(\epsilon)$ are defined as before. The estimated variance of the above gap estimator is then given by

$$S_{\text{Gap}}^2 = \frac{1}{M} s_L^2(M;\epsilon) + \frac{1}{T} s_U^2(\bar{x}, T;\epsilon).$$

5. Preliminary Computational Results

We have carried out numerical tests on SAA with regularization and SAA with NCP reformulation. The tests are carried out by implementing mathematical programming codes in MATLAB 6.5 installed in a PC with Windows XP Operating System. We use Matlab built-in solver *fmincon* for solving the regularized SAA problem (4.1) and smoothing NCP SAA problem (4.3). We use the Latin Hypercube Sampling (LHS) to generate samples.

We now report the numerical test results. In each of the following test problems, we choose different values for the regularization or the smoothing parameter ϵ and sample sizes N, M, N', and T. In the tables below, "reg" denotes the regularized SAA scheme and "ncp-min" denotes the smoothing NCP SAA method. $\tilde{v}_N(\epsilon)$ stands for the lower bound estimation $L_{N,M}(\epsilon)$ and $\tilde{\vartheta}(\bar{x}_N^j, \epsilon)$ represents the upper bound estimation where $\bar{x}_N^j$ is an optimal solution which estimates the smallest optimal value as discussed in Section 4. The 95%-confidence intervals around $\tilde{v}_N(\epsilon)$ and $\tilde{\vartheta}(\bar{x}_N^j, \epsilon)$ are also reported. For brevity, $Gap(\bar{x}_N^j)$ denotes $Gap_{N,M,N',T}(\bar{x}_N^j)$ and S_{Gap} represents the estimated variance of the optimality gap as follows:

$$S_{\text{Gap}} := \sqrt{\frac{1}{M} S_L^2(M;\epsilon) + \frac{1}{T} S_U^2(\bar{x}_N^j, T;\epsilon)}.$$

In order to compare with the performance of the proposed SAA methods of this paper, for each test problem, we also report the numerical results of the corresponding deterministic MPCC problem.

Example 5.1. *Consider the following problem*

$$\begin{array}{ll} \min & \mathbb{E}[(x_1-1)^2 + (x_2-1)^2 + x_1\xi_1 + x_2\xi_2 + \xi_1\xi_2] \\ \text{s.t.} & 0 \le x_1 \le 2, \quad 0 \le x_2 \le 2, \\ & 0 \le \mathbb{E}[F(x,\xi)] \perp \mathbb{E}[G(x,\xi)] \ge 0, \end{array}$$

where

$$F(x,\xi) = (x_1 - \xi_1 + \xi_2, x_2 + \xi_1 - \xi_2)^T, \;\; G(x,\xi) = (x_1 - \xi_1, x_2 - \xi_2)^T,$$

and ξ_1, ξ_2 are independent random variables having uniform distribution on $[0,1]$. The test results are presented in Table 5.1.1.

Table 5.1.1 Test results for Example 5.1

methods	N	M	N'	T	ϵ	$\tilde{v}_N(\epsilon)$ (95% conf. int.)	$\bar{x}_N^j$	$\tilde{\vartheta}(\bar{x}_N^j, \epsilon)$ (95% conf. int.)	$Gap(\bar{x}_N^j)$	S_{Gap}
reg	400	30	600	40	10^{-4}	1.2497 ± 0.0014	(0.5002, 0.5002)	1.2505 ± 0.0044	2.8549e-004	0.0028
ncp-min	400	30	600	40	10^{-4}	1.2502 ± 0.0012	(0.5000, 0.5000)	1.2503 ± 0.0051	1.6172e-004	0.0030
reg	1000	60	1500	60	10^{-6}	1.2501 ± 4.8539e-004	(0.5000, 0.5000)	1.2512 ± 0.0021	0.0011	0.0016
ncp-min	1000	60	1500	60	10^{-6}	1.2503 ± 5.5244e-004	(0.5000, 0.5000)	1.2506 ± 0.0028	2.5948e-004	0.0018

Note that in Example 5.1, we can integrate the underlying functions out and obtain a deterministic MPCC problem

$$\begin{array}{ll} \min & (x_1-1)^2+(x_2-1)^2+\dfrac{1}{2}(x_1+x_2)+\dfrac{1}{4} \\ \text{s.t.} & 0\le x_1\le 2, \quad 0\le x_2\le 2, \\ & 0\le \bar{F}(x)\perp \bar{G}(x)\ge 0, \end{array} \tag{5.1}$$

where $\bar{F}(x)=x$ and $\bar{G}(x)=(x_1-1/2, x_2-1/2)^T$. The test results are stated in Table 5.1.2.

