Journal of Computational and Applied Mathematics 235 (2011) 3870–3882

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Stochastic mathematical programs with hybrid equilibrium constraints[☆]

Yong-Chao Liu, Jin Zhang, Gui-Hua Lin [*]

School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China

ARTICLE INFO

MSC:
90C30
90C33

Keywords:
Stochastic mathematical program with hybrid equilibrium constraints
Sampling average approximation
NCP function
Convergence

ABSTRACT

This paper considers a stochastic mathematical program with hybrid equilibrium constraints (SMPHEC), which includes either "here-and-now" or "wait-and-see" type complementarity constraints. An example is given to describe the necessity to study SMPHEC. In order to solve the problem, the sampling average approximation techniques are employed to approximate the expectations and smoothing and penalty techniques are used to deal with the complementarity constraints. Limiting behaviors of the proposed approach are discussed. Preliminary numerical experiments show that the proposed approach is applicable.

1. Introduction

Mathematical program with equilibrium constraints (MPEC) is an optimization problem whose constraints include complementarity or variational inequality system. MPEC plays a very important role in many fields such as engineering design, economic equilibrium, transportation science and game theory. See [1–3] for more details about the MPEC theory, algorithms, and applications.

Since some elements may involve uncertain data in many practical problems, the stochastic MPEC (SMPEC) has drawn much attention in the recent literature; see the survey paper [4] and the references therein. There have been proposed two kinds of SMPECs in the literature: one is called a "here-and-now" model, in which both the upper-level decision and the lower-level decision are required to be made before a random event is observed. The other is called a "lower-level wait-and-see" model, in which the upper-level decision is made at once and the lower-level decision may be made after a random event is observed. See [5–16] for details about the recent developments in the here-and-now model and the lower-level wait-and-see model, respectively.

In this paper, we consider the following more general problem:

$$\begin{aligned} \min_{x,y,z(\xi)} \quad & \mathbb{E}[f(x,y,z(\xi),\xi)] \\ \text{s.t.} \quad & x \in X, \\ & 0 \le y \perp \mathbb{E}[G(x,y,z(\xi),\xi)] \ge 0, \\ & 0 \le z(\xi) \perp H(x,y,z(\xi),\xi) \ge 0, \quad \text{a.e. } \xi \in \Xi, \end{aligned} \tag{1.1}$$

where X is a nonempty closed subset of $\mathbb{R}^n$, 'a.e.' is the abbreviation for "almost every", $\xi : \Omega \to \mathbb{R}^d$ denotes a vector of random variables defined on the underlying probability space $(\Omega, \mathcal{F}, P)$ with support set $\Xi \subset \mathbb{R}^d$, $\mathbb{E}$ denotes the expectation operator, the functions $f : \mathbb{R}^{n+m+s} \times \Xi \to \Re$, $G : \mathbb{R}^{n+m+s} \times \Xi \to \mathbb{R}^m$ and $H : \mathbb{R}^{n+m+s} \times \Xi \to \mathbb{R}^s$ are all continuously differentiable in $(x, y, z(\xi))$ for almost every $\xi \in \Xi$ and locally Lipschitz continuous in ξ, and the symbol $\perp$

☆ This work was supported in part by NSFC Grant #10771025 and SRFDP Grant #20070141063.

* Corresponding author.
E-mail addresses: googleliu0717@yahoo.com.cn (Y.-C. Liu), footprint1986@hotmail.com (J. Zhang), lin_g_h@yahoo.com.cn (G.-H. Lin).

0377-0427/$ – see front matter Crown Copyright © 2011 Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2011.01.033

stands for orthogonality of the two vectors on both sides. Hereafter $z(\xi)$ means that z depends on ξ rather than a function of ξ.

As usual, we may regard x as upper-level decision variables and $(y, z(\xi))$ as lower-level decision variables. We call problem (1.1) a stochastic mathematical program with hybrid equilibrium constraints (SMPHEC), which means that the upper-level variables x and the lower-level variables y are made before the random event is known but the other lower-level variables will be made after the random event is known.

The rest of the paper is organized as follows. In Section 2, we give an example called picnic-vendor decision problem with multiple choices, which can be formulated as the model (1.1). In Section 3, we recall some definitions and preliminary results. In Section 4, we apply the smoothing and penalty techniques to present a sampling average approximation method for solving the SMPHEC (1.1) and, in Section 5, we study the limiting behaviors of the optimal solutions and stationary points of the approximation problems. We report some preliminary numerical results on some simple examples in Section 6.

2. Picnic-vendor decision problem with multiple choices

In this section, we give an example to illustrate the model (1.1). Different from the ones discussed in [5,7], the vendors have multiple choices in the example.

Consider a food company that wholesales picnic lunches to m vendors who sell lunches at different spots on every Sunday. The company and the vendors have the following contract:

- The vendors may order lunches either on Saturday at the price $x \in [a, b]$ or in the morning of Sunday at the price $k_0 x$, where both a and b are positive constants and $k_0 > 1$, and each vendor must buy no less than the amount $c > 0$.
- Even if there are any unsold lunches, the vendors cannot return them to the company but they can dispose of the unsold lunches with no cost.

Suppose that the ith vendor sells lunches to hikers at the price $k_i x$ with $k_i > k_0$. In general, the demands of lunches depend on the price and the weather on that day. Since the weather is uncertain, we treat it as a random variable ξ with a support set Ξ. Moreover, we denote by y_i and $u_i(\xi)$ the amounts of the ith vendor ordered on Saturday and Sunday, respectively. The company's objective is to maximize its total earnings and so its model can be formulated as

$$\begin{aligned} &\max \sum_{i=1}^{m}\Big(xy_i + k_0 x\mathbb{E}[u_i(\xi)]\Big) \\ &\text{s.t.} \quad a \le x \le b. \end{aligned} \tag{2.1}$$

Denote by $d_i(x, \xi)$ the demand at the ith spot. The model for the ith vendor is

$$\begin{aligned} &\max\ \mathbb{E}[k_i x \min\{d_i(x,\xi), y_i + u_i(\xi)\} - xy_i - k_0 x u_i(\xi)] \\ &\text{s.t.} \quad y_i \ge 0, \\ &\qquad u_i(\xi) \ge 0, \quad y_i + u_i(\xi) \ge c,\ \xi \in \Xi, \end{aligned}$$

which is equivalent to

$$\begin{aligned} &\min\ \mathbb{E}[-k_i x t_i(\xi) + xy_i + k_0 x u_i(\xi)] \\ &\text{s.t.} \quad y_i \ge 0, \qquad u_i(\xi) \ge 0, \quad y_i + u_i(\xi) \ge c, \\ &\qquad d_i(x,\xi) - t_i(\xi) \ge 0, \qquad y_i + u_i(\xi) - t_i(\xi) \ge 0, \quad \xi \in \Xi. \end{aligned} \tag{2.2}$$

We suppose for simplicity that $\Xi = \{\xi^1, \xi^2, \ldots, \xi^L\}$ and, for each ℓ, the probability p_ℓ of ξ^ℓ is positive. Note that, for any fixed x, problem (2.2) is a linear programming problem. As a result, (2.2) is equivalent to its Karush–Kuhn–Tucker (KKT) conditions, that is, there exist Lagrange multipliers $\alpha_i \in \mathbb{R}$, $\beta^i \in \mathbb{R}^L$, $\gamma^i \in \mathbb{R}^L$, $\delta^i \in \mathbb{R}^L$, and $\eta^i \in \mathbb{R}^L$ satisfying

$$\begin{pmatrix} x \\ k_0 xp \\ -k_i xp \end{pmatrix} - \begin{pmatrix} \alpha_i \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ \beta^i \\ 0 \end{pmatrix} - \begin{pmatrix} \sum_{l=1}^{L}\gamma_l^i \\ \gamma^i \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 0 \\ -\delta^i \end{pmatrix} - \begin{pmatrix} \sum_{l=1}^{L}\eta_l^i \\ \eta^i \\ -\eta^i \end{pmatrix} = 0, \tag{2.3}$$

where $p = (p_1, p_2, \ldots, p_L)^T$, and

$$0 \le \alpha_i \perp y_i \ge 0, \tag{2.4}$$

$$0 \le \beta_\ell^i \perp u_i(\xi^\ell) \ge 0, \tag{2.5}$$

$$0 \le \gamma_\ell^i \perp y_i + u_i(\xi^\ell) - c \ge 0, \tag{2.6}$$

$$0 \le \delta_\ell^i \perp d_i(x, \xi^\ell) - t_i(\xi^\ell) \ge 0, \tag{2.7}$$

$$0 \le \eta_\ell^i \perp y_i + u_i(\xi^\ell) - t_i(\xi^\ell) \ge 0 \tag{2.8}$$

for each ℓ. It follows from (2.3) that

$$\alpha_i = x - \sum_{l=1}^{L} \gamma_l^i - \sum_{l=1}^{L} \eta_l^i,$$
$$\beta^i = k_0 xp - \gamma^i - \eta^i,$$
$$\delta^i = (k_i - k_0)xp + \beta^i + \gamma^i.$$

