Optimization
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713646500

Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation

Alexander Shapiro [a]; Huifu Xu [b]
[a] School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, USA [b] School of Mathematics, University of Southampton, Highfield, Southampton, UK

Online Publication Date: 01 January 2008

To cite this Article Shapiro, Alexander and Xu, Huifu(2008)'Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation',Optimization,57:3,395 — 418
To link to this Article: DOI: 10.1080/02331930801954177
URL: http://dx.doi.org/10.1080/02331930801954177

Optimization
Vol. 57, No. 3, June 2008, 395–418

Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation

Alexander Shapiro[a]* and Huifu Xu[b]

[a]*School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, USA;*
[b]*School of Mathematics, University of Southampton, Highfield, Southampton, UK*

(*Received 13 July 2006; final version received 15 October 2007*)

In this article, we discuss the sample average approximation (SAA) method applied to a class of stochastic mathematical programs with variational (equilibrium) constraints. To this end, we briefly investigate the structure of both – the lower level equilibrium solution and objective integrand. We show almost sure convergence of optimal values, optimal solutions (both local and global) and generalized Karush–Kuhn–Tucker points of the SAA program to their true counterparts. We also study uniform exponential convergence of the sample average approximations, and as a consequence derive estimates of the sample size required to solve the true problem with a given accuracy. Finally, we present some preliminary numerical test results.

Keywords: stochastic programming; equilibrium constraints; Stackelberg–Nash–Cournot equilibrium; variational inequality; sample average approximation; exponential convergence; smoothing

1. Introduction

In this article, we consider the following optimization problem:

$$\begin{array}{ll} \underset{x\in\mathcal{X},\, y(\cdot)}{\text{Min}} & \mathbb{E}\,[F(x, y(\omega), \xi(\omega))] \\ \text{s.t.} & 0 \in H(x, y(\omega), \xi(\omega)) + \mathcal{N}_{\mathcal{C}}(y(\omega)), \quad \text{a.e. } \omega \in \Omega. \end{array} \tag{1.1}$$

Here $\mathcal{X}$ is a non-empty subset of $\mathbb{R}^n$, $F\colon \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}$ and $H\colon \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}^m$, $\xi(\omega)$ is a random vector defined on a probability space $(\Omega, \mathcal{F}, P)$, $\mathcal{C}$ is a non-empty convex closed subset of $\mathbb{R}^m$ and $\mathcal{N}_{\mathcal{C}}(y)$ denotes the normal cone to $\mathcal{C}$ at y,

$$\mathcal{N}_{\mathcal{C}}(y) := \left\{z \in \mathbb{R}^m : z^T(y' - y) \le 0, \quad \forall\, y' \in \mathcal{C}\right\}, \quad \text{if } y \in \mathcal{C},$$

and $\mathcal{N}_{\mathcal{C}}(y) := \emptyset$ if $y \notin \mathcal{C}$. By $\Xi \subset \mathbb{R}^d$ we denote the support of the distribution of $\xi(\omega)$, i.e. Ξ is the smallest closed subset of $\mathbb{R}^d$ such that $P\{\xi(\omega) \in \Xi\} = 1$.

Problem (1.1) can be viewed as a two-stage stochastic programming problem (of the 'here-and-now' type) with recourse and variational constraints. At the first-stage one is supposed to make a decision about value of vector x, restricted to the feasible set $\mathcal{X}$, *before*

*Corresponding author. Email: ashapiro@isye.gatech.edu

ISSN 0233–1934 print/ISSN 1029–4945 online

DOI: 10.1080/02331930801954177
http://www.informaworld.com

a realization of the random data $\xi(\omega)$ becomes available. At the second-stage, when x is fixed and a realization $\xi = \xi(\omega)$ becomes known, the following variational inequality (VI) is solved to calculate the second-stage decision vector y: find $y \in \mathcal{C}$ such that

$$H(x, y, \xi)^T(y' - y) \geq 0, \quad \forall\, y' \in \mathcal{C}. \tag{1.2}$$

(Note that this VI can also be written in the form $0 \in H(x, y, \xi) + \mathcal{N}_{\mathcal{C}}(y)$, and that if y is a solution of this VI, then necessarily $y \in \mathcal{C}$.) The second-stage problem (VI) depends on a realization of the random data $\xi(\omega)$ (and on x), and therefore its solution $y = y(\omega)$ is considered as a function of ω. Of course, it may happen that such solution is not unique or does not exist. The precise meaning of this model will be discussed in the next section. We refer to (1.1) as a stochastic mathematical program with equilibrium constraints (SMPEC) since the second-stage VI constraint is often interpreted as an equilibrium constraint.

Patriksson and Wynter [17] discussed a general class of SMPECs, similar to (1.1), except that in their formulation the objective function F does not depend on random variables while the set $\mathcal{C}$ in the variational constraint depends on both x and ω. They investigated existence of optimal solutions, Lipschitz continuity, convexity and differentiability of the object function, and proposed a numerical method for solving SMPECs with a finite number of scenarios. The discussion was extended by Evgrafov and Patriksson [5]. More recently Shapiro [27] considered a class of here-and-now type of SMPECs and investigated their basic properties such as measurability and continuity of the objective integrand, and consistency and rate of convergence on sample average approximation method for such problems. On the application aspect, Christiansen et al. [3] considered an SMPEC model for a class of stochastic bilevel programming problems in structural optimization, Xu [30] modelled a stochastic Stackelberg–Nash–Cournot equilibrium problem as a specific SMPEC. It could be mentioned that the two-stage SMPEC problem (1.1) is quite different from the stochastic variational inequality of the form

$$0 \in \mathbb{E}[H(y, \xi(\omega))] + \mathcal{N}_{\mathcal{C}}(y), \tag{1.3}$$

discussed in Gürkan et al. [8].

A particularly interesting case of SMPEC is when $\mathcal{C} := \mathbb{R}^m_+$. Consequently, (1.1) becomes a stochastic mathematical program with complementarity constraints (SMPCC). The latter was investigated by Lin et al. in [13] with a focus on the discrete distribution of $\xi(\omega)$. Xu [31] extended the discussion to the continuous distribution case and used an implicit programming approach to investigate such problems. A notable point in [31] is a deterministic discretization method in which a set of grid points in the support set of $\xi(\omega)$ are chosen and the stochastic complementarity constraint is replaced by a set of complementarity constraints with $\xi(\omega)$ taking value at these grid points. Error bounds for this type of deterministic discretization method are also discussed in [31].

It is well-known that even a crude discretization of the random data results in an exponential growth of the number of scenarios and quickly becomes unmanageable with increase in the number of random parameters. Therefore, that way or another, one needs to reduce the set of considered scenarios to a manageable level in order to solve the obtained SMPEC numerically. One approach to such a reduction is to use the

Monte Carlo sampling techniques. In that respect, the sample average approximation (SAA) method turned out to be surprisingly efficient.

The idea of the SAA method is simple indeed. A random (or rather pseudo-random) sample $\xi^1, \ldots, \xi^N$ of N realizations of the random vector $\xi(\omega)$ is generated and the involved expected value function is approximated by the corresponding sample average function. The SAA method and its variants, known under various names such as 'stochastic counterpart method', 'sample-path method', 'simulated likelihood method', etc., were discussed in the stochastic programming and statistics literature over the years. Statistical properties of SAA estimators were studied in [4,11,18,19,23,25], for example. Recently it was shown that (under certain regularity conditions) SAA estimators converge exponentially fast to their true counterparts with increase of the sample size N [15,30,31,34]. Theoretical findings were supported by numerical studies showing that the SAA method can be surprisingly efficient in solving large scale (even non-convex) stochastic programming problems (see, e.g. [8,15,16,21,29]).

In this article, we discuss an application of the SAA method to the SMPEC problem (1.1) with a focus on the case where the variational constraint has a unique solution for every $x \in \mathcal{X}$ and almost every realization of $\xi(\omega)$. This narrows the scope of the discussion, but allows one to obtain results somewhat sharper than those covered in [27]. Specifically, we briefly investigate differentiability properties of both the lower level equilibrium solution and objective integrand. Then we investigate almost sure convergence of optimal values, optimal solutions (both local and global) and generalized Karush–Kuhn–Tucker points of the SAA program to their true counterparts. We also study uniform exponential convergence of the sample average approximations, and consequently derive a probabilistic estimate of the sample size required for a sharp local optimizer of the true SMPEC to be a sharp local optimizer of a sample average approximation program. Finally, we outline implementation of the SAA method for SMPEC problems and present some initial numerical results.

The rest of this article is organized as follows. In Section 2, we present a detailed explanation of our model. As an example, we illustrate that a stochastic leader–followers problem can be fitted into the model. In Section 3, we investigate properties such as existence, uniqueness, Lipschitz continuity and directional differentiability of solutions of the parametric variational inequality (1.2). In Section 4, we discuss consistency and convergence of some estimators in our proposed sample average approximation scheme for (1.1). In Section 5, we show the exponential convergence of the sample average approximation scheme. Finally, in Section 6, we discuss some smoothing or regularization techniques and present preliminary numerical test results.

2. Preliminary discussions and an example

In this section, we make a more detailed discussion of model (1.1) and present an example. We then present some preliminary discussion of the SAA method for (1.1).

2.1 *A discussion of the SMPEC model*

Although our main emphasis, later on in this article, will rest on the case where the equilibrium constraint (1.2) has a unique solution, we believe that it will be helpful to discuss a precise meaning of model (1.1) in a general setting where (1.2) may have

multiple solutions. Of course, in some applications it happens that the solution of (1.2) is unique, but this cannot be guaranteed in general and it is important to understand an implication of existence of multiple equilibrium solutions in (1.1).

When the equilibrium constraint (1.2) has multiple solutions, a decision needs to be made as to which equilibrium solution should be taken into account, bearing in mind that different equilibrium will result in different objective values. In practice, such selection depends on the attitude and needs of the (first-stage) decision maker. We give below a precise mathematical formulation of (1.1) based on the following two equivalent approaches. One is to consider the optimization problem

$$\operatorname*{Min}_{y\in\mathbb{R}^m} \; F(x,y,\xi) \;\text{ s.t. }\; 0\in H(x,y,\xi)+\mathcal{N}_C(y), \tag{2.1}$$

where x and ξ are treated as parameters. Note that for given x and ξ, problem (2.1) is a deterministic mathematical program with equilibrium constraints (MPEC). Denote by $Q(x,\xi)$ the optimal value of problem (2.1). By the definition, $Q(x,\xi)=+\infty$ if the corresponding variational inequality (1.2) does not have a solution. Then the SMPEC (1.1) can be written as

$$\operatorname*{Min}_{x\in\mathcal{X}} \big\{f(x):=\mathbb{E}\,[Q(x,\xi(\omega))]\big\}, \tag{2.2}$$

where the expectation is taken with respect to the probability distribution P of the random vector $\xi(\omega)$.

