Set-Valued Anal (2011) 19:283–309
DOI 10.1007/s11228-010-0160-x

# Approximating Stationary Points of Stochastic Mathematical Programs with Equilibrium Constraints via Sample Averaging

**Huifu Xu · Jane J. Ye**

Received: 21 March 2010 / Accepted: 31 August 2010 /
Published online: 17 September 2010


**Abstract** We investigate sample average approximation of a general class of one-stage stochastic mathematical programs with equilibrium constraints. By using graphical convergence of unbounded set-valued mappings, we demonstrate almost sure convergence of a sequence of stationary points of sample average approximation problems to their true counterparts as the sample size increases. In particular we show the convergence of M(Mordukhovich)-stationary point and C(Clarke)-stationary point of the sample average approximation problem to those of the true problem. The research complements the existing work in the literature by considering a general equilibrium constraint to be represented by a stochastic generalized equation and exploiting graphical convergence of coderivative mappings.

**Keywords** SMPEC · Coderivative · Graphical convergence · M-stationary point · C-stationary point · Sample average approximation

**Mathematics Subject Classifications (2010)** 90C15 · 90C46 · 90C30 · 90C31 · 90C33

The work of the second author was partly supported by NSERC.

H. Xu (✉)
School of Mathematics, University of Southampton, Southampton, SO17 1BJ, UK
e-mail: h.xu@soton.ac.uk

J. J. Ye
Department of Mathematics and Statistics, University of Victoria,
Victoria, BC, Canada V8W 3R4
e-mail: janeye@uvic.ca

## 1 Introduction

In this paper we consider the following stochastic mathematical program with equilibrium constraints (SMPEC):

$$\begin{aligned}
&\min_{x,y} \mathbb{E}[f(x,y,\xi(\omega))] \\
&\text{s.t. } (x,y) \in C, \\
&\qquad 0 \in \mathbb{E}[F(x,y,\xi(\omega))] + \mathcal{N}_Y(y),
\end{aligned} \tag{1}$$

where $C$ is a nonempty closed subset of $\mathbb{R}^n \times \mathbb{R}^m$, $f : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}$ is locally Lipschitz continuous, $F : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}^m$ is continuously differentiable, $\xi : \Omega \to \Xi \subset \mathbb{R}^d$ is a vector of random variables defined on probability space $(\Omega, \mathcal{F}, P)$, $\mathbb{E}[\cdot]$ denotes the expected value with respect to the distribution of $\xi$, $\mathcal{N}_Y(y)$ is the limiting normal cone to a closed set $Y \subset \mathbb{R}^m$ at point $y$ with $\mathcal{N}_Y(y) := \emptyset$ if $y \notin Y$. The precise definition of the normal cone will be given in Section 2. In fact all results of this paper remain true if the normal cone operator $\mathcal{N}_Y$ is replaced by a closed set-valued mapping.

This is an one-stage two-level stochastic mathematical program with equilibrium constraints (SMPEC) where decision variables $x$ and $y$ must be chosen simultaneously before the realization of random variable $\xi$, and the relationship between upper level decision variable $x$ and lower level decision variable $y$ is governed by a constraint to be presented by a stochastic generalized equation that is often used to characterize an equilibrium in practical application. Note as a referee commented that in the case when $Y$ is convex, the limiting normal cone reduces to the classical normal cone in convex analysis and subsequently the generalized equation reduces to a variational inequality. Note also that when $\xi$ takes a single constant value, the problem becomes a deterministic MPEC. The latter has been extensively investigated over the past decade, see [19, 26, 28, 46] for comprehensive discussions on theory, numerical methods and applications of MPECs.

Analogous to ordinary stochastic programming models, SMPEC arises from practical needs in decision analysis and engineering where problem data involve some uncertainties. When a problem involves several decision makers who are in a hierarchical and/or competitive relationship, that is, some decision makers have control of or influence on others (e.g. principal-agent problem [21], Stackelberg leader-follower problem [38]), classical stochastic programming models are no longer applicable, a new framework of stochastic programming with bilevel (or multi-level) and/or equilibrium structure is needed where an equilibrium constraint is typically used to characterize a state of competition of parties with conflicting interests. Stochastic programming models as such were first considered by Christiansen et al. [5] for an optimal decision problem in structural engineering design and were studied more recently by Werner [39] to analyze optimal decision making/competition in the Norwegian telecommunication industry. In these problems, simplification of the random data through averaging may lead to an optimal decision to be achieved with small probability. From this point of view, SMPEC models should not be regarded as a generalization of deterministic MPEC, instead they are complements of deterministic MPECs where the latter cannot adequately describe the intrinsic stochastic nature of a decision making problem with equilibrium constraints.

There are two simple but important cases that SMPEC may recover by choosing a specific set $Y$. First, if $Y = \mathbb{R}^m_+$, then the SMPEC reduces to a stochastic mathematical program with complementarity constraints (SMPCC):

$$\begin{aligned} \min_{x,y}\ & \mathbb{E}[f(x, y, \xi(\omega))] \\ \text{s.t.}\ & (x, y) \in C, \\ & 0 \le y \perp \mathbb{E}[F(x, y, \xi(\omega))] \ge 0, \end{aligned} \tag{2}$$

where $\perp$ denotes the perpendicularity of two vectors. The SMPCC model has been well studied, see for instances [4, 14, 20] and references therein.

The other important special case of SMPEC (Eq. 1) is that $Y = \mathbb{R}^m$. In such a case, the model reduces to a classical stochastic programming problem:

$$\begin{aligned} \min\ & \mathbb{E}[f(x, y, \xi(\omega))] \\ \text{s.t.}\ & (x, y) \in C, \\ & \mathbb{E}[F(x, y, \xi(\omega))] = 0. \end{aligned} \tag{3}$$

At this point, it is important to note that our SMPEC model (Eq. 1) may include two-stage SMPECs by allowing $f$ to be nonsmooth (i.e. not continuously differentiable). To see this, let us consider the case when $f(x, y, \xi)$ is the optimal value function of the following second stage problem:

$$\begin{aligned} \min\nolimits_z\ & g(x, y, z, \xi) \\ \text{s.t.}\ & 0 \le z \perp G(x, y, z, \xi) \ge 0. \end{aligned} \tag{4}$$

It is well-known in parametric programming that the optimal value function of Eq. 4 is often Lipschitz continuous under some mild constraint qualifications (see Lucet and Ye [17, 18]). Subsequently, our model (Eq. 1) may cover the following mixed two-stage SMPEC (by exchanging the minimization w.r.t. $z(\cdot)$ with the expectation as in [29]):

$$\begin{aligned} \min\nolimits_{x,y,z(\cdot)}\ & \mathbb{E}[g(x, y, z(\omega), \xi(\omega))] \\ \text{s.t.}\ & (x, y) \in C, \\ & 0 \le z(\omega) \perp G(x, y, z(\omega), \xi) \ge 0 \text{ a.e.}, \\ & 0 \in \mathbb{E}[F(x, y, \xi)] + \mathcal{N}_Y(y), \end{aligned} \tag{5}$$

where $g$ is a smooth real-valued function and $G$ is a smooth vector valued function, the optimal decision on the second stage decision variable $z$ is chosen after the first stage decision variables $x, y$ are given and the uncertainty $\xi$ is realized. The SMPEC model (Eq. 5) may be used to analyze optimal decision of a firm which needs to make decisions at *present* and then at some stage in future in a noncollaborative oligopolistic market.

Here we include a brief literature review on the research of two-stage SMPECs. A general two-stage SMPEC model (without the expected equilibrium constraint) is first studied in [30]. It is regarded fundamentally as an extension of deterministic MPEC. The model has been consequently studied in [13]. Shapiro [35] seems to be the first to propose a Monte Carlo sampling method for solving a two-stage SMPEC. His focus is on the asymptotic consistency of optimal values and solutions obtained from the sample average approximation problem. The approach has been further studied in [36, 42] regarding the asymptotic convergence of statistical estimators of Clarke stationary points.

In this paper we are concerned with the sample average approximation method (SAA) for solving Eq. 1. SAA is a very popular method and is also known under different names such as sample path optimization, Monte Carlo method, or stochastic counterpart to name a few. The basic idea of the method is to use sample averages to approximate expected values. Specifically, let $\xi^1, \cdots, \xi^N$ be an independent and identically distributed (iid) sampling of $\xi$. Then we may consider the following sample average approximation of Eq. 1:

$$\begin{aligned} &\min \frac{1}{N}\sum_{k=1}^{N} f(x, y, \xi^k) \\ &\text{s.t. } (x, y) \in C, \\ &\qquad 0 \in \frac{1}{N}\sum_{k=1}^{N} F(x, y, \xi^k) + \mathcal{N}_Y(y). \end{aligned} \tag{6}$$

In a more recent development, sampling may be non-iid and hence may accommodate some popular numerical schemes such as the quasi-Monte Carlo method [27]. The main benefit of SAA is that it does not require any knowledge of the probability distribution of $\xi$ and/or the numerical approximation of a multiple integration.

Following the terminology in the literature of SAA methods, we call Eq. 1 the *true* problem and Eq. 6 the SAA problem. Our focus in this paper is to study the convergence of a sequence of stationary points of SAA problem (Eq. 6) to its true counterpart. The main contribution of this paper can be summarized as follows:

- We derive the first order necessary optimality conditions for both the true problem (Eq. 1) and the SAA problem (Eq. 6) in terms of the limiting subdifferentials and the coderivatives (see the definitions in Section 2) of the underlying functions under some standard constraint qualifications (no nonzero abnormal multipliers constraint qualifications). This gives a unified treatment of general equilibrium constraints including variational inequality constraints and complementarity constraints as a special case.
- We use the graphical convergence of a set-valued mapping as a main tool for analyzing the asymp totic convergence of a sequence of stationary points obtained from solving the sample average approximate problem (Eq. 6) and show under moderate conditions that w.p.1 a cluster point of such a sequence is a stationary point of the true problem. Graphical convergence of sample average random set-valued mapping is recently studied by Wets and Xu in [40]. Here we consider graphical convergence of co-derivative mapping of a normal cone. The analysis differs significantly from the existing asymptotic analysis in the literature which is essentially based on the uniform law of large number for random compact set-valued mappings [20, 41, 42].
- For the case of complementarity constraints, we derive almost sure (a.s. for short) convergence of M-stationary point and C-stationary point obtained respectively from solving the sample average problem (see e.g. [1, 12, 15] and the references therein for computational algorithms) to their true counterpart as the sample size increases. M-stationarity is generally sharper than C-stationarity, we include the latter as many MPECs solvers only give rise to C-stationary points.

The rest of this paper is organized as follows. In Section 2, we present some basic notions and results on variational analysis and investigate the outer semicontinuity of coderivative mappings. In Section 3, we develop first order optimality conditions for both the true problem and the SAA problem in terms of limiting subdifferentials and coderivatives under no nonzero abnormal multipliers constraint qualifications. In Section 4, we carry out convergence analysis for stationary points obtained from the SAA problem using the graphical convergence as a main tool. Finally in Section 5, we show the asymptotic convergence of a sequence of M-,C-stationary points of the SAA problem to their true counterpart when SMPEC reduces to SMPCC.