It is not difficult to see that $\delta_\ell^i > 0$ for each ℓ. This together with (2.7) yields that $t_i(\xi^\ell) = d_i(x, \xi^\ell)$ for each ℓ. Thus, the KKT conditions (2.3)–(2.8) are rewritten as

$$\left.\begin{array}{l} 0 \le y_i \perp x - \sum_{l=1}^{L} \gamma_l^i - \sum_{l=1}^{L} \eta_l^i \ge 0, \\ 0 \le u_i(\xi^\ell) \perp p_\ell k_0 x - \gamma_\ell^i - \eta_\ell^i \ge 0, \\ 0 \le \gamma_\ell^i \perp y_i + u_i(\xi^\ell) - c \ge 0, \\ 0 \le \eta_\ell^i \perp y_i + u_i(\xi^\ell) - d_i(x, \xi^\ell) \ge 0, \end{array}\right\} \quad \ell = 1, 2, \ldots, L.$$

In consequence, the model (2.1) for the company can be reformulated as

$$\begin{array}{ll} \max & \sum_{\ell=1}^{L} p_\ell \sum_{i=1}^{m} \Big(xy_i + k_0 x u_i(\xi^\ell) \Big) \\ \text{s.t.} & a \le x \le b, \\ & 0 \le y_i \perp \sum_{l=1}^{L} p_l \Big(x - \gamma^i(\xi^l) - \eta^i(\xi^l) \Big) \ge 0, \\ & 0 \le u_i(\xi^\ell) \perp k_0 x - \gamma^i(\xi^\ell) - \eta^i(\xi^\ell) \ge 0, \\ & 0 \le \gamma^i(\xi^\ell) \perp y_i + u_i(\xi^\ell) - c \ge 0, \\ & 0 \le \eta^i(\xi^\ell) \perp y_i + u_i(\xi^\ell) - d_i(x, \xi^\ell) \ge 0, \\ & i = 1, 2, \ldots, m, \ \ell = 1, 2, \ldots, L, \end{array}$$

which is obviously a special case of problem (1.1).

3. Preliminaries

Throughout this paper, we use $\|\cdot\|$ to denote the Euclidean norm and $B(x, \gamma)$ to denote the closed ball with center x and radius γ and, for simplicity, we use $\mathcal{B}$ to denote the closed unit ball. We next recall some definitions and results that will be used later on.

Let X be a closed subset of $\mathbb{R}^n$. A set-valued mapping $\mathcal{G} : X \to 2^{\mathbb{R}^m}$ is said to be *upper semi-continuous* at point $x \in X$ if for any neighborhood U of $\mathcal{G}(x)$, there exists $\eta > 0$ such that $\mathcal{G}(x') \subseteq U$ holds for any $x' \in B(x, \eta) \cap X$. Moreover, $\mathcal{G}$ is said to be *locally bounded* at x if there exists a neighborhood U of x such that $\bigcup_{x' \in U} \mathcal{G}(x')$ is bounded.

Consider now a random set-valued mapping $\mathcal{G}(\cdot, \xi(\cdot)) : X \times \Omega \to 2^{\mathbb{R}^n}$ (we are slightly abusing the notation $\mathcal{G}$). Let $\mathfrak{B}$ denote the space of nonempty, closed subsets of $\mathbb{R}^n$. $\mathcal{G}(x, \xi(\cdot))^{-1}$ is said to be $\mathcal{F}$-measurable if $\mathcal{G}(x, \xi(\cdot))^{-1}B$ is $\mathcal{F}$-measurable for every $B \in \mathfrak{B}$ [17]. In addition, $a(x, \xi(\omega)) \in \mathcal{G}(x, \xi(\omega))$ is said to be a *measurable selection* of the random set $\mathcal{G}(x, \xi(\omega))$, if $a(x, \xi(\omega))$ is measurable. The *expectation of* $\mathcal{G}(x, \xi(\omega))$, denoted by $\mathbb{E}[\mathcal{G}(x, \xi(\omega))]$, is defined as the collection of $\mathbb{E}[a(x, \xi(\omega))]$, where $a(x, \xi(\omega))$ is an integrable measurable selection. The expected value is also known as Aumann's integral [18].

Let $\mathcal{D}$ be a nonempty closed subset of $\mathbb{R}^n$. Given $z \in \mathcal{D}$, the *Clarke normal cone* is defined as the polar cone of the Clarke tangent cone $\mathcal{T}_D(z)$, that is,

$$\mathcal{N}_{\mathcal{D}}(z) = \left\{ \zeta \in \mathbb{R}^n : \zeta^T \eta \le 0, \forall \eta \in \mathcal{T}_{\mathcal{D}}(z) \right\},$$

where $T_{\mathcal{D}}(z) = \liminf_{t \to 0, \mathcal{D} \ni z' \to z} \frac{1}{t}(\mathcal{D} - z')$.

Let $\Gamma : \mathbb{R}^n \to \mathbb{R}^m$ be a locally Lipschitz continuous function. The *Clarke generalized Jacobian* of Γ at $x \in \mathbb{R}^n$ is defined as

$$\partial \Gamma(x) := \text{conv} \left\{ \lim_{y \in D, y \to x} \nabla \Gamma(y) \right\}, \tag{3.1}$$

where D denotes the set of points at which Γ is Fréchet differentiable, $\nabla \Gamma(y)$ denotes the usual Jacobian of Γ, and "conv" denotes the convex hull of a set. Note that the Clarke Jacobian $\partial \Gamma(\cdot)$ is upper semi-continuous [19].

Lemma 3.1 ([20]). *Let $\Gamma(x, \xi)$ be locally Lipschitz continuous in both x and ξ. Then the Clarke generalized gradient $\partial_x \Gamma(x, \xi)$ is measurable.*

Let $\Gamma : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^s$. We say $\Gamma(x, y)$ to be *uniformly strongly monotone* with respect to x if there exists a constant $\alpha > 0$ such that

$$(\Gamma(x', y) - \Gamma(x'', y))^T(x' - x'') \geq \alpha\|x' - x''\|^2, \quad x', x'' \in \mathbb{R}^n,\ y \in \mathbb{R}^m.$$

We say a sequence $\{\psi^N\}$ of extended real-valued functions defined on $\mathbb{R}^n$ to *converge continuously* to ψ, denoted by $\psi^N \xrightarrow{C} \psi$ for short, if $\psi^N(x^N) \to \psi(x)$ holds for any given $x \in \mathbb{R}^n$ and any sequence $\{x^N\}$ converging to x. Note that, if the limiting function $\psi(\cdot)$ is continuous, the continuous convergence coincides with uniform convergence on compact set.

A function $\phi : \mathbb{R}^2 \to \mathbb{R}$ is said to be an *NCP function* if

$$\phi(a, b) = 0 \Longleftrightarrow a \geq 0, \quad b \geq 0,\ ab = 0$$

holds for any $a, b \in \mathbb{R}$. See [21] for more details about NCP functions. In this paper, we consider the well-known Fischer–Burmeister (FB) function defined by

$$\phi_{\text{FB}}(a, b) := a + b - \sqrt{a^2 + b^2}.$$

Since the nonsmoothness of ϕ_{FB} may incur troubles in calculation, we will use the *smoothed FB function*

$$\phi(a, b; \mu) := a + b - \sqrt{a^2 + b^2 + \mu^2},$$

where μ is a positive scalar, to approximate $\phi(a, b; 0)$. It is obvious that $\phi(a, b, \mu)$ is globally Lipschitz continuous and continuously differentiable everywhere except $(0, 0, 0)$. Moreover, we have the following result.

Lemma 3.2 ([10]). *Let $\mu \geq 0$. Then, for any real numbers a_i and b_i, $i = 1, 2$, we have*

$$|\phi(a_1, b_1; \mu) - \phi(a_2, b_2; \mu)| \leq 2(|a_1 - a_2| + |b_1 - b_2|),$$
$$|\phi(a_1, b_1; \mu) - \phi(a_2, b_2; 0)| \leq 2(|a_1 - a_2| + |b_1 - b_2|) + \mu.$$

4. Approximation method for SMPHEC

Problem (1.1) has three main difficulties. Firstly, the variable $z(\cdot)$ depends on ξ, which means that there are infinitely many complementarity constraints in general. Secondly, it may be numerically too expensive to calculate the expected value. Thirdly, problem (1.1) fails to satisfy a standard constraint qualification at any feasible point [22]. In what follows, we consider a special case where the complementarity constraint $0 \leq z(\xi) \perp H(x, y, z(\xi), \xi) \geq 0$ has a unique solution for each ξ and (x, y). In order to solve problem (1.1), we will employ the well-known sampling average approximation (SAA) method to approximate the expectation and use the smoothed NCP function to approach the complementarities, that is, we will use a sequence of standard nonlinear programs to approximate the true problem (1.1).