The other way which leads to the formulation of (1.1) is as follows. Denote by $\mathfrak{S}(x,\xi)$ the set of solutions of the variational inequality (1.2). Note that the set $\mathfrak{S}(x,\xi)$ can be empty or contain more than one element. Then we can rewrite (1.1) as

$$\operatorname*{Min}_{x\in\mathcal{X},\, y(\cdot)\in\mathfrak{S}(x,\xi(\cdot))} \mathbb{E}\,[F(x,y(\omega),\xi(\omega))], \tag{2.3}$$

where the optimization in (2.3) is performed over all $x\in\mathcal{X}$ and all measurable selections $y(\omega)\in\mathfrak{S}(x,\xi(\omega))$. It should be noted that in this formulation, $y(\omega)$ lies in a functional space of measurable mappings $y:\Omega\to\mathbb{R}^m$.

Equivalence of formulations (2.2) and (2.3) is well documented in the stochastic programming literature (e.g. [20, Chapter 1, section 2.4]). In that respect note that $Q(x,\xi)$ can be viewed as the optimal value of the problem:

$$\operatorname*{Min}_{y\in\mathbb{R}^m} \; F(x,y,\xi) \;\text{ s.t. }\; y\in\mathfrak{S}(x,\xi). \tag{2.4}$$

In particular, if $\mathfrak{S}(x,\xi)=\{\bar{y}(x,\xi)\}$ is a singleton for any $x\in\mathcal{X}$ and $\xi\in\Xi$, then $Q(x,\xi)=F(x,\bar{y}(x,\xi),\xi)$ and problems (2.2) and (2.3) can be written in the form (2.6) below. If $\mathfrak{S}(x,\xi)$ contains more than one solution, then problem (2.3) will choose a solution which minimizes $F(x,\cdot,\xi(\omega))$. Only for such a choice of an element of $\mathfrak{S}(x,\xi)$, problems (2.2) and (2.3) are equivalent, and this gives the precise meaning of formulation (1.1).

Since formulation (1.1) suggests the minimal value of the objective in the case that the set $\mathfrak{S}(x,\xi)$ is not a singleton, it can be viewed as *optimistic*. A pessimistic decision maker

would try to hedge against a worst possible situation by maximizing $F(x, \cdot, \xi(\omega))$, instead of minimizing, over the set $\mathfrak{S}(x, \xi)$ of the second-stage equilibrium solutions. This leads to the following min–max formulation of the corresponding SMPEC:

$$\operatorname*{Min}_{x \in \mathcal{X}} \operatorname*{Max}_{y(\cdot) \in \mathfrak{S}(x, \xi(\cdot))} \mathbb{E}[F(x, y(\omega), \xi(\omega))]. \tag{2.5}$$

Of course, two formulations (2.3) and (2.5) are equivalent if the set $\mathfrak{S}(x, \xi)$ is a singleton for all $x \in \mathcal{X}$ and a.e. (P-almost every) $\xi \in \Xi$. Note also that the *pessimistic* formulation (2.5) is equivalent to (2.2) if $Q(x, \xi)$ is defined as the *maximal* (instead of minimal) value of $F(x, y, \xi)$ over $y \in \mathfrak{S}(x, \xi)$.

It is important to note that in model (1.1), $y(\cdot)$ is treated as the second-stage decision variable which is a solution of the problem depending on realization of the random vector $\xi(\omega)$ and the first-stage decision vector x. In particular, if for any $x \in \mathcal{X}$ and a.e. $\xi \in \Xi$, variational inequality (1.2) has a *unique* solution $\bar{y} = \bar{y}(x, \xi)$, i.e. $\mathfrak{S}(x, \xi) = \{\bar{y}(x, \xi)\}$, then the second-stage decision vector is an implicit function of x and ξ. In that case, problem (1.1) can be written as

$$\operatorname*{Min}_{x \in \mathcal{X}} \mathbb{E}[F(x, \bar{y}(x, \xi(\omega)), \xi(\omega))]. \tag{2.6}$$

This fits a stochastic leader–followers model as we shall explain in an example later.

It could also be mentioned that even under the condition that variational inequality (1.2) possesses unique solution for all $x \in \mathcal{X}$ and a.e. $\xi \in \Xi$, it could be quite a challenge to solve the corresponding SMPEC numerically. At this stage it seems to be hopeless to solve SMPEC problems numerically in absence of such condition. Therefore, the discussion of situations with possibly non-unique solutions is given mainly for conceptual purposes.

The above SMPEC model is somewhat different from a here-and-now SMPEC recourse problem considered by Lin et al. [13] and Lin and Fukushima [14]. In their model, both x and y are treated as *first*-stage decision variables (which may belong to different decision makers) and are subject to a complementarity constraint where y is treated as the prime variable and x and $\xi(\omega)$ are treated as parameters. At the second-stage, an additional recourse vector $z(\omega)$ is introduced to deal with the case when the complementarity constraint does not have a solution for some realizations of $\xi(\omega)$. A penalty is consequently added to the objective. This can be viewed as an extension of the classical stochastic programming recourse model to the SMPEC setting. Practical interpretations of this approach and the one considered in this article are different and could be applied to different situations.

Model (1.1) has a wide range of applications in areas such as decision analysis [30] and structural optimization [3]. Here, we give an example in decision analysis.

Example 2.1 Consider a two-stage stochastic leader–followers problem in an oligopoly market where $m+1$ firms compete to supply a homogeneous product in a non-competitive manner. The leader's decision problem can be formulated as follows:

$$\begin{aligned} \operatorname*{Max}_{x \in [0, U_0], y(\cdot)} \quad & \mathbb{E}\left[xp\left(x + \textstyle\sum_{i=1}^{m} y_i(\omega), \xi(\omega)\right)\right] - c_0(x) \\ \text{s.t.} \quad & y_i(\omega) \in \arg\operatorname*{max}_{y_i \in [0, U_i]} \left\{ y_i p\left(x + y_i + \sum_{k \neq i} y_k(\omega), \xi(\omega)\right) - c_i(y_i) \right\}, \ i = 1, \ldots, m. \end{aligned} \tag{2.7}$$

Here x denotes leader's supply with lower bound being normalized to 0 and upper bound $U_0 > 0$, and y_i denotes i-th follower's decision variable with lower bound 0 and upper bound $U_i > 0$, $p(q, \xi)$ denotes the inverse demand function under random shock $\xi = \xi(\omega)$, i.e. if the total supply to the market is q, then market price is $p(q, \xi)$ at scenario $\xi(\omega) = \xi$, $c_0(\cdot)$ denotes the leader's cost function and $c_i(\cdot)$ dhenotes follower i's cost function.

In this problem, the leader needs to make a decision before a realization of uncertainties in market demand to maximize his expected profit, based on his knowledge of the distribution of $\xi(\omega)$ and his belief on how followers will react to his supply in every demand scenario. That is, in the above formulation (2.7), the leader's decision variable x is viewed as the first-stage decision variable which should be made before a realization of the random data becomes known, while at the second-stage every follower will make his optimal decision $y_i(\omega)$ based on knowledge of leader's decision and a realization of the market demand.

Suppose that $p(\cdot, \xi)$ is concave, differentiable and strictly decreasing for every $\xi \in \Xi$, and functions $c_i(\cdot)$, $i = 1, \ldots, m$, are convex and differentiable. It follows from [36, Proposition 2.4] that for every $i \in \{1, \ldots, m\}$, the function inside the parentheses in the right-hand side of (2.7) is concave in $y_i \in \mathbb{R}_+$. Consequently, using KKT-optimality conditions, we can rewrite the followers' maximization problems as a variational inequality (1.2) with $\mathcal{C} := [0, U_1] \times \cdots \times [0, U_m]$ and $H(x, y, \xi) = (H_1(x, y, \xi), \ldots, H_m(x, y, \xi))$,

$$H_i(x, y, \xi) := -p\left(x + \sum_{i=1}^{m} y_i, \xi\right) - y_i p'\left(x + \sum_{i=1}^{m} y_i, \xi\right) + c_i'(y_i), \quad i = 1, \ldots, m. \tag{2.8}$$

Program (2.7) can be reformulated. Then, as a SMPEC of the form (1.1). In particular, if we assume that the corresponding variational inequality has a unique solution, i.e. followers have a unique Nash equilibrium for each feasible x and a.e. $\xi \in \Xi$, then the optimistic and pessimistic formulations are equivalent and the resulting SMPEC can be written in the form (2.6). Such uniqueness of solutions can be ensured under some convexity conditions on cost functions c_i and concavity condition on demand function p, in which case H satisfies a monotonicity condition. However, when the follower's equilibrium has multiple solutions, the leader must consider a proper selection from the equilibria in his decision making. In particular, if the leader is optimistic, then he will select an equilibrium that maximizes his profit (for the given x in each demand scenario $\xi(\omega) = \xi$). This will eventually lead to the optimistic formulation of the corresponding SMPEC as it was discussed above (by negating the leader's profit function).

In general, the leader's objective function, given by the expected profit, might not be a concave function of x. However, under some special circumstances such as a linear market demand function and linear cost function of followers, the leader's objective function is concave (see [31] for details).

2.2 *Sample average approximations of SMPECs problems*

Now suppose that $\xi(\omega)$ has a discrete distribution with a finite support, i.e. it can take a finite number $\xi_1, \ldots, \xi_K$ of possible values (called scenarios[1]) with respective probabilities $p_1, \ldots, p_K$. Then the expected value function $f(x)$, defined in (2.2), can be written as the finite summation $f(x) = \sum_{k=1}^{K} p_k Q(x, \xi_k)$. Also in that case mappings $y : \Omega \to \mathbb{R}^m$ can be

identified with vectors $(y_1, \ldots, y_K) \in \mathbb{R}^{mK}$, and hence formulation (2.3) implies that the SMPEC problem (1.1) (optimistic formulation) can be written as one large MPEC:

$$\begin{array}{ll} \underset{x, y_1, \ldots, y_K}{\text{Min}} & \sum_{k=1}^{K} p_k F(x, y_k, \xi_k) \\ \text{s.t.} & x \in \mathcal{X},\ 0 \in H(x, y_k, \xi_k) + \mathcal{N}_{\mathcal{C}}(y_k),\ k = 1, \ldots, K. \end{array} \tag{2.9}$$

The corresponding pessimistic formulation can be written as

$$\begin{array}{ll} \underset{x \in \mathcal{X}}{\text{Min}}\ \underset{y_1, \ldots, y_K}{\text{Max}} & \sum_{k=1}^{K} p_k F(x, y_k, \xi_k) \\ \text{s.t.} & 0 \in H(x, y_k, \xi_k) + \mathcal{N}_{\mathcal{C}}(y_k),\ k = 1, \ldots, K. \end{array} \tag{2.10}$$

As it was discussed earlier, two formulations (2.9) and (2.10) are equivalent if the constraint variational inequality has unique solution for all $x \in \mathcal{X}$ and a.e. ξ. Unless stated otherwise, in the subsequent analysis we refer to the optimistic formulation of the SMPEC model.