## 2 Preliminaries

Throughout this paper, we use the following notation. $\|\cdot\|$ denotes the Euclidean norm of a vector and a compact set of vectors. If $\mathcal{M}$ is a compact set of vectors, then

$$\|\mathcal{M}\| := \max_{M\in\mathcal{M}} \|M\|.$$

$d(x,\mathcal{D}) := \inf_{x'\in\mathcal{D}} \|x-x'\|$ denotes the distance from point $x$ to set $\mathcal{D}$. For two compact sets $\mathcal{C}$ and $\mathcal{D}$,

$$\mathbb{D}(\mathcal{C},\mathcal{D}) := \sup_{x\in\mathcal{C}} d(x,\mathcal{D})$$

denotes the deviation from set $\mathcal{C}$ to set $\mathcal{D}$ (in some references [10] also called *excess* of $\mathcal{C}$ over $\mathcal{D}$), and $\mathbb{H}(\mathcal{C},\mathcal{D})$ denotes the Hausdorff distance between the two sets, that is,

$$\mathbb{H}(\mathcal{C},\mathcal{D}) := \max\left(\mathbb{D}(\mathcal{C},\mathcal{D}),\mathbb{D}(\mathcal{D},\mathcal{C})\right).$$

We use $\mathcal{C}+\mathcal{D}$ to denote the Minkowski addition of the two sets, that is,

$$\{C+D : C\in\mathcal{C}, D\in\mathcal{D}\}.$$

We use $a^Tb$ to denote the scalar product of vectors $a$ and $b$, where $a^T$ denotes the transpose of vector $a$. When $A$ is a matrix, $A^Tb$ denotes matrix vector multiplication. We use $a\geq 0$ to denote the componentwise nonnegativity of a vector $a$, and $a\perp b$ the perpendicularity vectors $a$ and $b$. Specifically, $0\leq a\perp b\geq 0$ indicates the complementary relationship between $a$ and $b$, that is, the i.e., $a_i, b_i\geq 0$, $a_ib_i=0$ and one positive values forces the other to be zero for every pair of components.

For a set-valued mapping $\Phi:\mathbb{R}^m\to 2^{\mathbb{R}^q}$ (assigning to each $z\in\mathbb{R}^m$ a set $\Phi(z)\subset R^q$ which may be empty), we denote by gph $\Phi$ the graph of $\Phi$, i.e.,

$$\text{gph }\Phi := \{(z,v) :\in \mathbb{R}^m\times\mathbb{R}^q : v\in\Phi(z)\}.$$

conv $\mathcal{C}$ denotes the convex hull of a set $\mathcal{C}$. We denote by $B(x,\delta)$ the open ball with radius $\delta$ and center $x$, that is $B(x,\delta) := \{x' : \|x'-x\| < \delta\}$. When $\delta$ is dropped, $B(x)$ represents a neighborhood of point $x$. More specifically, we use $\mathcal{B}$ to denote a closed unit ball in a finite dimensional space.

2.1 Variational Analysis

Let $\Phi : \mathbb{R}^m \to 2^{\mathbb{R}^m}$ be a set-valued mapping. We denote by $\limsup_{x \to \bar{x}} \Phi(x)$ the *Painlevé–Kuratowski outer limit*:

$$\limsup_{x \to \bar{x}} \Phi(x) := \{v \in R^m : \exists \text{ seqences } x_k \to \bar{x}, v_k \to v \\ \text{with } v_k \in \Phi(x_k) \ \forall k = 1, 2, \ldots\}.$$

**Definition 2.1** (Normal cones) Let $\mathcal{C}$ be a nonempty closed subset of $\mathbb{R}^m$. Given $z \in \mathcal{C}$, the convex cone

$$\mathcal{N}^\pi_\mathcal{C}(z) := \{\zeta \in R^m : \exists \sigma > 0, \text{ such that } \zeta^T(z' - z) \le \sigma \|z' - z\|^2 \ \forall z' \in \mathcal{C}\}$$

is called the *proximal normal cone* to set $\mathcal{C}$ at point $z$. By convention, for $z \notin \mathcal{C}$, $\mathcal{N}^\pi_\mathcal{C}(z) = \emptyset$. The closed cone

$$\mathcal{N}_\mathcal{C}(z) := \limsup_{z' \to z} \mathcal{N}^\pi_\mathcal{C}(z')$$

is called the *limiting normal cone* (also known as Mordukhovich normal cone or basic normal cone) to $\mathcal{C}$ at point $z$.

The above construction of the limiting normal cone using the proximal normal cone was given by Mordukhovich in [23]. The limiting normal cone is in general smaller than the Clarke normal cone, and in the case when $\mathcal{C}$ is convex, the proximal normal cone, the limiting normal cone and the Clarke normal cone coincide with the normal cone in convex analysis, i.e.,

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

For set-valued mappings, the definition of a limiting normal cone leads to the definition of Mordukhovich's coderivative which was first introduced in [24].

**Definition 2.2** (Coderivatives of a set-valued mapping) Let $\Phi : \mathbb{R}^m \to 2^{\mathbb{R}^q}$ be a set-valued mapping and $(\bar{z}, \bar{v}) \in \text{gph}\Phi$. The *coderivative* of $\Phi$ at point $(\bar{z}, \bar{v})$ is defined as

$$D^*\Phi(\bar{z}, \bar{v})(\eta) := \left\{\zeta \in \mathbb{R}^m : (\zeta, -\eta) \in \mathcal{N}_{\text{gph}\,\Phi}(\bar{z}, \bar{v})\right\}.$$

By convention, for $(\bar{z}, \bar{v}) \notin \text{gph}\Phi$, $D^*\Phi(\bar{z}, \bar{v})(\eta) = \emptyset$.

A particularly interesting case relevant to our discussion later on is when $\Phi(z) = \mathcal{N}_\mathcal{C}(z)$ and $\mathcal{C}$ is a closed subset of $\mathbb{R}^m$. By the definition of coderivatives,

$$\zeta \in D^*\mathcal{N}_\mathcal{C}(\bar{z}, \bar{v})(\eta) \Longleftrightarrow (\zeta, -\eta) \in \mathcal{N}_{\text{gph}\,\mathcal{N}_\mathcal{C}}(\bar{z}, \bar{v}).$$

The calculation of the coderivative $D^*\Phi(\bar{z}, \bar{v})(\eta)$ depends on that of the limiting normal cone to the graph of the normal cone, denoted by $\mathcal{N}_{\text{gph}\,\mathcal{N}_\mathcal{C}}(\bar{z}, \bar{v})$. In the case when $\mathcal{C} = \mathbb{R}^m_+$, the following is well-known. The proof follows easily from the formula for the proximal normal cone in [44, Proposition 2.7] and the definition of the limiting normal cones.

**Proposition 2.3** *For any* $(\bar{z}, -\bar{v}) \in gph\, \mathcal{N}_{\mathbb{R}^m_+}$, *let*

$$L := L(\bar{z}, \bar{v}) := \{i \in \{1, 2, \ldots, m\} : \bar{z}_i > 0, \bar{v}_i = 0\},$$
$$I_+ := I_+(\bar{z}, \bar{v}) := \{i \in \{1, 2, \ldots, m\} : \bar{z}_i = 0, \bar{v}_i > 0\},$$
$$I_0 := I_0(\bar{z}, \bar{v}) := \{i \in \{1, 2, \ldots, m\} : \bar{z}_i = 0, \bar{v}_i = 0\}.$$

*Then*

$$\mathcal{N}_{gph\, \mathcal{N}_{\mathbb{R}^m_+}}(\bar{z}, -\bar{v}) = \{(\alpha, -\beta) \in \mathbb{R}^{2m} : \alpha_L = 0, \beta_{I_+} = 0, \quad \forall i \in I_0,\ \textit{either}\ \alpha_i < 0, \beta_i < 0\ \textit{or}\ \alpha_i\beta_i = 0\}.$$

For recent results on calculating the normal cone to the graph of a standard normal cone (coderivative of the standard normal cone mapping), see [7, 8] and [9, Section 3].

**Definition 2.4** (Subdifferentials) Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a lower semicontinuous function and finite at $x \in \mathbb{R}^n$. The proximal subdifferential of $f$ at $x$ is defined to be the set

$$\partial^\pi f(x) := \{\zeta \in \mathbb{R}^n : \exists \sigma > 0, \delta > 0 \text{ such that } f(y) \geq f(x) + \zeta^T (y - x) - \sigma \|y - x\|^2 \quad \forall y \in B(x, \delta)\} \tag{7}$$

and the limiting (Mordukhovich or basic [25]) subdifferential of $f$ at $x$ to be the set

$$\partial f(x) = \limsup_{x' \xrightarrow{f} x} \partial^\pi f(x'),$$

where $x' \xrightarrow{f} x$ signifies that $x'$ and $f(x')$ converge to $x$ and $f(x)$ respectively. When $f$ is Lipschitz continuous near $x$, the Clarke subdifferential (or generalized gradient, see p. 27 in [6]) of $f$ at $x$ is equal to conv $\partial f(x)$.

The limiting subdifferential is in general smaller than the Clarke subdifferential, and in the case when $f$ is convex and locally Lipschitz, the proximal subdifferential, the limiting subdifferential and the Clarke subdifferential coincide with the subdifferential in the sense of convex analysis [32]. In the case when $f$ is continuously differentiable, these subdifferentials reduce to classical gradient, denoted by $\nabla f(x)$, i.e., $\partial f(x) = \{\nabla f(x)\}$.

### 2.2 Outer Semicontinuity of Coderivative Mappings

The notion of the coderivative of a set-valued mapping is well studied and documented [25]. In this paper, we need some properties, namely outer semicontinuity, of a coderivative as a set-valued mapping for the convergence analysis. These properties are easy to derive but apparently not available in the literature. We give details of them in this subsection.

Let $\mathcal{C}$ be a closed subset of $\mathbb{R}^n$. Recall that a set-valued mapping $\Phi : \mathcal{C} \to 2^{\mathbb{R}^m}$ is said to be *closed* at $\bar{x}$ if for $x_k \subset \mathcal{C}$, $x_k \to \bar{x}$, $y_k \in \Phi(x_k)$ and $y_k \to \bar{y}$ implies $\bar{y} \in \Phi(\bar{x})$. If $\mathcal{C} = \mathbb{R}^n$, then $\Phi : \mathbb{R}^n \to 2^{\mathbb{R}^m}$ is closed if and only if its graph is a closed set.

**Definition 2.5** A set-valued mapping $\Phi : \mathbb{R}^n \to 2^{\mathbb{R}^m}$ is said to be *outer semicontinuous (osc)* at $\bar{x} \in X$ relative to $X \subset \mathbb{R}^n$ if $\limsup_{x \in X, x \to \bar{x}} \Phi(x) \subset \Phi(\bar{x})$ or equivalently $\limsup_{x \in X, x \to \bar{x}} \Phi(x) = \Phi(\bar{x})$. $\Phi$ is said to be osc at $\bar{x}$ if $X = \mathbb{R}^n$ in the above definition.

**Proposition 2.6** [33, Proposition 5.12] *Let $\Phi : \mathbb{R}^n \to 2^{\mathbb{R}^m}$ be a closed set-valued mapping and $\bar{x} \in X \subset \mathbb{R}^n$. Then $\Phi$ is osc at $\bar{x}$ relative to $X$ if and only if for every $\epsilon > 0$ and every $\rho > 0$ there exists a neighborhood $V$ of $\bar{x}$ such that*

$$\Phi(x) \cap \rho\mathcal{B} \subset \Phi(\bar{x}) + \epsilon\mathcal{B}$$

*for all $x \in V \cap X$.*

In what follows, we need to demonstrate that the coderivative of the normal cone of a closed set-valued mapping is closed and outer semicontinuous.