In what follows, we denote by

$$\Phi(x, y, z, \xi, \mu) := \begin{pmatrix} \phi\big(z_1, H_1(x, y, z, \xi); \mu\big) \\ \vdots \\ \phi\big(z_s, H_s(x, y, z, \xi); \mu\big) \end{pmatrix}.$$

Then we have the following result.

Lemma 4.1 ([13]). *Suppose that $H(x, y, z, \xi)$ is uniformly strongly monotone in z and uniformly locally Lipschitz continuous in (x, y, ξ). Then*

(i) *$\partial_z\Phi(x, y, z, \xi, \mu)$ is uniformly nonsingular and there exists a constant $\mu_0 > 0$ such that the system of equation*

$$\Phi(x, y, z, \xi, \mu) = 0$$

defines a unique implicit function $z(x, y, \xi; \mu)$ satisfying

$$\Phi(x, y, z(x, y, \xi; \mu), \xi, \mu) = 0$$

for any $x \in X$, $y \in \mathbb{R}^m$, $\xi \in \Xi$, and $\mu \in [-\mu_0, \mu_0]$;

(ii) *$z(x, y, \xi; \mu)$ is locally Lipschitz continuous with respect to (x, y, ξ, μ) on $X \times \mathbb{R}^m \times \Xi \times [-\mu_0, \mu_0]$;*

(iii) *$z(x, y, \xi; \mu)$ is continuously differentiable with respect to (x, y, μ) on $X \times \mathbb{R}^m \times ([-\mu_0, \mu_0] \setminus \{0\})$;*

(iv) *$z(x, y, \xi; \mu)$ is uniformly calm in μ at 0, that is, there exists a constant $C > 0$ such that*

$$\|z(x, y, \xi; \mu) - z(x, y, \xi; 0)\| \leq C|\mu|, \quad \mu \in [-\mu_0, \mu_0];$$

(v) *the Clarke generalized Jacobian $\partial_{(x,y)}z(x, y, \xi; 0)$ can be estimated as*

$$\partial_{(x,y)}z(x, y, \xi; 0) \subseteq -[\partial_z\Phi(x, y, z(x, y, \xi; 0), \xi, 0)]^{-1}\partial_{(x,y)}\Phi(x, y, z(x, y, \xi; 0), \xi, 0).$$

Note that the continuity of $z(x, y, \xi; \mu)$ implies its measurability. Using the above implicit function theorem and the FB function ϕ_{FB}, problem (1.1) can be rewritten as

$$\begin{aligned}&\min_{x,y}\ \mathbb{E}[f(x, y, z(x, y, \xi; 0), \xi)] \\ &\text{s.t.}\quad x \in X, \quad \Psi(x, y) = 0,\end{aligned} \tag{4.1}$$

where $z(x, y, \xi; 0)$ solves $\Phi(x, y, z, \xi, 0) = 0$ and

$$\Psi(x, y) := \begin{pmatrix} \phi_{\mathrm{FB}}\big(y_1, \mathbb{E}[G_1(x, y, z(x, y, \xi; 0), \xi)]\big) \\ \vdots \\ \phi_{\mathrm{FB}}\big(y_m, \mathbb{E}[G_m(x, y, z(x, y, \xi; 0), \xi)]\big) \end{pmatrix}.$$

We next employ the SAA method to approximate the expectations. Indeed the SAA method has been wildly used in dealing with SMPEC; see [6,11,16,13,23]. In general, for an integrable function $\psi : \Xi \to \mathbb{R}$, the SAA method estimates $\mathbb{E}[\psi(\xi)]$ by taking independently and identically distributed random samples $\{\xi^1, \ldots, \xi^N\}$ from Ξ and calculating $\frac{1}{N}\sum_{\ell=1}^{N} \psi(\xi^\ell)$. The strong law of large numbers guarantees that this procedure converges with probability one (abbreviated by "w.p.1"), i.e.,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{\ell=1}^{N} \psi(\xi^\ell) = \mathbb{E}[\psi(\xi)] := \int_{\Xi} \psi(\xi) \mathrm{d}\zeta(\xi), \tag{4.2}$$

where $\zeta(\xi)$ denotes the distribution function of ξ and N tends to $+\infty$.

As mentioned before, we use the smoothed function $\phi(\cdot, \cdot; \mu)$ to approximate $\phi(\cdot, \cdot; 0)$. Thus, by taking independently and identically distributed random samples $\{\xi^1, \ldots, \xi^N\}$ from Ξ and a sequence of smooth parameters $\{\mu_N\}$ satisfying $\mu_N \downarrow 0$ as $N \to \infty$ and employing a penalty technique, (4.1) can be approximated by the problem

$$\begin{aligned}&\min_{x,y}\ \tfrac{1}{N}\sum_{\ell=1}^{N} f(x, y, z(x, y, \xi^\ell; \mu_N), \xi^\ell) + \rho_N \|\Psi^N(x, y)\|^2 \\ &\text{s.t.}\quad x \in X,\end{aligned} \tag{4.3}$$

where $\rho_N > 0$ is a penalty parameter, $z(x, y, \xi^\ell; \mu_N)$ solves $\Phi(x, y, z, \xi^\ell, \mu_N) = 0$ for each ℓ, and

$$\Psi^N(x, y) := \begin{pmatrix} \phi\left(y_1, \dfrac{1}{N}\displaystyle\sum_{\ell=1}^{N} G_1(x, y, z(x, y, \xi^\ell; \mu_N), \xi^\ell); \mu_N\right) \\ \vdots \\ \phi\left(y_m, \dfrac{1}{N}\displaystyle\sum_{\ell=1}^{N} G_m(x, y, z(x, y, \xi^\ell; \mu_N), \xi^\ell); \mu_N\right) \end{pmatrix}.$$

Difference from (1.1), for any fixed N, (4.3) is a standard nonlinear problem and all the functions involved are continuously differentiable. There are lots of mature algorithms having been proposed for this kind of problems. This means that we can use various mature algorithms to solve the approximation problem (4.3). The key point is whether the solutions or stationary points obtained by solving (4.3) converge to the counterparts of problem (1.1) as N tends to infinity. In the rest of this section, we study the convergence properties of (4.3). For the simplicity of notation, we take $z(x, y, \xi; \mu)$ to be the solution of $\Phi(x, y, z, \xi, \mu) = 0$ in mind and let $\mathcal{F}$ denote the feasible region of (4.1).

5. Convergence analysis

Suppose that we obtain a sequence of optimal solutions or stationary points by solving (4.3). We study the tendency of them as N to infinity. The following lemmas will be used.

Lemma 5.1 *([24, Proposition 7 of Chapter 6]). Let $\Theta : \mathbb{R}^n \times \Xi \to \mathbb{R}^s$ be an integrable function and D be a nonempty compact subset of $\mathbb{R}^n$. Suppose that* (i) *the function $\Theta(\cdot, \xi)$ is continuous on D for almost every $\xi \in \Xi$;* (ii) *$\Theta(x, \cdot)$ is dominated by an integrable function;* (iii) *the samples are independently and identically distributed. Then the expected value function $\mathbb{E}[\Theta(x, \xi)]$ is finite valued and continuous on D and, with probability one, $\Theta^N(x) := \frac{1}{N}\sum_{\ell=1}^{N} \Theta(x, \xi^\ell)$ converges to $\mathbb{E}[\Theta(x, \xi)]$ uniformly on D.*

Lemma 5.2 *([24, Proposition 2 of Chapter 2]). Suppose that* (i) *$\Theta(x, \cdot)$ is measurable for all x in a neighborhood of x_0 and $\Theta(x_0, \cdot)$ is integrable;* (ii) *there exists an integrable function $\kappa(\xi)$ such that $\|\Theta(x', \xi) - \Theta(x'', \xi)\| \le \kappa(\xi)\|x' - x''\|$ holds for any x' and x'' close to x_0 and almost every $\xi \in \Xi$. If $\Theta(z, \xi)$ continuously differentiable for almost every $\xi \in \Xi$, then $\mathbb{E}[\Theta(z, \xi)]$ is continuously differentiable and $\nabla\mathbb{E}[\Theta(z, \xi)] = \mathbb{E}[\nabla_z\Theta(z, \xi)]$.*

5.1. Convergence of global optimal solutions

In order to show convergence of the global optimal solutions of the approximation problems, we choose the parameters ρ_N and μ_N to satisfy the following conditions:

$$\lim_{N\to\infty} \rho_N = +\infty, \qquad \limsup_{N\to\infty} \rho_N \mu_N < +\infty, \tag{5.1}$$

$$\lim_{N\to\infty} \sqrt{\rho_N}\left(\frac{1}{N}\sum_{\ell=1}^{N} G_i(x, y, z(x, y, \xi^\ell; 0), \xi^\ell) - \mathbb{E}[G_i(x, y, z(x, y, \xi; 0), \xi)]\right) = 0 \quad \text{w.p.1} \tag{5.2}$$

for $i = 1, 2, \ldots, m$. Note that the convergence in (4.2) is of order $O(N^{-1/2})$ in probability [25]. This means that $\left\{\sqrt{N}\left(\frac{1}{N}\sum_{\ell=1}^{N}\psi(\xi^\ell) - \mathbb{E}[\psi(\xi)]\right)\right\}$ is convergent in probability as $N \to +\infty$. Therefore, we may choose a suitable sequence $\{\rho_N\}$ such that (5.1) and (5.2) hold at least in probability. See [10] for more details about these assumptions.