It is well-known that even a crude discretization of the random data results in an exponential growth of the number of scenarios which makes the deterministic problem (2.9) unmanageable even for a moderate number of random parameters. As it was mentioned in the previous section, an approach to a drastic reduction of the number of scenarios is to use the Monte Carlo sampling techniques. In that respect, the sample average approximation (SAA) method turned out to be surprisingly efficient. The idea of the SAA method is simple indeed. A random (or rather pseudo-random) sample $\xi^1, \ldots, \xi^N$ of N realizations of the random vector $\xi(\omega)$ is generated and the expected value $\mathbb{E}[Q(x, \xi(\omega))]$ is approximated by the corresponding sample average $(1/N)\sum_{j=1}^{N} Q(x, \xi^j)$, and hence the (true) SMPEC problem (2.2) is approximated by the following, so-called sample average approximating (SAA), problem:

$$\underset{x \in \mathcal{X}}{\text{Min}} \left\{ \hat{f}_N(x) := \frac{1}{N} \sum_{j=1}^{N} Q(x, \xi^j) \right\}. \tag{2.11}$$

This SAA problem can also be written in the following equivalent form

$$\begin{array}{ll} \underset{x, y_1, \ldots, y_N}{\text{Min}} & \frac{1}{N} \sum_{j=1}^{N} F(x, y_j, \xi^j) \\ \text{s.t.} & x \in \mathcal{X},\ 0 \in H(x, y_j, \xi^j) + \mathcal{N}_{\mathcal{C}}(y_j),\ j = 1, \ldots, N. \end{array} \tag{2.12}$$

Note that after the random sample is generated, the obtained SAA problem (2.12) can be viewed as SMPEC with N scenarios $\{\xi^1, \ldots, \xi^N\}$ and equal probabilities $1/N$. We discuss statistical properties of solutions of the SAA problem later.

Note also that (2.12) is essentially a deterministic MPEC and we may use a suitable MPEC algorithm to solve it. We will discuss this in detail in Section 6. In comparison with (2.12), (2.11) is an implicit formulation where variables y_i, $i = 1, \ldots, N$ are 'eliminated'. This type of formulation is helpful for the convergence analysis in Sections 4 and 5.

3. Properties of solutions of variational inequalities

In order to proceed with the analysis, we need to understand the behaviour of the set $\mathfrak{S}(x,\xi)$ of solutions of the variational inequality (1.2), and eventually of the optimal value function $Q(x,\xi)$ of the second-stage problem (2.1). First, we have to ensure that $\mathfrak{S}(x,\xi)$ is non-empty for all $x \in \mathcal{X}$ and a.e. $\xi \in \Xi$ (this corresponds to the so-called relatively complete recourse condition in two-stage stochastic programming). A simple condition ensuring that (1.2) has a solution is that the mapping $H(x,\cdot,\xi)$ is continuous and the set $\mathcal{C}$ is non-empty and compact (recall that $\mathcal{C}$ is assumed to be convex throughout this article). This condition already suffices for application to the Stackelberg–Nash–Cournot Equilibrium problem considered in Example 2.1. For a more extensive discussion of existence of solutions of variational inequalities we may refer, e.g. to [9 section 2.2]. Uniqueness of solutions can be ensured by strict monotonicity of mapping $H(x,\cdot,\xi)$. This holds in Example 2.1 if every function inside the parentheses in the right-hand side of (2.7) is *strictly* concave in $y_i \in [0, U_i]$.

In order to simplify notation, we write now $H(y,\pi)$, or $H_\pi(y)$, for the mapping $H(x,y,\xi)$ and view $\pi := (x,\xi)$ as a parameter vector of the corresponding variational inequality:

$$0 \in H_\pi(y) + \mathcal{N}_{\mathcal{C}}(y). \tag{3.1}$$

We refer to (3.1) as $VI(\mathcal{C}, H_\pi)$ and consider a solution $\bar{y}(\pi) \in \mathfrak{S}(\pi)$ of $VI(\mathcal{C}, H_\pi)$. There is an extensive literature where continuity and differentiability properties of $\bar{y}(\pi)$, viewed as a function of π, are discussed (see, e.g. [1,7] for a thorough discussion of that topic). It will be beyond the scope of this article to give a comprehensive review of that theory. In what follows, we briefly discuss some results which will be relevant for our analysis.

We assume throughout the remainder of this article that the mapping $H(\cdot,\cdot)$ is continuous. Also we deal with the case where the solution set $\mathfrak{S}(\pi) = \{\bar{y}(\pi)\}$, of variational inequality $VI(\mathcal{C}, H_\pi)$, is a singleton for all $\pi \in \mathcal{X} \times \Xi$. (Note that by assuming $\mathfrak{S}(\pi)$ is a singleton we postulate that $\mathfrak{S}(\pi)$ consists of exactly one element, and hence is non-empty.) If we assume, further, that $\bar{y}(\pi)$ is uniformly bounded for all π in a neighbourhood of a given point $\bar{\pi}$, then $\bar{y}(\cdot)$ is continuous at $\bar{\pi}$. This follows easily by compactness arguments from the fact that the multifunction (point-to-set mapping) $y \mapsto \mathcal{N}_{\mathcal{C}}(y)$ is closed. Of course, if the set $\mathcal{C}$ is bounded, then $\bar{y}(\pi)$ are uniformly bounded for all $\pi \in \mathcal{X} \times \Xi$.

Recall that, for a given π, mapping $H_\pi : \mathbb{R}^m \to \mathbb{R}^m$ is said to be *strongly monotone*, on $\mathcal{C}$, if there is a constant $c > 0$ such that

$$(H_\pi(y') - H_\pi(y))^T (y' - y) \geq c\|y' - y\|^2, \quad \forall\, y', y \in \mathcal{C}. \tag{3.2}$$

We say that $H_\pi(y)$ is *uniformly* strongly monotone if the constant $c > 0$ in (3.2) is independent of $\pi \in \mathcal{X} \times \Xi$. As far as Lipschitz continuity is concerned we have the following relatively simple result (see, e.g. [7, Theorem 2.3.3]).

PROPOSITION 3.1 *Suppose that $H_\pi(\cdot) = H(\cdot,\pi)$ is continuous for all π, $H(y,\cdot)$ is Lipschitz continuous with Lipschitz constant independent of y, and $H_\pi(\cdot)$ is uniformly strongly monotone on $\mathcal{C}$. Then $\mathfrak{S}(\pi) = \{\bar{y}(\pi)\}$ is a singleton, and the mapping $\bar{y}(\cdot)$ is Lipschitz continuous.*

It is more tricky to establish directional differentiability of $\bar{y}(\pi)$. We assume now that the set $\mathcal{C}$ is given in the form

$$\mathcal{C} := \{y \in \mathbb{R}^m : G(y) \in \mathcal{Q}\}, \tag{3.3}$$

where $G : \mathbb{R}^m \to \mathbb{R}^\ell$ is a continuously differentiable mapping and $\mathcal{Q} \subset \mathbb{R}^\ell$ is a polyhedral convex cone. In particular, if $\mathcal{Q} := \{0_q\} \times (-\mathbb{R}^p_+) \subset \mathbb{R}^\ell$ and $G(y) = (g_1(y), \ldots, g_\ell(y))$, where $\ell = q + p$, then the corresponding set $\mathcal{C}$ is defined by q equality and p inequality constraints, that is,

$$\mathcal{C} = \big\{y \in \mathbb{R}^m : g_i(y) = 0,\ i = 1, \ldots, q;\ g_i(y) \le 0,\ i = q+1, \ldots, \ell\big\}. \tag{3.4}$$

It is said that Robinson's constraint qualification, for the system $G(y) \in \mathcal{Q}$, holds at a point $y_0 \in \mathcal{C}$, if

$$[\nabla G(y_0)]\mathbb{R}^m + T_{\mathcal{Q}}(G(y_0)) = \mathbb{R}^\ell. \tag{3.5}$$

Here $\nabla G(y)$ denotes the Jacobian matrix of $G(\cdot)$ at y, and $T_{\mathcal{Q}}(z)$ denotes the tangent cone to $\mathcal{Q}$ at $z \in \mathcal{Q}$. Note that for $\mathcal{Q} := \{0_q\} \times (-\mathbb{R}^p_+)$, Robinson's constraint qualification coincides with the Mangasarian–Fromovitz Constraint Qualification (MFCQ).

Denote by $\Lambda(y, \pi)$ the set of all Lagrange multipliers λ satisfying condition:

$$V(y, \pi, \lambda) = 0, \quad \lambda \in \mathcal{N}_{\mathcal{Q}}(G(y)), \tag{3.6}$$

where

$$V(y, \pi, \lambda) := H(y, \pi) + [\nabla G(y)]^T \lambda = H(y, \pi) + \sum_{i=1}^{\ell} \lambda_i \nabla g_i(y). \tag{3.7}$$

We have that if $y \in \mathfrak{S}(\pi)$ and Robinson's constraint qualification holds at y, then $\Lambda(y, \pi)$ is non-empty and bounded.

Let us now fix a point π_0 and suppose that $\mathfrak{S}(\pi_0) = \{y_0\}$. By

$$\mathcal{C}_0 := \big\{d \in \mathbb{R}^m : H_{\pi_0}(y_0)^T d = 0,\ d \in T_{\mathcal{C}}(y_0)\big\} \tag{3.8}$$

we denote the so-called critical cone to $\mathcal{C}$ at $y_0 \in \mathcal{C}$, and for a given direction $d \in \mathbb{R}^m$ consider

$$\Lambda^*(d) := \arg \max_{\lambda \in \Lambda(y_0, \pi_0)} \sum_{i=1}^{\ell} \lambda_i d^T [\nabla^2 g_i(y_0)] d.$$

We can now formulate the following result about directional differentiability of $\bar{y}(\cdot)$ at π_0, [26].

Proposition 3.2 *Suppose that*: (i) $H(\cdot, \cdot)$ *is continuously differentiable and* $G(\cdot)$ *is twice continuously differentiable*, (ii) *the set* $\mathfrak{S}(\pi) = \{\bar{y}(\pi)\}$ *is a singleton and* $\bar{y}(\pi)$ *tends to* y_0 *as* $\pi \to \pi_0$, (iii) *Robinson's constraint qualification holds at the point* y_0, (iv) *for any* δ *the system*

$$0 \in \nabla_y H(y_0, \pi_0) d + \nabla_\pi H(y_0, \pi_0)\delta + \cup_{\lambda \in \Lambda^*(d)} \left\{ \sum_{i=1}^{\ell} \lambda_i [\nabla^2 g_i(y_0)] d \right\} + \mathcal{N}_{\mathcal{C}_0}(d) \tag{3.9}$$

has unique solution $\bar{d} = \bar{d}(\delta)$. *Then* $\bar{y}(\cdot)$ *is directionally differentiable at* π_0 *and* $\bar{y}'(\pi_0, \delta) = \bar{d}(\delta)$.

Let us consider some particular cases of the above result. If $\Lambda(y_0, \pi_0) = \{\bar{\lambda}\}$ is a singleton, then system (3.9) becomes

$$0 \in \nabla_y V(y_0, \pi_0, \bar{\lambda})d + \nabla_\pi V(y_0, \pi_0, \bar{\lambda})\delta + \mathcal{N}_{\mathcal{C}_0}(d), \tag{3.10}$$

where $V(y, \pi, \lambda)$ is defined in (3.7). If the mapping $G(y)$ is affine, then $\nabla^2 G(y) = 0$, and hence in that case systems (3.9) becomes

$$0 \in \nabla_y H(y_0, \pi_0)d + \nabla_\pi H(y_0, \pi_0)\delta + \mathcal{N}_{\mathcal{C}_0}(d). \tag{3.11}$$

Of course, if $\bar{y}(\cdot)$ is directionally differentiable at π_0 and $\bar{y}'(\pi_0, \delta)$ is linear in δ, then $\bar{y}(\cdot)$ is Gâteaux differentiable at π_0, and if, moreover, $\bar{y}(\cdot)$ is locally Lipschitz continuous, then Fréchet differentiability follows. We may refer to [7, Section 5.4] for a detailed discussion of differentiability properties of $\bar{y}(\cdot)$ when the set $\mathcal{C}$ is defined by a finite number of constraints in the form (3.4).