**Proposition 2.7** *Let $\Phi : \mathbb{R}^m \to 2^{\mathbb{R}^q}$ be a set-valued mapping with a closed graph and $\mathcal{C}$ be a closed subset of $\mathbb{R}^n$. Then the following statements hold.*

(i) *The graph of the coderative mapping $D^*\Phi(\cdot, \cdot)(\cdot)$ is closed.*
(ii) *The graph of the coderivative mapping $D^*\mathcal{N}_{\mathcal{C}}(\cdot, \cdot)(\cdot)$ is closed.*

*Proof*

Part (i) Suppose that $\zeta_k \to \zeta$ and $\eta_k \to \eta$, where $\zeta_k \in D^*\Phi(z_k, v_k)(\eta_k)$ and

$$(z_k, v_k) \in \text{gph } \Phi, \quad (z_k, v_k) \to (z, v).$$

By the definition of coderivative,

$$\zeta_k \in D^*\Phi(z_k, v_k)(\eta_k) \Longleftrightarrow (\xi_k, -\eta_k) \in \mathcal{N}_{\text{gph } \Phi}(z_k, v_k).$$

Since the normal cone mapping has a closed graph, by Proposition 6.6 in [33] or discussions in [25, p. 11] $(\zeta, -\eta) \in \mathcal{N}_{\text{gph } \Phi}(z, v)$, which implies $\zeta \in D^*\Phi(z, v)(\eta)$. This demonstrates the closedness of the coderivative $D^*\Phi(\cdot, \cdot)(\cdot)$.

Part (ii) Since $\mathcal{C}$ is closed, by Proposition 6.6 in [33] or discussions in [25, p. 11], the graph of the normal cone mapping gph $\mathcal{N}_{\mathcal{C}}(\cdot)$ is closed. The conclusion follows by applying part (i) to $\Phi := \mathcal{N}_{\mathcal{C}}$. □

With the forgoing closedness property, we are able to derive the outer semicontinuity of a coderivative mapping.

**Proposition 2.8** *Let $\Phi : \mathbb{R}^m \to 2^{\mathbb{R}^q}$ be a set-valued mapping with a closed graph and $\mathcal{C}$ be a closed subset of $\mathbb{R}^n$. Then the following statements hold.*

(i) *$D^*\Phi(\cdot, \cdot)(\cdot)$ is osc on $\mathbb{R}^n$.*

(ii) *Let* $(z, v) \in gph\, \Phi$ *be fixed and* $\eta \in \mathbb{R}^q$. *Then for every* $\rho > 0$ *and* $\epsilon > 0$, *there exists a neighborhood* $V$ *of* $(z, v, \eta)$ *such that*

$$D^*\Phi(z', v')(\eta') \cap \rho\mathcal{B} \subset D^*\Phi(z, v)(\eta) + \epsilon\mathcal{B},$$

*for all* $(z', v', \eta') \in V$, *where* $\mathcal{B}$ *denotes the closed unit ball in* $\mathbb{R}^q$.

(iii) *Let* $(z, v) \in gph\, \mathcal{N}_C$ *be fixed. Then for any* $\rho > 0$ *and* $\epsilon > 0$, *there exists a neighborhood* $V$ *of* $(z, v, \eta)$ *such that*

$$D^*\mathcal{N}_C(z', v')(\eta') \cap \rho\mathcal{B} \subset D^*\mathcal{N}_C(z, v)(\eta) + \epsilon\mathcal{B},$$

*for all* $(z', v', \eta') \in V$, *where* $\mathcal{B}$ *denotes the closed unit ball in* $\mathbb{R}^n$.

*Proof* Part (i). By Proposition 2.7, the graph of co-derivative mapping $D^*\Phi(\cdot, \cdot)(\cdot)$ is closed. Through [33, Theorem 5.7 (a)], this is equivalent to the outer semicontinuity of $D^*\Phi(\cdot, \cdot)(\cdot)$. Part (ii) follows from Proposition 2.6. Part (iii) is no more than an application of Part (ii) to $\Phi(z) := \mathcal{N}_C(z)$. □

2.3 Expectation of Random Set-Valued Mappings

Let $\mathcal{X}$ be a closed subset of $\mathbb{R}^n$ and $S(\cdot, \cdot) : \mathcal{X} \times \mathbb{R}^d \to 2^{\mathbb{R}^m}$ be a set-valued mapping. and $\xi$ be a random vector defined on a probability space $(\Omega, \mathcal{F}, P)$ and taking values in $\mathbb{R}^d$. Let $x \in \mathcal{X}$ be fixed and consider the measurability of the set-valued mapping $S(x, \xi(\cdot)) : \Omega \to 2^{\mathbb{R}^m}$. Let $\mathfrak{B}$ denote the space of nonempty, closed subsets of $\mathbb{R}^m$. Then $S(x, \xi(\cdot))$ can be viewed as a single valued mapping from $\Omega$ to $\mathfrak{B}$. Using [33, Theorem 14.4], we know that $S(x, \xi(\cdot))$ is measurable if and only if for every $B \in \mathfrak{B}$, $S(x, \xi(\cdot))^{-1}B$ is $\mathcal{F}$-measurable.

Recall that $A(x, \xi(\omega)) \in S(x, \xi(\omega))$ is said to be a *measurable selection* of the random set $S(x, \xi(\omega))$, if $A(x, \xi(\omega))$ is measurable. Measurable selections exist, see [2] and references therein. The *expectation of* $S(x, \xi(\omega))$, denoted by $\mathbb{E}[S(x, \xi(\omega))]$, is defined as the collection of $\mathbb{E}[A(x, \xi(\omega))]$, where $A(x, \xi(\omega))$ is an integrable measurable selection of $S(x, \xi(\omega))$. The expected value is also known as Aumann's integral [10] as it was first studied comprehensively by Aumann in [3]. $\mathbb{E}[S(x, \xi(\omega))]$ is regarded as well defined if $\mathbb{E}[S(x, \xi(\omega))] \in \mathfrak{B}$ is nonempty. A sufficient condition of this is $\mathbb{E}[\|S(x, \xi(\omega))\|] := \mathbb{E}[\mathbb{H}(0, S(x, \xi(\omega)))] < \infty$, see [2, 22]. In such a case, $S$ is said to be *integrably bounded* [3, 10].

## 3 Optimality Conditions

In this section, we investigate the first order necessary optimality conditions for both the true problem (Eq. 1) and the SAA problem (Eq. 6) in terms of the limiting subdifferentials and the coderivatives of the underlying functions under standard constraint qualifications.

3.1 First Order Necessary Conditions of the True Problem

Our aim here is to derive the first order optimality conditions for the true problem and the SAA problem. To proceed the discussion, we need to make some standard assumptions on the underlying functions and constraint qualifications.

**Assumption 3.1** *Let $f(x, y, \xi)$ and $F(x, y, \xi)$ be defined as in Eq. 1.*

(a) $\mathbb{E}[f(x, y, \xi)]$ *and* $\mathbb{E}[F(x, y, \xi)]$ *are well-defined for every* $x \in \mathbb{R}^n$ *and* $y \in \mathbb{R}^m$;
(b) *there exists a positive function* $\kappa(\xi)$ *such that*

$$\|f(x', y', \xi) - f(x, y, \xi)\| \leq \kappa(\xi)(\|x' - x\| + \|y' - y\|), \ \forall x', x \in \mathbb{R}^n, \ \ y', y \in \mathbb{R}^m$$

*for almost every* $\xi$, *where* $\mathbb{E}[\kappa(\xi)] < \infty$;
(c) $F$ *is continuously differentiable w.r.t.* $x, y$ *for almost every* $\xi$ *and* $\nabla F(x, y, \xi)$ *is integrably bounded, that is, there exists a positive function* $\kappa(\xi)$ *such that* $\|\nabla F(x, y, \xi)\| \leq \kappa(\xi)$ *for almost every* $\xi$, *where* $\mathbb{E}[\kappa(\xi)] < \infty$, *where* $\nabla F(x, y, \xi)$ *denotes the Jacobian of* $F$ *with respect to* $(x, y)$.

Under Assumption 3.1 (a) and (b), the Aumann's integral $\mathbb{E}[\partial f(x, y, \xi)]$, where $\partial f(x, y, \xi)$ denotes the limiting subdifferential of $f$ with respect to $(x, y)$, is well-defined because the limiting subdifferential is contained in the Clarke subdifferential and the latter is well-defined [43]. Moreover, under Assumption 3.1 (b) and (c), $\mathbb{E}[F(x, y, \xi)]$ is continuously differentiable, and $\nabla\mathbb{E}[F(x, y, \xi)] = \mathbb{E}[\nabla F(x, y, \xi)]$.

We now consider constraint qualifications. A well-known constraint qualification in the literature of deterministic MPECs is the so-called no nonzero abnormal multiplier constraint qualification (NNAMCQ) introduced by Ye et al. [47]. In a number of papers such as [45, 47, 48], the first order necessary optimality conditions are derived under NNAMCQ for deterministic optimization problems with variational inequaity constraints where, in our context, $Y$ needs to be a closed convex set. It is easy to observe, however, that these results on optimality still hold if $\mathcal{N}_Y$ is replaced by any closed set-valued mapping, in other words, $Y$ is not necessarily convex. This motivates us to use NNAMCQ to derive the first order optimality conditions of Eq. 1 and its sample average approximation.

**Definition 3.2** We call $\eta \in \mathbb{R}^m$ a vector of *abnormal multipliers* of true problem (Eq. 1) at $(x, y)$ if it satisfies

$$0 \in \mathbb{E}[\nabla F(x, y, \xi)]^T\eta + \{0_n\} \times D^*\mathcal{N}_Y(y, -\mathbb{E}[F(x, y, \xi)])(\eta) + \mathcal{N}_C(x, y).$$

We say that the no nonzero abnormal multiplier constraint qualification (NNAMCQ) for problem 1 holds at a feasible solution $(x, y)$ if there does not exist a nonzero abnormal multiplier at $(x, y)$. Here and later on the differential operators "$\partial$" and "$\nabla$" are with respect to $(x, y)$ and $0_n$ denotes the zero vector in $\mathbb{R}^n$.

Note that by virtue of [45, Theorem 4.4], NNAMCQ is a sufficient condition for the pseudo-Lipschitz continuity (also called the Aubin property) of the set of solutions to the perturbed generalized equation defined as follows:

$$\mathcal{F}(q) := \{(x, y) \in C : q \in \mathbb{E}[F(x, y, \xi)] + \mathcal{N}_Y(y)\}.$$

The condition is not necessary under general circumstance. To see this, consider the case when $F$ is merely locally Lipschitz (not continuously differentiable) and there is no abstract constraint $(x, y) \in C$ (i.e. $\mathcal{N}_C(x, y) = \{0_{n+m}\}$). By the sum rule for coderivatives,

$$D^*(\mathbb{E}[F]+\mathcal{N}_Y)((\bar{x}, \bar{y}), 0)(\eta) \subseteq \partial\mathbb{E}[F(\bar{x}, \bar{y}, \xi)]^T\eta + \{0\} \times D^*\mathcal{N}_Y(\bar{y}, -\mathbb{E}[F(\bar{x}, \bar{y}, \xi)])(\eta). \tag{8}$$

The Aubin property of $\Sigma(q)$ requires

$$0 \in D^*(\mathbb{E}[F] + \mathcal{N}_Y)((\bar{x}, \bar{y}), 0)(\eta) \Longrightarrow \eta = 0.$$

In general this does not imply NNAMCQ because the inclusion in Eq. 8 may be strict. Similar arguments apply to the case when there is an abstract constraint $(x, y) \in C$.

