MATHEMATICS OF OPERATIONS RESEARCH
Vol. 36, No. 4, November 2011, pp. 670–694
ISSN 0364-765X | EISSN 1526-5471 | 11 | 3604 | 0670

http://dx.doi.org/10.1287/moor.1110.0513

Penalized Sample Average Approximation Methods for Stochastic Mathematical Programs with Complementarity Constraints

Yongchao Liu
Department of Mathematics, Dalian Maritime University, 116026 Dalian, China,
googleliu0717@yahoo.com.cn

Huifu Xu
School of Mathematics, University of Southampton, Southampton SO17 1BJ, United Kingdom,
h.xu@soton.ac.uk, http://www.personal.soton.ac.uk/hx/

Jane J. Ye
Department of Mathematics and Statistics, University of Victoria, Victoria, British Columbia V8W 3R4, Canada,
janeye@uvic.ca, http://www.math.uvic.ca/faculty/janeye/

This paper considers a one-stage stochastic mathematical program with a complementarity constraint (SMPCC), where uncertainties appear in both the objective function and the complementarity constraint, and an optimal decision on both upper- and lower-level decision variables must be made before the realization of the uncertainties. A partially exactly penalized sample average approximation (SAA) scheme is proposed to solve the problem. Asymptotic convergence of optimal solutions and stationary points of the penalized SAA problem is carried out. It is shown under some moderate conditions that the statistical estimators obtained from solving the penalized SAA problems converge almost surely to its true counterpart as the sample size increases. Exponential rate of convergence of estimators is also established under some additional conditions.

Key words: MPEC-metric regularity; partial exact penalization; M-stationary point
MSC2000 subject classification: Primary: 90C15, 90C33; secondary: 49J53, 90C31
OR/MS subject classification: Primary: programming, stochastic, complementarity; secondary: mathematics, sets, programming, nondifferentiable
History: Received May 28, 2010; revised December 11, 2010, and March 4, 2011. Published online in *Articles in Advance* October 14, 2011.

1. Introduction. In this paper, we study the following stochastic mathematical program with complementarity constraints (SMPCC):

$$\begin{aligned} \min_{x,y} \quad & \mathbb{E}[f(x,y,\xi(\omega))] \\ \text{s.t.} \quad & (x,y)\in D, \\ & 0\le \mathbb{E}[F(x,y,\xi(\omega))]\perp y\ge 0, \end{aligned} \tag{1}$$

where D is a nonempty closed subset of $\mathbb{R}^{n+m}$, $f\colon \mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^d\to\mathbb{R}$, $F\colon \mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^d\to\mathbb{R}^m$ are continuously differentiable, $\xi\colon \Omega\to\Xi\subset\mathbb{R}^d$ is a vector of random variables defined on the probability space $(\Omega,\mathcal{F},P)$, $\mathbb{E}[\cdot]$ denotes the expected value with respect to the probability distribution of ξ, and $a\perp b$ means that vector a is perpendicular to vector b. For the simplicity of discussion, we assume that the expected values of the underlying random functions are well defined for every $(x,y)\in D$. SMPCC is also known as stochastic mathematical program with equilibrium constraints (SMPEC) in that the complementarity constraint often represents an equilibrium in practical applications. As far as we are concerned, there are essentially two classes of SMPECs being investigated up to date: one-stage SMPECs where both upper- and lower-level decision variables must be chosen before realization of uncertainties and two-stage SMPECs where lower-level decision variables must be chosen after upper-level decision is made and the uncertainties are realized (Patriksson and Wynter [24]). Obviously, our model is a one-stage SMPEC. Like deterministic mathematical program with equilibrium constraints (MPECs), SMPEC models have many applications in economics, engineering, networks and management sciences, see, for instances, Christiansen et al. [5], Werner [35], Werner and Wang [36], Tomasgard et al. [34], and the references therein.

In this paper, we are concerned with numerical methods for solving SMPCC (1). Observe that if we know the distribution of ξ and can obtain a closed form of $\mathbb{E}[f(x,y,\xi)]$ and $\mathbb{E}[F(x,y,\xi)]$, then SMPCC (1) reduces to a deterministic mathematical program with complementarity constraints (MPCC) and subsequently we can solve it by an existing numerical method for deterministic MPECs. In practice, the distribution of ξ is often unknown or it is numerically too expensive to calculate the expected values. Instead it might be possible to obtain a sample of the random vector ξ from past data. This motivates one to find an approximate optimal solution to (1) on the basis of the sampling information.

A well-known approximation method in stochastic programming based on sampling is sample average approximation (SAA). That is, if we have an independent identically distributed (iid) sample $\xi^1, \ldots, \xi^N$ of random vector ξ, then we may use the sample average $(1/N)\sum_{k=1}^N f(x, y, \xi^k)$ and $(1/N)\sum_{k=1}^N F(x, y, \xi^k)$ to approximate $\mathbb{E}[f(x, y, \xi)]$ and $\mathbb{E}[F(x, y, \xi)]$. This kind of statistical approximation is guaranteed by the classical law of large numbers in statistics. Consequently, we may consider the following approximate SMPCC problem:

$$\begin{aligned} \min_{x,y} \quad & \frac{1}{N}\sum_{k=1}^N f(x, y, \xi^k) \\ \text{s.t.} \quad & (x, y) \in D, \\ & 0 \le \frac{1}{N}\sum_{k=1}^N F(x, y, \xi^k) \perp y \ge 0. \end{aligned} \tag{2}$$

We call SMPCC (1) the *true* problem and (2) the SAA problem. SAA method is a popular method in stochastic programming and it has been applied to solve SMPECs over the past few years although most of the applications are focused on two-stage SMPECs, see, for instance, Shapiro [32], Shapiro and Xu [33], Xu and Meng [38], and the references therein.

The SMPCC model (1) and its SAA (2) are not new either. Indeed, Birbil et al. [2] studied the model and applied the sample-path optimization (SPO) method (Robinson [27]) to obtain some approximation results. SPO is essentially SAA although the former is slightly more general. More recently, Meng and Xu [18] discussed the SAA problem (2) and obtained exponential convergence of weak stationary points of SAA problem (2); that is, for any $\epsilon > 0$, there exist constants $c(\epsilon) > 0$ and $k(\epsilon) > 0$ and positive integer $N(\epsilon) \ge 0$ such that

$$\text{Prob}(\|x^N - x^*\| \ge \epsilon) \le c(\epsilon)e^{-k(\epsilon)N}$$

for $N \ge N(\epsilon)$, where x^* and x^N denote the weak stationary points to the true problem (1) and the SAA problem (2), respectively.

The results obtained in Birbil et al. [2] and Meng and Xu [18] are mainly for weak stationary points and they are obtained under very strong assumptions such as upper-level strict complementarity condition, lower-level strict complementarity condition, or strong regularity condition. It is well known in the MPEC literature that the weak stationary condition is usually too weak and most of numerical algorithms aim at finding at least Clarke [6] stationary points (see Definition 2.14 for definition and relationships of various stationary points of MPEC). Moreover most algorithms for solving MPECs require a very strong constraint qualification called MPEC-linear independence constraint, qualification (LICQ); see Liu et al. [14] for a discussion on this issue. For stochastic MPECs, it is difficult to prove the convergence by SAA to a MPEC stationary point that are stronger than the weak stationary point since the degenerate index set $\{i: (1/N)\sum_{k=1}^N F(x, y, \xi^k) = 0,\ y_i = 0\}$ changes as the sample size N changes, and all MPEC stationary points, except the weak stationary point, depend on this index set. In this paper, we resolve these issues by using partial exact penalty method, a technique recently proposed by Liu et al. [14] for deterministic MPCCs under MPEC metric regularity (see Definition 2.11), which is a much weaker constraint qualification than MPEC-LICQ.

Specifically, we introduce a new decision vector z and reformulate the true problem (1) as

$$\begin{aligned} \min_{x,y,z} \quad & \mathbb{E}[f(x, y, \xi(\omega))] \\ \text{s.t.} \quad & (x, y, z) \in D \times \mathbb{R}^m, \\ & \mathbb{E}[F(x, y, \xi(\omega))] - z = 0, \\ & 0 \le z \perp y \ge 0 \end{aligned} \tag{3}$$

and the SAA problem (2) as

$$\begin{aligned} \min_{x,y,z} \quad & \frac{1}{N}\sum_{k=1}^N f(x, y, \xi^k) \\ \text{s.t.} \quad & (x, y, z) \in D \times \mathbb{R}^m, \\ & \frac{1}{N}\sum_{k=1}^N F(x, y, \xi^k) - z = 0, \\ & 0 \le z \perp y \ge 0. \end{aligned} \tag{4}$$

We then consider a partial exact penalization of the reformulated true problem (3):

$$\begin{aligned} \min_{x,y,z} \quad & \psi(x,y,z,\rho) := \mathbb{E}[f(x,y,\xi(\omega))] + \rho\|\mathbb{E}[F(x,y,\xi(\omega)] - z\|_1 \\ \text{s.t.} \quad & (x,y,z) \in D \times \mathbb{R}^m, \\ & 0 \le z \perp y \ge 0, \end{aligned} \tag{5}$$

where ρ is a positive constant, and a partial exact penalization of the reformulated SAA problem (4):

$$\begin{aligned} \min_{x,y,z} \quad & \psi_N(x,y,z,\rho_N) := \frac{1}{N}\sum_{k=1}^N f(x,y,\xi^k) + \rho_N\left\|\frac{1}{N}\sum_{k=1}^N F(x,y,\xi^k) - z\right\|_1, \\ \text{s.t.} \quad & (x,y,z) \in D \times \mathbb{R}^m, \\ & 0 \le z \perp y \ge 0, \end{aligned} \tag{6}$$

where ρ_N is a positive number. Here and later on $\|\cdot\|_1$ denotes the 1-norm of a vector. The following are three main benefits to consider the partial penalization:

(a) Since the original problem (1) does not satisfy usual Mangasarian-Fromovitz constraint qualification (MFCQ), we cannot establish a full exact penalization of all constraints.[1] Partial exact penalization is, however, feasible under MPEC-generalized Mangasarian-Fromovitz constraint qualification (GMFCQ) (Outrata [22], Ye [40]) or even weaker constraint qualification such as MPEC-metric regularity or equivalently MPEC-no nonzero abnormal multiplier constraint qualification (NNAMCQ) to be defined in §2, as we keep the complementarity constraint in (5).

(b) For the convergence result although we only require MPEC metric regularity for the original problem (1), and hence MPEC-LICQ may fail for the original problem (1), the MPEC-LICQ is satisfied at every feasible point of the penalized problem regardless of structure of the original problem. From a numerical perspective, this is very important as the stability of many existing numerical methods such as the nonlinear program (NLP) regularization method (Scholtes [30]) depend on MPEC-LICQ. Indeed, this is a key motivation for Liu et al. [14] to consider the partial exact penalization in Liu et al. [14].

(c) The constraints of both problems (5) and (6) are independent of sampling and this will significantly simplify our convergence analysis.