Table 5.1.2 Numerical results for problem (5.1)

methods	ϵ	x^*	optimal value
reg	10^{-4}	(0.5002, 0.5002)	1.2498
ncp-min	10^{-4}	(0.5000, 0.5000)	1.2500
reg	10^{-6}	(0.5000, 0.5000)	1.2500
ncp-min	10^{-6}	(0.5000, 0.5000)	1.2500

Example 5.2. *Consider the following problem*

$$\begin{array}{ll} \min & \mathbb{E}[(x_1-1)^2+(x_2-1)^2-x_1+x_2+x_1\xi_1+x_2\xi_2+\xi_1\xi_2] \\ \text{s.t.} & -2\le x_1\le 2, \quad -2\le x_2\le 2, \\ & 0\le \mathbb{E}[F(x,\xi)]\perp \mathbb{E}[G(x,\xi)]\ge 0, \end{array}$$

where

$$F(x,\xi)=(x_1-x_2-\xi_1+\xi_2, x_2-1+\xi_1-\xi_2)^T, \;\; G(x,\xi)=(x_1-1-\xi_1, x_2-\xi_2)^T,$$

and ξ_1, ξ_2 are independent random variables having a truncated normal distribution $N(\mu,\sigma^2)$ on [-0.5, 0.5] with $\mu=0$ and $\sigma=3$. The test results are presented in Table 5.2.1.

Table 5.2.1 Test results for Example 5.2

methods	N	M	N'	T	ϵ	$\tilde{v}_N(\epsilon)$ (95% conf. int.)	$\bar{x}_N^j$	$\tilde{\vartheta}(\bar{x}_N^j, \epsilon)$ (95% conf. int.)	$Gap(\bar{x}_N^j)$	S_{Gap}
reg	400	40	600	40	10^{-5}	0.0697 ± 0.0056	(1.0280, 0.9570)	-0.0052 ± 0.0047	0.0645	0.0029
ncp-min	400	40	600	40	10^{-5}	-0.0169 ± 0.0076	(1.0000, 0.9786)	-0.0058 ± 0.0055	0.0111	0.0057
reg	800	60	700	60	10^{-7}	-0.0568 ± 0.0118	(1.0233, 0.9642)	-0.0089 ± 0.0008	0.0478	0.0013
ncp-min	800	60	700	60	10^{-7}	-0.0388 ± 0.0040	(0.0759, 0.0427)	-0.0012 ± 0.0027	0.0376	0.0029

Note that in Example 5.2, we can integrate the underlying functions out and obtain a deterministic MPCC problem

$$\begin{array}{ll} \min & (x_1-1)^2+(x_2-1)^2-x_1+x_2 \\ \text{s.t.} & -2\le x_1\le 2, \quad -2\le x_2\le 2, \\ & 0\le \bar{F}(x)\perp \bar{G}(x)\ge 0, \end{array} \tag{5.2}$$

where $\bar{F}(x)=(x_1-x_2, x_2-1)^T$ and $\bar{G}(x)=(x_1-1, x_2)^T$. The test results are presented in Table 5.2.2.

Table 5.2.2 Numerical results for problem (5.2)

methods	ϵ	x^*	optimal value
reg	10^{-5}	(1.0032, 1.0000)	-0.0032
ncp-min	10^{-5}	(1.0000, 1.0000)	-5.0000e-006
reg	10^{-7}	(1.0003, 1.0000)	-3.1615e-004
ncp-min	10^{-7}	(1.0000, 1.0000)	-2.4425e-015

Example 5.3. *Consider the following problem*

$$\begin{aligned} \min \quad & \mathbb{E}[(x_1-1)^2+(x_2-1)^2+x_1x_2+\xi_1+\xi_2] \\ \text{s.t.} \quad & 0\le x_1\le 2, \quad 0\le x_2\le 2, \\ & 0\le \mathbb{E}[F(x,\xi)] \perp \mathbb{E}[G(x,\xi)]\ge 0, \end{aligned}$$

where

$$F(x,\xi)=(x_1+x_2-\xi_1,x_2-\xi_2)^T,\ \ G(x,\xi)=(x_1-\xi_1-\xi_2,x_2-\xi_1-\xi_2)^T,$$

and ξ_1, ξ_2 are independent random variables having uniform distribution on $[0,1]$. The test results are stated in Table 5.3.1.