**Theorem 3.3** (First order optimality condition of the true problem) *Let Assumption* 3.1 *hold and* $(x, y)$ *be a local optimal solution*[1] *to the true problem (Eq.* 1*). Suppose that NNAMCQ for problem (Eq.* 1*) holds at* $(x, y)$*. Then there exists a vector of multipliers* $\eta$ *such that the following first order optimality conditions hold:*

$$0 \in \mathbb{E}[\partial f(x, y, \xi)] + \mathbb{E}[\nabla F(x, y, \xi)]^T\eta + \{0_n\} \times D^*\mathcal{N}_Y(y, -\mathbb{E}[F(x, y, \xi)])(\eta) + \mathcal{N}_C(x, y). \tag{9}$$

*Proof* Under Assumption 3.1 (a) and (b), $\mathbb{E}[f(x, y, \xi)]$ is Lipschitz continuous w.r.t. $x$ and $y$, hence the limiting subdifferential $\partial\mathbb{E}[f(x, y, \xi)]$ is well defined. Moreover, it follows from Theorem 2.9 (ii) in [43] that $\mathbb{E}[\partial f(x, y, \xi)]$ is also well-defined. Under the NNAMCQ, we can use [45, Corollary 4.8(5)] to show that there exists a vector of multipliers $\eta$, such that

$$0 \in \partial\mathbb{E}[f(x, y, \xi)] + \nabla\mathbb{E}[F(x, y, \xi)]^T\eta + \{0_n\} \times \nabla D^*\mathcal{N}_Y(y, -\mathbb{E}[F(x, y, \xi)])(\eta) + \mathcal{N}_C(x, y).$$

The rest follows from Theorem 2.9 (ii) in [43], that is, $\partial\mathbb{E}[f(x, y, \xi)] \subseteq \mathbb{E}[\partial f(x, y, \xi)]$. □

For the convergence analysis in Section 4, we will require a slightly weaker version of the first order optimality conditions than Eq. 9 defined as follows:

$$0 \in \mathbb{E}[\text{conv}\, \partial f(x, y, \xi)] + \mathbb{E}[\nabla F(x, y, \xi)]^T\eta + \{0_n\} \times D^*\mathcal{N}_Y(y, -\mathbb{E}[F(x, y, \xi)])(\eta) + \mathcal{N}_C(x, y). \tag{10}$$

Here 'weak' is in the sense that a stationary point defined by Eq. 9 is a stationary point defined by Eq. 10, but not vice versa in general. This is because

$$\partial\mathbb{E}[f(x, y, \xi)] \subset \mathbb{E}[\partial f(x, y, \xi)] \subset \mathbb{E}[\text{conv}\, \partial f(x, y, \xi)]$$

and the equality holds in the first inclusion when $f$ is Clarke regular at $x$ for almost every $\xi$ and the equality holds in the second inclusion when the probability

[1] For the simplicity of discussion, we assume throughout this paper that the true problem (Eq. 1) has an optimal solution. Sufficient conditions for the existence of optimal solutions are covered by those in the deterministic case (see e.g. [28, Proposition 1.1]) because the expected values of the underlying random functions are deterministic.

space $(\Omega, \mathcal{F}, P)$ is non-atomic or the atoms are convex. The latter is known as *Aumann's identity*, see Aumann's pioneering work [3] and a comprehensive review of Aumann's integral by Hess [10].

### 3.2 First Order Necessary Conditions of the SAA Problem

We now move on to discuss the first order optimality conditions of the SAA problem (Eq. 6). Let us start by looking at the feasible set of the SAA problem (Eq. 6). For the feasibility result in this section, we assume that $C = X \times \mathbb{R}^m$, where $X$ is a closed subset of $\mathbb{R}^n$. Let $\mathcal{G}(x)$ denote the set of solutions of the parametric generalized equation

$$0 \in \mathbb{E}[F(x, y, \xi)] + \mathcal{N}_Y(y)$$

for give $x \in X$ and $\mathcal{F} = \{x \in X : \mathcal{G}(x) \neq \emptyset\}$. Let

$$\hat{F}^N(x, y) := \frac{1}{N} \sum_{k=1}^{N} F(x, y, \xi^k) \tag{11}$$

and $\mathcal{G}^N(x)$ denote the set of solutions of the parametric generalized equation

$$0 \in \hat{F}_N(x, y) + \mathcal{N}_Y(y) \tag{12}$$

and $\mathcal{F}^N = \{x \in X : \mathcal{G}^N(x) \neq \emptyset\}$. Assume that $\mathcal{F}$ is nonempty, that is, the feasible set of the true problem is nonempty. We discuss the conditions under which the feasible set $\mathcal{F}^N$ is nonempty.

**Definition 3.4** A set-valued mapping $\Phi : \mathbb{R}^m \to 2^{\mathbb{R}^m}$ is said to be *subinvertible* at $(y^*, 0)$, if one has $0 \in \Phi(y^*)$ and there exists a compact convex neighborhood $U$ of $y^*$ in $\mathbb{R}^m$, a positive constant $\epsilon > 0$, and a nonempty convex-valued mapping $G : \epsilon B \rightrightarrows U \subset \mathbb{R}^m$ such that the graph of $G$, denoted by gph $G$, is closed, the point $y^*$ belongs $G(0)$, and $G(z)$ is contained in $\Phi^{-1}(z)$ for all $z \in \epsilon B$.

An instance of $\Phi(y)$ being subinvertible at $(y^*, 0)$ is that there exists a continuous selection $g(z)$ of $\Phi^{-1}(z)$ on a compact neighborhood of 0 such that $g(0) = y^*$.

The concept of subinvertibility was proposed by King and Rockafellar [11] for the study of existence of a perturbed generalized equation. Here we use the notion to derive the nonemptiness of the feasible set $\mathcal{F}^N$.

**Proposition 3.5** *Let*

$$\Phi(x, y) := \mathbb{E}[F(x, y, \xi)] + \mathcal{N}_Y(y)$$

*and $x \in X$. Assume: (a) Assumption 3.1 holds; (b) for every $x \in X$, there exists $y^* \in \mathcal{G}(x)$ such that $\Phi(x, \cdot)$ is subinvertible at $(y^*, 0)$ in $\mathbb{R}^m$. Then*

(i) *w.p.1 $\mathcal{F}^N$ is nonempty for $N$ sufficiently large;*
(ii) *for any scalar $r > 0$ such that $r\mathcal{B} \cap \mathcal{F} \neq \emptyset$, one has*

$$\lim_{N \to \infty} \mathcal{F}^N \cap r\mathcal{B} = \mathcal{F} \cap r\mathcal{B}$$

*w.p.1.*

*Proof*

Part (i) Let $x \in X$. Under the subinvertibility condition, there exists $y^* \in \mathcal{G}(x)$ such that $\Phi(x, \cdot)$ is subinvertible at $(y^*, 0)$ in $\mathbb{R}^m$. It follows from [11, Proposition 3.1] that there exists a compact neighborhood $U$ of $y^*$ and a small positive number $\epsilon$ such that when

$$\sup_{y \in U} \left\| \mathbb{E}[F(x, y, \xi)] - \hat{F}_N(x, y) \right\| \leq \epsilon, \tag{13}$$

$\mathcal{G}^N(x) \cap U \neq \emptyset$. On the other hand, for the given $\epsilon$ and under condition (a), the classical law of large numbers (see e.g. [34, Lemma A]) ensures that Eq. 13 holds w.p.1 for $N$ sufficiently large. The conclusion follows.

Part (ii) Equation 13 holds if $U$ is replaced by any compact subset of $Y$. The conclusion then follows from Lemma 4.2 in [41]. □

*Remark 3.6* To see how strong the subinvertibility condition is in Proposition 3.5, let us consider an example where $\Phi$ is derived from the first order optimality conditions of the following one stage stochastic minimization problem:

$$\begin{aligned} &\min_y \ \mathbb{E}[h(x, y, \xi)] \\ &\text{s.t. } y \in \mathcal{C}, \\ &\qquad \mathbb{E}[g(x, y, \xi)] \in \mathcal{K}^o, \end{aligned} \tag{14}$$

where $\mathcal{C}$ is a polyhedral set in $\mathbb{R}^m$ and $\mathcal{K}^o$ is the polar of a polyhedral cone $\mathcal{K}$ in $\mathbb{R}^s$. The problem is varied from a parametric nonlinear programming problem considered by King and Rockafellar in [11] and earlier by Robinson in [31]. We assume that for each fixed $x \in \mathbb{R}^n$ and $\xi \in \mathbb{R}^d$, $h(x, \cdot, \xi)$ and $g(x, \cdot, \xi)$ are twice continuously differentiable in a neighborhood of a considered point $y^* \in \mathcal{C}$ and their first and second order derivatives are integrably bounded. Let $\mu \in \mathbb{R}^s$ and denote by

$$F(x, y, \mu, \xi) := (\nabla_y h(x, y, \xi) + \nabla_y g(x, y, \xi)^T \mu, -g(x, y, \xi)).$$

Then the first order optimality conditions of Eq. 14 can be written as

$$0 \in \Phi(x, y, \mu) := \mathbb{E}[F(x, y, \mu, \xi)] + \mathcal{N}_{\mathcal{C} \times \mathcal{K}}(y, \mu). \tag{15}$$

Let $(y^*, \mu^*)$ be a stationary pair which satisfies Eq. 15. Assume that Eq. 15 satisfies Robinson's regularity condition, that is,

$$0 \in \text{int}\, \{\mathbb{E}[g(x, y^*, \xi) + \nabla_y g(x, y^*, \xi)][\mathcal{C} - y^*] - \mathcal{K}^o\},$$

where "int" denotes the interior of a set, and the second order sufficient condition is satisfied at $(y^*, \mu^*)$, namely, for all $u \in T_{\mathcal{C}}(y^*)$ with $u$ nonzero and

$$\mathbb{E}[\nabla_y g(x, y^*, \xi)]u \in T_{\mathcal{K}^o}(\mathbb{E}[g(x, y^*, \xi)]) \text{ and } \mathbb{E}[\nabla_y h(x, y^*, \xi)]u = 0$$

one has

$$u^T \nabla_y^2 L(x, y^*, \mu^*) u > 0,$$

where $T_{\mathcal{C}}$ denotes the tangent cone in convex analysis and

$$L(x, y, \mu) := \mathbb{E}[h(x, y, \xi(\omega))] + \mathbb{E}[g(x, y, \xi(\omega))]^T \mu$$

denotes the Lagrangian function. It follows from [11, Proposition 7.1] that $\Phi(x, y, \mu)$ is subinvertible at $((y^*, \mu^*), 0)$.

In what follows, we derive the first order necessary conditions for the SAA problem (Eq. 6).

**Proposition 3.7** *Suppose that for each $N$, $(x^N, y^N, \eta^N)$ is a solution to the following system:*

$$0 \in \nabla \hat{F}^N(x^N, y^N)^T\eta + \{0_n\} \times D^*\mathcal{N}_Y\left(y^N, -\hat{F}^N(x^N, y^N)\right)(\eta) + \mathcal{N}_C(x^N, y^N), \quad (16)$$

*Suppose also that $(x, y)$ is a cluster point of $\{(x^N, y^N)\}$, Assumption 3.1 is satisfied and NNAMCQ holds at $(x, y)$. Then for $N$ sufficiently large, $\eta^N = 0$ w.p.1.*

We delay the proof of this proposition to the appendix as it requires some technical results related to graphical convergence which will be detailed in Section 4.

Following Proposition 3.7, we can derive, for $N$ sufficiently large, the following first order optimality conditions of the SAA problem (Eq. 6) hold with probability one.