In this paper, we analyze the convergence analysis of optimal solutions and stationary points of (6) as sample size increases, assuming (6) can be solved by a deterministic MPCC solver, which can effectively deal with the nonsmoothness in the objective. We do so by showing the existence of bounded penalty parameters in both (5) and (6) and this is indeed another departure from the existing research in the literature of SMPECs (Lin et al. [12, 13]). Moreover, we consider a smoothing method proposed by Liu et al. [14] to tackle the nonsmoothness. That is, we consider a smooth partial penalty problem of (4):

$$\begin{aligned} \min_{x,y,z} \quad & \hat{\psi}_N(x,y,z,\rho_N,\delta_N) \\ \text{s.t.} \quad & (x,y,z) \in D \times \mathbb{R}^m, \\ & 0 \le z \perp y \ge 0, \end{aligned} \tag{7}$$

where

$$\hat{\psi}_N(x,y,z,\rho_N,\delta_N) := \frac{1}{N}\sum_{k=1}^N f(x,y,\xi^k) + \rho_N\sum_{i=1}^m\sqrt{\left(\frac{1}{N}\sum_{k=1}^N F_i(x,y,\xi^k) - z_i\right)^2 + \delta_N}$$

and $\delta_N \downarrow 0$ is a smoothing parameter. Since the problem is smooth and the MPEC-LICQ holds, existing MPEC solvers can be used to solve the problem or at least to find some stationary points.

The rest of the paper is organized as follows. In §2, we review some definitions and preliminary results in variational analysis and MPECs. In §3, we discuss the relationship between the problems (1) and (5) and boundedness of penalty parameters of (6). In §4, we investigate the uniform convergence of the objective function and its subdifferential of the penalty problem (6). In §5, we use the uniform convergence results established in §4 to analyze the convergence of optimal solutions and stationary points obtained from solving the SAA penalty problem (6). Finally, in §6, we provide some numerical tests on the smoothed SAA penalty problem (7) along with some convergence analysis.

[1] Here, full penalization means the whole complementarity constraint in (1) is penalized in the form of $|\mathbb{E}[F(x,y,\xi(\omega))]^T y| + \|y_-\|_1 + \|\mathbb{E}[F(x,y,\xi(\omega))]_-\|_1$ or the complementarity constraint $0 \le z \perp y \ge 0$ in (5) is also penalized to the objective in the form of $|y^T z| + \|y_-\|_1 + \|z_-\|_1$, where $a_- = \min(0,a)$ for a real number a and the minimum is taken componentwise when a is a vector.

2. Preliminaries. Throughout this paper, we use the following notation. $\|\cdot\|$ denotes the Euclidean norm of a vector, a matrix and a compact set of vectors/matricies in a finite-dimensional space. When $\mathscr{M}$ is a compact set,

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

$\|\cdot\|_1$ denotes the 1-norm of a vector. We use $d(x,\mathscr{D}) := \inf_{x'\in\mathscr{D}} \|x-x'\|$ to denote the distance from a point x to a set $\mathscr{D}$. For two compact sets $\mathscr{C}$ and $\mathscr{D}$,

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

denotes the deviation of $\mathscr{C}$ from $\mathscr{D}$. Equivalently,

$$\mathbb{D}(\mathscr{C},\mathscr{D}) = \inf\{t \geq 0\colon \mathscr{C} \subseteq \mathscr{D} + t\mathscr{B}\},$$

where $\mathscr{B}$ denotes the closed unit ball in the corresponding finite-dimensional space here and throughout the paper. We use $\mathscr{C}+\mathscr{D}$ to denote the Minkowski addition of the two sets; that is, $\{C+D\colon C\in\mathscr{C},\ D\in\mathscr{D}\}$; $B(x,\gamma)$ the closed ball with center x and radius γ; a^Tb the scalar product of vectors a and b, where a^T denotes the transpose of vector a. If A is a matrix, A^Tb denotes matrix vector multiplication. When f is real-valued function, $\nabla f(x)$ denotes the gradient of f at x (which is a column vector) and when f is a vector-valued function, $\nabla f(x)$ represents the classical Jacobian of f at x, where the transpose of the gradient of the j-th component of f forms the j-th row of the Jacobian.

2.1. Variational analysis. Let X be a finite-dimensional space and $\Gamma\colon X \rightrightarrows \mathbb{R}^n$ be a set-valued mapping. We say that Γ is *upper semicontinuous* at a point $x\in X$ if for any $\epsilon>0$, there exists a number $\delta>0$ (which may depend on x) such that

$$\Gamma(x') \subseteq \Gamma(x) + \epsilon\mathscr{B}, \quad \forall\, x' \in x+\delta\mathscr{B}.$$

Definition 2.1 (Uniform Upper Semicontinuity). Let X be a finite-dimensional space and $\Gamma\colon X \rightrightarrows \mathbb{R}^n$ be a set-valued mapping. We say Γ is *uniformly upper semicontinuous* over a set $\mathscr{X}\subseteq X$ if for any given $\epsilon>0$, there exists a number $\delta>0$ such that

$$\sup_{x\in\mathscr{X}} \mathbb{D}(\Gamma(x'),\Gamma(x)) \leq \epsilon, \quad \forall\, x' \in (x+\delta\mathscr{B})\cap\mathscr{X}.$$

Note that the uniform upper semicontinuity depends on set $\mathscr{X}$. Consider function $f(x_1,x_2)=|x_1-x_2|$ defined on $\mathbb{R}^2$. The function is continuously differentiable in $\mathbb{R}^2$ except at line $x_1=x_2$. It is easy to see that the (usual convex) subdifferential mapping $\partial f(x_1,x_2)$ is uniformly upper semicontinuous on line $x_1=x_2$ but not in any neighborhood of this line. When $\mathscr{X}$ consists of a finite number of points, the uniform upper semicontinuity is equivalent to pointwise upper semicontinuity. We need the concept in Lemma 4.1.

Definition 2.2 (Normal Cone; See Chapter 1 in Mordukhovich [21]). Let $\mathscr{D}$ be a nonempty closed subset of $\mathbb{R}^n$. Given $z\in\mathscr{D}$, the convex cone

$$N^{\pi}_{\mathscr{D}}(z) := \{\zeta\in\mathbb{R}^n\colon \exists\sigma>0 \text{ such that } \zeta^T(z'-z) \leq \sigma\|z'-z\|^2,\ \forall\, z'\in\mathscr{D}\}$$

is called the *proximal normal cone* to set $\mathscr{D}$ at point z. By convention, for $z\notin\mathscr{D}$, $\mathscr{N}^{\pi}_{\mathscr{D}}(z)=\varnothing$. The closed cone

$$\mathscr{N}_{\mathscr{D}}(z) := \limsup_{z'\to z} \mathscr{N}^{\pi}_{\mathscr{D}}(z')$$

is called the *limiting normal cone* (also known as Mordukhovich [21] normal cone or basic normal cone) to $\mathscr{D}$ at point z.

Note that the limiting normal cone is, in general, smaller than the Clarke [6] normal cone, which is defined as the polar cone of the Clarke tangent cone $\mathscr{T}^c_D(z)$; that is,

$$\mathscr{N}^c_{\mathscr{D}}(z) = \{\zeta\in\mathbb{R}^n\colon \zeta^T\eta \leq 0,\ \forall\,\eta\in\mathscr{T}^c_{\mathscr{D}}(z)\},$$

where $T^c_{\mathscr{D}}(z) := \liminf_{t\to 0,\,\mathscr{D}\ni z'\to z}(1/t)(\mathscr{D}-z')$. In the case when $\mathscr{D}$ is convex, the proximal normal cone, the limiting normal cone and the Clarke [6] normal cone coincide, see Chapter 1 in Mordukhovich [21] and part B of Chapter 6 in Rockafellar and Wets [28].

The following expressions for the limiting normal cone can be easily derived, see e.g., Ye [39, Proposition 3.7].

PROPOSITION 2.3. *Let $\mathcal{W} = \{(y,z) \in \mathbb{R}^m \times \mathbb{R}^m\colon 0 \le z \perp y \ge 0\}$. The limiting normal cone of $\mathcal{W}$ at $(y,z) \in \mathcal{W}$ is*

$$\mathcal{N}_{\mathcal{W}}(y,z) = \left\{ (u,v) \in \mathbb{R}^m \times \mathbb{R}^m\colon \begin{array}{ll} u_i = 0, & \text{if } y_i > 0, \\ v_i = 0, & \text{if } z_i > 0, \\ \text{either } u_i v_i = 0, \text{ or } u_i < 0,\ v_i < 0, & \text{if } y_i = z_i = 0. \end{array} \right\}$$

DEFINITION 2.4 (SUBDIFFERENTIALS). Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a lower semicontinuous function and finite valued at $x \in \mathbb{R}^n$. The *proximal subdifferential* of f at x is the set

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

and the *limiting* (*Mordukhovich or basic Mordukhovich* [21]) *subdifferential* of f at x is 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 convex hull of $\partial f(x)$ coincides with the *Clarke subdifferential* (Clarke calls it generalized gradient, see p. 27 in Clarke [6]), denoted by $\partial^c f(x)$; that is, $\operatorname{conv} \partial f(x) = \partial^c f(x)$.

The definition of limiting normal cone leads to the definition of coderivative of a set-valued mapping.

DEFINITION 2.5. Let X, Y be finite-dimensional spaces and $\Gamma\colon X \rightrightarrows Y$ a set-valued mapping with a closed graph. Let $(\bar{x},\bar{y}) \in \operatorname{gph}\Gamma := \{(x,y) \mid y \in \Gamma(x)\}$. The set-valued mapping $D^*\Gamma(\bar{x} \mid \bar{y})$ from Y to X defined by

$$D^*\Gamma(\bar{x} \mid \bar{y})(\eta) = \{\xi \in X\colon (\xi, -\eta) \in \mathcal{N}_{\operatorname{gph}\Gamma}((\bar{x},\bar{y}))\}$$

is called the *coderivative* of Γ at point $(\bar{x},\bar{y})$. The symbol $D^*\Gamma(\bar{x})$ is used when Γ is single valued at $\bar{x}$ and $\bar{y} = \Gamma(\bar{x})$. See §1.2 in Mordukhovich [21] for a detailed discussion of the notion.

In a special case when a set-valued mapping is single valued, the coderivative is related to the limiting subdifferential in the following way.

PROPOSITION 2.6 (MORDUKHOVICH [20, PROPOSITION 2.11]). *Let X, Y be finite-dimensional spaces and $\Gamma\colon X \to Y$ be a single valued and Lipschitz continuous near $\bar{x}$. Then $D^*\Gamma(\bar{x})(\eta) = \partial\langle \eta, \Gamma\rangle(\bar{x})$ for all $\eta \in Y$.*

The following sum rule will be useful.