Theorem 5.1. *Suppose that $\nabla_z G_i(x, y, z, \xi)$, $1 \le i \le m$, and $f(x, y, z, \xi)$ are uniformly dominated by the integrable function $\delta(\xi)$ on $X \times \mathbb{R}^m \times \mathbb{R}^s$. Let (x^N, y^N) be a global optimal solution of problem (4.3) for each N and (x^*, y^*) be an accumulation point of $\{(x^N, y^N)\}$. Then (x^*, y^*) is a global optimal solution of problem (4.1) with probability one.*

Proof. Taking a subsequence if necessary, we assume for the simplicity of notation that (x^N, y^N) tends to (x^*, y^*). Since (x^N, y^N) is an optimal solution of problem (4.3), we have

$$\begin{aligned}&\frac{1}{N}\sum_{\ell=1}^{N} f(x^N, y^N, z(x^N, y^N, \xi^\ell; \mu_N), \xi^\ell) + \rho_N \left\|\Psi^N(x^N, y^N)\right\|^2 \\ &\quad \le \frac{1}{N}\sum_{\ell=1}^{N} f(\bar{x}, \bar{y}, z(\bar{x}, \bar{y}, \xi^\ell; \mu_N), \xi^\ell) + \rho_N \left\|\Psi^N(\bar{x}, \bar{y})\right\|^2\end{aligned} \tag{5.3}$$

for any $(\bar{x}, \bar{y}) \in \mathcal{F}$ and hence

$$\begin{aligned}&\left\|\Psi^N(x^N, y^N)\right\|^2 - \left\|\Psi^N(\bar{x}, \bar{y})\right\|^2 \\ &\quad \le \rho_N^{-1}\left(\frac{1}{N}\sum_{\ell=1}^{N} f(\bar{x}, \bar{y}, z(\bar{x}, \bar{y}, \xi^\ell; \mu_N), \xi^\ell) - \frac{1}{N}\sum_{\ell=1}^{N} f(x^N, y^N, z(x^N, y^N, \xi^\ell; \mu_N), \xi^\ell)\right).\end{aligned} \tag{5.4}$$

Noting that (x^N, y^N, μ_N) tends to $(x^*, y^*, 0)$, we can choose a compact set U containing the whole sequence $\{(x^N, y^N, \mu_N)\}$ for all N sufficiently large. By the assumptions and Lemma 5.1, the sequences

$$\left\{\frac{1}{N}\sum_{\ell=1}^{N} f(x, y, z(x, y, \xi^\ell; \mu), \xi^\ell)\right\}, \qquad \left\{\frac{1}{N}\sum_{\ell=1}^{N} G(x, y, z(x, y, \xi^\ell; \mu), \xi^\ell)\right\}$$

converge to $\mathbb{E}[f(x, y, z(x, y, \xi^\ell; \mu), \xi^\ell)]$ and $\mathbb{E}[G(x, y, z(x, y, \xi^\ell; \mu), \xi^\ell)]$ uniformly on U with probability one respectively. Then, with probability one,

$$\frac{1}{N}\sum_{\ell=1}^{N} f(x^N, y^N, z(x^N, y^N, \xi^\ell; \mu_N), \xi^\ell) \xrightarrow{N\to\infty} \mathbb{E}[f(x^*, y^*, z(x^*, y^*, \xi; 0), \xi)], \tag{5.5}$$

$$\frac{1}{N}\sum_{\ell=1}^{N} G(x^N, y^N, z(x^N, y^N, \xi^\ell; \mu_N), \xi^\ell) \xrightarrow{N\to\infty} \mathbb{E}[G(x^*, y^*, z(x^*, y^*, \xi; 0), \xi)]. \tag{5.6}$$

Note that $\Psi(\bar{x}, \bar{y}) = 0$ by $(\bar{x}, \bar{y}) \in \mathcal{F}$. Taking a limit on both sides of (5.4), we have w.p.1

$$\begin{aligned}\left\|\Psi(x^*, y^*)\right\|^2 &= \left\|\Psi(x^*, y^*)\right\|^2 - \|\Psi(\bar{x}, \bar{y})\|^2 \\ &\le \lim_{N\to\infty} \rho_N^{-1}\left(\mathbb{E}[f(x^*, y^*, z(x^*, y^*, \xi; 0), \xi) - f(\bar{x}, \bar{y}, z(\bar{x}, \bar{y}, \xi; 0), \xi)]\right) \\ &= 0\end{aligned}$$

with probability one. As a result, (x^*, y^*) is feasible to (4.1).

Let $(\bar{x}, \bar{y})$ be an arbitrary point of $\mathcal{F}$. Then, with probability one,

$$
\begin{aligned}
&\rho_N \left\|\Psi^N(\bar{x}, \bar{y})\right\|^2 \\
&= \rho_N \sum_{i=1}^{m} \left[\phi\left(\bar{y}_i, \frac{1}{N}\sum_{\ell=1}^{N} G_i(\bar{x}, \bar{y}, z(\bar{x}, \bar{y}, \xi^\ell; \mu_N), \xi^\ell); \mu_N\right) - \phi_{\rm FB}(\bar{y}_i, \mathbb{E}[G_i(\bar{x}, \bar{y}, z(\bar{x}, \bar{y}, \xi; 0), \xi)])\right]^2 \\
&\le \rho_N \sum_{i=1}^{m} \left(2\left|\frac{1}{N}\sum_{\ell=1}^{N} G_i(\bar{x}, \bar{y}, z(\bar{x}, \bar{y}, \xi^\ell; \mu_N), \xi^\ell) - \mathbb{E}[G_i(\bar{x}, \bar{y}, z(\bar{x}, \bar{y}, \xi; 0), \xi)]\right| + \mu_N\right)^2 \\
&\le \rho_N \sum_{i=1}^{m} \Bigg(2\left|\frac{1}{N}\sum_{\ell=1}^{N} \left(G_i(\bar{x}, \bar{y}, z(\bar{x}, \bar{y}, \xi^\ell; \mu_N), \xi^\ell) - G_i(\bar{x}, \bar{y}, z(\bar{x}, \bar{y}, \xi^\ell; 0), \xi^\ell)\right)\right| \\
&\quad + 2\left|\frac{1}{N}\sum_{\ell=1}^{N} G_i(\bar{x}, \bar{y}, z(\bar{x}, \bar{y}, \xi^\ell; 0), \xi^\ell) - \mathbb{E}[G_i(\bar{x}, \bar{y}, z(\bar{x}, \bar{y}, \xi; 0), \xi)]\right| + \mu_N\Bigg)^2 \\
&\le \sum_{i=1}^{m} \Bigg[2\sqrt{\rho_N}\left|\frac{1}{N}\sum_{\ell=1}^{N} \delta(\xi^\ell)\cdot(z(\bar{x}, \bar{y}, \xi^\ell; \mu_N) - z(\bar{x}, \bar{y}, \xi^\ell; 0))\right| \\
&\quad + 2\sqrt{\rho_N}\left|\frac{1}{N}\sum_{\ell=1}^{N} G_i(\bar{x}, \bar{y}, z(\bar{x}, \bar{y}, \xi^\ell; 0), \xi^\ell) - \mathbb{E}[G_i(\bar{x}, \bar{y}, z(\bar{x}, \bar{y}, \xi; 0), \xi)]\right| + \sqrt{\rho_N}\mu_N\Bigg]^2 \\
&\le \sum_{i=1}^{m} \Bigg[2\sqrt{\rho_N}\left|\frac{1}{N}\sum_{\ell=1}^{N} \delta(\xi^\ell)\cdot \hat{C}\cdot\mu_N\right| \\
&\quad + 2\sqrt{\rho_N}\left|\frac{1}{N}\sum_{\ell=1}^{N} G_i(\bar{x}, \bar{y}, z(\bar{x}, \bar{y}, \xi^\ell; 0), \xi^\ell) - \mathbb{E}[G_i(\bar{x}, \bar{y}, z(\bar{x}, \bar{y}, \xi; 0), \xi)]\right| + \sqrt{\rho_N}\mu_N\Bigg]^2 \\
&\xrightarrow{N\to\infty} 0. \qquad (5.7)
\end{aligned}
$$

Here, the first inequality follows from Lemma 3.2, the third inequality follows from the mean value theorem and the assumption that $\nabla_z G_i(x, y, z, \xi)$ is dominated by $\delta(\xi)$ for each $i = 1, \ldots, m$, the fourth inequality follows from (iv) of Lemma 4.1, and the limit follows from (5.1)–(5.2) and the fact that δ is integrable. In consequence, taking a limit in (5.3), we have from (5.5) and (5.7) that

$$\mathbb{E}[f(x^*, y^*, z(x^*, y^*, \xi; 0), \xi)] \le \mathbb{E}[f(\bar{x}, \bar{y}, z(\bar{x}, \bar{y}, \xi; 0), \xi)]$$

with probability one. This indicates that (x^*, y^*) is an optimal solution of problem (4.1). □

5.2. Convergence of stationary points

In general, it is very difficult to get a global optimal solution since MPEC problems are generally nonconvex due to their combinatorial nature of the constraints, whereas computation of stationary points is relatively easy. Therefore, it is necessary to investigate the limiting behavior of stationary points.