As it was already mentioned in Section 2, we have that if $\mathfrak{S}(x, \xi) = \{\bar{y}(x, \xi)\}$ is a singleton, then $Q(x, \xi) = F(x, \bar{y}(x, \xi), \xi)$, and hence continuity and differentiability properties of $\bar{y}(x, \xi)$ can be used to derive differentiability properties of $Q(x, \xi)$. This, in turn, can be translated into the corresponding properties of the expected value function $f(x)$. For a further discussion of this topic we may refer to [20, Chapter 2] and [27].

4. Convergence of SAA estimators

In this section, we present a convergence analysis for the SAA program (2.11). The reason that we deal with (2.12), rather than (2.12), is that we can use properties of the optimal value function $Q(x, \xi)$ in a direct way. We denote by $\hat{v}_N$ and $\hat{S}_N$ the optimal value and the set of optimal solutions, respectively, of problem (2.11). Note that $\hat{v}_N$ and $\hat{S}_N$ are functions of the generated sample and hence should be treated as random variables. In this section, we discuss convergence properties of $\hat{v}_N$ and $\hat{S}_N$ viewed as statistical estimators of their counterparts v^* and S^* of the true problem (2.2). We assume in this section that the random sample $\xi^1, \ldots, \xi^N$ is i.i.d.

4.1 *Consistency of SAA estimators*

Following the standard statistical terminology we say that $\hat{v}_N$ is a *consistent* estimator of v^* if $\hat{v}_N$ converges with probability one (w.p.1) to v^* as $N \to \infty$. We also discuss consistency of SAA estimators $\hat{x}_N \in \hat{S}_N$ of optimal solutions.

It is known that the sample average function $\hat{f}_N(x)$ converges w.p.1 to the expected value $f(x)$ *uniformly* in $x \in \mathcal{X}$, i.e.

$$\lim_{N\to\infty} \left\{ \sup_{x\in\mathcal{X}} \left| \hat{f}_N(x) - f(x) \right| \right\} = 0, \quad \text{w.p.1}, \tag{4.1}$$

if the following conditions hold (e.g. [19, Section 2.6]):

(A1) The set $\mathcal{X}$ is compact.

(A2) For a.e. $\xi \in \Xi$, the function $Q(\cdot, \xi)$ is continuous on $\mathcal{X}$.

(A3) $\{Q(x,\cdot)\}_{x\in\mathcal{X}}$, is dominated by an integrable function, i.e. there is a P-integrable function $\phi(\xi)$ such that $\sup_{x\in\mathcal{X}}|Q(x,\xi)|\leq\phi(\xi)$ for a.e. $\xi\in\Xi$.

The above assumption (A2) holds, e.g. if the solution $\bar{y}(\cdot,\xi)$ is unique and continuous for a.e. ξ. The assumption (A3) holds, e.g. if $|Q(x,\xi)|\leq c$ for all $(x,\xi)\in\mathcal{X}\times\Xi$ and some $c\in\mathbb{R}$. This, in turn, holds if $Q(\cdot,\cdot)$ is continuous and $\mathcal{X}$ and Ξ are compact. We have that, under the above conditions (A1)–(A3), $\hat{v}_N$ converges w.p.1 to v^* and for any (measurable) selection $\hat{x}_N\in\hat{S}_N$ the distance $\mathrm{dist}(\hat{x}_N,S^*)\to 0$ w.p.1 as $N\to\infty$. That is, under mild regularity conditions, we have that any (globally) optimal solution $\hat{x}_N$ of the SAA problem converges w.p.1 to the true optimal set S^*. This settles the question of consistency of *globally* optimal solutions of the SAA problem. However, in the present case the true and SAA problems typically are non-convex and can have many local optima and stationary points which are even not locally optimal. Moreover, the function $Q(\cdot,\xi)$ may not be everywhere differentiable. Under some additional conditions it is also possible to show convergence of locally optimal solutions. Consider the following conditions.

(A4) There exists a P-integrable function $\kappa(\xi)$ such that for an open set $\mathcal{V}\subset\mathbb{R}^n$ containing the set $\mathcal{X}$, and all x, $x'\in\mathcal{V}$ and $\xi\in\Xi$ it holds that

$$|Q(x',\xi)-Q(x,\xi)|\leq\kappa(\xi)\|x'-x\|. \tag{4.2}$$

(A5) For every (fixed) $x\in\mathcal{X}$, the function $Q(\cdot,\xi)$ is continuously differentiable at x w.p.1.

Condition (A4) means that $Q(\cdot,\xi)$ is Lipschitz continuous, with constant $\kappa(\xi)$, on an open set $\mathcal{V}$ containing $\mathcal{X}$. Recall that if the solution $\bar{y}(\cdot,\xi)$ is unique, then $Q(x,\xi)=F(x,\bar{y}(x,\xi),\xi)$, and hence Lipschitz continuity of $Q(\cdot,\xi)$ follows from Lipschitz continuity of $\bar{y}(\cdot,\xi)$. Note that if the set $\mathcal{X}$ is bounded, then condition (A4) implies that $\{Q(x,\cdot)\}_{x\in\mathcal{X}}$ is dominated by function $\phi(\xi):=Q(\bar{x},\xi)+D\kappa(\xi)$, where $\bar{x}$ is a point of $\mathcal{X}$ and D is the diameter of $\mathcal{X}$. Therefore, if $\mathcal{X}$ is bounded and $Q(\bar{x},\cdot)$ is P-integrable for some $\bar{x}\in\mathcal{X}$, then condition (A4) implies condition (A3).

Let us now discuss condition (A5). For $\xi\in\Xi$, consider the set $\Delta(\xi)$ of such $x\in\mathcal{V}$ that $Q(\cdot,\xi)$ is differentiable at x (the set $\mathcal{V}$ is specified in condition (A4)). Since $Q(\cdot,\xi)$ is Lipschitz continuous on $\mathcal{V}$, the set $\mathcal{V}\setminus\Delta(\xi)$ has Lebesgue measure zero, and hence $\Delta(\xi)$ is dense in $\mathcal{V}$. It is said that $Q(\cdot,\xi)$ is continuously differentiable at x if $x\in\Delta(\xi)$ and for any sequence $x_k\in\Delta(\xi)$ converging to x it follows that $\nabla_x Q(x_k,\xi)\to\nabla_x Q(x,\xi)$ as $k\to\infty$. It is also possible to formulate this in the following form. Consider Clarke's generalized gradient $\partial Q(x,\xi)$ of $Q(\cdot,\xi)$ at $x\in\mathcal{V}$, which is formed by the convex hull of all limits of the form $\lim_{k\to\infty}\nabla_x Q(x_k,\xi)$ with $\Delta(\xi)\ni x_k\to x$. Then $Q(\cdot,\xi)$ is continuously differentiable at x iff $\partial Q(x,\xi)$ is a singleton. Therefore, the above condition (A5) means that $\partial Q(x,\xi)$ is a singleton for a.e. $\xi\in\Xi$.

Recall that if the solution $\bar{y}(x,\xi)$ is unique, then $Q(x,\xi)=F(x,\bar{y}(x,\xi),\xi)$. Therefore, if for a given realization ξ of the random data vector, $\bar{y}(\cdot,\xi)$ is differentiable at x, then $Q(\cdot,\xi)$ is differentiable at x. We have that under various regularity conditions, $\bar{y}(\cdot,\xi)$ is continuously differentiable at a given point x (cf., [7, Theorem 5.4.6] and the discussion of Section 3). Let $\Xi(x)$ denote the set of $\xi\in\Xi$ such that $\bar{y}(\cdot,\xi)$ is continuously differentiable at x. Then condition (A5) holds if, for any $x\in\mathcal{X}$, the Lebesgue measure of $\Xi\setminus\Xi(x)$ is zero and the probability distribution of $\xi(\omega)$ has a density (i.e. is absolutely continuous with respect to the Lebesgue measure). For a more detailed discussion of this subject in the SMPCC setting, see, e.g. [31, Section 2].

For two sets $A, B \subset \mathbb{R}^n$ we denote by

$$\mathbb{D}(A, B) := \sup_{x \in A}[\text{dist}(x, B)]$$

the deviation of A from B. We have the following result [23].

PROPOSITION 4.1 *Suppose that the above conditions* (A3)–(A5) *hold. Then*: (a) *the expected value function* $f(x)$ *is continuously differentiable on* $\mathcal{X}$, (b) *for all* $x \in \mathcal{X}$,

$$\nabla f(x) = \mathbb{E}[\nabla_x Q(x, \xi(\omega))], \tag{4.3}$$

(c) $\partial \hat{f}_N(x)$ *converges to* $\nabla f(x)$ *w.p.1 uniformly on any compact subset* C *of* $\mathcal{X}$, *that is*,

$$\lim_{N \to \infty} \sup_{x \in C} \mathbb{D}(\partial \hat{f}_N(x), \nabla f(x)) = 0, \quad w.p.1, \tag{4.4}$$

where $\partial \hat{f}_N(x)$ *is Clarke's generalized gradient of* $\hat{f}_N$ *at* x *and*

$$\mathbb{D}(\partial \hat{f}_N(x), \nabla f(x)) := \sup_{z \in \partial \hat{f}_N(x)} \|z - \nabla f(x)\|$$

is the deviation of the set $\partial \hat{f}_N(x)$ *from the* (*singleton*) *set* $\{\nabla f(x)\}$.

Suppose that the above condition (A4) holds, and hence $f(x)$ and $\hat{f}_N(x)$ are Lipschitz continuous on $\mathcal{V}$. We say that a point $\bar{x} \in \mathcal{X}$ is a *stationary* point of the (true) problem (2.2) if

$$0 \in \partial f(\bar{x}) + \mathcal{N}_{\mathcal{X}}(\bar{x}). \tag{4.5}$$

Of course, if $f(x)$ is continuously differentiable at $\bar{x}$, i.e. $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$ is a singleton, then the generalized gradient $\partial f(\bar{x})$ in (4.5) is reduced to the usual gradient vector $\nabla f(\bar{x})$. In the similar way, a stationary point of the corresponding SAA problem (2.11) by replacing $f(x)$ in (4.5) with $\hat{f}_N(x)$ is defined. If the set $\mathcal{X}$ is convex, then $\mathcal{N}_{\mathcal{X}}(\bar{x})$ denotes the usual normal cone. For non-convex sets there are several possible concepts of normal cones. For example, one can define $\mathcal{N}_{\mathcal{X}}(\bar{x})$ as the polar of the contingent (Bouligand) cone to $\mathcal{X}$ at $\bar{x}$. For convex set $\mathcal{X}$ these two concepts of normal cones are equivalent. Also, suppose that the set $\mathcal{X}$ is defined by a finite number of constraints

$$\mathcal{X} := \{x : b_i(x) = 0,\ i = 1, \ldots, r;\ b_i(x) \leq 0,\ i = r+1, \ldots, s\}, \tag{4.6}$$

where $b_i(x)$, $i = 1, \ldots, s$, are continuously differentiable functions and let $\mathcal{I}(x)$ be the index set of active at x inequality constraints. Then we can define $\mathcal{N}_{\mathcal{X}}(\bar{x})$ as the cone generated by vectors $\nabla b_i(\bar{x})$, $i = 1, \ldots, r$, with unconstrained coefficients and vectors $\nabla b_i(\bar{x})$, $i \in (\bar{x})$, with non-negative coefficients. Under the Mangasarian–Fromovitz constraint qualification, the such defined normal cone coincides with the Bouligand normal cone, and (4.5) represents a necessary condition for $\bar{x}$ to be a locally optimal solution of the problem (2.2).