PROPOSITION 2.7 (SEE MORDUKHOVICH [20, COROLLARY 4.4]). *Let X, Y be finite-dimensional spaces and $\phi\colon X \to Y$ be strictly differentiable near $\bar{x}$ and Ω is closed. Let $\Gamma(x) := -\phi(x) + \Omega$ and $0 \in \Gamma(\bar{x})$. Then*

$$D^*\Gamma(\bar{x} \mid 0)(\eta) = -\nabla\phi(\bar{x})^T\eta \quad \textit{for } -\eta \in N_\Omega(\phi(\bar{x})).$$

DEFINITION 2.8 (METRIC REGULARITY). Let X, Y be finite-dimensional spaces and $\Gamma\colon X \rightrightarrows Y$ be a set-valued mapping. Let $(\bar{x},\bar{y}) \in \operatorname{gph}\Gamma$. Γ is said to be *metrically regular* at $\bar{x}$ for $\bar{y}$ if there exist constants $\kappa > 0$, $\delta > 0$ such that

$$d(x, \Gamma^{-1}(y)) \le \kappa d(y, \Gamma(x)), \quad \forall\,(x,y) \in (\bar{x},\bar{y}) + \delta\mathcal{B}. \tag{8}$$

The *regularity modulus* of Γ at $\bar{x}$ for $\bar{y}$ is the value

$$\operatorname{reg}\Gamma(\bar{x} \mid \bar{y}) := \inf\{\kappa \in (0,\infty) \mid (8) \text{ holds}\} \in [0,\infty]. \tag{9}$$

PROPOSITION 2.9 ((ESTIMATE FOR LIPSCHITZ PERTURBATIONS) DONTCHEV ET AL. [7]). *Consider any set-valued mapping $\Gamma\colon X \rightrightarrows Y$ and any $(\bar{x},\bar{y}) \in \operatorname{gph}\Gamma$ at which $\operatorname{gph}\Gamma$ is locally closed. Consider also a mapping $G\colon X \to Y$. If $\operatorname{reg}\Gamma(\bar{x} \mid \bar{y}) < \kappa < \infty$ and $\operatorname{lip} G(\bar{x}) < \lambda < \kappa^{-1}$, then*

$$\operatorname{reg}(\Gamma + G)(\bar{x} \mid \bar{y} + G(\bar{x})) < \frac{\kappa}{1 - \lambda\kappa}, \tag{10}$$

where $\operatorname{lip} G(\bar{x})$ denotes the Lipschitz modulus of a single-valued mapping G at $\bar{x}$, i.e.,

$$\operatorname{lip} G(\bar{x}) := \limsup_{x, x' \to \bar{x},\, x, x' \neq \bar{x}} \frac{\|G(x') - G(x)\|}{\|x' - x\|}. \tag{11}$$

PROPOSITION 2.10 ((MORDUKHOVICH'S CRITERIA FOR METRIC REGULARITY) MORDUKHOVICH [19, COROLLARY 5.4]). *Let Γ be a set-valued mapping with closed graph and $(\bar{x},\bar{y}) \in \operatorname{gph}\Gamma$. Then Γ is metrically regular at $\bar{x}$ for $\bar{y}$ if and only if*

$$D^*\Gamma(\bar{x} \mid \bar{y})(\eta) = \{0\} \implies \eta = 0.$$

2.2. MPEC constraint qualification and stationarity. Consider now the following deterministic MPCC:

$$\begin{aligned} \min\ & f(x) \\ \text{s.t.}\ & x \in X, \\ & 0 \le G(x) \perp H(x) \ge 0, \end{aligned} \tag{12}$$

where X is a closed subset of $\mathbb{R}^n$, $f\colon \mathbb{R}^n \to \mathbb{R}$, $G\colon \mathbb{R}^n \to \mathbb{R}^m$, and $H\colon \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable. When X is represented by a system of smooth equalities and inequalities, it is well known that the classical MFCQ fails at any feasible solution, see Ye et al. [41, Proposition 1.1]. Since the classical MFCQ when X is a system of smooth equalities and inequalities is equivalent to the metric regularity of the set-valued mapping

$$\Gamma(x) := \begin{pmatrix} G(x) \\ H(x) \\ G(x)^T H(x) \\ -x \end{pmatrix} + \begin{pmatrix} \mathbb{R}^m_- \\ \mathbb{R}^m_- \\ \{0\} \\ X \end{pmatrix},$$

the above set-valued mapping is not metrically regular at any feasible point of MPCC. However, the following weaker metric regularity may hold.

Definition 2.11. (MPEC-Metric Regularity). Let x be a feasible point of problem (12) and

$$\Gamma(x) := -(G(x), H(x), x) + \mathscr{W} \times X,$$

where $\mathscr{W} := \{(y, z)\colon 0 \le z \perp y \ge 0\}$. Problem (12) is said to satisfy *MPEC-metric regularity* at x if the set-valued mapping $\Gamma(x)$ is metrically regular at x for 0.

In §3 (Theorem 3.4), we will consider MPEC metric regularity of (1), which is in the sense of the metric regularity of the set-valued mapping

$$\Gamma(x, y) := -(\mathbb{E}[F(x, y, \xi)], y, (x, y)) + \mathscr{W} \times D, \tag{13}$$

where $\mathscr{W} := \{(u, v) \in \mathbb{R}^m \times \mathbb{R}^m\colon 0 \le u \perp v \ge 0\}$.

The metric regularity is, however, not easy to verify by definition. By using Mordukhovich's criteria [19] for metric regularity (Proposition 2.10), the sum rule for coderivatives (Proposition 2.7) and the expression for the normal cone of $\mathscr{W}$ (Proposition 2.3), one can show that MPEC-metric regularity is indeed equivalent to a much easier to verify condition called MPEC-NNAMCQ in the case where the functions involved are smooth (and is weaker when the functions involved are nonsmooth but Lipschitz continuous).

Definition 2.12 (MPEC-NNAMCQ). Let x be a feasible point of problem (12). We say that MPEC-NNAMCQ holds at x if there exist no nonzero vectors $(\lambda, \beta) \in \mathbb{R}^m \times \mathbb{R}^m$ such that

$$\begin{gathered} 0 \in \nabla G(x)^T \lambda + \nabla H(x)^T \beta + \mathcal{N}_X(x), \\ \lambda_i = 0, \quad \text{if } G_i(x) > 0; \qquad \beta_i = 0, \quad \text{if } H_i(x) > 0, \\ \lambda_i < 0, \quad \beta_i < 0, \quad \text{or} \quad \lambda_i \beta_i = 0, \quad \text{if } G_i(x) = H_i(x) = 0, \end{gathered}$$

where subscript i denotes the i-th component of a vector.

Note that MPEC-NNAMCQ is weaker than the generalized MPEC-MFCQ (MPEC-GMFCQ) in the literature of deterministic MPECs. In the case when x falls into the interior of set X, the two conditions are equivalent, see Ye [39, 40] for the definition of MPEC-GMFCQ and the proof of the equivalence.

To accommodate a nonfeasible point obtained from a numerical algorithm, we also need the following extended MPEC-NNAMCQ, which was introduced in Liu et al. [14]. The extended MPEC-NNAMCQ coincides with MPEC-NNAMCQ at a feasible point.

Definition 2.13 (Extended MPEC-NNAMCQ). Consider the reformulation of the problem (12):

$$\begin{aligned} \min\ & f(x) \\ \text{s.t.}\ & x \in X, \\ & z = G(x), \\ & y = H(x), \\ & 0 \le z \perp y \ge 0. \end{aligned} \tag{14}$$

A point (x, y, z) is said to be a *weak feasible point* of problem (14) if $x \in X$ and $0 \le z \perp y \ge 0$. We say problem (12) satisfies the *extended MPEC-NNAMCQ* at (x, y, z) if (x, y, z) is a weak feasible point to (14) and there exist no nonzero vectors $(\lambda, \beta) \in \mathbb{R}^m \times \mathbb{R}^m$ such that

$$\begin{gathered} 0 \in \nabla G(x)^T \lambda + \nabla H(x)^T \beta + \mathcal{N}_X(x), \\ \lambda_i = 0, \quad \text{if } y_i > 0; \qquad \beta_i = 0, \quad \text{if } z_i > 0, \\ \lambda_i < 0, \quad \beta_i < 0, \quad \text{or} \quad \lambda_i \beta_i = 0, \quad \text{if } y_i = z_i = 0. \end{gathered}$$

For the convenience of reference, we state briefly MPEC stationarity in the following definition.

Definition 2.14 (MPEC W-, C-, M-, S- Stationary Conditions). Let x be a feasible point of problem (12). We say that x is a weak (W-) stationary point of (12) if there exist no nonzero vectors $(\lambda, \beta) \in \mathbb{R}^m \times \mathbb{R}^m$ such that

$$0 \in \nabla f(x) + \nabla G(x)^T \lambda + \nabla H(x)^T \beta + \mathcal{N}_X(x), \tag{15}$$

$$\lambda_i = 0, \quad \text{if } G_i(x) > 0; \qquad \beta_i = 0, \quad \text{if } H_i(x) > 0. \tag{16}$$

We say that x is a Clarke (C-), Mordukhovich (M-), Strong (S-) stationary point of (12) if there exist no nonzero vectors $(\lambda, \beta) \in \mathbb{R}^m \times \mathbb{R}^m$ such that (15)–(16) hold and

$$\begin{gathered} \lambda_i \beta_i \ge 0 \quad \text{if } G_i(x) = H_i(x) = 0, \\ \lambda_i < 0, \quad \beta_i < 0, \quad \text{or} \quad \lambda_i \beta_i = 0, \quad \text{if } G_i(x) = H_i(x) = 0, \\ \lambda_i \le 0, \quad \beta_i \le 0 \quad \text{if } G_i(x) = H_i(x) = 0, \end{gathered}$$

respectively.

The following relationship between MPEC stationary points is well known:

$$\text{S-stationary} \implies \text{M-stationary} \implies \text{C-stationary} \implies \text{W-stationary}.$$

3. Exact penalization of the true problem. In this section, we investigate the exact penalty parameter ρ for problem (5) and the relationships between (5) and (1) in terms of optimal solutions and stationary points. This is to pave the way for our discussion on the existence of exact penalty parameter ρ_N of SAA problem (6) and convergence analysis of optimal solutions and stationary points of the problem.

3.1. Exact penalty parameters. We start by discussing sufficient conditions for the existence of a bounded penalty parameter for problem (5). To this end, we derive error bound for a general system of equalities and inequality and its perturbation.

Let $g^N\colon \mathbb{R}^n \to \mathbb{R}^l$ and $h^N\colon \mathbb{R}^n \to \mathbb{R}^m$, $N = 1, 2, 3, \ldots$, be two sequences of continuously differentiable mappings and $\mathscr{C}$ be a closed subset of $\mathbb{R}^n$. Assume that $g^N(x)$, $h^N(x)$, $\nabla g^N(x)$, and $\nabla h^N(x)$ converge, respectively, to $g(x)$, $h(x)$, $\nabla g(x)$, and $\nabla h(x)$ uniformly over set $\mathscr{C}$ as $N \to \infty$. Denote by

$$\begin{gathered} S := \{x \in \mathbb{R}^n\colon g(x) \le 0,\ h(x) = 0,\ x \in \mathscr{C}\}, \\ S^N := \{x \in \mathbb{R}^n\colon g^N(x) \le 0,\ h^N(x) = 0,\ x \in \mathscr{C}\}. \end{gathered}$$

Let

$$\Gamma(x) := \begin{pmatrix} -g(x) \\ -h(x) \\ -x \end{pmatrix} + \begin{pmatrix} \mathbb{R}^m_- \\ \{0\} \\ \mathscr{C} \end{pmatrix},$$

and

$$\Gamma^N(x) := \begin{pmatrix} -g^N(x) \\ -h^N(x) \\ -x \end{pmatrix} + \begin{pmatrix} \mathbb{R}^m_- \\ \{0\} \\ \mathscr{C} \end{pmatrix}.$$

The system $\{g(x) \le 0,\ h(x) = 0,\ x \in \mathscr{C}\}$ is said to be *metrically regular* at a feasible point $\bar{x} \in S$ if the set-valued mapping $\Gamma(x)$ is metrically regular at $\bar{x}$ for $\bar{y} = 0$.