Table 5.3.1 Test results for Example 5.3

methods	N	M	N'	T	ϵ	$\tilde{v}_N(\epsilon)$ (95% conf. int.)	$\bar{x}_N^j$	$\tilde{\vartheta}(\bar{x}_N^j,\epsilon)$ (95% conf. int.)	$Gap(\bar{x}_N^j)$	S_{Gap}
reg	500	40	700	40	10^{-5}	2.000 ± 3.0543e-005	(0.9999, 0.9999)	2.0007 ± 0.0037	0.0007	0.0022
ncp-min	500	40	700	40	10^{-5}	2.0000 ± 2.7410e-005	(0.9999, 0.9999)	2.0003 ± 2.7410e-005	0.0003	0.0024
reg	1000	60	1500	60	10^{-7}	2.0000 ± 8.0645e-006	(1.0000, 1.0000)	2.0005 ± 0.0021	0.0005	0.0013
ncp-min	1000	60	1500	60	10^{-7}	2.0000 ± 9.0249e-006	(1.0000, 1.0000)	2.0032 ± 0.0025	0.0032	0.0015

Note that in Example 5.3, we can integrate the underlying functions out and obtain a deterministic MPCC problem

$$\begin{aligned} \min \quad & (x_1-1)^2+(x_2-1)^2+x_1x_2+1 \\ \text{s.t.} \quad & 0\le x_1\le 2, \quad 0\le x_2\le 2, \\ & 0\le \bar{F}(x) \perp \bar{G}(x)\ge 0, \end{aligned} \tag{5.3}$$

where $\bar{F}(x)=(x_1+x_2-1/2,x_2-1/2)^T$ and $\bar{G}(x)=(x_1-1,x_2-1)^T$. The test results are presented in Table 5.3.2.

Table 5.3.2 Numerical results for problem (5.3)

methods	ϵ	x^*	optimal value
reg	10^{-5}	(1.0000, 1.0000)	2.0000
ncp-min	10^{-5}	(1.0000, 1.0000)	2.0000
reg	10^{-7}	(1.0000, 1.0000)	2.0000
ncp-min	10^{-7}	(1.0000, 1.0000)	2.0000

Example 5.4. *Consider the following problem*

$$\begin{aligned} \min \quad & \mathbb{E}[2(x_1-1)^2+(x_2-1)^2+2x_1x_2+\xi_1^2+\xi_2^2] \\ \text{s.t.} \quad & -3\le x_1\le 3, \quad -3\le x_2\le 3, \\ & 0\le \mathbb{E}[F(x,\xi)] \perp \mathbb{E}[G(x,\xi)]\ge 0, \end{aligned}$$

where

$$F(x,\xi)=(x_1+x_2+\xi_1+\xi_2,2x_2-1+\xi_1+\xi_2)^T,\ \ G(x,\xi)=(x_1-\xi_1,x_2-\xi_2)^T,$$

and ξ_1, ξ_2 are independent random variables having a truncated normal distribution $N(\mu,\sigma^2)$ on [-0.3, 0.3] with $\mu=0,\sigma=5$. The test results are stated in Table 5.4.2.

Table 5.4.1 Test results for Example 5.4

methods	N	M	N'	T	ϵ	$\tilde{v}_N(\epsilon)$ (95% conf. int.)	$\bar{x}_N^j$	$\tilde{\vartheta}(\bar{x}_N^j, \epsilon)$ (95% conf. int.)	$Gap(\bar{x}_N^j)$	S_{Gap}
reg	400	40	500	40	10^{-5}	2.2441 ± 0.0016	(0.0050, 0.4976)	2.2561 ± 4.7576e-005	0.0120	0.0009
ncp-min	400	40	500	40	10^{-5}	2.2367 ± 0.0021	(0.0081, 0.4958)	2.2568 ± 5.6762e-005	0.0201	0.0013
reg	800	60	1500	60	10^{-7}	2.2463 ± 0.0010	(0.0045, 0.4969)	2.2559 ± 2.4513e-005	0.0096	0.0006
ncp-min	800	60	1500	60	10^{-7}	2.2437 ± 0.0011	(0.0050, 0.4980)	2.2558 ± 2.4171e-005	0.0121	0.0007

Note that in Example 5.4, we can integrate the underlying functions out and obtain a deterministic MPCC problem

$$\begin{array}{ll} \min & 2(x_1-1)^2+(x_2-1)^2+2x_1x_2 \\ \text{s.t.} & -3 \le x_1 \le 3, \quad -3 \le x_2 \le 3, \\ & 0 \le \bar{F}(x) \perp \bar{G}(x) \ge 0, \end{array} \tag{5.4}$$

where $\bar{F}(x) = (x_1+x_2, 2x_2-1)^T$ and $\bar{G}(x) = (x_1, x_2)^T$. The test results are presented in Table 5.4.2.