We assume that the considered normal cone is such that the point-to-set mapping $x \mapsto \mathcal{N}_{\mathcal{X}}(x)$ is *closed* on $\mathcal{X}$, i.e. if $\mathcal{X} \ni x_k \to x$, $y_k \in \mathcal{N}_{\mathcal{X}}(x_k)$ and $y_k \to y$, then $y \in \mathcal{N}_{\mathcal{X}}(x)$. The above concepts of normal cones satisfy this property. Denote by Σ^* and $\hat{\Sigma}_N$ the sets of stationary points of the true (2.2) and SAA (2.11) problems, respectively. By Proposition 4.1 we have the following result (cf., [25, Section 7]).

PROPOSITION 4.2 *Suppose that the assumptions* (A1) *and* (A3)–(A5) *hold and the point-to-set mapping* $x \mapsto \mathcal{N}_{\mathcal{X}}(x)$ *is closed on* $\mathcal{X}$. *Then* $\mathbb{D}(\hat{\Sigma}_N, \Sigma^*) \to 0$ *w.p.1 as* $N \to \infty$.

Let us emphasize that in the SMPEC setting the value function $Q(\cdot, \xi)$, and hence the sample average function $\hat{f}_N(\cdot)$, typically are not everywhere differentiable. Nevertheless, in case the random data vector $\xi(\omega)$ has a continuous distribution, the condition (A5) may hold as we discussed earlier. Together with condition (A4) this implies that the expected value function $f(x)$ is continuously differentiable. This is a well-known 'smoothing' property of the expectation operator. Of course, if $f(\cdot)$ is continuously differentiable at $\bar{x}$, then the stationarity condition (4.5) means that $0 \in \nabla f(\bar{x}) + \mathcal{N}_{\mathcal{X}}(\bar{x})$, which is quite standard. On the other hand, the stationarity concept (4.5) applied to the SAA problem could be quite loose. The result of Proposition 4.2 shows that under mild regularity conditions we may expect that stationary points of the SAA problem (where stationarity is understood in a broad sense based on Clarke's generalized gradients) converge w.p.1 to stationary points of the true problem with increase of the sample size N.

5. Exponential convergence

In the preceding section, we discussed convergence of the SAA estimators. However, those results have not addressed an important issue which is interesting from the conceptual and computational points of view. That is, how large the sample size should be to achieve a desired accuracy of SAA estimators? In this section, we address this issue. It should be emphasized that the derived estimates of the sample size are too conservative for practical applications. Nevertheless, these estimates give an important insight into theoretical complexity of the considered problems.

Let us start with the following general result about uniform convergence. The following analysis is self contained and is based on a relatively elementary application of the upper bound of Cramér's Large Deviation Theorem. Consider a function $h : \mathcal{X} \times \Xi \to \mathbb{R}$ and the corresponding expected value function $f(x) := \mathbb{E}[h(x, \xi(\omega))]$, where the expectation is taken with respect to the probability distribution P of the random vector $\xi(\omega)$. Recall that $\mathcal{X}$ is a closed subset of $\mathbb{R}^n$ and $\Xi \subset \mathbb{R}^d$ denotes the support of the probability distribution P. We assume that for every $x \in \mathcal{X}$ the expectation $f(x)$ is well defined, i.e. $h(x, \cdot)$ is measurable and P-integrable. In our applications we take, of course, $h(x, \xi) := Q(x, \xi)$. However, the following results may have an independent interest and it is not essential at the moment to assume that the considered function is given by the optimal value of the second-stage problem.

Let $\xi^1, \ldots, \xi^N$ be an i.i.d sample of the random vector $\xi(\omega)$, and consider the corresponding sample average function $\hat{f}_N(x) := (1/N) \sum_{j=1}^N h(x, \xi^j)$. We discuss now uniform exponential rates of convergence of $\hat{f}_N(x)$ to $f(x)$. We denote by

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

the moment generating function of the random variable $h(x, \xi(\omega)) - f(x)$. Let us make the following assumptions.

(C1) For every $x \in X$ the moment generating function $M_x(t)$ is finite valued for all t in a neighbourhood of zero.

(C2) There exists a (measurable) function $\kappa : \Xi \to \mathbb{R}_+$ and constant $\gamma > 0$ such that

$$|h(x', \xi) - h(x, \xi)| \leq \kappa(\xi)\|x' - x\|^{\gamma} \tag{5.1}$$

for all $\xi \in \Xi$ and all $x', x \in \mathcal{X}$.

(C3) The moment generating function $M_{\kappa}(t)$ of $\kappa(\xi(\omega))$ is finite valued for all t in a neighbourhood of zero.

Assumption (C1) (assumption (C3)) means that probability distribution of the random variable $h(x, \xi(\omega))$ (random variable $\kappa(\xi(\omega))$) dies exponentially fast in the tails. In particular, it holds if this random variable has a distribution supported on a bounded subset of $\mathbb{R}$. For $h(x, \xi) = Q(x, \xi)$ and $\gamma = 1$, assumption (C2) basically is the same as the assumption (A4) used in the previous section. In some interesting cases the power constant γ could be less than one, this is why we consider the more general setting here.

THEOREM 5.1 *Suppose that conditions* (C1)–(C3) *hold and the set* $\mathcal{X}$ *is compact. Then for any* $\varepsilon > 0$ *there exist positive constants* $C = C(\varepsilon)$ *and* $\beta = \beta(\varepsilon)$, *independent of* N, *such that*

$$\text{Prob}\left\{\sup_{x \in \mathcal{X}} \left|\hat{f}_N(x) - f(x)\right| \geq \varepsilon\right\} \leq C(\varepsilon)e^{-N\beta(\varepsilon)}. \tag{5.2}$$

Proof By Cramér's Large Deviation (LD) Theorem we have that for any $x \in \mathcal{X}$ and $\varepsilon > 0$ it holds that

$$\text{Prob}\{\hat{f}_N(x) - f(x) \geq \varepsilon\} \leq \exp\{-NI_x(\varepsilon)\}, \tag{5.3}$$

where

$$I_x(z) := \sup_{t \in \mathbb{R}} \left\{zt - \log M_x(t)\right\} \tag{5.4}$$

is the LD rate function of random variable $h(x, \xi(\omega)) - f(x)$. Similarly

$$\text{Prob}\{\hat{f}_N(x) - f(x) \leq -\varepsilon\} \leq \exp\{-NI_x(-\varepsilon)\},$$

and hence

$$\text{Prob}\{\left|\hat{f}_N(x) - f(x)\right| \geq \varepsilon\} \leq \exp\left\{-NI_x(\varepsilon)\right\} + \exp\left\{-NI_x(-\varepsilon)\right\}. \tag{5.5}$$

By assumption (C1) we have that both $I_x(\varepsilon)$ and $I_x(-\varepsilon)$ are positive for every $x \in \mathcal{X}$.

For a $\nu > 0$, let $\bar{x}_1, \ldots, \bar{x}_M \in \mathcal{X}$ be such that for every $x \in \mathcal{X}$ there exists $\bar{x}_i$, $i \in \{1, \ldots, M\}$, such that $\|x - \bar{x}_i\| \leq \nu$, i.e. $\{\bar{x}_1, \ldots, \bar{x}_M\}$ is a ν-net in $\mathcal{X}$. We can choose this net in such a way that $M \leq [O(1)D/\nu]^n$, where $D := \sup_{x', x \in \mathcal{X}} \|x' - x\|$ is the diameter of $\mathcal{X}$ and $O(1)$ is a generic constant. By (5.1) we have that

$$|f(x') - f(x)| \leq L\|x' - x\|^{\gamma}, \tag{5.6}$$

where $L := \mathbb{E}[\kappa(\xi(\omega))]$ is finite by assumption (C3). Moreover,

$$\left|\hat{f}_N(x') - \hat{f}_N(x)\right| \leq \hat{\kappa}_N\|x' - x\|^{\gamma}, \tag{5.7}$$

where $\hat{\kappa}_N := N^{-1}\sum_{j=1}^{N}\kappa(\xi^j)$. Again, because of condition (C3), by Cramér's LD Theorem we have that for any $L' > L$ there is a positive constant λ such that

$$\text{Prob}\{\hat{\kappa}_N \geq L'\} \leq \exp\{-N\lambda\}. \tag{5.8}$$

Consider $Z_i := \hat{f}_N(\bar{x}_i) - f(\bar{x}_i)$, $i = 1, \ldots, M$. We have that the event $\{\max_{1\leq i\leq M}|Z_i| \geq \varepsilon\}$ is equal to the union of the events $\{|Z_i| \geq \varepsilon\}$, $i = 1, \ldots, M$, and hence

$$\text{Prob}\left(\max_{1\leq i\leq M}|Z_i| \geq \varepsilon\right) \leq \sum_{i=1}^{M}\text{Prob}\big(|Z_i| \geq \varepsilon\big).$$

Together with (5.5) this implies that

$$\text{Prob}\left(\max_{1\leq i\leq M}|\hat{f}_N(\bar{x}_i) - f(\bar{x}_i)| \geq \varepsilon\right) \leq 2\sum_{i=1}^{M}\exp\big\{-N[I_{\bar{x}_i}(\varepsilon) \wedge I_{\bar{x}_i}(-\varepsilon)]\big\}.$$

For an $x \in \mathcal{X}$ let $i(x) \in \arg\min_{1\leq i\leq M}\|x - \bar{x}_i\|$. By construction of the ν-net we have that $\|x - \bar{x}_{i(x)}\| \leq \nu$ for every $x \in \mathcal{X}$. Then

$$\begin{aligned}|\hat{f}_N(x) - f(x)| &\leq |\hat{f}_N(x) - \hat{f}_N(\bar{x}_{i(x)})| + |\hat{f}_N(\bar{x}_{i(x)}) - f(\bar{x}_{i(x)})| + |f(\bar{x}_{i(x)}) - f(x)| \\ &\leq \hat{\kappa}_N\nu^\gamma + |\hat{f}_N(\bar{x}_{i(x)}) - f(\bar{x}_{i(x)})| + L\nu^\gamma.\end{aligned}$$