PROPOSITION 3.1. *Suppose that the system* $\{g(x) \le 0, h(x) = 0, x \in \mathscr{C}\}$ *satisfies metric regularity at a feasible point* $\bar{x} \in S$ *with regularity modulus* κ*. Then, there exists a neighborhood of* $\bar{x}$*, denoted by* $U_{\bar{x}}$ *such that*

(i) *the system* $\{g(x) \le 0, h(x) = 0, x \in \mathscr{C}\}$ *satisfies metric regularity at every point* $x \in U_{\bar{x}} \cap S$ *for* 0;

(ii) *there exists* N_0 *such that for* $N \ge N_0$*, the system* $\{g^N(x) \le 0, h^N(x) = 0, x \in \mathscr{C}\}$ *satisfies metric regularity at* $\bar{x}$ *for* 0*, i.e., there exist positive constants* κ *and* δ *such that*

$$d(x, (\Gamma^N)^{-1}(y)) \le 2\kappa d(y, \Gamma^N(x)), \quad \forall (x, y) \in U_{\bar{x}} \times \delta\mathscr{B}; \tag{17}$$

(iii) *the statements in parts* (i) *and* (ii) *hold when the metric regularity is replaced by NNAMCQ; that is, there exists no nonzero vectors* $\lambda \in \mathbb{R}^l_+$ *and* $\beta \in \mathbb{R}^m$ *such that*

$$\begin{cases} 0 \in \nabla g(\bar{x})^T \lambda + \nabla h(\bar{x})^T \beta + \mathscr{N}_{\mathscr{C}}(\bar{x}), \\ 0 \le \lambda \perp -g(\bar{x}) \ge 0. \end{cases}$$

PROOF. Part (i). By the definition of metric regularity, there exists an open neighborhood of $\bar{x}$ and a positive constant δ such that

$$d(x, \Gamma^{-1}(y)) \le \kappa d(y, \Gamma(x)), \quad \forall (x, y) \in U_{\bar{x}} \times \delta\mathscr{B}.$$

Let $\tilde{x} \in U_{\bar{x}} \cap S$. Then, there is a neighborhood of $\tilde{x}$, denoted by $U_{\tilde{x}}$ such that

$$U_{\tilde{x}} \times \delta\mathscr{B} \subset U_{\bar{x}} \times \delta\mathscr{B}.$$

Therefore

$$d(x, \Gamma^{-1}(y)) \le \kappa d(y, \Gamma(x)), \quad \forall (x, y) \in U_{\tilde{x}} \times \delta\mathscr{B},$$

which means that $\Gamma(x)$ is metrically regular at $\tilde{x}$ for 0. This shows part (i).

We now prove part (ii). For each fixed N, let

$$G^N(x) := \begin{pmatrix} g(x) - g^N(x) \\ h(x) - h^N(x) \\ 0 \end{pmatrix}.$$

It is easy to derive the Lipschitz modulus of G^N at x:

$$\operatorname{lip} G^N(x) = \sqrt{\|\nabla g(x) - \nabla g^N(x)\| + \|\nabla h(x) - \nabla h^N(x)\|}.$$

By assumption, $\nabla g^N(x)$ and $\nabla h^N(x)$ converge to $\nabla g(x)$ and $\nabla h(x)$, respectively, uniformly over set $\mathscr{C}$ as $N \to \infty$. This implies $\operatorname{lip} G^N(x) \to 0$ as $N \to \infty$. By Proposition 2.9,

$$\operatorname{reg}(\Gamma + G^N)(\bar{x} \mid G^N(\bar{x})) \le 2\kappa,$$

and hence (17) holds for N sufficiently large.

Part (iii) follows from Mordukhovich's criteria [19] for metric regularity (Proposition 2.10) and the sum rule for coderivatives (Proposition 2.7). It is also covered by a recent result by Ioffe and Outrata [9, Proposition 3.5]. □

Using Proposition 3.1, we are able to derive a local error bound for the feasible set of the systems defined in the proposition.

PROPOSITION 3.2. *Let* S *and* S^N *be defined as in Proposition* 3.1*,* $x^N \in S^N$*, and* $x^N \to \bar{x}$*. Then* $\bar{x} \in S$*. Moreover, if the system* $\{g(x) \le 0, h(x) = 0, x \in \mathscr{C}\}$ *satisfies metric regularity at point* $\bar{x}$ *for* 0 *with regularity modulus* κ*, then*

(i) *there exist positive constants* κ *and* δ *such that*

$$d(x, S) \le \kappa(\|g(x)_+\|_1 + \|h(x)\|_1), \quad \forall x \in \mathscr{C} \cap B(\bar{x}, \delta);$$

(ii) *there exists a constant* $\delta > 0$ *such that for* N *sufficiently large*

$$d(x, S^N) \le 2\kappa(\|g^N(x)_+\|_1 + \|h^N(x)\|_1), \quad \forall x \in \mathscr{C} \cap B(\bar{x}, \delta); \tag{18}$$

(iii) *statements* (i) *and* (ii) *hold when the metric regularity is replaced by NNAMCQ.*

Proof. The statement that $\bar{x} \in S$ follows from the uniform convergence of $\{g^N(x)\}$, $\{h^N(x)\}$ on set $\mathscr{C}$. The metric regularity of the set-valued mapping Γ at $\bar{x}$ for 0 means that there exist positive constants κ and δ such that

$$d(x, \Gamma^{-1}(y)) \leq \kappa d(y, \Gamma(x)), \quad \forall (y, x) \in (0, \bar{x}) + \delta\mathscr{B}.$$

Taking $y = 0$ in the above, we have

$$d(x, \Gamma^{-1}(0)) \leq \kappa d(0, \Gamma(x)) = \kappa d\left(\begin{pmatrix} g(x) \\ h(x) \\ x \end{pmatrix}, \begin{pmatrix} \mathbb{R}^m_- \\ \{0\} \\ \mathscr{C} \end{pmatrix}\right) \leq \kappa(\|g(x)_+\|_1 + \|h(x)\|_1)$$

for all $x \in (\bar{x} + \delta\mathscr{B}) \cap \mathscr{C}$. This shows part (i).

Part (ii). In the same manner, we can derive from (17) that

$$d(x, (\Gamma^N)^{-1}(0)) \leq 2\kappa d(0, \Gamma^N(x)) = 2\kappa d\left(\begin{pmatrix} g^N(x) \\ h^N(x) \\ x \end{pmatrix}, \begin{pmatrix} \mathbb{R}^m_- \\ \{0\} \\ \mathscr{C} \end{pmatrix}\right) \leq 2\kappa(\|g^N(x)_+\|_1 + \|h^N(x)\|_1)$$

for all $x \in (\bar{x} + \delta\mathscr{B}) \cap \mathscr{C}$.

Part (iii) follows from the equivalence of the metric regularity and NNAMCQ as shown in the proof of Proposition 3.1(iii). The proof is complete. □

The technical result in part (i) of Proposition 3.2 is needed for establishing a relationship between optimal solutions to the true problem (1) and its penalization (5), which will be detailed in Theorem 3.4. Part (ii) of the proposition will be needed to address issues of the SAA problems (2) and (6) to be detailed in Theorem 3.5.

It is often easier to show that a solution to an original optimization problem is a solution of its exactly penalized counterpart than the reverse. However, the reverse argument is equally important if not more since one often hopes to solve the original problem by solving its penalized counterpart. In what follows, we derive the equivalence between problems (3) and (5) under some moderate conditions.

Assumption 3.3. *Let $f(x, y, \xi)$ and $F(x, y, \xi)$ be defined as in* (1).

(a) *The Lipschitz modulus of f and F w.r.t. (x, y) are bounded by an integrable function $\kappa_1(\xi) \geq 0$.*

(b) *$\nabla_{(x,y)} f(x, y, \xi)$ and $\nabla_{(x,y)} F(x, y, \xi)$ are Lipschitz continuous w.r.t. (x, y) and their Lipschitz modulus are bounded by an integrable function $\kappa_2(\xi) \geq 0$.*

Theorem 3.4. *Let $(\bar{x}, \bar{y})$ be a local optimal solution to problem* (1). *Suppose that the MPEC-NNAMCQ* (*or equivalently MPEC-metric regularity*) *holds at $(\bar{x}, \bar{y})$. Under Assumption* 3.3(a),

(i) *there exists a constant $\rho^* > 0$ such that $(\bar{x}, \bar{y}, \bar{z})$, where $\bar{z} = \mathbb{E}[F(\bar{x}, \bar{y}, \xi)]$ is a local optimal solution of* (5) *if $\rho \geq \rho^*$;*

(ii) *if, in addition, D is a compact set and MPEC-NNAMCQ* (*or equivalently MPEC-metric regularity*) *holds at every optimal solution of problem* (1), *then there exists $\bar{\rho}$ such that for any $\rho > \bar{\rho}$, the sets of optimal solutions of problems* (3) *and* (5), *denoted by S_{opt} and S^{ρ}_{opt}, respectively, coincide.*

Proof. Part (i). Let $\Gamma(x, y)$ be defined by (13). The MPEC-metric regularity of (1) at $(\bar{x}, \bar{y})$ means that $\Gamma(x, y)$ is metric regular at $(\bar{x}, \bar{y})$ for 0. However, we will not use the metric regularity argument in the proof, instead, we use the equivalent argument MPEC-NNAMCQ, which is relatively easy to handle. By Definition 2.12, MPEC-NNAMCQ for problem (1) at $(\bar{x}, \bar{y})$ states that there are no nonzero vectors $(\lambda, \beta) \in \mathbb{R}^m \times \mathbb{R}^m$ such that

$$\begin{gathered} 0 \in \nabla \mathbb{E}[F(\bar{x}, \bar{y}, \xi)]^T \lambda + (0, \beta) + \mathscr{N}_D(\bar{x}, \bar{y}), \\ \lambda_i = 0, \quad \text{if } \mathbb{E}[F_i(\bar{x}, \bar{y}, \xi)] > 0; \qquad \beta_i = 0 \quad \text{if } \bar{y}_i > 0, \\ \lambda_i < 0, \quad \beta_i < 0, \quad \text{or} \quad \lambda_i \beta_i = 0 \quad \text{if } \mathbb{E}[F_i(\bar{x}, \bar{y}, \xi)] = \bar{y}_i = 0. \end{gathered}$$

It is easy to see that this condition is equivalent to the nonexistence of nonzero vectors $(\lambda, \beta^y, \beta^z) \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^m$ such that

$$\begin{gathered} 0 \in \nabla \mathbb{E}[F(\bar{x}, \bar{y}, \xi)]^T \lambda \times \{-\lambda\} + (0, \beta^y, \beta^z) + \mathscr{N}_D(\bar{x}, \bar{y}) \times \{0\}, \\ \beta^z_i = 0 \quad \text{if } \mathbb{E}[F_i(\bar{x}, \bar{y}, \xi)] > 0; \qquad \beta^y_i = 0 \quad \text{if } \bar{y}_i > 0, \\ \beta^z_i < 0, \quad \beta^y_i < 0, \quad \text{or} \quad \beta^z_i \beta^y_i = 0 \quad \text{if } \mathbb{E}[F_i(\bar{x}, \bar{y}, \xi)] = \bar{y}_i = 0, \end{gathered}$$

which happens to be the MEPC-NNAMCQ for problem (3) at $(\bar{x}, \bar{y}, \bar{z})$ with $\bar{z} = \mathbb{E}[F(\bar{x}, \bar{y}, \xi)]$. By the expression of the limiting normal cone in Proposition 2.3, it is easy to check that the above condition is equivalent to

$$[\{0\} \times \mathcal{N}_{\mathcal{W}}(\bar{y}, \bar{z})] \cap [-\mathcal{N}_D(\bar{x}, \bar{y}) \times \{0\}] = \{0\} \tag{19}$$

and there is no nonzero vector $\lambda \in \mathbb{R}^m$ such that

$$0 \in \nabla \mathbb{E}[F(\bar{x}, \bar{y}, \xi)]^T \lambda \times \{-\lambda\} + \{0\} \times \mathcal{N}_{\mathcal{W}}(\bar{y}, \bar{z}) + \mathcal{N}_D(\bar{x}, \bar{y}) \times \{0\},$$

where $\mathcal{W} = \{(y, z) \in \mathbb{R}^m \times \mathbb{R}^m \colon 0 \leq z \perp y \geq 0\}$. By virtue of Mordukhovich [21, Theorem 3.4] and (19), we have

$$\mathcal{N}_{(\mathbb{R}^n \times \mathcal{W}) \cap (D \times \mathbb{R}^m)}(\bar{x}, \bar{y}, \bar{z}) \subset \{0\} \times \mathcal{N}_{\mathcal{W}}(\bar{y}, \bar{z}) + \mathcal{N}_D(\bar{x}, \bar{y}) \times \{0\}.$$