Definition 5.1 (*[20]*). $(x^*, y^*) \in \mathcal{F}$ is said to be a *generalized Clarke stationary point* of problem (4.1) if there exists $\alpha^* \in \mathbb{R}^m$ such that

$$0 \in \partial\mathbb{E}[f(x^*, y^*, z(x^*, y^*, \xi; 0), \xi)] + \partial\Psi(x^*, y^*)\alpha^* + (\mathcal{N}_X(x^*), 0)^T.$$

$(x^*, y^*) \in \mathcal{F}$ is said to be a *weak generalized Clarke stationary point* of problem (4.1) if there exists $\alpha^* \in \mathbb{R}^m$ such that

$$
\begin{aligned}
0 \in\; & \mathbb{E}[\nabla_{(x,y)} f(x^*, y^*, z(x^*, y^*, \xi; 0), \xi) + \mathcal{A}(x^*, y^*, \xi, 0)\nabla_z f(x^*, y^*, z(x^*, y^*, \xi; 0), \xi)] \\
& + \left\{\mathbb{E}[\nabla_{(x,y)} G(x^*, y^*, z(x^*, y^*, \xi; 0), \xi) + \mathcal{A}(x^*, y^*, \xi, 0)\nabla_z G(x^*, y^*, z(x^*, y^*, \xi; 0), \xi)]\mathscr{B}\right. \\
& \left. + (0, \mathscr{A})^T\right\}\alpha^* + (\mathcal{N}_X(x^*), 0)^T, \qquad (5.8)
\end{aligned}
$$

where

$$\mathcal{A}(x, y, \xi, 0) := -\partial_{(x,y)}\Phi(x, y, z(x, y, \xi; 0), \xi)[\partial_z \Phi(x, y, z(x, y, \xi; 0), \xi)]^{-1}, \qquad (5.9)$$

$\mathscr{A}$ and $\mathscr{B}$ denote the sets of the diagonal matrices $\mathrm{diag}(a_1, \ldots, a_m)$ and $\mathrm{diag}(b_1, \ldots, b_m)$, respectively, with

$$\begin{cases} a_i = 1 - \epsilon \\ b_i = 1 - \delta \end{cases} \quad (\text{where } \epsilon^2 + \delta^2 \le 1) \qquad (5.10)$$

if $y_i^* = \mathbb{E}[G_i(x^*, y^*, z(x^*, y^*, \xi; 0), \xi)] = 0$ and, otherwise,

$$\begin{cases} a_i = 1 - \dfrac{y_i^*}{\sqrt{(y_i^*)^2 + (\mathbb{E}[G_i(x^*, y^*, z(x^*, y^*, \xi; 0), \xi)])^2}}, \\ b_i = 1 - \dfrac{\mathbb{E}[G_i(x^*, y^*, z(x^*, y^*, \xi; 0), \xi)]}{\sqrt{(y_i^*)^2 + (\mathbb{E}[G_i(x^*, y^*, z(x^*, y^*, \xi; 0), \xi)])^2}}. \end{cases} \tag{5.11}$$

Under some conditions, generalized Clarke stationary points are weak generalized Clarke stationary points [24,20], but the converse may not be true.

Definition 5.2. $(x^N, y^N) \in X \times \mathbb{R}^m$ is said to be a *stationary point* of problem (4.3) if

$$\begin{aligned} 0 \in \frac{1}{N}\sum_{\ell=1}^{N}\Bigg(& \nabla_{x,y} f(x^N, y^N, z(x^N, y^N, \xi^\ell; \mu_N), \xi^\ell) + \mathcal{A}(x^N, y^N, \xi^\ell, \mu_N)\nabla_z f(x^N, y^N, z(x^N, y^N, \xi^\ell; \mu_N), \xi^\ell) \\ & + \Big[\big(\mathcal{A}(x^N, y^N, \xi^\ell, \mu_N)\nabla_z G(x^N, y^N, z(x^N, y^N, \xi^\ell; \mu_N), \xi^\ell) + \nabla_{(x,y)} G(x^N, y^N, z(x^N, y^N, \xi^\ell; \mu_N), \xi^\ell)\big) B^N \\ & + (0, A^N)^T\Big] \times 2\rho_N \Psi^N(x^N, y^N)\Bigg) + (\mathcal{N}_X(x^N), 0)^T, \end{aligned} \tag{5.12}$$

where

$$\mathcal{A}(x, y, \xi, \mu) := -\nabla_{(x,y)}\Phi(x, y, z(x, y, \xi; \mu), \xi)[\nabla_z \Phi(x, y, z(x, y, \xi; \mu), \xi)]^{-1},$$

$A^N := \mathrm{diag}(a_1^N, \dots, a_m^N)$ and $B^N := \mathrm{diag}(b_1^N, \dots, b_m^N)$ are matrices with

$$\begin{cases} a_i^N = 1 - \dfrac{y_i^N}{\sqrt{(y_i^N)^2 + \left(\frac{1}{N}\sum_{\ell=1}^{N} G_i(x^N, y^N, z(x^N, y^N, \xi^\ell; \mu_N), \xi^\ell)\right)^2 + \mu_N^2}}, \\ b_i^N = 1 - \dfrac{\frac{1}{N}\sum_{\ell=1}^{N} G_i(x^N, y^N, z(x^N, y^N, \xi^\ell; \mu_N), \xi^\ell)}{\sqrt{(y_i^N)^2 + \left(\frac{1}{N}\sum_{\ell=1}^{N} G_i(x^N, y^N, z(x^N, y^N, \xi^\ell; \mu_N), \xi^\ell)\right)^2 + \mu_N^2}}. \end{cases} \tag{5.13}$$

Since f, G, H, and Φ are all locally Lipschitz continuous in (x, y, ξ), the Clarke generalized gradients $\partial_{(x,y)} f$, $\partial_{(x,y)} G$, $\partial_{(x,y)} H$, and $\partial_{(x,y)}\Phi$ are all measurable.

Theorem 5.2. *Suppose that $\nabla_{(x,y)} H_i(x, y, z, \xi)$, $i = 1, \dots, s$, are dominated by an integrable function $\delta(\xi)$ uniformly on $X \times \mathbb{R}^m \times \mathbb{R}^s$. Then*

(i) *$\mathcal{A}(x, y, \xi, \mu)$ is upper semi-continuous with respect to (x, y, μ) on $X \times \mathbb{R}^m \times [-\mu_0, \mu_0]$ for almost every $\xi \in \Xi$, where μ_0 is given in Lemma 4.1;*

(ii) *there exists an integrable function $\tau(\xi)$ such that*

$$\|\mathcal{A}(x, y, \xi, \mu)\| := \sup_{A \in \mathcal{A}(x,y,\xi,\mu)} \|A\| \le \tau(\xi) \tag{5.14}$$

holds for any $(x, y, \mu) \in X \times \mathbb{R}^m \times [-\mu_0, \mu_0]$ and almost every $\xi \in \Xi$.

Proof. (i) When $\mu \neq 0$, we have

$$\begin{aligned} \mathcal{A}(x, y, \xi, \mu) &= -\nabla_{(x,y)}\Phi(x, y, z(x, y, \xi; \mu), \xi, \mu)[\nabla_z \Phi(x, y, z(x, y, \xi; \mu), \xi, \mu)]^{-1} \\ &= -\nabla_{(x,y)} H(x, y, z(x, y, \xi; \mu), \xi)\,\bar{B}[\nabla_z \Phi(x, y, z(x, y, \xi; \mu), \xi, \mu)]^{-1}, \end{aligned}$$

where $\bar{B} := \mathrm{diag}(b_1, \dots, b_s)$ with

$$b_i = 1 - \frac{H_i(x, y, z(x, y, \xi; \mu), \xi)}{\sqrt{(y_i)^2 + [H_i(x, y, z(x, y, \xi; \mu), \xi)]^2 + \mu^2}}.$$

Since $H(x, y, z, \xi)$ and $\phi(a, b; \mu)$ are continuously differentiable in (x, y, z) and (a, b) respectively, $\mathcal{A}(x, y, \xi, \mu)$ is continuous in (x, y, μ) for almost every $\xi \in \Xi$.