Now let us take a ν-net with such ν that $L\nu^\gamma = \varepsilon/4$, i.e. $\nu := [\varepsilon/(4L)]^{1/\gamma}$. Then

$$\text{Prob}\left\{\sup_{x\in\mathcal{X}}|\hat{f}_N(x) - f(x)| \geq \varepsilon\right\} \leq \text{Prob}\left\{\hat{\kappa}_N\nu^\gamma + \max_{1\leq i\leq M}|\hat{f}_N(\bar{x}_i) - f(\bar{x}_i)| \geq \frac{3\varepsilon}{4}\right\}.$$

Moreover, by (5.8) we have that

$$\text{Prob}\left\{\hat{\kappa}_N\nu^\gamma \geq \frac{\varepsilon}{2}\right\} \leq \exp\{-N\lambda\}$$

for some $\lambda > 0$, and hence

$$\begin{aligned}\text{Prob}\left\{\sup_{x\in\mathcal{X}}|\hat{f}_N(x) - f(x)| \geq \varepsilon\right\} &\leq \exp\{-N\lambda\} + \text{Prob}\left\{\max_{1\leq i\leq M}|\hat{f}_N(\bar{x}_i) - f(\bar{x}_i)| \geq \frac{\varepsilon}{4}\right\} \\ &\leq \exp\{-N\lambda\} + 2\sum_{i=1}^{M}\exp\left\{-N\left[I_{\bar{x}_i}\left(\frac{\varepsilon}{4}\right) \wedge I_{\bar{x}_i}\left(\frac{-\varepsilon}{4}\right)\right]\right\}. \qquad (5.9)\end{aligned}$$

Since the above choice of the ν-net does not depend on the sample (although it depends on ε), and both $I_{\bar{x}_i}(\varepsilon/4)$ and $I_{\bar{x}_i}(-\varepsilon/4)$ are positive, $i = 1, \ldots, M$, we obtain that (5.9) implies (5.2), and hence completes the proof. ■

It follows from (5.2) that

$$\limsup_{N\to\infty}\left[\frac{1}{N}\log\,\text{Prob}\left\{\sup_{x\in\mathcal{X}}|\hat{f}_N(x) - f(\mathcal{X})| \geq \varepsilon\right\}\right] \leq -\beta(\varepsilon). \tag{5.10}$$

Now let us strengthen condition (C1) to the following condition:

(C4) There exists constant $\sigma > 0$ such that for any $x \in \mathcal{X}$, the following inequality holds:

$$M_x(t) \leq \exp\{\sigma^2 t^2/2\}, \quad \forall\, t \in \mathbb{R}. \tag{5.11}$$

Note that if random variable $h(x, \xi(\omega)) - f(x)$ has normal distribution with variance σ^2, then its moment generating function is equal to the right-hand side of (5.11). In that case, the above inequality (5.11) holds as equality.

Suppose, further, for the sake of simplicity that $\gamma = 1$. It follows from condition (5.11) that $\log M_x(t) \leq \sigma^2 t^2/2$, and hence

$$I_x(z) \geq \frac{z^2}{2\sigma^2}, \quad \forall\, z \in \mathbb{R}. \tag{5.12}$$

Consequently, inequality (5.9) implies

$$\text{Prob}\left\{\sup_{x\in\mathcal{X}} |\hat{f}_N(x) - f(x)| \geq \varepsilon\right\} \leq \exp\{-N\lambda\} + 2M \exp\left\{-\frac{N\varepsilon^2}{32\sigma^2}\right\}, \tag{5.13}$$

where $M = [\text{O}(1)D/\nu)]^n = [\text{O}(1)DL/\varepsilon]^n$.

It follows from (5.13) that for $\alpha \in (0, 1)$ and the sample size

$$N \geq \frac{O(1)\sigma^2}{\varepsilon^2}\left[n \log\left(\frac{O(1)DL}{\varepsilon}\right) + \log\left(\frac{1}{\alpha}\right)\right], \tag{5.14}$$

we have that

$$\text{Prob}\left\{\sup_{x\in\mathcal{X}} |\hat{f}_N(x) - f(x)| \geq \varepsilon\right\} \leq \alpha. \tag{5.15}$$

It follows from the above estimates that if $\hat{x}_N$ is an $\varepsilon/2$-minimizer of $\hat{f}_N(x)$ over $\mathcal{X}$ and the sample size N satisfies (5.14), then $\hat{x}_N$ is an ε-minimizer of $f(x)$ over $\mathcal{X}$ with probability at least $1 - \alpha$. In other words, we have the following.

> By solving the SAA problem with accuracy $\varepsilon/2$ and the sample size satisfying (5.14), we are guaranteed that the obtained solution solves the true problem to accuracy ε with probability at least $1 - \alpha$.

For a further discussion of this topic we may refer to [28].

5.1 *Convergence of directional derivatives*

The above results can also be applied to establishing rates of convergence of directional derivatives of $\hat{f}_N(x)$ at a given point $\bar{x} \in \mathcal{X}$. We assume in the remainder of this section that condition (C2) holds with constant $\gamma = 1$, and that:

(C5) For every $\xi \in \Xi$, the function $h_\xi(\cdot) = h(\cdot, \xi)$ is directionally differentiable at $\bar{x}$.

Conditions (C2)–(C5) imply that $f(\cdot)$ is Lipschitz continuous in a neighbourhood of $\bar{x}$, $f(\cdot)$ is directionally differentiable at $\bar{x}$, its directional derivative $f'(\bar{x}, \cdot)$ is Lipschitz

continuous, and $f'(\bar{x}, \cdot) = \mathbb{E}[\eta(\cdot, \xi(\omega))]$, where $\eta(\cdot, \xi) := h'_\xi(\bar{x}, \cdot)$ (e.g. [20, Proposition 2, p. 66]). Here we also have that $\hat{f}'_N(x, \cdot) = \hat{\eta}_N(\cdot)$, where

$$\hat{\eta}_N(d) := \frac{1}{N}\sum_{j=1}^{N} \eta(d, \xi^j), \quad d \in \mathbb{R}^n, \tag{5.16}$$

and $\mathbb{E}[\hat{\eta}_N(d)] = f'(\bar{x}, d)$, for all $d \in \mathbb{R}^n$. Note that condition (C2) implies that $|\eta(d, \xi)| \le \kappa(\xi)\|d\|$ for any $d \in \mathbb{R}^n$ and $\xi \in \Xi$, and hence together with condition (C3) this implies that the moment generating function of $\eta(d, \xi(\omega))$ is finite valued in a neighbourhood of zero. Therefore, Theorem 5.1 implies the following result. (By $S^{n-1} := \{d \in \mathbb{R}^n : \|d\| = 1\}$ we denote the unit sphere taken with respect to a norm $\|\cdot\|$ on $\mathbb{R}^n$.)

THEOREM 5.2 *Suppose that conditions* (C2)–(C5) *hold, with constant* $\gamma = 1$ *in condition* (C2). *Then for any* $\varepsilon > 0$ *there exist positive constants* $C = C(\varepsilon)$ *and* $\beta = \beta(\varepsilon)$, *independent of* N, *such that*

$$\text{Prob}\left\{\sup_{d \in S^{n-1}} \left|\hat{f}'_N(\bar{x}, d) - f'(\bar{x}, d)\right| > \varepsilon\right\} \le C(\varepsilon)e^{-N\beta(\varepsilon)}. \tag{5.17}$$

The above result for directional derivatives can be translated into estimates of convergence of sample average subdifferentials. Recall that

$$\mathbb{H}(A, B) := \max\left\{\mathbb{D}(A, B), \mathbb{D}(B, A)\right\} = \max\left\{\sup_{x \in A} \text{dist}(x, B), \sup_{x \in B} \text{dist}(x, A)\right\}$$

is called the Hausdorff distance between sets $A, B \subset \mathbb{R}^n$. We have that if A and B are compact convex subsets of $\mathbb{R}^n$, then

$$\mathbb{H}(A, B) = \sup_{d \in S^{n-1}} \left|\sigma_A(d) - \sigma_B(d)\right|, \tag{5.18}$$

where $\sigma_A(d) := \sup_{z \in A} d^T z$ denotes the support function of the set A. Note also that if $\phi : \mathbb{R}^n \to \mathbb{R}$ is a real valued positively homogeneous convex function, then there exists a (uniquely defined) convex compact set $A \subset \mathbb{R}^n$ such that $\phi(\cdot) = \sigma_A(\cdot)$. If a function $f : \mathbb{R}^n \to \mathbb{R}$ is directionally differentiable at a point x and $f'(x, \cdot)$ is convex, we denote by $\partial f(x)$ the corresponding convex set such that $f'(x, \cdot) = \sigma_{\partial f}(x)(\cdot)$. In case the function f is convex, $\partial f(x)$ coincides with the subdifferential of f at x in the usual sense of convex analysis. If f is regular, at x, in the sense of Clarke [2, Definition 2.3.4], then $\partial f(x)$ coincides with the generalized gradient of Clarke.

The following result is a consequence of Theorem 5.2.

THEOREM 5.3 *Suppose that conditions* (C2)–(C5) *hold, with constant* $\gamma = 1$ *in condition* (C2), *and for every* $\xi \in \Xi$ *the directional derivative* $h'_\xi(\bar{x}, d)$ *is convex in* $d \in \mathbb{R}^n$. *Then* $f'(\bar{x}, \cdot)$ *is convex, and for any* $\varepsilon > 0$ *there exist positive constants* $C = C(\varepsilon)$ *and* $\beta = \beta(\varepsilon)$, *independent of* N, *such that*

$$\text{Prob}\left\{\mathbb{H}\left(\partial \hat{f}_N(\bar{x}), \partial f(\bar{x})\right) > \varepsilon\right\} \le C(\varepsilon)e^{-N\beta(\varepsilon)}. \tag{5.19}$$

Now suppose that $h(x, \xi) \equiv Q(x, \xi)$, the set $\mathcal{X}$ is convex and $\bar{x}$ is a sharp local minimizer of the true problem, i.e. there is a constant $\varepsilon > 0$ such that

$$f'(\bar{x}, d) \ge \varepsilon\|d\|, \quad \forall\, d \in T_{\mathcal{X}}(\bar{x}), \tag{5.20}$$

where $T_{\mathcal{X}}(\bar{x})$ denotes the tangent cone to $\mathcal{X}$ at $\bar{x} \in \mathcal{X}$ (note that here we do not assume $f'(\bar{x}, \cdot)$ is convex). Then, under conditions (C2)–(C5), it follows from (5.17) that

$$\text{Prob}\left\{\hat{f}'_N(\bar{x}, d) > 0, \ \forall d \in T_{\mathcal{X}}(\bar{x}) \setminus \{0\}\right\} \geq 1 - C(\varepsilon)e^{-N\beta(\varepsilon)}. \tag{5.21}$$

That is, the probability of the event that $\bar{x}$ is a sharp local minimizer of the corresponding SAA problem approaches one exponentially fast with increase of the sample size N. We can write this result as follows.