**Theorem 3.8** (First order optimality condition of the SAA problem) *Let $(x^N, y^N)$ be a local optimal solution to the SAA problem (Eq. 6) and $(x, y)$ be a cluster point of $\{(x^N, y^N)\}$. Suppose that Assumption 3.1 is satisfied and NNAMCQ holds at $(x, y)$. Then there exists a vector of multipliers, denoted by $\eta^N$, such that $(x^N, y^N, \eta^N)$ satisfies the following first order optimality conditions w.p.1:*

$$0 \in \mathcal{A}^N(x, y) + \nabla \hat{F}^N(x, y)^T\eta + \{0_n\} \times D^*\mathcal{N}_Y\left(y, -\hat{F}^N(x, y)\right)(\eta) + \mathcal{N}_C(x, y), \quad (17)$$

*where*

$$\mathcal{A}^N(x, y) = \frac{1}{N}\sum_{k=1}^{N} \partial f(x, y, \xi^k).$$

*Moreover the sequence $\{\eta^N\}$ must be bounded. We call a solution $(x^N, y^N, \eta^N)$ to the above first order optimality condition a* stationary pair *of the SAA problem (Eq. 6).*

Similar to Proposition 3.7, the proof of this theorem requires some technical details of the graphical convergence to be detailed in Section 4, therefore we delay it to the Appendix.

## 4 Convergence Analysis

In this section, we analyze the convergence of the stationary point $(x^N, y^N)$ defined by the first order optimality condition (Eq. 17) as sample size $N$ increases. For the simplicity of notation, let

$$\begin{aligned}\mathcal{H}(x, y, \eta) := \mathbb{E}[\mathrm{conv}\partial f(x, y, \xi)] + \mathbb{E}[\nabla F(x, y, \xi)]^T\eta \\ +\{0_n\} \times D^*\mathcal{N}_Y(y, -\mathbb{E}[F(x, y, \xi)])(\eta) + \mathcal{N}_C(x, y) \quad (18)\end{aligned}$$

and

$$\mathcal{H}^N(x, y, \eta) := \mathcal{A}^N(x, y) + \nabla \hat{F}^N(x, y)^T \eta$$
$$+\{0_n\} \times D^*\mathcal{N}_Y\left(y, -\hat{F}^N(x, y)\right)(\eta) + \mathcal{N}_C(x, y). \tag{19}$$

The first order optimality conditions (Eqs. 10 and 17) can be written as

$$0 \in \mathcal{H}(x, y, \eta) \tag{20}$$

and

$$0 \in \mathcal{H}^N(x, y, \eta) \tag{21}$$

respectively. Let $(x^*, y^*, \eta^*)$ be a cluster point of sequence $\{x^N, y^N, \eta^N\}$ w.p.1. Our aim here is to show that $0 \in \mathcal{H}(x^*, y^*, \eta^*)$ w.p.1. A well-known sufficient condition is that $\mathcal{H}^N$ converges uniformly to $\mathcal{H}$ over a compact set which contains the sequence $\{(x^N, y^N, \eta^N)\}$ w.p.1 and the considered cluster point $(x^*, y^*, \eta^*)$. Unfortunately this kind of uniform convergence is difficult to establish in that the underlying set-valued mappings are unbounded.

To deal with the challenge, we resort to the concept of graphical convergence of set-valued mappings. In this context, if $\mathcal{H}^N$ converges to $\mathcal{H}$ graphically at the considered cluster points, then the cluster points satisfy the first order optimality conditions (Eq. 10) of the true problem. This kind of fundamental result belongs to Rockfellar and Wets who dealt with a general deterministic set-valued mapping $\mathcal{H}^N$, see [33, Theorem 5.37]. More recently, Wets and Xu [40] extended the result to the case when $\mathcal{H}^N$ is the sample average of an unbounded random set-valued mapping, see [40, Theorem 5.7]. Here we use [33, Theorem 5.37] instead of [40, Theorem 5.7] in that $\mathcal{H}^N$ is not the sample average of a random set-valued mapping. To this end, we recall some basic definitions and results related to graphical convergence. The materials are taken from Rockafellar and Wets' book [33, Chapter 5].

**Definition 4.1** For a sequence of set-valued mappings $S^\nu : \mathbb{R}^n \to 2^{\mathbb{R}^m}$, the *pointwise outer limit* and *pointwise inner limit* are the mappings $(p - \limsup_\nu S^\nu)(x)$ and $(p - \liminf_\nu S^\nu)(x)$ defined at each point $x$ by

$$(p - \limsup_\nu S^\nu)(x) := \limsup_\nu S^\nu(x),$$

and

$$(p - \liminf_\nu S^\nu)(x) := \liminf_\nu S^\nu(x)$$

respectively. When the pointwise outer and inner limits agree, the pointwise limit $(p - \liminf_\nu S^\nu)(x)$ is said to exist and it is denoted by $\lim_\nu S^\nu(x)$.

**Definition 4.2** For a sequence of set-valued mappings $S^\nu : \mathbb{R}^n \to 2^{\mathbb{R}^m}$, the *graphical outer limit*, denoted by $(g - \limsup_\nu S^\nu)$, is the mapping having as its graph of the set $\limsup_\nu(\text{gph } S^\nu)$:

$$\text{gph } (g - \limsup_\nu S^\nu) = \limsup_\nu(\text{gph } S^\nu).$$

The *graphical inner limit*, denoted by $(g - \liminf_\nu S^\nu)$, is the mapping having as its graph of the set $\liminf_\nu (gphS^\nu)$:

$$gph(g - \liminf_\nu S^\nu) = \liminf_\nu (gphS^\nu).$$

When the graphical outer and inner limits agree, the graphical limit $(g - \lim_\nu S^\nu)$ is said to exist.

**Definition 4.3** A sequence of set-valued mappings $S^\nu : \mathbb{R}^n \to 2^{\mathbb{R}^m}$ is said to be *equi-outer semicontinuous* ( equi-osc for brevity) at $\bar{x}$ relative to $X$ (a set containing $\bar{x}$) if for every $\epsilon > 0$ and $\rho > 0$, there exists a neighborhood $V$ of $\bar{x}$ such that

$$S^\nu(x) \cap \rho\mathcal{B} \subset S^\nu(\bar{x}) + \epsilon\mathcal{B}, \quad \forall x \in V \cap X$$

for all $\nu$ and *asymptotically equi-osc* if the above relationship holds for all $\nu$ sufficiently large.

**Lemma 4.4** *Let $A$, $B \subset \mathbb{R}^m$ be two closed sets and $A$ bounded. Then for any number $\rho > 0$ there exists $\rho' > 0$ such that*

$$(A + B) \cap \rho\mathcal{B} \subset A \cap \rho'\mathcal{B} + B \cap \rho'\mathcal{B}. \tag{22}$$

This is perhaps a well known elementary result, see for instance [40].

**Proposition 4.5** *Let the set-valued mappings $S^\nu : \mathbb{R}^n \to 2^{\mathbb{R}^m}$ and $T^\nu : \mathbb{R}^n \to 2^{\mathbb{R}^m}$ be asymptotically equi-osc at $\bar{x} \in \mathbb{R}^n$. Suppose that either $S^\nu(x)$ or $T^\nu(x)$ is contained in a compact subset of $\mathbb{R}^m$ for $x$ close to $\bar{x}$ when $\nu$ is sufficiently large. Then the sum of the set valued mappings $S^\nu + T^\nu$ is asymptotically equi-osc at $\bar{x}$.*

*Proof* By Lemma 4.4, for any $\rho > 0$ there exists $\rho' > 0$ such that

$$(S^\nu(x) + T^\nu(x)) \cap \rho\mathcal{B} \subset S^\nu(x) \cap \rho'\mathcal{B} + T^\nu(x) \cap \rho'\mathcal{B}$$

for $\nu$ being sufficiently large. Since $S^\nu$, $T^\nu$ are equi-osc at $\bar{x}$, for every $\varepsilon > 0$ and $\rho' > 0$, there exists $V$, a neighborhood of $\bar{x}$ such that

$$S^\nu(x) \cap \rho'\mathcal{B} \subset S^\nu(\bar{x}) + \varepsilon\mathcal{B},$$
$$T^\nu(x) \cap \rho'\mathcal{B} \subset T^\nu(\bar{x}) + \varepsilon\mathcal{B},$$

for all $x \in V$ and $\nu$ being sufficiently large. Hence it follows that

$$(S^\nu(x) + T^\nu(x)) \cap \rho\mathcal{B} \subset S^\nu(\bar{x}) + T^\nu(\bar{x}) + 2\varepsilon\mathcal{B}$$

for all $x \in V$ and $\nu$ being sufficiently large. □

**Proposition 4.6** *Let $S^\nu : \mathbb{R}^n \to 2^{\mathbb{R}^m}$, $T : \mathbb{R}^m \to 2^{\mathbb{R}^p}$ be set-valued mappings. Let $x \in \mathbb{R}^n$ be fixed. Suppose: (a) $S^\nu$ is pointwise convergent at $x$ and there exists a neighborhood $V$ of $x$ and $\rho > 0$ such that $S^\nu(V) \subset \rho\mathcal{B}$ for $\nu$ sufficiently large; (b) $T$*

*is osc in* $\rho\mathcal{B}$. *Then the sequence of the composite mappings* $\{T \circ S^\nu\}$ *is asymptotically equi-osc at* $x \in \mathbb{R}^n$ *if one of the following conditions holds:*

(c) $S^\nu$ *is graphical-osc at* $x$ *and* $T$ *is uniformly osc at* $S(x)$ *in the sense that for any* $\epsilon > 0$ *and* $\rho' > 0$, *there exists* $\gamma > 0$ *such that*

$$T(y + \gamma\mathcal{B}) \cap \rho'\mathcal{B} \subset T(y) + \epsilon\mathcal{B}$$

*for every* $y \in S(x)$, *where* $S(x) = \limsup_{\nu\to\infty} S^\nu(x)$;

(d) $S^\nu$ *is equi-osc at* $x$ *and* $T$ *is uniformly osc at* $S(x)$;

(e) $T$ *is uniformly osc in a neighborhood of* $S(x)$, *there is an open or closed neighborhood* $V$ *of* $x$ *such that* $S^\nu$ *is continuous in* $V$ *and uniformly convergent to* $S$ *over* $V$;

(f) *there is an open or closed neighborhood* $V$ *of* $x$ *such that* $S^\nu$ *is single valued, continuous in* $V$ *and uniformly convergent to* $S$ *over* $V$.

*Proof* By Assumption, $S^\nu$ is pointwise convergent at $x$. Therefore, (d) implies (c) via Theorem 5.40 in Rockafellar and Wets [33]. Hence it suffices to prove the conclusion for (c), (e) and (f).

Assume condition (c). By [33, Eq. 5(9), p. 168 and Theorem 4.10 (b)], for every $\gamma > 0$, there exists a $\delta > 0$ such that $x + \delta\mathcal{B} \subset V$ and under (a)

$$S^\nu(x + \delta\mathcal{B}) = S^\nu(x + \delta\mathcal{B}) \cap \rho\mathcal{B} \subset S(x) + \gamma\mathcal{B} \tag{23}$$

for all $\nu$ sufficiently large. Since $T$ is uniformly osc in a neighborhood of $S(x)$ by assumption, for any $\epsilon > 0$ and $\rho' > 0$, we can set $\gamma$ sufficiently small such that $S(x) + \gamma\mathcal{B} \subset \rho\mathcal{B}$ and

$$T(y') \cap \rho'\mathcal{B} \subset T(y) + \epsilon\mathcal{B}$$

for all $y' \in y + \gamma\mathcal{B}$ and $y \in S(x)$, which implies

$$T(S(x) + \gamma\mathcal{B}) \cap \rho'\mathcal{B} \subset T(S(x)) + \epsilon\mathcal{B}. \tag{24}$$

Combining Eqs. 23 and 24, we have that for any $\epsilon > 0$ and $\rho' > 0$, there exists $\delta > 0$ such that

$$T(S^\nu(x + \delta\mathcal{B})) \cap \rho'\mathcal{B} \subset T(S(x) + \gamma\mathcal{B}) \cap \rho'\mathcal{B} \subset T(S(x)) + \epsilon\mathcal{B},$$

which yields the equi-osc of $T \circ S^\nu$ at $x$.