When $\mu = 0$, we have

$$\mathcal{A}(x, y, \xi, 0) := -\partial_{(x,y)}\Phi(x, y, z(x, y, \xi; 0), \xi, 0)[\partial_z \Phi(x, y, z(x, y, \xi; 0), \xi, 0)]^{-1}.$$

It follows from (i) of Lemma 4.1 that $\partial_z \Phi(x, y, z(x, y, \xi; 0), \xi)$ is uniformly nonsingular, that is, there exists a constant $\nu > 0$ such that

$$\|[\partial_z \Phi(x, y, z(x, y, \xi; 0), \xi, 0)]^{-1}\| \le \nu \tag{5.15}$$

for any $(x, y) \in X \times \mathbb{R}^m$ and $\xi \in \Xi$. Note that $\phi(a, b; \mu)$ is Lipschitz continuous with respect to (a, b, μ) and $H(x, y, z, \xi)$ is continuously differentiable with respect to (x, y, z). Then, $\Phi(x, y, z(x, y, \xi; 0), \xi)$ is locally Lipschitz continuous with respect to (x, y) and hence $\partial_{(x,y)}\Phi(x', y', z(x', y', \xi; 0), \xi)$ is bounded for all (x', y', μ') closing to $(x, y, 0)$. Consequently, there exists a neighborhood $U_{(x,y,0)}$ of $(x, y, 0)$ such that the closure of $\cup_{(x',y',\mu')\in U_{(x,y,0)}} \mathcal{A}(x', y', \xi, \mu')$ is a compact set. By the definition of Clarke generalized Jacobian, $\mathcal{A}(\cdot, \cdot, \xi, \cdot)$ is closed at $(x, y, 0)$. Therefore, we have from [26, Theorem 5] that $\mathcal{A}(\cdot, \cdot, \xi, \cdot)$ is upper semi-continuous on $X \times \mathbb{R}^m \times [-\mu_0, \mu_0]$.

(ii) When $\mu \neq 0$, we have

$$\begin{aligned}
\|\mathcal{A}(x, y, \xi, \mu)\| &= \left\|-\nabla_{(x,y)}H(x, y, z(x, y, \xi; \mu), \xi)\bar{B}[\nabla_z \Phi(x, y, z(x, y, \xi; \mu), \xi)]^{-1}\right\| \\
&\le \left\|\nabla_{(x,y)}H(x, y, z(x, y, \xi; \mu), \xi)\right\| \cdot \left\|\bar{B}\right\| \cdot \|[\nabla_z \Phi(x, y, z(x, y, \xi; \mu), \xi)]^{-1}\| \\
&\le s\nu \left\|\bar{B}\right\| \delta(\xi) \\
&\le 2s\nu\delta(\xi)
\end{aligned} \tag{5.16}$$

for every $(x, y, \mu) \in X \times \mathbb{R}^m \times [-\mu_0, \mu_0]$ and almost every $\xi \in \Xi$. Here the second inequality follows from (5.15) and the fact that $\nabla_{(x,y)}H_i(x, y, z, \xi)$, $i = 1, \ldots, s$, are dominated by $\delta(\xi)$ and the last inequality follows from $|b_i| \le 2$ for $i = 1, \ldots, m$.

When $\mu = 0$, we have

$$\begin{aligned}
\mathcal{A}(x, y, \xi, 0) &= -\partial_{(x,y)}\Phi(x, y, z(x, y, \xi; 0), \xi, 0)[\partial_z \Phi(x, y, z(x, y, \xi; 0), \xi, 0)]^{-1} \\
&= -\nabla_{(x,y)}H(x, y, z(x, y, \xi; 0), \xi)\overline{\mathcal{B}}[\partial_z \Phi(x, y, z(x, y, \xi; 0), \xi, 0)]^{-1},
\end{aligned}$$

where $\overline{\mathcal{B}}$ is the set of the matrices $\mathrm{diag}(b_1, \ldots, b_s)$ with

$$b_i = \begin{cases} 0 \le c \le 2, & \text{if } z_i(x, y, \xi; 0) = H_i(x, y, z(x, y, \xi; 0), \xi) = 0, \\ 1 - \dfrac{H_i(x, y, z(x, y, \xi; 0), \xi)}{\sqrt{(y_i)^2 + [H_i(x, y, z(x, y, \xi; 0), \xi)]^2}}, & \text{otherwise.} \end{cases}$$

In a similar way to (5.16), we can show

$$\|\mathcal{A}(x, y, \xi, 0)\| \le 2s\nu\delta(\xi)$$

for every $(x, y, \mu) \in X \times \mathbb{R}^m \times [-\mu_0, \mu_0]$ and almost every $\xi \in \Xi$. Letting $\tau(\xi) = 2s\nu\delta(\xi)$, we have (5.14) immediately. □

We next show the main convergence result of this subsection. The following lemma is useful.

Lemma 5.3 ([20]). *Let V be a compact set and $\mathcal{G}(v, \xi) : V \times \Xi \to 2^{\mathbb{R}^m}$ be a measurable and compact set-valued mapping that is upper semi-continuous with respect to v on V for almost every ξ. Let $\xi^1, \ldots, \xi^N$ be independently and identically distributed random samples and*

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

Suppose that $\mathcal{G}(v, \xi)$ is dominated by an integrable function $\delta(\xi)$. Consider the stochastic generalized equation

$$0 \in \mathbb{E}[\mathrm{conv}\,\mathcal{G}(v, \xi)], \tag{5.17}$$

and its sample average approximation

$$0 \in \mathcal{G}_N(v). \tag{5.18}$$

Suppose that both (5.17) and (5.18) have nonempty solution sets. Let v^N be a solution of (5.18). Then with probability one, an accumulation point of $\{v^N\}$ is a solution of (5.17).

A probability measure space $(\Omega, \mathcal{C}, P)$ is called nonatomic if, for any C_1 and C_2 in $\mathcal{C}$, $P(C_1) > 0$ implies that there exists C_2 satisfying $C_2 \subset C_1$ and $0 < P(C_2) < P(C_1)$. By Theorem 5.4 of [27], $\mathbb{E}[\mathrm{conv}\,\mathcal{G}(z, \xi)] = \mathbb{E}[\mathcal{G}(z, \xi)]$ if the involved probability measure in Lemma 5.3 is nonatomic.

Theorem 5.3. *Suppose that $\nabla_{(x,y)}H_i(x, y, z, \xi)$, $i = 1, \dots, s$, $\nabla_{(x,y,z)}G_j(x, y, z, \xi)$, $j = 1, \dots, m$, and $\nabla_{(x,y,z)}f(x, y, z, \xi)$ are all dominated by the integrable function $\delta(\xi)$ on $X \times \mathbb{R}^m \times \mathbb{R}^s$, and the involved probability measure space is nonatomic. Let (x^N, y^N) be a stationary point of problem (4.3) for each N. Suppose that there exists a constant π such that*

$$\frac{1}{N}\sum_{\ell=1}^{N} f(x^N, y^N, z(x^N, y^N, \xi^\ell; \mu_N), \xi^\ell) + \rho_N \left\| \Psi^N(x^N, y^N) \right\|^2 \le \pi \tag{5.19}$$

holds for each N. Then, any accumulation point of $\{(x^N, y^N)\}$ is almost surely a weak generalized Clarke stationary point of (4.1).

Proof. Taking a subsequence if necessary, we assume for the simplicity of notation that (x^N, y^N) tends to (x^*, y^*). It follows from (5.19) that

$$\left\| \Psi^N(x^N, y^N) \right\|^2 \le \rho_N^{-1} \left(\pi - \frac{1}{N}\sum_{\ell=1}^{N} f(x^N, y^N, z(x^N, y^N, \xi^\ell; \mu_N), \xi^\ell) \right).$$

Taking a limit on both sides of the formula above, we have by (5.5)–(5.6) that $\Psi(x^*, y^*) = 0$ with probability one, which means that (x^*, y^*) is feasible to problem (4.1). In a similar way to (5.5)–(5.6), we can show that, with probability one,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{\ell=1}^{N} \nabla_{(x,y,z)} f(x^N, y^N, z(x^N, y^N, \xi^\ell; \mu_N), \xi^\ell) = \mathbb{E}[\nabla_{(x,y,z)} f(x^*, y^*, z(x^*, y^*, \xi; 0), \xi)], \tag{5.20}$$

$$\lim_{N\to\infty} \frac{1}{N}\sum_{\ell=1}^{N} \nabla_{(x,y,z)} G(x^N, y^N, z(x^N, y^N, \xi^\ell; \mu_N), \xi^\ell) = \mathbb{E}[\nabla_{(x,y,z)} G(x^*, y^*, z(x^*, y^*, \xi; 0), \xi)]. \tag{5.21}$$

Since (x^N, y^N) tends to (x^*, y^*), we can choose a compact neighborhood U^* of $(x^*, y^*, 0)$ that contains the whole sequence $\{(x^N, y^N, \mu_N)\}$ for N sufficiently large. Note that $\mathcal{A}(\cdot, \cdot, \xi, \cdot) : U^* \to 2^{\mathbb{R}^{(n+m+s)}}$ is a random compact set-valued mapping. It follows from Theorem 5.2 and Lemma 5.3 that

$$\lim_{N\to\infty} \frac{1}{N}\sum_{\ell=1}^{N} \mathcal{A}(x^N, y^N, \xi^\ell, \mu_N) \in \mathbb{E}[\mathcal{A}(x^*, y^*, \xi, 0)] \quad \text{w.p.1.} \tag{5.22}$$

By the definitions, $\{A^N\}$ and $\{B^N\}$ converge to $\mathscr{A}$ and $\mathscr{B}$ respectively. By (5.5) and (5.19), the sequence $\{\rho_N \|\Psi^N(x^N, y^N)\|\}$ is bounded with probability one. Therefore, we may assume that

$$\alpha^* := \lim_{N\to\infty} 2\rho_N \left\| \Psi^N(x^N, y^N) \right\|. \tag{5.23}$$

Taking a limit in (5.12), we have by (5.20)–(5.23) and upper semi-continuity of Clarke normal cone that (5.8) holds with probability one. That is, w.p.1 (x^*, y^*) is a weak generalized Clarke stationary point of (1.1), where α^* is the corresponding Lagrange multiplier. □

6. Preliminary numerical results

We have tested the proposed method on some examples of SMPHEC. The examples are constructed based on the ones given in [16].