Hence MPEC-NNAMCQ for problem (1) at $(\bar{x}, \bar{y})$ implies that there is no nonzero vectors $\lambda \in \mathbb{R}^m$ such that

$$0 \in \nabla \mathbb{E}[F(\bar{x}, \bar{y}, \xi)]^T \lambda \times \{-\lambda\} + \mathcal{N}_{(\mathbb{R}^n \times \mathcal{W}) \cap (D \times \mathbb{R}^m)}(\bar{x}, \bar{y}, \bar{z}).$$

We now use Proposition 3.2(i) to prove the first claim. Let $h(x, y, z) = \mathbb{E}[F(x, y, \xi)] - z$, $\mathcal{C} := (\mathbb{R}^n \times \mathcal{W}) \cap (D \times \mathbb{R}^m)$ and $S := \{(x, y, z) \colon h(x, y, z) = 0, (x, y, z) \in \mathcal{C}\}$. We have just shown that MPEC-NNAMCQ for problem (1) at $(\bar{x}, \bar{y})$ implies that the NNAMCQ of the system $\{h(x, y, z) = 0, (x, y, z) \in \mathcal{C}\}$ holds at $(\bar{x}, \bar{y}, \bar{z})$ with $\bar{z} = \mathbb{E}[F(\bar{x}, \bar{y}, \xi)]$. By Proposition 3.2(i) and (iii), there exist a constant $\tilde{\rho} > 0$ and a neighborhood of $(\bar{x}, \bar{y}, \bar{z})$, denoted by $U_{(\bar{x}, \bar{y}, \bar{z})}$ such that

$$d((x, y, z), S) \leq \tilde{\rho}(\|h(x, y, z)\|_1, \quad \forall\, (x, y, z) \in \mathcal{C} \cap U_{(\bar{x}, \bar{y}, \bar{z})}.$$

In terms of the terminology of Liu et al. [14], $\|h(x, y, z)\|_1$ is a partial error function on set $\mathcal{C}$ around $(\bar{x}, \bar{y}, \bar{z})$ with modulus $\tilde{\rho}$. Since $(\bar{x}, \bar{y}, \bar{z})$ is a local minimizer of (3), by the principle of partial exact penalization (Liu et al. [14, Theorem 3.3]), $(\bar{x}, \bar{y}, \bar{z})$ is also a local minimizer of (5) for $\rho \geq \tilde{\rho}\kappa$, where κ is the Lipschitz modulus of function $\mathbb{E}[f(x, y, \xi)]$. Note that under Assumption 3.3, such a κ exists. This shows the existence of a positive constant $\rho^* = \tilde{\rho}\kappa$ such that for any $\rho \geq \rho^*$, $(\bar{x}, \bar{y}, \bar{z})$ with $\bar{z} = \mathbb{E}[F(\bar{x}, \bar{y}, \xi)]$, is a local optimal solution of (5).

Part (ii). Since D is a compact set and $\mathbb{E}[f(x, y, \xi)]$ is continuous, then both optimal solution sets S_{opt} and S_{opt}^{ρ} are nonempty. We first show the existence of a constant $\bar{\rho} > 0$ such that for any $\rho \geq \bar{\rho}$, $S_{\text{opt}}^{\rho} \subseteq S_{\text{opt}}$. Assume for a contradiction that this is not true. Then, for any $\rho_k > 0$, there exists $(x(\rho_k), y(\rho_k), z(\rho_k)) \in S_{\text{opt}}^{\rho_k}$ such that $(x(\rho_k), y(\rho_k), z(\rho_k)) \notin S_{\text{opt}}$. Let $\rho_k \to \infty$. The compactness of D implies that the sequence $\{(x(\rho_k), y(\rho_k), z(\rho_k))\}$ is bounded where the boundedness of z_k follows from the equality constraint $z = \mathbb{E}[F(x, y, \xi)]$. By drawing a subsequence if necessary, we assume for the simplicity of notation that $((x(\rho_k), y(\rho_k), z(\rho_k)) \to (x^*, y^*, z^*)$. Let $(\bar{x}, \bar{y}, \bar{z}) \in S_{\text{opt}}$. Since $(x(\rho_k), y(\rho_k), z(\rho_k)) \in S_{\text{opt}}^{\rho_k}$, we have

$$\psi(x(\rho_k), y(\rho_k), z(\rho_k), \rho_k) \leq \psi(\bar{x}, \bar{y}, \bar{z}, \rho_k) = \mathbb{E}[f(\bar{x}, \bar{y}, \xi)],$$

which implies that

$$\rho_k \|\mathbb{E}[F(x(\rho_k), y(\rho_k), \xi)] - z(\rho_k)\|_1 \leq \mathbb{E}[f(\bar{x}, \bar{y}, \xi)] - \mathbb{E}[f(x(\rho_k), y(\rho_k), \xi)].$$

Taking a limit on both sides of the formula above, we obtain

$$0 \leq \mathbb{E}[f(\bar{x}, \bar{y}, \xi)] - \mathbb{E}[f(x^*, y^*, \xi)]$$

and $\|\mathbb{E}[F(x^*, y^*, \xi)] - z^*\|_1 = 0$, which means (x^*, y^*, z^*) is an optimal solution of (3) and (x^*, y^*) is an optimal solution of (1). Under the assumption that problem (1) satisfies MPEC-NNAMCQ at an optimal solution point (x^*, y^*), it follows from the proof of part (i), there exists a positive constant $\hat{\rho}$ such that (x^*, y^*, z^*) is a local minimizer of $\psi(x, y, z, \rho)$ for all $\rho \geq \hat{\rho}\kappa$, where κ is the Lipschitz modulus of function $\mathbb{E}[f(x, y, \xi)]$. Since $((x(\rho_k), y(\rho_k), z(\rho_k)) \to (x^*, y^*, z^*)$ and $(x(\rho_k), y(\rho_k), z(\rho_k)) \in S_{\text{opt}}^{\rho_k}$, we may find a neighborhood of (x^*, y^*, z^*), denoted by U such that both $((x(\rho_k), y(\rho_k), z(\rho_k))$ and (x^*, y^*, z^*) are minima of $\psi(x, y, z, \rho_k)$ over the set $U \cap \mathcal{F}^{\rho}$, where $\mathcal{F}^{\rho}$ denotes the feasible region of the penalized problem (5) for all $\rho_k \geq \hat{\rho}\kappa$. Consequently,

$$\begin{aligned}
\psi((x(\rho_k), y(\rho_k), z(\rho_k), \rho_k) &= \psi(x^*, y^*, z^*, \rho_k) = \mathbb{E}[f(x^*, y^*, \xi)] \\
&= \psi(x^*, y^*, z^*, (\rho_k + \hat{\rho}\kappa)/2) \\
&\leq \mathbb{E}[f(x(\rho_k), y(\rho_k), \xi)] + \frac{\rho_k + \hat{\rho}\kappa}{2} \|\mathbb{E}[F(x(\rho_k), y(\rho_k), \xi)] - z(\rho_k)\|_1,
\end{aligned}$$

which implies that

$$\rho_k \|\mathbb{E}[F(x(\rho_k), y(\rho_k), \xi)] - z(\rho_k)\|_1 \le \frac{\rho_k + \hat{\rho}\kappa}{2} \|\mathbb{E}[F(x(\rho_k), y(\rho_k), \xi)] - z(\rho_k)\|_1.$$

For $\rho_k > \hat{\rho}\kappa$, the above inequality implies that $\mathbb{E}[F(x(\rho_k), y(\rho_k), \xi)] = z(\rho_k)$, and hence

$$(x(\rho_k), y(\rho_k), z(\rho_k)) \in S_{\text{opt}},$$

which contradicts the fact that $(x(\rho_k), y(\rho_k), z(\rho_k)) \notin S_{\text{opt}}$. This shows $S^{\rho}_{\text{opt}} \subseteq S_{\text{opt}}$ for all $\rho > \bar{\rho} := \hat{\rho}\kappa$.

We are now ready to show that for any $\rho \ge \bar{\rho}$, $S_{\text{opt}} = S^{\rho}_{\text{opt}}$. Let $(\tilde{x}, \tilde{y}, \tilde{z}) \in S_{\text{opt}}$ and $(x(\rho), y(\rho), z(\rho)) \in S^{\rho}_{\text{opt}}$. Then, for any $\rho > \bar{\rho}$, since $S^{\rho}_{\text{opt}} \subseteq S_{\text{opt}}$, we have $(x(\rho), y(\rho), z(\rho)) \in S_{\text{opt}}$. Therefore $\psi(x(\rho), y(\rho), z(\rho), \rho) = \psi(\tilde{x}, \tilde{y}, \tilde{z}, \rho)$, which means $(\tilde{x}, \tilde{y}, \tilde{z})$ is also an optimal solution of problem (5). The proof is complete. □

We make a few comments on Theorem 3.4.

First, we implicitly assume that the true problem (1) has an optimal solution. It might be interesting to discuss conditions under which such an optimal solution exists. Observe that both $\mathbb{E}[f(x, y, \xi)]$ and $\mathbb{E}[F(x, y, \xi)]$ are deterministic functions and (1) is essentially a deterministic MPEC. Outrata et al. [23] presented a detailed discussion about conditions for a deterministic MPEC to have an optimal solution, see discussions in §§1.3 and 4.2 in Outrata et al. [23] for details. In this context, if there exists a point $(x_0, y_0) \in D$ such that

$$(\mathbb{E}[F(x_0, y, \xi)] - \mathbb{E}[F(x_0, y_0, \xi)])^T (y - y_0)/\|y - y_0\| \to \infty \quad \text{for } (x_0, y) \in D,\ \|y\| \to \infty, \tag{20}$$

then (1) has a feasible solution. Moreover, if there exists a feasible solution and the lower-level set of $\mathbb{E}[f(x, y, \xi)]$ at this feasible solution is bounded, then the optimal solution of (1) exists. Sufficient conditions for (20): D is compact or there exists a nonnegative integrable function $\sigma(\xi)$ with $\mathbb{E}[\sigma(\xi)] > 0$ such that

$$(F(x_0, y, \xi) - F(x_0, y_0, \xi))^T (y - y_0)/\|y - y_0\| \ge \sigma(\xi)\|y - y_0\|^2.$$

From this (taking expectation on both sides of the inequality), we immediately obtain

$$(\mathbb{E}[F(x_0, y, \xi)] - \mathbb{E}[F(x_0, y_0, \xi)])^T (y - y_0)/\|y - y_0\| \ge \mathbb{E}[\sigma(\xi)]\|y - y_0\|^2,$$

which implies the strong monotonicity of $\mathbb{E}[F(x_0, \cdot, \xi)]$. It is also possible to derive some weaker conditions using Outrata et al. [23, Proposition 1.1], but this is beyond the focus of this paper.