THEOREM 5.4 *Suppose that the set $\mathcal{X}$ is convex and, for $h(x, \xi) \equiv Q(x, \xi)$, conditions* (C2)–(C5) *hold, with $\gamma = 1$ in condition* (C2). *Let $\bar{x} \in \mathcal{X}$ be a sharp locally optimal solution of the true problem. Then with probability approaching one exponentially fast with increase of the sample size N, $\bar{x}$ is a sharp locally optimal solution of the corresponding SAA problem.*

The main additional assumption of the above Theorem 5.4 is the assumption of sharpness of locally optimal solution $\bar{x}$ of the true problem. As compared with general results of the first part of this section, this assumption holds only in rather specific situations (cf. [24]). However, the assumption may suit the context of SMPEC. This can be intuitively explained as follows. In SMPEC, particularly in SMPCC [30,31], the components of $\bar{y}(\cdot, \xi)$ are generally piecewise smooth. It is quite typical that smooth pieces are joined at a 'kink' which is reasonably sharp. This type of kink may become a sharp local minimizer or a sharp local maximizer of $Q(x, \xi)$ and eventually that of $f(x)$. In other words, we may expect that $f(x)$ contains many sharp local minimizers in SMPEC. In any case, convergence results of this section may have an independent interest.

6. Some algorithmic aspects of SAA

In the preceding section, we have established theoretical results about convergence of the SAA method for SMPEC. In this section, we discuss some algorithmic details for solving the SAA problem (2.12) and give simple numerical examples.

First, let us consider the following approximation scheme for solving the SAA problems. Suppose that we can construct a function $Q_\tau : \mathcal{X} \times \Xi \to \mathbb{R}$, depending on the parameter $\tau > 0$, such that $Q_\tau(x, \xi)$ converges to $Q(x, \xi)$ as $\tau \downarrow 0$ uniformly in $(x, \xi) \in \mathcal{X} \times \Xi$. Then for a generated sample ξ^j, $j = 1, \ldots, N$, and a sequence $\tau_N \downarrow 0$, we approximate the corresponding SAA problem by the problem

$$\min_{x \in \mathcal{X}} \left\{\tilde{f}_{\tau_N}(x) := \frac{1}{N}\sum_{j=1}^{N} Q_{\tau_N}(x, \xi^j)\right\}. \tag{6.1}$$

Because of the uniform convergence of $Q_{\tau_N}(x, \xi)$ to $Q(x, \xi)$ we have that for any $\varepsilon > 0$ the inequality

$$\sup_{x \in \mathcal{X}} \left|\tilde{f}_{\tau_N}(x) - \hat{f}_N(x)\right| < \varepsilon$$

holds for all N large enough (independently of the sample). Together with (4.1) this implies

$$\lim_{N \to \infty} \left\{\sup_{x \in \mathcal{X}} \left|\tilde{f}_{\tau_N}(x) - f(x)\right|\right\} = 0, \quad \text{w.p.1.} \tag{6.2}$$

Therefore, under assumptions (A1)–(A3), we obtain consistency of the optimal value and (global) optimal solutions of problem (6.1). If, further, for every $\xi \in \Xi$ and $\tau > 0$ the function $Q_\tau(\cdot, \xi)$ is continuously differentiable and $\mathbb{D}(\{\nabla \tilde{f}_{\tau_N}(x)\}, \partial \hat{f}_N(x))$ converges uniformly to zero (independently of the sample), then under assumptions (A1) and (A3)–(A5), consistency of the stationary points of problem (6.1) follows.

In order to motivate such approximations let us consider the following construction. Assume that the sets $\mathcal{C}$ and $\mathcal{X}$ have a specific structure, that is, $\mathcal{C}$ can be expressed as in (3.3) with $\mathcal{Q} := -\mathbb{R}^\ell_+$ and $\mathcal{X}$ can be expressed as in (4.6). Consequently, we can rewrite (2.12) as

$$\begin{array}{ll} \underset{\substack{x, y_1, \ldots, y_N \\ \lambda_1, \ldots, \lambda_N}}{\text{Min}} & \frac{1}{N} \sum_{j=1}^{N} F(x, y_j, \xi^j) \\ \text{s.t.} & b_i(x) = 0, \ i = 1, \ldots, r, \\ & b_i(x) \leq 0, \ i = r+1, \ldots, s, \\ & H(x, y_j, \xi^j) + [\nabla G(y_j)]^T \lambda_j = 0, \ j = 1, \ldots, N, \\ & 0 \leq \lambda_j \perp -G(y_j) \geq 0, j = 1, \ldots, N. \end{array} \tag{6.3}$$

This is a deterministic mathematical program with complementarity constraints. From the algorithmic point of view, complementarity constraints may cause some computational difficulties. There are several ways to deal with these difficulties. For instance, we may use Scholtes' regularization method [22] to relax the j-th complementarity constraint by replacing it with a perturbed system of inequalities

$$\lambda_j \geq 0, \quad G(y_j) \leq 0, -g_i(y_j)\lambda_{ji} \leq \tau, \ i = 1, \ldots, \ell, \tag{6.4}$$

and solve the following regularized SAA scheme

$$\begin{array}{ll} \underset{\substack{x, y_1, \ldots, y_N \\ \lambda_1, \ldots, \lambda_N}}{\text{Min}} & \frac{1}{N} \sum_{j=1}^{N} F(x, y_j, \xi^j) \\ \text{s.t.} & b_i(x) = 0, \ i = 1, \ldots, r, \\ & b_i(x) \leq 0, \ i = r+1, \ldots, s, \\ & H(x, y_j, \xi^j) + [\nabla G(y_j)]^T \lambda_j = 0, \ j = 1, \ldots, N, \\ & \lambda_j \geq 0, \ -G(y_j) \geq 0, j = 1, \ldots, N, \\ & -g_i(y_j)\lambda_{ji} \leq \tau_N, \ i = 1, \ldots, \ell, j = 1, \ldots, N. \end{array} \tag{6.5}$$

Here $g_i(\cdot)$ denotes the i-th component of $G(\cdot)$, λ_{ji} denotes the i-th component of the vector of Lagrange multipliers λ_j, and $\tau_N \downarrow 0$. Observe that (6.5) can be viewed as a two-stage program: in the first-stage, minimization is taken with respect to x and in the second-stage, minimization is taken with respect to y_j for $\xi = \xi^j$, $j = 1, \ldots, N$. Let $Q_{\tau_N}(x, \xi^j)$ be the optimal value the second-stage problem, $j = 1, \ldots, N$. Then by the above discussion we have that, under mild regularity conditions, consistency of the globally (locally) optimal solutions of the regularized SAA problem (6.5) holds.

Another way to deal with the complementarity constraints is to use an NCP function such as

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

or

$$\psi(a, b, c) := a + b - \sqrt{a^2 + b^2 + c^2},$$

so that the complementarity constraint can be replaced by

$$\psi\big(g_i(y_j), -\lambda_{ji}, \tau\big) = 0,\ i = 1, \ldots, \ell. \tag{6.6}$$

In the MPEC literature, this is known as a smoothing NCP method (see, e.g. a discussion in [6,9,10]). Consequently, we may solve the following smoothed SAA scheme

$$\begin{array}{ll} \underset{\substack{x, y_1, \ldots, y_N \\ \lambda_1, \ldots, \lambda_N}}{\text{Min}} & \dfrac{1}{N}\displaystyle\sum_{j=1}^{N} F(x, y_j, \xi^j) \\ \text{s.t.} & b_i(x) = 0,\ i = 1, \ldots, r, \\ & b_i(x) \le 0,\ i = r+1, \ldots, s, \\ & H(x, y_j, \xi^j) + [\nabla G(y_j)]^T \lambda_j = 0,\ j = 1, \ldots, N, \\ & \psi\big(g_i(y_j), -\lambda_{ji}, \tau\big) = 0,\ i = 1, \ldots, \ell, j = 1, \ldots, N. \end{array} \tag{6.7}$$

Convergence results analogous to the regularized SAA scheme are expected for (6.7). We omit the details.

6.1 *Some preliminary tests*

In this section, we report results of a numerical implementation of the SAA method based on smoothing schemes discussed above. Our focus is on the regularization scheme (6.5) and smoothed NCP scheme (6.7). Since there are very few SMPEC source problems in the literature, we have to make them by ourselves. It is not difficult to construct an SMPEC problem, but it is more tricky to have a problem with solutions which can be calculated in a closed form. We would like to have a problem with an explicitly known solution in order to verify numerical results.

We carried out tests in Matlab 6.5 installed in a PC with Windows XP operating system. This is because MATLAB 6.5 has a built-in optimization solver function *fmincon* for general non-linear smooth optimization problems and our programs (6.5) and (6.7) are non-linear smooth optimization problems for $\tau > 0$. In order to speed up rates of convergence, we used the Latin Hypercube Sampling (LHS) to generate samples. The parameter τ is set $1/N$ for both programs where N is the sample size.

To see the performance of the SAA method, we have also carried out tests for the Deterministic Discretization Approximation (DDA) method [31] for (6.3) which incorporates a regularization scheme (6.4) and a smoothing NCP scheme (6.6) and compare the results with those from SAA schemes.

Note that *fmincon* is an iterative algorithm. In the implementation, the starting point is set to $0.5e$, where e denotes a vector with unit components. The dimension of e depends on the sample (grid) size and problem size.

Throughout the tests, we recorded CPU times, number of function evaluations (fun eva.#), number of iterations (iter.#) and approximate optimal solution (opt.val.) and optimal value at the solution (opt. val.) and these quantities are displayed in the tables of test results. The sample size N in SAA related methods and the size of grid points (also denoted by N) in DDA related methods are both set to 100.

The NCP smoothing we used throughout the tests is based on

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

which is a smoothing of the min-function $\min(a,b)$.

Example 6.1 Consider

$$\begin{array}{ll}\text{Min} & f(x) := \mathbb{E}\left[(x_1-1)^2+(x_2-1)^2+\bar{y}_1(x,\xi)^2+\bar{y}_2(x,\xi)^2\right]\\ \text{s.t.} & x_1\in[0,2], x_2\in[0,2],\end{array}$$

where $x=(x_1,x_2)$, $\xi=(\xi_1,\xi_2)$, and $\bar{y}_1(x,\xi)$ and $\bar{y}_2(x,\xi)$ are solutions of the complementarity constraints:

$$0\le(y_1,y_2)\perp(y_1-x_1+\xi_1,y_2-x_2+\xi_2)\ge 0.$$

We assume that ξ_1,ξ_2 are independent random variables, both having uniform distribution on the interval $[0,1]$.

It is not difficult to verify that the complementarity problem here has a unique solution $\bar{y}(x,\xi)=(\bar{y}_1(x,\xi),\bar{y}_2(x,\xi))$, where

$$\bar{y}_i(x,\xi)=\begin{cases}x_i-\xi_i, & \text{if } \xi_i\le x_i,\\ 0, & \text{otherwise,}\end{cases}$$

for $i=1,2$. Consequently, we have

$$\begin{aligned}f(x_1,x_2) &= (x_1-1)^2+(x_2-1)^2+\int_0^{x_1}(x_1-t_1)^2\mathrm{d}t_1+\int_0^{x_2}(x_2-t_2)^2\mathrm{d}t_2\\ &= (x_1-1)^2+(x_2-1)^2+\frac{x_1^3}{3}+\frac{x_2^3}{3}.\end{aligned}$$

Therefore, the exact solution is $(-1+\sqrt{3},-1+\sqrt{3})=(0.7320,0.7320)$ and the optimal value $v^*=0.4051$. The test results are displayed in Table 1.