Example 6.1. Consider the problem

$$\begin{aligned}
&\min\ f(x, y, z) := \sum_{i=1}^{2} \mathbb{E}[(x_i - 1)^2 + (y_i - 1)^2 + z_i^2] \\
&\text{s.t.}\quad x_1 \in [0, 10], \quad x_2 \in [0, 5], \\
&\qquad\ 0 \le y_1 \perp \mathbb{E}[z_1 + x_1 + \xi_1] \ge 0, \\
&\qquad\ 0 \le y_2 \perp \mathbb{E}[z_2 - x_2 + \xi_2] \ge 0, \\
&\qquad\ 0 \le z_i \perp z_i - x_i + \xi_i \ge 0, \quad i = 1, 2,
\end{aligned}$$

where the random variables ξ_1 and ξ_2 are independent and uniformly distributed on [0, 1]. Note that, for any fixed x and ξ, the last two complementarity constraints imply that

$$z_i(x, \xi) = \max\{x_i - \xi_i, 0\}, \quad i = 1, 2.$$

It follows that

$$f(x, y, z(x, \xi)) = \sum_{i=1}^{2}\left((x_i - 1)^2 + (y_i - 1)^2 + \frac{x_i^3}{3}\right)$$
$$\mathbb{E}[z_1(x, \xi) + x_1 + \xi_1] = \frac{x_1^2}{2} + x_1 + \frac{1}{2},$$
$$\mathbb{E}[z_2(x, \xi) - x_2 + \xi_2] = \frac{x_2^2}{2} - x_2 + \frac{1}{2},$$

and hence

$$\begin{cases} 0 \le y_1 \perp \mathbb{E}[z_1(x, \xi) + x_1 + \xi_1] \ge 0 \\ 0 \le y_2 \perp \mathbb{E}[z_2(x, \xi) - x_2 + \xi_2] \ge 0 \end{cases} \iff \begin{cases} y_1 = 0, \\ 0 \le y_2 \perp (x_2 - 1)^2 \ge 0. \end{cases}$$

By straightforward analysis, we have that the optimal solution is

$$x^* = \begin{pmatrix} -1 + \sqrt{3} \\ 1 \end{pmatrix}, \qquad y^* = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad z^*(\xi) = \begin{pmatrix} \max\{-1 + \sqrt{3} - \xi_1, 0\} \\ 1 - \xi_2 \end{pmatrix}$$

and the optimal value is 1.535898.

Example 6.2. Consider the problem

$$\begin{aligned} \min\ & \sum_{i=1}^{3} \mathbb{E}[(x_i - 1)^2 + (y_i - 1)^2 + z_i^2] \\ \text{s.t.}\ & x_1 \in [0, 10], \qquad x_2 \in [0, 5], \qquad x_3 \in [0, 7], \\ & 0 \le y_1 \perp \mathbb{E}[z_1 + x_1 + \xi_1] \ge 0, \\ & 0 \le y_2 \perp \mathbb{E}[z_2 - x_2 + \xi_2] \ge 0, \\ & 0 \le y_3 \perp \mathbb{E}[z_3 + x_3 + \xi_3] \ge 0, \\ & 0 \le z_i \perp z_i - x_i + \xi_i \ge 0, \quad i = 1, 2, 3, \end{aligned}$$

where the random variables ξ_1, ξ_2, and ξ_3 are independent and uniformly distributed on $[0, 1]$. In a similar way to Example 6.1 , we can get that the optimal solution is

$$x^* = \begin{pmatrix} -1 + \sqrt{3} \\ 1 \\ -1 + \sqrt{3} \end{pmatrix}, \qquad y^* = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \qquad z^*(\xi) = \begin{pmatrix} \max\{-1 + \sqrt{3} - \xi_1, 0\} \\ 1 - \xi_2 \\ \max\{-1 + \sqrt{3} - \xi_3, 0\} \end{pmatrix}$$

and the optimal value is 2.738463.

Example 6.3. Consider the problem

$$\begin{aligned} \min\ & \sum_{i=1}^{2} \mathbb{E}[(x_i - 1)^2 + (y_i - 1)^2 + z_i^2] \\ \text{s.t.}\ & x_1 \in [0, 2], \qquad x_2 \in [0, 5], \\ & 0 \le y_1 \perp \mathbb{E}[z_1 + \xi_1 + \xi_2] \ge 0, \\ & 0 \le y_2 \perp \mathbb{E}[z_2 - x_2 + \xi_2] \ge 0, \\ & 0 \le z_1 \perp z_1 - x_1 + \xi_1 - \xi_2 \ge 0, \\ & 0 \le z_2 \perp z_2 - x_2 + \xi_2 \ge 0, \end{aligned}$$

where the random variables ξ_1 and ξ_2 are independent and uniformly distributed on $[0, 1]$. Note that, for any fixed x and ξ, the last two complementarity constraints imply that

$$z_1(x, \xi) = \max\{x_1 - \xi_1 + \xi_2, 0\}, \qquad z_2(x, \xi) = \max\{x_2 - \xi_2, 0\}.$$

Then, by straightforward calculus, we can show that the problem is equivalent to

$$\begin{aligned} \min\ & f(x, y) \\ \text{s.t.}\ & x_1 \in [0, 2], \qquad x_2 \in [0, 5], \qquad y_1 = 0, \\ & 0 \le y_2 \perp (x_2 - 1)^2 \ge 0, \end{aligned}$$

Table 1
Numerical results for Example 6.1.

N	ρ_N	μ_N	*Iter*	*Opt.Sol*		*Opt.val*
				x^N	y^N	
100	200	10^{-2}	24	(0.742622, 1.012852)	$(3.3 \times 10^{-5}, 0.994175)$	1.531074
200	400	10^{-3}	22	(0.734633, 0.993295)	$(3.3 \times 10^{-7}, 1.000085)$	1.497065
400	800	10^{-4}	30	(0.730337, 0.996850)	$(3.3 \times 10^{-9}, 1.000425)$	1.508721
600	1600	10^{-5}	24	(0.735411, 0.998209)	$(3.0 \times 10^{-11}, 0.998459)$	1.509286
1200	3200	10^{-6}	30	(0.732355, 0.998397)	$(-7.0 \times 10^{-11}, 0.999168)$	1.524088
2400	6400	10^{-7}	33	(0.733752, 0.999900)	$(6.2 \times 10^{-13}, 1.000067)$	1.532532

Table 2
Numerical results for Example 6.2.

N	ρ_N	μ_N	*Iter*	*Opt.Sol*		*Opt.val*
				x^N	y^N	
100	200	10^{-2}	32	(0.748955, 0.980532, 0.738435)	$(3.3 \times 10^{-5}, 0.995699, 3.3 \times 10^{-5})$	2.731662
200	400	10^{-3}	45	(0.759120, 0.993629, 0.736650)	$(3.2 \times 10^{-7}, 0.998971, 3.3 \times 10^{-7})$	2.717602
400	800	10^{-4}	55	(0.741465, 0.997210, 0.737922)	$(1.5 \times 10^{-9}, 1.000141, 2.0 \times 10^{-10})$	2.708110
600	1600	10^{-5}	55	(0.733683, 0.998472, 0.735687)	$(-3.7 \times 10^{-10}, 1.000040, -1.1 \times 10^{-10})$	2.712522
1200	3200	10^{-6}	62	(0.739799, 0.999596, 0.740623)	$(4.8 \times 10^{-11}, 1.000626, -1.1 \times 10^{-9})$	2.722092
2400	6400	10^{-7}	47	(0.730216, 0.999507, 0.733022)	$(9.7 \times 10^{-12}, 1.000025, 1.6 \times 10^{-11})$	2.733244

Table 3
Numerical results for Example 6.3.