Let us now make a few comments on the second part of Theorem 3.4. The compactness of D and the continuity of $\mathbb{E}[F(x, y, \xi)]$ imply that the feasible set of (3), denoted by $\mathcal{F}$, is bounded. Moreover, for any fixed $\rho > 0$, it is easy to see that $\psi(x, y, z, \rho) \to \infty$ as $\|z\| \to \infty$, which means that there exists a compact $\mathcal{K} \subseteq \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ such that $S^{\rho}_{\text{opt}} \subseteq \mathcal{K}$. We can choose $\mathcal{K}$ sufficiently large such that $S^{\rho}_{\text{opt}} \subset \operatorname{int} \mathcal{K}$, where "int" denotes the interior of a set. Theorem 3.4(ii) states that $S^{\rho}_{\text{opt}} = S_{\text{opt}}$ for a sufficiently large ρ. Following the terminology of Pillo and Grippo [25, Definition 1], such a problem (5) is a *weak exact penalty problem* of (3) for large enough ρ.

It might be interesting to ask whether there exists a penalty parameter $\check{\rho}$ such that for $\rho \ge \check{\rho}$, a local minimizer of problem (5) is a local optimal solution of problem (3) (in the terminology of Pillo and Grippo [25], problem (5) is an *exact penalty* problem of (3)). Unfortunately, we are unable to show the existence of such a parameter because of the complication resulting from partial penalization and nonexistence of the interior of the set of feasible solutions to problem (3), nor can we find a counterexample. We leave this to interested readers to explore. The issue seems to be relatively easier to address when a full penalization (including all constraints) is considered. For instance, Burke [3] derived an exact penalization scheme for a general nonlinear constrained minimization problem under some regularity conditions (of the constraint system) and derived a relationship between local optimal solutions of the penalized problem and their original counterparts, see Burke [3, Corollary 2.3.1].

Let us now use part (ii) of Proposition 3.2 and Liu et al. [14, Theorem 3.3] to derive a relationship between the set of local minimizers of the SAA problem (4) and its penalization (6). The following result states that under some moderate conditions, there exists a bounded sequence of penalty parameters such that a local minimizer of (4) is also a local minimizer of the penalty problem (6).

THEOREM 3.5. *Let (x^N, y^N, z^N) be a local optimal solution to problem* (4) *and (x^*, y^*, z^*) be a limit point of sequence $\{(x^N, y^N, z^N)\}$. Under Assumption* 3.3(a) *(x^*, y^*) is a feasible point of problem* (1)*. If, in addition, MPEC-NNAMCQ* (*or equivalently MPEC-metric regularity*) *holds at (x^*, y^*), then there exists a bounded sequence of penalty parameters $\{\rho_N\}$ such that (x^N, y^N, z^N) is a local optimal solution of* (6).

Proof. We give a sketch of the proof although it is similar to that of Theorem 3.4(i). The feasibility of (x^*, y^*) can be easily verified using Assumption 3.3, which ensures the uniform convergence of the underlying functions. In the proof of Theorem 3.4, we have shown that the MPEC-NNAMCQ at (x^*, y^*) implies the NNAMCQ of the system $\{h(x, y, z) = 0, (x, y, z) \in \mathscr{C}\}$ at (x^*, y^*, z^*) with $z^* = \mathbb{E}[F(x^*, y^*, \xi)]$, where $h(x, y, z) := \mathbb{E}[F(x, y, \xi)] - z$ and $\mathscr{C}$ is defined as in the proof of Theorem 3.4(i). Let $h^N(x, y, z) := (1/N)\sum_{k=1}^{N} F(x, y, \xi^k) - z$ and $S^N := \{(x, y, z): h^N(x, y, z) = 0, (x, y, z) \in \mathscr{C}\}$. By Proposition 3.2(ii) (without $g^N(x)$ here), there exist a bounded sequence of positive numbers $\{\rho_N\}$ and a neighborhood of (x^N, y^N, z^N), denoted by $U_{(x^N, y^N, z^N)}$ such that

$$d((x, y, z), S^N) \le \rho^N \|h^N(x, y, z)\|_1, \quad \forall (x, y, z) \in \mathscr{C} \cap U_{(x^N, y^N, z^N)}.$$

Applying the principle of partial exact penalization (Liu et al. [14, Theorem 3.3]), the inequality above implies (x^N, y^N, z^N) is also a local minimizer of (6) for $\rho \ge \rho_N \kappa_N$, where κ_N converges to the Lipschitz modulus of function $\mathbb{E}[f(x, y, \xi)]$ under Assumption 3.3. □

From numerical perspective, Theorem 3.5 is more useful than part (i) of Theorem 3.4 in that for a given problem, since the distribution ξ is usually unknown in practice, it is often difficult to estimate ρ^*. Through the proof of Proposition 3.2, Theorem 3.5 provides a practical way to set/estimate the penalty parameter ρ_N. Note also that we are short of claiming in Theorem 3.5 that a local optimal solution (x^N, y^N, z^N) to the penalized SAA problem (6) is a local optimal solution to problem (4), although this is obvious if the former has a unique local optimal solution or the local optimal solution to the former falls into the feasible set of the latter.

3.2. Stationary points. It is well known that MPEC problems are notoriously nonconvex because of their combinatorial nature of constraints, which means that we may often obtain a local optimal solution or even a stationary point rather than a global optimal solution in numerical computation. This motivates us to study stationary points of problems (5) and (1) and their relationships. Here, we focus on M-stationary points although our discussion can be extended to Clarke [6] stationary points.

The proposition below states the relationship between M-stationary points of (3) and (1).

Proposition 3.6. *If (x, y, z) is an M-stationary point of problem (3), then (x, y) is an M-stationary of problem (1). Conversely, if (x, y) is an M-stationary point of problem (1), then (x, y, z) is an M-stationary point of problem (3) with $z = \mathbb{E}[F(x, y, \xi)]$.*

Proof. Let (x, y) be an M-stationary point of problem (1). Then, there exist multipliers $(\lambda, \beta) \in \mathbb{R}^m \times \mathbb{R}^m$ such that

$$\begin{cases} 0 \in \nabla\mathbb{E}[f(x, y, \xi)] + \nabla\mathbb{E}[F(x, y, \xi)]^T\lambda + \mathcal{N}_D(x, y) + \{(0, \beta)\}, \\ (\beta, \lambda) \in \mathcal{N}_{\mathscr{W}}(y, \mathbb{E}[F(x, y, \xi)]), \end{cases}$$

where $\mathscr{W} = \{(y, z) \in \mathbb{R}^m \times \mathbb{R}^m: 0 \le z \perp y \ge 0\}$ and the limiting normal cone $\mathcal{N}_{\mathscr{W}}(y, z)$ is defined as in Proposition 2.3. Let (x, y, z) be an M-stationary point of the reformulated problem (3). Then, there exist multipliers $(\lambda, \beta_y, \beta_z) \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^m$ such that

$$\begin{cases} 0 \in \nabla\mathbb{E}[f(x, y, \xi)] + \nabla\mathbb{E}[F(x, y, \xi)]^T\lambda + \mathcal{N}_D(x, y) + \{(0, \beta_y)\}, \\ 0 = -\lambda + \beta_z, \\ (\beta_y, \beta_z) \in \mathcal{N}_{\mathscr{W}}(y, z). \end{cases} \tag{21}$$

The equivalence of the two set of stationary points is straightforward. □

The next proposition describes the relationship between the M-stationary points of (3) and the penalization problem (5).

Proposition 3.7. *If $(x, y, z; \beta_y, \beta_z)$ is an M-stationary pair of problem (5) and $\rho > \|\beta_z\|_1$, then (x, y, z) is an M-stationary point of (3). Conversely, let $(x, y, z; \lambda, \beta_y, \beta_z)$ be an M-stationary pair of problem (3). If $\rho \ge \|\lambda\|_1$, then (x, y, z) is an M-stationary point of problem (5).*

Proof. Problem (5) ⇒ Problem (3). By definition, $(x, y, z; \beta_y, \beta_z)$ satisfies

$$0 \in \partial_{(x,y,z)}\psi(x, y, z, \rho) + \mathcal{N}_D(x, y) \times \{0\} + \{(0, \beta_y, \beta_z)\}. \tag{22}$$

Observe first that norm $\|\cdot\|_1$ is a convex function and $\mathbb{E}[f(x,y,\xi)]$ and $\mathbb{E}[F(x,y,\xi)]-z$ are continuously differentiable functions. By the sum rule (Mordukhovich [21, Proposition 1.107]) and the chain rule (Rockafellar and Wets [28, Theorem 10.6]) or (Clarke [6, Theorem 2.3.10]), we have

$$\partial_{(x,y,z)}\psi(x,y,z,\rho)=\nabla_{(x,y,z)}\mathbb{E}[f(x,y,\xi)]+\rho\nabla_{(x,y,z)}(\mathbb{E}[F(x,y,\xi)]-z)^T\mathcal{G}(x,y,z),$$

where $\mathcal{G}(x,y,z)$ is the set of vectors $b=(b_1,\ldots,b_i,\ldots,b_m)^T$ with

$$b_i=\begin{cases}1 & \text{if } \mathbb{E}[F_i(x,y,\xi)]-z_i>0,\\ [-1,1] & \text{if } \mathbb{E}[F_i(x,y,\xi)]-z_i=0,\\ -1 & \text{if } \mathbb{E}[F_i(x,y,\xi)]-z_i<0.\end{cases} \tag{23}$$

Consequently, (22) can be equivalently written as

$$\begin{cases}0\in\nabla\mathbb{E}[f(x,y,\xi)]+\rho\nabla\mathbb{E}[F(x,y,\xi)]^T\mathcal{G}(x,y,z)+\mathcal{N}_D(x,y)+\{(0,\beta_y)\},\\ 0\in-\rho\mathcal{G}(x,y,z)+\beta_z,\\ (\beta_y,\beta_z)\in\mathcal{N}_{\mathcal{W}}(y,z).\end{cases} \tag{24}$$

In what follows, we show that an M-stationary point (x,y,z) satisfying (24) is an M-stationary point of (3) defined by (21). Let $b^*\in\mathcal{G}(x,y,z)$ be such that

$$\begin{cases}0\in\nabla\mathbb{E}[f(x,y,\xi)]+\rho\nabla\mathbb{E}[F(x,y,\xi)]^Tb^*+\mathcal{N}_D(x,y)+\{(0,\beta_y)\},\\ 0=-\rho b^*+\beta_z,\\ (\beta_y,\beta_z)\in\mathcal{N}_{\mathcal{W}}(y,z).\end{cases} \tag{25}$$

Then $(x,y,z;\rho b^*,\beta_y,\beta_z)$ satisfies (21). To show that it is an M-stationary pair of problem (3), it suffices to prove that (x,y,z) is a feasible point of problem (3) for $\rho>\|\beta_z\|_1$. Assume for a contradiction that there exists an index $1\le i_0\le m$ such that $\mathbb{E}[F_{i_0}(x,y,\xi)]-z_{i_0}\neq 0$. Then we must have $|b^*_{i_0}|=1$. By (25), $0=-\rho b^*_{i_0}+[\beta_z]_{i_0}$. Then $\rho=|\rho b^*_{i_0}|=|[\beta_z]_{i_0}|\le\|\beta_z\|_1$, which contradicts the fact that $\rho>\|\beta_z\|_1$.