Example 6.2 Consider

$$\begin{array}{ll}\text{Min} & f(x) := \mathbb{E}\left[(x_1-1)^2+(x_2-1)^2+(x_3-1)^2+\bar{y}_1(x,\xi)^2+\bar{y}_2(x,\xi)^2+\bar{y}_3(x,\xi)^2\right]\\ \text{s.t.} & x_1\in[0,2], x_2\in[0,2], x_3\in[0,2],\end{array}$$

where $\bar{y}_1(x,\xi)$, $\bar{y}_2(x,\xi)$ and $\bar{y}_3(x,\xi)$ are solutions of:

$$0\le(y_1,y_2,y_3)\perp(y_1-x_1+\xi_1,y_2-x_2+\xi_2,y_3-x_3+\xi_3)\ge 0,$$

Table 1. Numerical results for Example 6.1.

Methods	N	CPU	fun eva.#	iter.#	opt.val.	opt. solutions
SAA-REG	100	43.4060	3875	18	0.4053	(0.7315, 0.7315)
SAA-NCP	100	14.9840	1836	8	0.4054	(0.7321, 0.7317)
DDA-REG	100	61.0150	3059	14	0.4351	(0.7257, 0.7257)
DDA-NCP	100	7.2180	1019	4	0.4351	(0.7255, 0.7255)

Table 2. Numerical results for Example 6.2.

Methods	N	CPU	funt eva.#	iter.#	opt.val.	opt. solutions
SAA-REG	100	145.5310	6004	20	0.6079	(0.7318, 0.7320, 0.7320)
SAA-NCP	100	27.3750	2745	8	0.6089	(0.7316, 0.7321, 0.7318)

and ξ_1, ξ_2, ξ_3 are independent random variables each having uniform distribution on $[0, 1]$. This example is modified from Example 6.1 to make a case that the random vector ξ has three components. By a similar analysis, we obtain

$$f(x) = (x_1 - 1)^2 + (x_2 - 1)^2 + (x_3 - 1)^2 + \frac{x_1^3}{3} + \frac{x_2^3}{3} + \frac{x_3^3}{3}$$

and deduce that the global optimal solution is $(-1 + \sqrt{3}, -1 + \sqrt{3}, -1 + \sqrt{3})$ and optimal value $v^* = 0.6077$. The test results are displayed in Table 2. No tests have been carried out for DDA-Reg and DDA-NCP as the set of grid points is too large.

Example 6.3 Consider

$$\begin{array}{ll} \text{Min} & f(x) := \mathbb{E}\left[(x_1 - 1)^2 + (x_2 - 1)^2 + \bar{y}_1(x, \xi)^2 + \bar{y}_2(x, \xi)^2\right] \\ \text{s.t.} & x_1 \in [0, 2], x_2 \in [0, 2], \end{array}$$

where $\bar{y}_1(x, \xi_1, \xi_2)$ and $\bar{y}_2(x, \xi_1, \xi_2)$ are solutions of:

$$0 \le (y_1, y_2) \perp (y_1 - x_1 + \xi_1 - \xi_2, y_2 - x_2 + \xi_2) \ge 0,$$

and ξ_1, ξ_2 are independent random variables, both having uniform distribution on $[0, 1]$. This example is also modified from Example 6.1 by changing the first complementarity constraint so that it depends on both ξ_1 and ξ_2.

It is not difficult to verify that the complementarity problem has unique solution $\bar{y}(x, \xi) = (\bar{y}_1(x, \xi), \bar{y}_2(x, \xi))$, where

$$\bar{y}_1(x, \xi) = \begin{cases} x_1 - \xi_1 + \xi_2, & \text{if } x_1 - \xi_1 + \xi_2 \ge 0, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\bar{y}_2(x, \xi) = \begin{cases} x_2 - \xi_2, & \text{if } \xi_2 \le x_2, \\ 0, & \text{otherwise.} \end{cases}$$

Consequently, we can work out the explicit expression of the objective function

$$f(x) = \begin{cases} -\frac{1}{6}x_1^4 + \frac{1}{12}(1 + x_1)^4 + (x_1 - 1)^2 + (x_2 - 1)^2 + \frac{1}{3}x_2^3, & \text{if } x_1 \in [0, 1], x_2 \in [0, 2], \\ -\frac{1}{6}x_1^4 + \frac{1}{12}[(1 + x_1)^4) + (x_1 - 1)^4] + (x_1 - 1)^2 + (x_2 - 1)^2 + \frac{1}{3}x_2^3, & \text{if } x_1 \in [1, 2], x_2 \in [0, 2]. \end{cases}$$

Table 3. Numerical results for Example 6.3.

Methods	N	CPU	fun eva.#	iter.#	opt.val.	opt. solutions
SAA-REG	100	56.8120	4079	19	0.8661	(0.4882, 0.7312)
SAA-NCP	100	19.4840	3062	14	0.8558	(0.4915, 0.7322)
DDA-REG	100	71.2500	3674	17	0.9117	(0.4838, 0.7252)
DDA-NCP	100	21.2030	2652	12	0.9117	(0.4838, 0.7254)

Through a simple analysis, we can obtain an approximate optimal solution $(0.48887, -1+\sqrt{3})$. Note that the first coordinate is not exact as we obtained it by solving $\partial f(x)/\partial x_1 = 0$ in Maple. The numerical test results are displayed in Table 3.

Our preliminary results show that SAA related methods perform better than DDA related methods. From the smoothing perspective, we find that NCP related methods perform better than regularization related methods. This is perhaps to do with the fact that our complementarity constraints are all linear and strongly monotone with respect to y. Of course, to draw a more definitive conclusion, more tests are required. This could be a subject of further investigation.

Acknowledgements

We would like to thank Fanwen Meng for his help with the numerical tests, and an anonymous referee for constructive comments which helped to improve the presentation. Research of A. Shapiro was partly supported by the NSF awards DMS-0510324 and DMI-0619977. The work of H. Xu is supported by the United Kingdom Engineering and Physical Sciences Research Council grant GR/S90850/01.

Note

1. We use subscript ξ_k to denote a scenario, while superscript ξ^j stands for a generated sample point.

References

[1] J.F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer, New York, 2000.
[2] F.H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.
[3] S. Christiansen, M. Patriksson, and L. Wynter, *Stochastic bilevel programming in structral optimization*, Struct. Optim. 21 (2001), pp. 361–371.
[4] J. Dupacova and R.J.-B. Wets, *Asymtotic behaviour of statistical estimators and of optimal solutions of stochastic optimization problems*, Ann. Stat. 16 (1988), pp. 1517–1549.
[5] A. Evgrafov and M. Patriksson, *On the existence of solutions to stochastic mathematical programs with equilibrium constraints*, Journal of Optimization Theory and Applications 121 (2004), pp. 67–76.
[6] A. Fischer, *A special Newton-type optimization method*, Optimization 24 (1992), pp. 269–284.
[7] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer, New York, 2003.
[8] G. Gürkan, A.Y. Özge, and S.M. Robinson, *Sample-path solution of stochastic variational inequalities*, Math. Prog., Ser. A 84 (1999), pp. 313–333.

[9] H. Jiang and D. Ralph, *Smooth SQP methods for mathematical programs with non-linear complementarity constraints*, SIAM J. Optim. 10 (2000), pp. 779–808.
[10] C. Kanzow, *Some non-interior continuation methods for linear complementarity problems*, SIAM J. Matrix Anal. Appl. 17 (1996), pp. 851–868.
[11] A.J. King and R.J.-B. Wets, *Epi-consistency of convex stochastic programs*, Statistics 34 (1991), pp. 83–92.
[12] A.J. Kleywegt, A. Shapiro, and T. Homem-de-Mello, *The sample average approximation method for stochastic discrete optimization*, SIAM J. Optim. 12 (2001), pp. 479–502.
[13] G.-H. Lin, X. Chen, and M. Fukushima, *Smoothing implicit programming approaches for stochastic mathematical programs with linear complementarity constraints*, to appear in Math. Program. Available at http://www.amp.i.kyoto-u.ac.jp/tecrep/index-e.html
[14] G.-H. Lin and M. Fukushima, *A Class of stochastic mathematical programs with complementarity constraints: reformulations and algorithms*, J. Indus. Manage. Optim. 1 (2005), pp. 99–122.
[15] J. Linderoth, A. Shapiro, and S. Wright, *The empirical behaviour of sampling methods for stochastic programming*, Ann. Oper. Res. 142 (2006), pp. 215–241.
[16] W.K. Mak, D.P. Morton, and R.K. Wood, *Monte Carlo bounding techniques for determining solution quality in stochastic programs*, Oper. Res. Lett. 24 (1999), pp. 47–56.
[17] M. Patriksson and L. Wynter, *Stochastic mathematical programs with equilibrium constraints*, Oper. Res. Lett. 25 (2005), pp. 159–167.
[18] S.M. Robinson, *Analysis of sample-path optimization*, Math. Oper. Res. 21 (1996), pp. 513–528.
[19] R.Y. Rubinstein and A. Shapiro, *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Method*, John Wiley and Sons, New York, 1993.
[20] A. Ruszczyński and A. Shapiro, eds. *Stochastic Programming*, Handbook in OR & MS, Vol. 10, North-Holland Publishing Company, Amsterdam, 2003.
[21] T. Santoso, et al., *A stochastic programming approach for supply chain network design under uncertainty*, Eur. J. Oper. Res. 167 (2005), pp. 96–115.
[22] S. Scholtes, *Convergence properties of a regularization scheme for mathematical programs with complementarity constraints*, SIAM J. Optim. 11 (2001), pp. 918–936.
[23] A. Shapiro, *Asymptotic properties of statistical estimators in stochastic programming*, Ann. Stat. 17 (1989), pp. 841–858.
[24] A. Shapiro and T. Homem-de-Mello, *On rate of convergence of Monte Carlo approximations of stochastic programs*, SIAM J. Optim. 11 (2000), pp. 70–86.
[25] A. Shapiro, *Monte Carlo sampling methods*, in *Stochastic Programming*, A. Rusczyński and A. Shapiro, eds., Handbooks in OR & MS, Vol. 10, North-Holland Publishing Company, Amsterdam, 2003.
[26] A. Shapiro, *Sensitivity analysis of parameterized variational inequalities*, Math. Oper. Res. 30 (2005), pp. 76–91.
[27] A. Shapiro, *Stochastic mathematical programs with equilibrium constraints*, J. Optim. Theory Appl. 128 (2006), pp. 223–243.
[28] A. Shapiro and A. Nemirovski, *On complexity of stochastic programming problems*, in *Continuous Optimization: Current Trends and Applications*, V. Jeyakumar and A.M. Rubinov, eds., Springer, New York, 2005, pp. 111–144.
[29] B. Verweij, et al., *The sample average approximation method applied to stochastic routing problems: a computational study*, Comput. Optim. Appl. 24 (2003), pp. 289–333.
[30] H. Xu, *An MPCC approach for stochastic Stackelberg–Nash–Cournot equilibrium*, Optimization 54 (2005), pp. 27–57.
[31] H. Xu, *An implicit programming approach for a class of stochastic mathematical programs with equilibrium constraints*, SIAM J. Optim. 16 (2006), pp. 670–696.