N	ρ_N	μ_N	*Iter*	*Opt.Sol*		*Opt.val*
				x^N	y^N	
100	200	10^{-2}	18	(0.542437, 0.994594)	$(2.5 \times 10^{-5}, 0.995368)$	1.980037
200	400	10^{-3}	25	(0.503722, 0.998290)	$(2.5 \times 10^{-9}, 0.999816)$	1.977099
400	800	10^{-4}	35	(0.478069, 0.998976)	$(1.4 \times 10^{-9}, 0.999705)$	1.991928
600	1600	10^{-5}	22	(0.500111, 0.998526)	$(-1.4 \times 10^{-10}, 1.000354)$	1.971583
1200	3200	10^{-6}	31	(0.493165, 0.998809)	$(2.8 \times 10^{-13}, 0.998754)$	1.980111
2400	6400	10^{-7}	33	(0.493582, 0.999614)	$(-2.7 \times 10^{-11}, 1.000089)$	1.972298

Table 4
Initial points used in the tests.

	x	y
Example 6.1	(0.731772, 1.495137)	(0.836858, 1.443392)
Example 6.2	(0.625683, 0.769499, 0.566540)	(1.291231, 1.225757, 1.443821)
Example 6.3	(0.751742, 1.138252)	(0.527568, 1.472677)

where

$$f(x,y) = \begin{cases} -\dfrac{1}{12}x_1^4 + \dfrac{1}{12}(1+x_1)^4 + \dfrac{1}{3}x_2^3 + \displaystyle\sum_{i=1}^{2}((x_i-1)^2 + (y_i-1)^2), & x_1 \in [0,1], \\ -\dfrac{1}{6}x_1^4 + \dfrac{1}{12}[(1+x_1)^4 + (x_1-1)^4] + \dfrac{1}{3}x_2^3 + \displaystyle\sum_{i=1}^{2}((x_i-1)^2 + (y_i-1)^2), & x_1 \in [1,2]. \end{cases}$$

We further can get that the optimal solution is

$$x^* = \begin{pmatrix} \dfrac{1}{6}\sqrt{141} - \dfrac{3}{2} \\ 1 \end{pmatrix}, \qquad y^* = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \qquad z^*(\xi) = \begin{pmatrix} \max\left\{\dfrac{1}{6}\sqrt{141} - \dfrac{3}{2} - \xi_1 + \xi_2, 0\right\} \\ 1 - \xi_2 \end{pmatrix}$$

and the optimal value is 1.999226.

The computational results for Examples 6.1–6.3 are reported in Tables 1–3, respectively. In our experiments, we employed the random number generator `rand` in Matlab 7.1 to generate the initial points, which are shown in Table 4, and the independently and identically distributed random samples from Ξ and we solved the approximation problems by the solver `fmincon` in Matlab 7.1. In the tables, *iter* denotes the number of iterations returned by `fmincon`, *Opt.Sol* denotes the approximate optimal solutions and *Opt.Val* denotes the optimal values. The results shown in the tables reveal that the proposed method was able to solve the examples successfully.

7. Conclusions

We have presented a new SMPEC model, which includes either "here-and-now" or "wait-and-see" type complementarity constraints. An example has been employed to describe the necessity of the study of the new model. In order to solve the new model, we employed the SAA techniques to approximate the expectations and used the smoothing and penalty techniques to deal with the complementarity constraints. Convergence theory of the proposed approach has been established. Actually, the FB function employed in the previous sections may be replaced by other NCP functions.

On the other hand, we have noticed that Anitescu and Birge discussed the following stochastic optimization problems with mixed expectations and per-scenario constraints (SOESC) in the recent paper [28]:

$$\begin{aligned} \min_{x,z(\xi)} \quad & \mathbb{E}[f(x, z(\xi), \xi)] \\ \text{s.t.} \quad & \mathbb{E}[G(x, z(\xi), \xi)] = 0, \\ & H(x, z(\xi), \xi) = 0, \quad \text{a.e. } \xi \in \Xi. \end{aligned}$$

Obviously, by using some NCP function in dealing with the complementarity constraints, problem (1.1) can be rewritten as a special case of SOESC, in which the constraints related to the complementarity constraints are generally nonsmooth. However, all functions involved in [28] are assumed to be smooth. Therefore, problem (1.1) is essentially different from the problem studied in [28].

Acknowledgements

The authors are grateful to the anonymous referees for their helpful suggestions and comments.

References

[1] M. Fukushima, G.H. Lin, Smoothing methods for mathematical programs with equilibrium constraints, in: Proceedings of the ICKS'04, IEEE Computer Society, 2004, pp. 206–213.
[2] Z.Q. Luo, J.S. Pang, D. Ralph, Mathematical Programs with Equilibrium Constraints, Cambridge University Press, Cambridge, UK, 1996.
[3] J. Outrata, M. Kocvara, J. Zowe, Nonsmooth Approach to Optimization Problems with Equilibrium Constraints: Theory, Applications, and Numerical Results, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.
[4] G.H. Lin, M. Fukushima, Stochastic equilibrium problems and stochastic mathematical programs with equilibrium constraints: a survey, Pacific Journal of Optimization 6 (2010) 455–482.
[5] G.H. Lin, X. Chen, M. Fukushima, Smoothing implicit programming approaches for stochastic mathematical programs with linear complementarity constraints, Technical Report 2003-006, Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, Kyoto, Japan, 2003.
[6] S.I. Birbil, G Gürkan, O. Listes, Solving stochastic mathematical programs with complementarity constraints using simulation, Mathematics of Operations Research 31 (2006) 739–760.
[7] G.H. Lin, X. Chen, M. Fukushima, Solving stochastic mathematical programs with equilibrium constraints via approximation and smoothing implicit programming with penalization, Mathematical Programming 116 (2009) 343–368.
[8] G.H. Lin, M. Fukushima, A class of stochastic mathematical programs with complementarity constraints: formulations and algorithms, Journal of Industrial and Management Optimization 1 (2005) 99–122.
[9] G.H. Lin, M. Fukushima, Regularization method for stochastic mathematical programs with complementarity constraints, European Series of Applied and Industrial Mathematics (ESAIM): Control, Optimisation and Calculus of Variations 11 (2005) 252–265.
[10] G.H. Lin, H. Xu, M. Fukushima, Monte Carlo and quasi-Monte Carlo sampling methods for a class of stochastic mathematical programs with equilibrium constraints, Mathematical Methods of Operations Research 67 (2008) 423–441.
[11] F. Meng, H. Xu, Exponential convergence of sample average approximation methods for a class of stochastic mathematical programs with complementarity constraints, Journal of Computational Mathematics 24 (2006) 733–748.
[12] H. Xu, An MPCC approach for stochastic Stackelberg–Nash–Cournot equilibrium, Optimization 54 (2005) 27–57.
[13] H. Xu, An implicit programming approach for a class of stochastic mathematical programs with linear complementarity constraints, SIAM Journal on Optimization 16 (2006) 670–696.
[14] M. Patriksson, L. Wynter, Stochastic mathematical programs with equilibrium constraints, Operations Research Letters 25 (1999) 159–167.
[15] A. Shapiro, Stochastic mathematical programs with equilibrium constraints, Journal of Optimization Theory and Applications 128 (2006) 223–243.
[16] A. Shapiro, H. Xu, Stochastic mathematical programs with equilibrium constraints, modeling and sample average approximation, Optimization 57 (2008) 395–418.
[17] R.T. Rockafellar, R.J-B. Wets, Variational Analysis, Springer, Berlin, 1998.
[18] R.J. Aubin, H. Frankowska, Set-Valued Analysis, Birkhauser, Boston, 1990.
[19] F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley, New York, 1983.
[20] H. Xu, D. Zhang, Smooth sample average approximation of stationary points in nonsmooth stochastic optimization and applications, Mathematical Programming 119 (2009) 371–401.
[21] D. Sun, L. Qi, On NCP-functions, Computational Optimization and Applications 13 (1999) 201–220.
[22] Y. Chen, M. Florian, The nonlinear bilevel programming problem: formulations, regularity and optimality conditions, Optimization 32 (1995) 193–209.
[23] H. Xu, F. Meng, Convergence analysis of sample average approximation methods for a class of stochastic mathematical programs with equality constraints, Mathematics of Operations Research 32 (2007) 648–669.
[24] A. Ruszczynski, A. Shapiro, Stochastic Programming, Handbook in Operations Research and Management Science, Elsevier, 2003.
[25] P. Hall, J.S. Marron, Local minima in cross-validation functions, Journal of the Royal Statistical Society, Series B 53 (1991) 245–252.
[26] W.W. Hogan, Point-to-sets maps in mathematical programming, SIAM Review 15 (1973) 591–603.
[27] C. Hess, Set-valued integration and set-valued probability theory: an overview, in: Handbook of Measure Theory, North-Holland, Amsterdam, 2002, pp. 617–673.
[28] M. Anitescu, J. Birge, Convergence of stochastic average approximation for stochastic optimization problems with mixed expectation and per-scenario constraints, Preprint ANL/MCS 1562–1108.