Let us now assume (e). By [33, Proposition 5.46 (b)], the uniform convergence implies that $S^\nu$ converges graphically to $S$ over $V$. The rest follows from (c).

Finally assume (f). As far as $S^\nu$ is concerned, this is a special case of (e). So the graphical of $S^\nu$ to $S$ over $V$ is guaranteed. As for mapping $T$, we have removed the assumption of its uniform outer semicontinuity over set $S(x)$. However, we do not really need this assumption because $S(x)$ is single-valued and Eq. 24 holds. The proof is complete. □

*Remark 4.7* In the proof of Theorem 4.8, we only need Proposition 4.6 under conditions (a), (b) and (f) with $S^\nu$ being single valued. The other conditions (c), (d) and (e) are presented not only for the proof of the conclusion under (e) but also for general interest beyond the scope of this paper.

We are now ready to state our main convergence results.

**Theorem 4.8** *Let $\mathcal{H}^N : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \to 2^{\mathbb{R}^n}$ and $\mathcal{H} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \to 2^{\mathbb{R}^n}$ be defined as in Eqs. 19 and 18, respectively. Suppose that Assumption* 3.1 *holds and there exists $\kappa_1(\xi) > 0$ such that*

$$\|\nabla F(x', y', \xi) - \nabla F(x, y, \xi)\| \le \kappa_1(\xi)(\|x' - x\| + \|y' - y\|),\ \forall (x', y'), (x, y) \in C,$$

*where $\mathbb{E}[\kappa_1(\xi)] < \infty$. Then the following assertions hold.*

(i) *$\mathcal{H}^N$ is asymptotically equi-osc at any point $(x, y, \eta) \in C \times \mathbb{R}^m$ with $y \in \mathcal{G}(x)$, and*

$$(g - \limsup_N \mathcal{H}^N)(x, y, \eta) = (p - \limsup_N \mathcal{H}^N)(x, y, \eta) \subset \mathcal{H}(x, y, \eta). \tag{25}$$

(ii) *Let $\{(x^N, y^N)\}$ be a sequence of stationary points satisfying Eq. 17 and $(x^*, y^*)$ a cluster point of $\{(x^N, y^N)\}$ w.p.1. Assume that w.p.1 the sequence is contained in a compact subset $\mathcal{W}$ of $C$ and NNAMCQ holds at $(x^*, y^*)$. Then $(x^*, y^*)$ satisfies Eq. 10 w.p.1.*

*Proof*

Part (i) Let $(x, y, \eta) \in C \times \mathbb{R}^m$ be fixed. By Theorem 5.40 [33], it suffices to prove that

$$(p - \limsup_N \mathcal{H}^N)(x, y, \eta) \subset \mathcal{H}(x, y, \eta) \tag{26}$$

and $\mathcal{H}^N$ is equi-osc at $(x, y, \eta)$. Note that since $y \in \mathcal{G}(x)$, $(y, -\mathbb{E}[F(x, y, \xi)]) \in \text{gph}\, \mathcal{N}_Y$, which implies

$$D^*\mathcal{N}_Y(y, -\mathbb{E}[F(x, y, \xi)])(\eta) \neq \emptyset,$$

and further that $\mathcal{H}(x, y, \eta) \neq \emptyset$. For clarity, we divide the rest of the proof into four steps.

Step 1 We prove Eq. 26. Under Assumption 3.1 and Lipschitz continuity of $\nabla F$, it follows by the classical uniform law of large numbers (see e.g. [34, Sections 2.6 and 6.2]),

$$\lim_{N\to\infty} \hat{F}^N(x', y') = \mathbb{E}[F(x', y', \xi)],$$

$$\lim_{N\to\infty} \hat{F}^N(x', y')^T \eta' = \mathbb{E}[\nabla F(x', y', \xi)]^\top \eta',$$

w.p.1 uniformly for $(x', y', \eta')$ over any compact set, where $\hat{F}^N$ denotes the sample average of $F(x, \xi)$ as defined in Eq. 11. In particular, for fixed $(x, y, \eta)$, we have

$$\begin{aligned}(p - \limsup_N \nabla \hat{F}^N)(x, y)^T \eta &= \lim_{N\to\infty} \nabla \hat{F}^N(x, y)^\top \eta \\ &= \mathbb{E}[\nabla F(x, y, \xi)]^\top \eta, \end{aligned} \tag{27}$$

w.p.1. In what follows, we show

$$\lim_{N\to\infty} D^*\mathcal{N}_Y(y, -\hat{F}^N(x,y))(\eta) \subset D^*\mathcal{N}_Y(y, -\mathbb{E}[F(x,y,\xi)])(\eta). \tag{28}$$

The conclusion is trivial if $D^*\mathcal{N}_Y(y, -\hat{F}^N(x,y))(\eta) = \emptyset$ because this would imply $\mathcal{H}^N(x,y,\eta) = \emptyset$. Therefore, we assume $D^*\mathcal{N}_Y(y, -\hat{F}^N(x,y))(\eta) \neq \emptyset$. Let $\zeta^N \in D^*\mathcal{N}_Y(y, -\hat{F}^N(x,y))(\eta)$ and without loss of generality that $\zeta^N \to \zeta$. By the definition of coderivative,

$$(\zeta^N, -\eta) \in \mathcal{N}_{\text{gph}\,\mathcal{N}_Y}(y, -\hat{F}^N(x,y)).$$

Since $\hat{F}^N(x,y)$ converges to $\mathbb{E}[F(x,y,\xi)]$ w.p.1 and both $\mathcal{N}_Y$ and $\mathcal{N}_{\text{gph}\,\mathcal{N}_Y}$ have a closed graph (by virtue of Proposition 6.6 in [33] twice), then

$$(\zeta, -\eta) \in \mathcal{N}_{\text{gph}\,\mathcal{N}_Y}(y, -\mathbb{E}[F(x,y,\xi)])$$

which implies Eq. 28.

On the other hand, under Assumption 3.1, Artstein and Vitale's law of large numbers for random set-valued mappings [2] implies that for fixed $x, y$,

$$\lim_{N\to\infty} \mathcal{A}^N(x,y) = \mathbb{E}[\text{conv}\,\partial f(x,y,\xi)] \tag{29}$$

w.p.1. Combining Eqs. 27–29, we obtain Eq. 26.

Step 2 For the fixed $(x,y)$, we show that $\mathcal{A}^N(x,y)$ is equi-outer semicontinuous at $(x,y)$. This is indeed covered by a similar result in Wets and Xu [40] for the sample average of an integrably bounded, closed set-valued mapping. We omit the details.

Step 3 We show that the set-valued mapping $D^*\mathcal{N}_Y(\cdot, -\hat{F}^N(\cdot,\cdot))(\eta)$ is equi-osc at $(x,y)$. Recall that in Proposition 2.8 (iii) we proved that the set-valued mapping $D^*\mathcal{N}_Y(\cdot,\cdot)(\cdot)$ is osc. On the other hand, at Step 1, we have shown that $\hat{F}^N(x',y')$ is uniformly convergent to $\mathbb{E}[F(x',y',\xi)]$ in any compact subset of $X \times Y$. Moreover, from the proofs in Steps 1–2, we can observe that $\mathcal{A}^N(x',y')$, $\hat{F}^N(x',y')$ and $\nabla \hat{F}^N(x',y')$ may be contained in a compact set for all $(x',y')$ close to $(x,y)$ w.p.1 when $N$ is sufficiently large. All these allow us to use Proposition 4.6 (f) to claim that the set-valued mapping $D^*\mathcal{N}_Y(y', -\hat{F}^N(x',y'))(\eta')$ is equi-osc at $(x,y,\eta)$, and hence by virtue of Proposition 4.5 the asymptotic equi-outer semi-continuity of $\mathcal{H}^N$ at $(x,y,\eta)$ in that $\mathcal{H}^N$ is the sum of three set-valued mappings where two of them are contained in a compact set w.p.1 for $N$ sufficiently large.

Step 4 The asymptotic equi-outer semi-continuity of $\mathcal{H}^N$ at $(x,y,\eta)$ and Eq. 26 implies Eq. 25 through Theorem 5.40 in [33].

Part (ii) follows from the boundedness of $\eta^N$ under NNAMCQ at $(x^*, y^*)$ (through Theorem 3.8 and (i)), Eq. 26 and Theorem 5.37 in [33]. The proof is complete. □

*Remark 4.9* In the literature of stochastic programming, sample average approximation is said to use the so-called empirical probability measure for discretizing/approximating the original probability measure. That is, if we define

$$P_N := \frac{1}{N}\sum_{k=1}^{N} \mathbb{1}_{\xi^k}(\omega)$$

where $\xi^1, \cdots, \xi^N$ is an independent and identically distributed sampling of $\xi$ and $\mathbb{1}_{\xi^k}(\omega)$ denotes the probability mass at $\xi^k$, then

$$\mathbb{E}_{P_N}[f(x, y, \xi)] = \frac{1}{N}\sum_{k=1}^{N} f(x, y, \xi^k).$$

The empirical probability measure converges to $\mathbb{E}$ in distribution w.p.1 [37]. Under the condition that the Lipschitz modulus of $f(x, y, \xi)$ is bounded by a constant (independent of $\xi$), it follows from [16, Lemma 5.1] that

$$\lim_{N\to\infty} \sup_{(x,y)\in\mathcal{X}} \mathbb{H}(\partial\mathbb{E}_{P_N}[f(x, y, \xi)], \partial\mathbb{E}[f(x, y, \xi)]) = 0 \qquad \text{w.p.1}$$

for any compact set $\mathcal{X} \subset C$. This implies that if we obtain a stationary pair $(x^N, y^N, \eta^N)$ from solving Eq. 6 which satisfies

$$\begin{aligned} 0 \in \partial\left(\frac{1}{N}\sum_{k=1}^{N} f(x, y, \xi^k)\right) + \nabla \hat{F}^N(x, y)^T\eta \\ +\{0_n\} \times D^*\mathcal{N}_Y\left(y, -\hat{F}^N(x, y)\right)(\eta) + \mathcal{N}_C(x, y), \end{aligned} \tag{30}$$

then under the conditions of Theorem 4.8, we can claim that w.p.1 a cluster point $(x^*, y^*, \eta^*)$ of sequence $\{(x^N, y^N, \eta^N)\}$ satisfies the following first order optimality conditions:

$$\begin{aligned} 0 \in \partial\mathbb{E}[f(x, y, \xi)] + \nabla\mathbb{E}[F(x, y, \xi)]^T\eta + \{0_n\} \times D^*\mathcal{N}_Y(y, -\mathbb{E}[F(x, y, \xi)])(\eta) \\ +\mathcal{N}_C(x, y). \end{aligned} \tag{31}$$

Obviously first order optimality conditions (Eqs. 30 and 31) are stronger than their weak counterparts (Eqs. 17 and 9) in that

$$\partial\left(\frac{1}{N}\sum_{k=1}^{N} f(x, y, \xi^k)\right) \subset \frac{1}{N}\sum_{k=1}^{N} \partial f(x, y, \xi^k) = \mathcal{A}^N(x, y),$$

(see [25, Theorems 2.33 and 3.36]) and $\partial\mathbb{E}[f(x, y, \xi)] \subset \mathbb{E}[\partial f(x, y, \xi)]$.