Problem (3) $\Rightarrow$ Problem (5). Let $(x,y,z;\lambda,\beta_y,\beta_z)$ be an M-stationary pair of problem (3). Then

$$0\in\partial_{(x,y,z)}\psi(x,y,z,\rho)+\mathcal{N}_D(x,y)\times\{0\}+\{0\}\times\mathcal{N}_{\mathcal{W}}(y,z).$$

Let $\rho\ge\|\lambda\|_1$ and $b^*=\lambda/\rho$. Then $b^*_i\in[-1,1]$ for each $i=1,\ldots,m$, and hence $b^*\in\mathcal{G}(x,y,z)$ since $\mathbb{E}[F(x,y,\xi)]-z=0$. Subsequently, $(x,y,z;\beta_y,\beta_z)$ satisfies (24). The proof is complete. $\square$

4. Uniform convergence. To facilitate the convergence analysis of statistical estimators of optimal solutions and stationary points obtained from solving (6) in the following section, we investigate, in this section, the uniform convergence of the function $\psi_N(x,y,z,\rho_N)$ and its limiting subdifferential to their true counterpart. To this end, we need some technical results related to SAA of the limiting subdifferential of the composition of a locally Lipschitz continuous function and the expected value of a random vector-valued function.

Let $Q(w)$: $\mathbb{R}^m\to\mathbb{R}$ be a locally Lipschitz continuous function and $H(v,\xi)$: $\mathbb{R}^n\times\mathbb{R}^d\to\mathbb{R}^m$ be a continuous function, which is continuously differentiable with respect to v for almost every $\xi\in\Xi$. Let ξ be a random vector with support set $\Xi\subset\mathbb{R}^d$. We consider the following composite function of Q and the expected value of H:

$$G(v):=Q(\mathbb{E}[H(v,\xi)]).$$

Let $\xi^1,\ldots,\xi^N$ be an iid sampling of ξ. Denote by

$$H_N(v):=\frac{1}{N}\sum_{k=1}^N H(v,\xi^k)\qquad\text{and}\qquad G_N(v):=Q(H_N(v)).$$

Under some moderate conditions, it is well known that the classical law of large numbers of random function guarantees that $H_N(v)$ converges to $\mathbb{E}[H(v,\xi)]$ uniformly over any compact subset of $\mathbb{R}^n$. This implies the same convergence for $G_N(v)$ to $G(v)$, see, for instance, Xu [37, §3]. What is less known is the uniform convergence of their (approximate) subdifferentials as a set-valued mapping. The lemma below addresses this.

LEMMA 4.1. *Let $W \subseteq \mathbb{R}^m$ and $V \subseteq \mathbb{R}^n$ be compact sets. Let Q: $\mathbb{R}^m \to \mathbb{R}$ be a locally Lipschitz continuous function and $\mathcal{A}Q$ an abstract subdifferential operator of Q that is compact set valued and uniformly upper semicontinuous on W. Let $\mathcal{A}G_N(v) := \nabla H_N(v)^T \mathcal{A}Q(H_N(v))$ and $\mathcal{A}G(v) := \mathbb{E}[\nabla_v H(v,\xi)]^T \mathcal{A}Q(\mathbb{E}[H(v,\xi)])$ and*

$$\mathcal{V} := \{v \in V\colon \mathbb{E}[H(v,\xi)] \in W\}.$$

Suppose: (a) *$H_N(v)$ and $\nabla H_N(v)$ converge to $\mathbb{E}[H(v,\xi)]$ and $\mathbb{E}[\nabla_v H(v,\xi)]$ uniformly over $\mathcal{V}$, respectively,* (b) *$H(v,\xi)$ and $\nabla H(v,\xi)$ are integrable bounded. Then*

$$\lim_{N\to\infty} \sup_{v\in\mathcal{V}} \mathbb{D}(\mathcal{A}G_N(v), \mathcal{A}G(v)) = 0.$$

The proof is rather standard, we move it to the appendix.

REMARK 4.2. It might be helpful to make some comments on the uniform upper semicontinuity of the abstract subdifferential operator in Lemma 4.1. There are two cases. One is that $\mathcal{A}Q$ is the limiting or Clarke [6] subdifferential, whereas W is a discrete set, which consists of a finite number of points. In this case, the uniform upper semicontinuity comes from the usual pointwise upper semicontinuity of the subdifferential operators. The other case is when Q is convex and $\mathcal{A}Q$ is the ϵ-convex subdifferential defined as follows:

$$\partial_\epsilon Q(w) = \{\zeta \in \mathbb{R}^m\colon Q(w') \geq Q(w) + \zeta^T(w - w') - \epsilon\},$$

where ϵ is a fixed positive number, see Hiriart-Urruty and Lemaréchal [8]. It is well known that $\partial_\epsilon Q(\cdot)$ is convex, compact set-valued and Hausdorff continuous, see Hiriart-Urruty and Lemaréchal [8, Theorem 4.1.3]. In this paper, we consider the case that $Q = \|\cdot\|_1$, which is a convex function.

Let $\mathcal{F}$ denote the feasible set of (3); that is,

$$\mathcal{F} := \{(x,y,z)\colon \mathbb{E}[F(x,y,\xi)] - z = 0;\ 0 \leq z \perp y \geq 0,\ (x,y,z) \in D \times \mathbb{R}^m\}.$$

The proposition below presents the uniform convergence of $\partial_{(x,y,z)}\psi_N(x,y,z,\rho_N)$ to $\partial_{(x,y,z)}\psi(x,y,z,\rho)$ over $\mathcal{F}$ under Assumption 3.3 as simple size N increases.

PROPOSITION 4.3 (UNIFORM ALMOST SURE CONVERGENCE). *Under Assumption 3.3,*

(i) *w.p.*1 $(1/N)\sum_{k=1}^N f(x,y,\xi^k)$ *and* $(1/N)\sum_{k=1}^N F(x,y,\xi^k)$ *converge to* $\mathbb{E}[f(x,y,\xi)]$ *and* $\mathbb{E}[F(x,y,\xi)]$ *uniformly over any compact set in* $\mathbb{R}^n \times \mathbb{R}^m$ *as* $N \to \infty$, *respectively*; $(1/N)\sum_{k=1}^N \nabla_{(x,y)} f(x,y,\xi^k)$ *and* $(1/N)\sum_{k=1}^N \nabla_{(x,y)} F(x,y,\xi^k)$ *converge to* $\mathbb{E}[\nabla_{(x,y)} f(x,y,\xi)]$ *and* $\mathbb{E}[\nabla_{(x,y)} F(x,y,\xi)]$, *respectively, w.p.*1 *uniformly over any compact set in* $\mathbb{R}^n \times \mathbb{R}^m$ *as* $N \to \infty$;

(ii) *if* $\rho_N \to \rho$ *as* $N \to \infty$, *then* $\psi_N(x,y,z,\rho_N)$ *converges to* $\psi(x,y,z,\rho)$ *w.p.*1 *uniformly over any compact subset of* $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$;

(iii) *if D is a compact set and* $\rho_N \to \rho$ *as* $N \to \infty$, *then*

$$\lim_{N\to\infty} \sup_{(x,y,z)\in\mathcal{F}} \mathbb{D}\big(\partial_{(x,y,z)}\psi_N(x,y,z,\rho_N), \partial_{(x,y,z)}\psi(x,y,z,\rho)\big) = 0 \quad \textit{w.p.}1.$$

PROOF. Part (i) can be easily proved by virtue of Ruszczynski and Shapiro [29, §6, Proposition 7]. Part (ii) follows from part (i), the Lipschitz continuity of $\|\cdot\|_1$ and the fact that $\rho_N \to \rho$. Our focus is on part (iii) and we use Lemma 4.1 to prove it. To this end, we verify the conditions of the lemma. Let $v := (x,y,z)$ and $Q(\cdot) := \|\cdot\|_1$. Define

$$\begin{gathered} H(v,\xi) := F(x,y,\xi) - z, \qquad H_N(v) := \frac{1}{N}\sum_{k=1}^N H(v,\xi^k), \\ G(v) := Q(\mathbb{E}[H(v,\xi)]), \qquad G_N(v) := Q(H_N(v)). \end{gathered} \tag{26}$$

Since Q is a convex function and its domain is the whole space $\mathbb{R}^m$, it is locally Lipschitz, which implies that the only horizon subgradient of Q is zero, i.e., $\partial^\infty Q(v) = \{0\}$ for any $v \in \mathbb{R}^m$. Moreover, the convexity of Q implies that it is Clarke [6] regular. Hence, by the chain rule (Rockafellar and Wets [28, Theorem 10.6]), or (Clarke [6, Theorem 2.3.10]) G and G_N are Clarke regular and

$$\partial G(v) = \mathbb{E}[\nabla_v H(v,\xi)]^T \partial Q(\mathbb{E}[H(v,\xi)])$$

and

$$\partial G_N(v) = \nabla H_N(v)^T \partial Q(H_N(v)).$$

Note that in this case, the limiting subdifferential coincides with Clarke [6] subdifferential. Let $W := \{0\}$, $\mathcal{A}Q := \partial Q$, and $\mathcal{V} := \mathcal{F}$. Since W is a singleton, the uniform upper semicontinuity of $\mathcal{A}Q$ over W reduces trivially to (pointwise) upper semicontinuity of the set-valued mapping at point 0. On the other hand, since the feasible set $\mathcal{F}$ is a compact set under the compactness of D, the uniform convergence of $H_N(v)$ and $\nabla H_N(v)$ to $\mathbb{E}[H(v,\xi)]$ and $\mathbb{E}[\nabla_v H(v,\xi)]$ over $\mathcal{V}$ follows from part (i). This verifies all of the conditions in Lemma 4.1, and, consequently, yields

$$\lim_{N\to\infty} \sup_{v\in\mathcal{V}} \mathbb{D}(\partial G_N(v), \partial G(v)) = 0.$$

The rest follows straightforwardly from the fact that

$$\begin{aligned}
&\mathbb{D}(\partial_{(x,y,z)}\psi_N(x,y,z,\rho_N), \partial_{(x,y,z)}\psi(x,y,z,\rho)) \\
&\quad \le c\left\| \frac{1}{N}\sum_{k=1}^N \nabla_{(x,y,z)} f(x,y,\xi^k) - \mathbb{E}[\nabla_{(x,y,z)} f(x,y,\xi)]\right\| + \mathbb{D}(\rho_N \partial G_N(v), \rho\partial G(v))
\end{aligned}$$

and the uniform convergence of $(1/N)\sum_{k=1}^N \nabla_{(x,y,z)} f(x,y,\xi^k)$ to $\mathbb{E}[\nabla_{(x,y,z)} f(x,y,\xi)]$ over $\mathcal{F}$. $\square$

It is important to note that in part (iii) of Proposition 4.3, the uniform convergence is established only on a compact feasible set $\mathcal{F}$: we need compactness for the uniform convergence of $(1/N)\cdot\sum_{k=1}^N \nabla_{(x,y,z)} f(x,y,\xi^k)$ and $(1/N)\sum_{k=1}^N F(x,y,\xi^k)$, and the feasibility to secure the uniform convergence of $\partial_{(x,y,z)}(\|(1/N)\sum_{k=1}^N F(x,y,\xi^k) - z\|_1)$ because $(1/N)\sum_{k=1}^N F(x,y,\xi^k) - z$ uniformly converges to 0 under the feasibility condition. In general, $\partial\|\cdot\|_1$ is not uniformly upper semicontinuous on a set containing a point where the 1-norm is not differentiable.