Table 5.4.2 Numerical results for problem (5.4)

methods	ϵ	x^*	optimal value
reg	10^{-5}	(0.0000, 0.5000)	2.2499
ncp-min	10^{-5}	(0.0000, 0.5000)	2.2500
reg	10^{-7}	(0.0000, 0.5000)	2.2500
ncp-min	10^{-7}	(0.0000, 0.5000)	2.2500

From the preliminary computational experience, we can see that sample average approximation methods appear to be effective and promising for solving stochastic mathematical programs with equilibrium constraints. The numerical results also show that regularized SAA methods and smoothing NCP SAA methods have different performances for different problems. We are not able to conclude which method is better.

References

[1] Z. Artstein and R. J-B Wets, Consistency of minimizers and the SLLN for stochastic programs, *Journal of Convex Analysis*, **2** (1995), 1–17.

[2] F. Bastin, C. Cirillo, P. L. Toint, Convergence theory for nonconvex stochastic programming with an application to mixed logit, *Mathematical Programming, Series B*, **108** (2006), 207-234.

[3] S.I. Birbil, G. Gürkan, and O. Listes, Simulation-based solution of stochastic mathematical programs with complementarity constraints: Sample-path analysis, Econometric Institute Report EI 2004-03, Erasmus University Rotterdam, 2004.

[4] F. Facchinei and J.S. Pang, Finite-Dimensional Variational Inequalities and Complementarity Problems, Springer-Verlag, New York, 2003.

[5] A.J. King and R.J.B. Wets, Epi-consistency of convex stochastic programs, *Stochastics Reports*, **34** (1991), 83–92.

[6] Z.Q. Luo, J.S. Pang, and D. Ralph, Mathematical Programs with Equilibrium Constraints, Cambridge University Press, Cambridge, 1996.

[7] F. Meng and H. Xu, A Regularized Sample Average Approximation Method for Stochastic Mathematical Programs with Nonsmooth Equality Constraints, to appear in *SIAM Journal on Optimization*.

[8] F. Meng, D. Ralph and H. Xu, Regularization methods for stochastic MPCC with multiple-equilibria, manuscript, School of Mathematics, University of Southampton, Southampton, UK, 2005.

[9] E.L. Plambeck, B.R. Fu, S.M. Robinson, and R. Suri, Sample-path optimization of convex stochastic performance functions, *Mathematical Programming*, **75** (1996), 137-176.

[10] S.M. Robinson, Analysis of sample-path optimization, *Mathematics of Operations Research*, **21** (1996), 513-528.

[11] R.Y. Rubinstein and A. Shapiro, Discrete Events Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Methods, John Wiley and Sons, New York, 1993.

[12] H. Scheel and S. Scholtes, Mathematical programs with equilibrium constraints: stability, optimality, and sensitivity, *Mathematics of Operations Research*, **25** (2000), 1–22.

[13] S. Scholtes, Convergence properties of a regularization scheme for mathematical programs with complementarity constraints, *SIAM Journal on Optimization*, **11** (2001), 918–936.

[14] S. Scholtes, Nonconvex structure in nonlinear programming, *Mathematics of Operations Research*, **11** (2001), 918–936.

[15] A. Shapiro, Stochastic Programming by Monte Carlo Simulation Methods, Published electronically in: Stochastic Programming E-Print Series, 2000.

[16] A. Shapiro, Monte Carlo sampling methods, in A. Rusczyński and A. Shapiro, editors, *Stochastic Programming*, Handbooks in OR & MS, Vol. 10, North-Holland Publishing Company, Amsterdam, 2003.

[17] A. Shapiro, Stochastic mathematical programs with equilibrium constraints, *Journal of Optimization Theory and Applications,* **128** (2006), 223-243.

[18] A. Shapiro and T. Homem-de-Mello, On rate of convergence of Monte Carlo approximations of stochastic programs, *SIAM Journal on Optimization*, **11** (2000), 70-86.

[19] A. Shapiro and H. Xu, Stochastic mathematical programs with equilibrium constraints, modeling and sample average approximation, http://www.optimization-online.org/ DB_HTML / 2005/ 01/ 1046.html, 2005.

[20] H. Xu, An implicit programming approach for a class of stochastic mathematical programs with equilibrium constraints, *SIAM Journal on Optimization*, **16** (2006), 670-696.

[21] H. Xu and F. Meng, Convergence analysis of sample average approximation methods for a class of stochastic mathematical programs with equality constraints, manuscript, School of Mathematics, University of Southampton, Southampton, UK, 2005, To appear in *Mathematics of Operations Research*.