## 5 Complementarity Case

In this section we focus on the complementarity case where $Y = \mathbb{R}^n_+$. In this case, problem 1 reduces to a SMPCC (Eq. 2) and the SAA problem (Eq. 6) can be written as

$$\begin{aligned} &\min \frac{1}{N} \sum_{k=1}^{N} f(x, y, \xi^k) \\ &\text{s.t.} (x, y) \in C, \\ &\quad 0 \le y \perp \hat{F}^N(x, y) \ge 0. \end{aligned} \tag{32}$$

We discuss the asymptotic convergence of M-stationary point and C-stationary point of the SAA problem.

Throughout this section, "$\partial^c$" denotes the Clarke subdifferential operator or the Clarke Jacobian.

**Definition 5.1** (Stationary points for the true problem) Let $(x, y)$ be a feasible solution of the true problem SMPCC (Eq. 2). We say that $(x, y)$ is an *M-stationary point* of the true problem (Eq. 2) if there exists $\eta, \zeta$ such that

$$\begin{cases} 0 \in \mathbb{E}[\partial^c f(x, y, \xi)] + \nabla \mathbb{E}[F(x, y, \xi)]^T \eta + (0, \zeta) + \mathcal{N}_C(x, y), \\ \zeta_L = 0, \eta_{I_+} = 0, \end{cases} \tag{33}$$

and

$$\forall i \in I_0, \text{ either } \zeta_i < 0, \eta_i < 0, \text{ or } \zeta_i \eta_i = 0, \tag{34}$$

where

$$\begin{aligned} L &:= L(x, y) := \{i : y_i > 0, \mathbb{E}[F(x, y, \xi)] = 0\}, \\ I_+ &:= I_+(x, y) := \{i : y_i = 0, \mathbb{E}[F(x, y, \xi)] > 0\}, \\ I_0 &:= I_0(x, y) := \{i : y_i = 0, \mathbb{E}[F(x, y, \xi)] = 0\}. \end{aligned}$$

$(x, y)$ is called a *C-stationary point* if condition (34) is replaced by the following:

$$\forall i \in I_0, \ \zeta_i \eta_i \ge 0.$$

**Definition 5.2** (Stationary points of the SAA problem) Let $(x, y)$ be a feasible solution of the SAA problem (Eq. 6). We say that $(x, y)$ is an M-stationary point of the SAA problem (Eq. 6) if there exist $\eta$ and $\zeta$ such that

$$\begin{cases} 0 \in \mathcal{A}^N(x, y) + \nabla \hat{F}^N(x, y)^T \eta + (0, \zeta) + \mathcal{N}_C(x, y), \\ \zeta_{L^N} = 0, \eta_{I_+^N} = 0, \end{cases} \tag{35}$$

and

$$\forall i \in I_0^N, \text{ either } \zeta_i < 0, \eta_i < 0, \text{ or } \zeta_i \eta_i = 0, \tag{36}$$

where the index sets are defined as:

$$L^N := L^N(x, y) := \{i : y_i > 0, \hat{F}_i^N(x, y) = 0\},$$
$$I_+^N := I_+^N(x, y) := \{i : y_i = 0, \hat{F}_i^N(x, y) > 0\},$$
$$I_0^N := I_0^N(x, y) := \{i : y_i = 0, \hat{F}_i^N(x, y) = 0\}.$$

$(x, y)$ is called a *C-stationary point* if condition 36 is replaced by the following:

$$\forall i \in I_0^N,\ \zeta_i \eta_i \geq 0.$$

In what follows, we analyze the convergence of M-stationary points and C-stationary point of the SAA problem (Eq. 32) as sample size $N$ increases. Note that from Eq. 10 and Theorem 3.8, we obtain through Proposition 2.3 the M-stationary condition for the true problem and the M-stationary condition for the complementarity problem (Eq. 32), respectively. The convergence results therefore follow from Theorem 4.8.

In order to study the asymptotic convergence of C-stationary points of the SAA problem, we reformulate the true problem (Eq. 2) as follows:

$$\begin{aligned} &\min_{x,y} \mathbb{E}[f(x, y, \xi(\omega))] \\ &\text{s.t. } (x, y) \in C, \\ &\qquad \min\{y, \mathbb{E}[F(x, y, \xi(\omega))]\} = 0, \end{aligned} \tag{37}$$

where min is taken componentwise. The sample average approximation of Eq. 37 can be written as:

$$\begin{aligned} &\min \frac{1}{N} \sum\nolimits_{k=1}^{N} f(x, y, \xi^k) \\ &\text{s.t.} (x, y) \in C, \\ &\qquad \min\{y, \hat{F}^N(x, y)\} = 0. \end{aligned} \tag{38}$$

To facilitate use of the notation, let $\psi(a, b) := \min(a, b)$ for real numbers $a$ and $b$,

$$\Psi(x, y) = \min\{y, \mathbb{E}[F(x, y, \xi(\omega))]\}$$

and

$$\hat{\Psi}^N(x, y) = \min\{y, \hat{F}^N(x, y)\}.$$

Since $\mathbb{E}[F]$ is continuously differentiable and $\psi(a, b)$ is globally Lipschitz continuous, $\Psi(x, y)$ is locally Lipschitz continuous. Similar argument applies to $\hat{\Psi}^N(x, y)$. This implies the Clarke generalized Jacobian of $\Psi(x, y)$ and $\hat{\Psi}^N(x, y)$ are well defined. Applying the chain rule [6, Corollary, p. 75] to $\Psi(x, y)$, we obtain

$$\partial^c \Psi(x, y) = \left\{ (D_a, D_b) \begin{pmatrix} 0_{m\times n}, I_{m\times m} \\ \nabla \mathbb{E}[F(x, y, \xi)] \end{pmatrix} : (d_i^a, d_i^b) \in \partial^c \psi(y_i, \mathbb{E}[F_i(x, y, \xi)]), \right.$$
$$\left. i = 1, \cdots, m \right\},$$

where $D_a = diag(d_1^a, \cdots, d_m^a) \in \mathbb{R}^{m\times m}$ denotes the diagonal matrix with the $(i,i)$-th entry being $d_i^a$, for $i = 1, \cdots, m$, $D_b = diag(d_1^b, \cdots, d_m^b) \in \mathbb{R}^{m\times m}$ denotes the diagonal matrix with the $(i,i)$-th entry being $d_i^b$, for $i = 1, \cdots, m$, and $I_{m\times m}$ denotes the identity matrix in $\mathbb{R}^{m\times m}$. Likewise, we have

$$\partial^c \hat{\Psi}^N(x, y) = \left\{ (D_a, D_b) \begin{pmatrix} 0_{m\times n}, I_{m\times m} \\ \nabla \hat{F}^N(x, y) \end{pmatrix} : (d_i^a, d_i^b) \in \partial^c \psi(y_i, \hat{F}_i^N(x, y)), i = 1, \cdots, m \right\}.$$

Using the Clarke generalized Jacobians, we can write down the weak first order optimality conditions of true problem (Eq. 37):

$$0 \in \mathbb{E}[\partial^c f(x, y, \xi)] + \partial^c \Psi(x, y)\eta + \mathcal{N}_C(x, y), \tag{39}$$

and first order optimality conditions for the SAA problem (Eq. 38):

$$0 \in \mathcal{A}^N f(x, y) + \partial^c \hat{\Psi}^N(x, y)\eta + \mathcal{N}_C(x, y). \tag{40}$$

It is easy to see that feasible solution $(x, y)$ satisfying Eqs. 39 and 40 is a C-stationary point of the true problem and the SAA problem, respectively.

**Theorem 5.3** *Let $\{(x^N, y^N)\}$ be a sequence of C-stationary points and $(x^*, y^*)$ a cluster point w.p.1. Suppose: (a) Assumption* 3.1 *hold, (b) w.p.1 the sequence $\{(x^N, y^N)\}$ is contained in a compact subset of $C$ and $(x^*, y^*) \in C$, (c) NNAMCQ for problem* 37 *holds at $(x^*, y^*)$. Then w.p.1 $(x^*, y^*)$ is a C-stationary point of the true problem (Eq.* 2*).*

*Proof* We use Lemma 4.2 in [41] to prove the result. Let $\mathcal{W}$ be a compact subset of $C$ such that w.p.1 $\{(x^N, y^N)\}$ is contained in $\mathcal{W}$ for $N$ sufficiently large. Under condition (b), $(x^*, y^*) \in \mathcal{W}$. Let $\eta^N$ be a corresponding Lagrangian multiplier. Under NNAMCQ at $(x^*, y^*)$, it is not difficult to prove that there exists a compact subset $\mathcal{V}$ in $\mathbb{R}^m$ such that the $\{\eta_N\} \subset \mathcal{V}$ w.p.1 for $N$ sufficiently large. To see this, we note that $\partial f(x, y, \xi)$ and $\partial F(x, y, \xi)$ are integrably bounded under Assumption 3.1 and therefore $\mathcal{A}^N f(x, y)$ and $\partial \hat{\Psi}^N(x, y)$ are uniformly bounded for $(x, y) \in \mathcal{W}$ w.p.1. If $\eta^N$ is unbounded w.p.1, then it follows from Eq. 40 that

$$0 \in \frac{1}{\|\eta^N\|}\mathcal{A}^N f(x^N, y^N) + \partial \hat{\Psi}^N(x^N, y^N)\frac{\eta^N}{\|\eta^N\|} + \mathcal{N}_C(x^N, y^N).$$

By taking a subsequence if necessary, we may assume that $(x^N, y^N) \to (x^*, y^*)$ as $N \to \infty$ w.p.1. Driving $N$ to infinity, we have from the equation above

$$0 \in \Psi(x^*, y^*)\tilde{\eta} + \mathcal{N}_C(x^*, y^*),$$

where $\tilde{\eta}$ is a nonzero unit vector. This contradicts NNAMCQ at $(x^*, y^*)$. Therefore it suffices to show that for every $\delta > 0$, there exists an integer $\bar{N} > 0$ such that

$$\sup_{(x,y,\eta)\in\mathcal{W}\times\mathcal{V}} \mathbb{D}\big(\mathcal{A}^N f(x, y) + \partial \hat{\Psi}^N(x, y)\eta, \mathbb{E}[\partial f(x, y, \xi)] + \partial \Psi(x, y)\eta\big) \le \delta \tag{41}$$

w.p.1 for $N > \bar{N}$. Under Assumption 3.1, we can easily show, similar to the proof of Theorem 4.8 at Step 2, that for any $\delta_1 > 0$,

$$\sup_{(x,y,\eta)\in\mathcal{W}\times\mathcal{V}} \mathbb{D}\big(\mathcal{A}^N f(x, y), \mathbb{E}[\partial f(x, y, \xi)]\big) \le \delta_1$$

w.p.1 for $N \geq N_0$. On the other hand, since $\hat{F}^N(x, y)$ converges to $\mathbb{E}[F(x, y, \xi)]$ uniformly and the Clarke subdifferential $\partial\psi$ is osc, for any $\delta_2 > 0$, there exists $N_1 \geq N_0$ such that

$$\sup_{(x,y,\eta)\in\mathcal{W}\times\mathcal{V}} \mathbb{D}\big(\partial\hat{\Psi}^N(x, y)\eta, \partial\Psi(x, y)\eta\big) \leq \delta_2$$

w.p.1 for $N \geq N_1$. Equation 41 follows by choosing $\delta_1, \delta_2$ such that $\delta_1 + \delta_2 \leq \delta$. □

**Acknowledgements** We are grateful to Professor Boris Mordukhovich for pointing out the closedness of co-derivative mappings defined in finite dimensional spaces and an anonymous referee for the constructive comments which significantly helped improve the presentation of the paper.

## Appendix

Proof of Proposition 3.7

Observe first that if $\eta^N$ is a solution to the system (Eq. 16), then $t\eta^N$ is also a solution for any $t > 0$ because

$$\begin{aligned} &D^*\mathcal{N}_Y\left(y^N, -\frac{1}{N}\sum_{k=1}^N F(x^N, y^N, \xi^k)\right)(t\eta^N) \\ &\quad = tD^*\mathcal{N}_Y\left(y^N, -\frac{1}{N}\sum_{k=1}^N F(x^N, y^N, \xi^k)\right)(\eta^N) \end{aligned}$$

and $t\mathcal{N}_C(x, y) = \mathcal{N}_C(x, y)$. Therefore we may normalize $\eta^N$ to length 1. Assume without loss of generality that $\{\eta^N\} \to \hat{\eta}$ as $N \to \infty$. Let

$$\mathcal{H}^N(x, y, \eta) := \nabla\hat{F}^N(x, y)^T\eta + \{0_n\} \times D^*\mathcal{N}_Y\left(y, -\hat{F}^N(x, y)\right)(\eta) + \mathcal{N}_C(x, y)$$

and

$$\mathcal{H}(x, y, \eta) := \mathbb{E}[\nabla F(x, y, \xi)]^T\eta + \{0_n\} \times D^*\mathcal{N}_Y(y, -\mathbb{E}[F(x, y, \xi)])(\eta) + \mathcal{N}_C(x, y).$$

Analogous to the proof of Theorem 4.8, we can show that the graphical convergence of $\mathcal{H}^N$ to $\mathcal{H}$ at the cluster point $(x, y, \hat{\eta})$ (note that the graphical convergence in Theorem 4.8 does not require any result in Propositions 3.7 and 2.8), and hence the condition $0 \in \mathcal{H}(x^N, y^N, \eta^N)$ implies that w.p. 1, $0 \in \mathcal{H}(x, y, \hat{\eta})$ with $\|\hat{\eta}\| = 1$, a contradiction to NNAMCQ at $(x, y)$. □

Proof of Theorem 3.8

By Proposition 3.7, NNAMCQ holds at $(x^N, y^N)$ w.p.1 when $N$ is sufficiently large. By [47, Theorem 3.2(c)] or [45, Corollary 4.8(5)], $(x^N, y^N, \eta^N)$ must satisfy the following conditions:

$$0 \in \partial \left[ \frac{1}{N} \sum_{k=1}^{N} \partial f(x, y, \xi^k) \right] + \frac{1}{N} \sum_{k=1}^{N} \nabla F(x, y, \xi^k)^T \eta^N$$
$$+\{0_n\} \times D^* \mathcal{N}_Y \left( y, -\frac{1}{N} \sum_{k=1}^{N} F(x, y, \xi^k) \right) (\eta^N) + \mathcal{N}_C(x, y).$$

The desired first order necessary optimality condition follows by applying the sum rule [26, Theorems 2.33 and 3.36] to the limiting subdifferential of $\frac{1}{N} \sum_{k=1}^{N} f(x, y, \xi^k)$, that is,

$$\partial \left[ \frac{1}{N} \sum_{k=1}^{N} f(x, y, \xi^k) \right] \subset \frac{1}{N} \sum_{k=1}^{N} \partial f(x, y, \xi^k)$$

and the equality hold when Clarke regularity holds.

To prove the boundedness of $\eta^N$, assume for the sake of a contradiction that $\eta^N \to \infty$. Dividing the first order optimality conditions by $\|\eta^N\|$ and taking limits as $N \to \infty$, we get a contradiction as desired to the NNAMCQ at $(x, y)$ similar to the proof of Proposition 3.7. □

## References


1. Anitescu, M., Tseng, P., Wright, S.J.: Elastic-mode algorithms for mathematical programs with equilibrium constraints: global convergence and stationarity properties, Math. Program. **110**, 337–371 (2007)
2. Artstein, Z., Vitale, R.A.: A strong law of large numbers for random compact sets. Ann. Probab. **3**, 879–882 (1975)
3. Aumann, R.J.: Integrals of set-valued functions. J. Math. Anal. Appl. **12**, 1–12 (1965)
4. Birbil, S.I., Gurkan, G., Listes, O.L.: Solving stochastic mathematical programs with complementarity constraints using simulation. Math. Oper. Res. **31**, 739–760 (2006)
5. Christiansen, S., Patriksson, M., Wynter, L.: Stochastic bilevel programming in structral optimization. Preprint, PRISM, Université de Versailles, France (1999)
6. Clarke, F.H.: Optimization and Nonsmooth Analysis. Wiley, New York (1983)
7. Henrion, R., Outrata, J.: On calculating the normal to a finite union of convex polyhedra. Optimization **57**, 57–78 (2008)
8. Henrion, R., Outrata, J., Surowiec, T.: On the co-derivative of normal cone mappings to inequality systems. Nonlinear Anal. **71**, 1213–1226 (2009)
9. Henrion, R., Römisch, W.: On M-stationary point for a stochastic equilibrium problem under equilibrium constraints in electricity spot market modeling. Appl. Math. **52**, 473–494 (2007)
10. Hess, C.: Set-valued integration and set-valued probability theory: an overview. In: Handbook of Measure Theory, vols. I and II, pp. 617–673. North-Holland, Amsterdam (2002)
11. King, A.J., Rockafellar, R.T.: Sensitivity analysis for nonsmooth generalized equations. Math. Program. **55**, 193–212 (1992)
12. Lin, G.H., Guo, L., Ye, J.J.: Solving mathematical programs with equilibrium constraints as constrained equations. Preprint (2010)
13. Lin, G.H., Chen, X., Fukushima, M.: Solving stochastic mathematical programs with equilibrium constraints via approximation and smoothing implicit programming with penalization. Math. Program. Ser. B **116**, 343–368 (2009)

14. Lin, G.H., Xu, H., Fukushima, M.: Monte Carlo and quasi-Monte Carlo sampling methods for a class of stochastic mathematical programs with equilibrium constraints. Math. Methods Oper. Res. **67**, 423–441 (2008)
15. Liu, G.S., Ye, J.J., Zhou, J.: Partial exact penalty for mathematical programs with equilibrium constraints. Set-valued Anal. **16**, 785–804 (2008)
16. Liu, Y., Xu, H., Lin, G.H.: Stability analysis of two stage stochastic mathematical programs with complementarity constraints via NLP-regularization. Optimization Online. http://www.optimization-online.org/DB_HTML/2010/03/2575.html (2010). Accessed March 2010
17. Lucet, Y., Ye, J.J.: Sensitivity analysis of the value function for optimization problems with variational inequality constraints. SIAM J. Control Optim. **40**, 699–723 (2001)
18. Lucet, Y., Ye, J.J.: Erratum: sensitivity analysis of the value function for optimization problems with variational inequality constraints. SIAM J. Control Optim. **41**, 1315–1319 (2002)
19. Luo, Z.Q., Pang, J.-S., Ralph, D.: Mathematical Programs with Equilibrium Constraints. Cambridge University Press, Cambridge (1996)
20. Meng, F., Xu, H.: Exponential convergence of sample average approximation methods for a class of stochastic mathematical programs with complementarity constraints. J. Comput. Math. **24**, 733–748 (2006)
21. Mirrlees, J.: The theory of moral hazard and unobservable behaviour: part I. Rev. Econ. Stud. **66**, 3–21 (1999)
22. Molchanov, I.: Theory of Random Sets. In: J. Gani, et al (eds.) Probability and Its Applications. Springer (2005)
23. Mordukhovich, B.S.: Maximum principle in problems of time optimal control with nonsmooth constraints. J. Appl. Math. Mech. **40**, 960–969 (1976). English translation
24. Mordukhovich, B.S.: Metric approximation and necessary optimality conditions for general classes of nonsmooth extremal problems. Sov. Math., Dokl. **22**, 526–530 (1980)
25. Mordukhovich, B.S.: Variational Analysis and Generalized Differentiation, I: Basic Theory. Grundlehren Series (Fundamental Principles of Mathematical Sciences), vol. 330. Springer (2006)
26. Mordukhovich, B.S.: Variational Analysis and Generalized Differentiation, II: Applications. Grundlehren Series (Fundamental Principles of Mathematical Sciences), vol. 331. Springer (2006)
27. Niederreiter, H.: Random Number Generation and Quasi-Monte-Carlo Methods. SIAM, Philadelphia (1992)
28. Outrata, J., Kocvara, M., Zowe, J.: Nonsmooth Approach to Optimization Problems with Equilibrium Constraints–Theory, Applications and Numerical Constraints. Kluwer (1998)
29. Outrata, J., Römisch, W.: On optimization conditions for some nonsmooth optimization problems over $L^p$ spaces. J. Optim. Theory Appl. **126**, 411–438 (2005)
30. Patriksson, M., Wynter, L.: Stochastic mathematical programs with equilibrium constraints. Oper. Res. lett. **25**, 159–167 (1999)
31. Robinson, S.M.: Generalized equations and their solutions, part II: applications to nonlinear programming. Math. Program. Study **19**, 200–221 (1982)
32. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1970)
33. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin (1998)
34. Rubinstein, R.Y., Shapiro, A.: Discrete Events Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Methods. Wiley, New York (1993)
35. Shapiro, A.: Stochastic mathematical programs with equilibrium constraints. J. Optim. Theory Appl. **128**, 223–243 (2006)
36. Shapiro, A., Xu, H.: Stochastic mathematical programs with equilibrium constraints, modeling and sample average approximation. Optimization **57**, 395–418 (2008)
37. Shorack, G.R., Wellner, J.A.: Empirical Processes with Applications to Statistics. Wiley (1986)
38. Von Stackelberg, H.: Marktform and Gleichgewicht. Springer, Berlin (1934).
39. Werner, A.S.: Bilevel Stochastic Programming Problems: Analysis and Application to Telecommunications. Ph.D dissertation, Norwegian University of Science and Technology (2004)
40. Wets, R.J.-B., Xu, H.: Graphical convergence of sample average random set-valued mappings. Preprint (2008)
41. Xu, H.: Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming. J. Math. Anal. Appl. **368**, 692–710 (2010)
42. Xu, H., Meng, F.: Convergence analysis of sample average approximation methods for a class of stochastic mathematical programs with equality constraints. Math. Oper. Res. **32**, 648–668 (2007)

43. Xu, H., Ye, J.J.: Necessary optimality conditions for two-stage stochastic programs with equilibrium constraints. SIAM J. Optim. **20**, 1685–1715 (2010)
44. Ye, J.J.: Optimality conditions for optimization problems with complementarity constraints. SIAM J. Optim. **9**, 374–387 (1999)
45. Ye, J.J.: Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints. SIAM J. Optim. **10**, 943–962 (2000)
46. Ye, J.J.: Necessary and sufficient optimality conditions for mathemtical programs with equilibrium constraints. J. Math. Anal. Appl. **307**, 305–369 (2005)
47. Ye, J.J., Ye, X.Y.: Necessary optimality conditions for optimization problems with variational inequality constraints. Math. Oper. Res. **22**, 977–997 (1997)
48. Ye, J.J., Zhu, Q.J.: Multiobjective optimization problem with variational inequality constraints. Math. Program., Ser. A **96**, 139–160 (2003)