We move on to investigate the uniform exponential convergence rate of $\psi_N(x,y,z,\rho_N)$ to $\psi(x,y,z,\rho)$ as well as its subdifferentials.

Assumption 4.4. *Let $\mathcal{S}$ be a compact set of $\mathbb{R}^n \times \mathbb{R}^m$ and $\vartheta\colon \mathcal{S}\times\Xi \to \mathbb{R}$ denote an element (component in the case of a vector-valued function or a matrix-valued function) in the set of functions $\{f(x,y,\xi), F(x,y,\xi), \nabla_{(x,y)} f(x,y,\xi), \nabla_{(x,y)} F(x,y,\xi)\}$. $\vartheta(w,\xi)$ possesses the following properties:*

(a) *for every $w\in\mathcal{S}$, the moment-generating function*

$$M(t) := \mathbb{E}[e^{t(\vartheta(w,\xi)-\mathbb{E}[\vartheta(w,\xi)])}]$$

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

(b) *there exist a (measurable) function $\kappa\colon \mathbb{R}^d \to \mathbb{R}_+$ and constant $\gamma > 0$ such that*

$$|\vartheta(w',\xi) - \vartheta(w,\xi)| \le \kappa(\xi)\|w'-w\|^\gamma$$

for all $\xi\in\Xi$ and all $w', w\in\mathcal{S}$;

(c) *the moment-generating function $M_\kappa(t)$ of $\kappa(\xi)$ is finite valued for all t in a neighborhood of zero.*

Assumption 4.4(a) means that the random variable $\vartheta(w,\xi)$ does not have a heavy-tail distribution. In particular, it holds if this random variable has a distribution supported on a bounded subset. Assumption 4.4(b) requires $\vartheta(w,\xi)$ to be globally Hölder continuous with respect to w. Note that this assumption is weaker than Assumption 3.3. Assumption 4.4(c) is satisfied if $\mathbb{E}[\kappa(\xi)]$ is finite.

Theorem 4.5 (Uniform Exponential Convergence). *Let $\mathcal{K}$ be a compact subset of $\mathbb{R}^n\times\mathbb{R}^m\times\mathbb{R}^m$ and Assumption 4.4 hold on the orthogonal projection of $\mathcal{K}$ on (x,y) plane. Suppose $\rho_N\to\rho$ as $N\to\infty$. Then,*

(i) *for any $\alpha > 0$, there exist positive constants $c_2(\alpha)$ and $k_2(\alpha)$, independent of N such that*

$$\mathrm{Prob}\left\{\sup_{(x,y,z)\in\mathcal{K}} |\psi_N(x,y,z,\rho_N) - \psi(x,y,z,\rho)| \ge \alpha\right\} \le c_2(\alpha)e^{-Nk_2(\alpha)}$$

for N sufficiently large;

(ii) *let ϵ be a positive number and $\partial_\epsilon\|\cdot\|_1$ denote the ϵ-convex subdifferential of $\|\cdot\|_1$, let*

$$\mathcal{A}_\epsilon(x,y,z,\rho) = \mathbb{E}[\nabla_{(x,y,z)} f(x,y,\xi)] + \mathcal{A}_\epsilon(x,y,z,\rho) \tag{27}$$

and

$$\mathcal{A}_\epsilon^N(x,y,z,\rho_N) = \frac{1}{N}\sum_{k=1}^N \nabla_{(x,y,z)} f(x,y,\xi^k) + \mathcal{A}_\epsilon^N(x,y,z,\rho_N), \tag{28}$$

where

$$\mathcal{A}_\epsilon(x,y,z,\rho) = \rho\mathbb{E}[\nabla_{(x,y,z)} H(x,y,z,\xi)]^T \partial_\epsilon \|\mathbb{E}[H(x,y,z,\xi)]\|_1$$

and

$$\mathcal{A}_\epsilon^N(x,y,z,\rho_N) = \rho_N \nabla_{(x,y,z)} H_N(x,y,z) \partial_\epsilon \|H_N(x,y,z)\|_1,$$

H and H_N are defined by (26). *Then, for any $\alpha > 0$, there exist positive constants $c_3(\alpha)$ and $k_3(\alpha)$ independent of N such that*

$$\operatorname{Prob}\left\{\sup_{(x,y,z)\in\mathcal{K}} \mathbb{D}(\mathcal{A}_\epsilon^N(x,y,z,\rho_N), \mathcal{A}_\epsilon(x,y,z,\rho)) \ge \alpha\right\} \le c_3(\alpha)e^{-Nk_3(\alpha)}$$

for N sufficiently large.

Proof. Part (i). The claim follows straightforwardly from Shapiro and Xu [33, Theorem 5.1] under Assumption 4.4 and $\rho_N \to \rho$. We omit the details.

Part (ii). Observe first that for any compact sets $A, B, C, D \subseteq \mathbb{R}^m$,

$$\mathbb{D}(A+C, B+D) \le \mathbb{D}(A,B) + \mathbb{D}(C,D). \tag{29}$$

Using the inequality, we have

$$\begin{aligned}
&\operatorname{Prob}\left\{\sup_{(x,y,z)\in\mathcal{K}} \mathbb{D}(\mathcal{A}_\epsilon^N(x,y,z,\rho_N), \mathcal{A}_\epsilon(x,y,z,\rho)) \ge \alpha\right\}\\
&\quad\le \operatorname{Prob}\left\{\sup_{(x,y,z)\in\mathcal{K}} \mathbb{D}\left(\frac{1}{N}\sum_{k=1}^N \nabla_{(x,y,z)} f(x,y,\xi^k),\ \mathbb{E}[\nabla_{(x,y,z)} f(x,y,\xi)]\right) \ge \frac{\alpha}{2}\right\}\\
&\qquad+ \operatorname{Prob}\left\{\sup_{(x,y,z)\in\mathcal{K}} \mathbb{D}(\mathcal{A}_\epsilon^N(x,y,z,\rho_N), \mathcal{A}_\epsilon(x,y,z,\rho)) \ge \frac{\alpha}{2}\right\}.
\end{aligned}$$

By Shapiro and Xu [33, Theorem 5.1], the first term at the right-hand side of the formula above converges to zero at an exponential rate. It suffices to show the second term at the right-hand side of the formula above converges to zero at an exponential rate. By (44),

$$\begin{aligned}
&\operatorname{Prob}\left\{\sup_{(x,y,z)\in\mathcal{K}} \mathbb{D}(\mathcal{A}_\epsilon^N(x,y,z,\rho_N), \mathcal{A}_\epsilon(x,y,z,\rho)) \ge \frac{\alpha}{2}\right\}\\
&\quad\le \operatorname{Prob}\left\{\sup_{(x,y,z)\in\mathcal{K}} \|\rho_N \nabla H_N(x,y,z) - \rho\nabla\mathbb{E}[H(x,y,z,\xi)]\| \ge \frac{\alpha}{\sup_{(x,y,z)\in\mathcal{K}} 4\|\partial_\epsilon\|H_N(x,y,z)\|_1\|}\right\}\\
&\qquad+ \operatorname{Prob}\left\{\sup_{(x,y,z)\in\mathcal{K}} \mathbb{D}(\partial_\epsilon\|H_N(x,y,z)\|_1, \partial_\epsilon\|\mathbb{E}[H(x,y,z,\xi)]\|_1) \ge \frac{\alpha}{\sup_{(x,y,z)\in\mathcal{K}} 4\rho\|\nabla\mathbb{E}[H(x,y,z,\xi)]\|}\right\},
\end{aligned}$$

where $\|\mathcal{M}\| = \sup_{M\in\mathcal{M}}\|M\|$ for a compact set $\mathcal{M}$. Note first that $\|\partial_\epsilon\|H_N(x,y,z)\|_1\|$ is bounded by integer m (problem dimension) for any $(x,y,z)\in\mathcal{K}$. The exponential rate of convergence of the first term, at the right-hand side of the formula above, follows from the fact that $\nabla H_N(x,y,y)$ converges to $\nabla\mathbb{E}[H(x,y,z,\xi)]$ over $\mathcal{K}$ at an exponential rate and $\rho_N\to\rho$. We omit the details. Let us look at the second term on the right-hand side of the formula. Under Assumption 4.4, $\|\nabla\mathbb{E}[H(x,y,z,\xi)]\|$ is bounded on compact set $\mathcal{K}$. On the other hand, $\partial_\epsilon\|\cdot\|_1$ is Hausdorff continuous on $\mathbb{R}^n$ and H_N converges to $\mathbb{E}[H]$ uniformly over $\mathcal{K}$ at the exponential rate, which implies that $\partial_\epsilon\|H_N(x,y,z)\|_1$ converges to $\partial_\epsilon\|\mathbb{E}[H(x,y,z,\xi)]\|_1$ uniformly over $\mathcal{K}$ at the exponential rate. The rest is straightforward. □

Note that the exponential rate of convergence stated in Theorem 4.5 relies on the Hausdorff continuity of the ϵ-convex subdifferential $\partial_\epsilon\|\cdot\|$. This is indeed one of the main reasons that we consider the approximate first-order optimality condition in §5.2. See also the comments at the end of §5.

5. Asymptotic convergence analysis. In the preceding section, we have investigated the uniform convergence of sample average random functions. We are now ready to use them to study the convergence of the statistical estimators obtained from solving (6).

5.1. Optimal solutions. Observe that the penalized SAA problem (6) and the penalized true problem (5) have the same feasible set, and in Proposition 4.3, we have proved that the objective function of (6), $\psi_N(x, y, z, \rho_N)$, converges uniformly to the objective function of (5), $\psi(x, y, z, \rho)$, on any compact subset of $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$. This paves the way for investigating the convergence of optimal solutions through the standard perturbation analysis. Note that a point (x, y) is an optimal solution of problem (1) if and only if (x, y, z) is an optimal solution of problem (3) with $z = \mathbb{E}[F(x, y, \xi)]$.

Theorem 5.1. *Let $\{(x^N, y^N, z^N)\}$ be a sequence of optimal solutions of problem* (6) *and Assumption* 3.3 *hold. Let $\rho_N \to \rho$. Then,*

(i) *w.p.*1 *any accumulation point of the sequence $\{(x^N, y^N, z^N)\}$, denoted by (x^*, y^*, z^*), is an optimal solution of the true penalty problem* (5) *with penalty parameter equal to ρ;*

(ii) *if, in addition,* (a) *D is a compact set,* (b) *MPEC-NNAMCQ holds at every optimal solution of problem* (1), (c) *Assumption* 4.4 *holds,* (d) *$\rho > \bar{\rho}$, where $\bar{\rho}$ is given in Theorem* 3.4, *then for any $\alpha > 0$, there exist positive constants $c(\alpha)$ and $k(\alpha)$ independent of N such that*

$$\text{Prob}\{d((x^N, y^N, z^N), S_{\text{opt}}) \geq \alpha\} \leq c(\alpha)e^{-Nk(\alpha)}$$

for N sufficiently large, where S_{opt} denotes the set of optimal solutions to (3).

Proof. Part (i). The conclusion follows by an application of the uniform convergence of $\psi_N(x, y, z, \rho_N)$ to $\psi(x, y, z, \rho)$ as stated in Proposition 4.3(ii) and Xu [37, Lemma 4.1].

Part (ii). The exponential rate of convergence of (x^N, y^N, z^N) to S_{opt}^{ρ} follows from Theorem 4.5(i) and Xu [37, Lemma 4.1]. Moreover, by Theorem 3.4(ii), $S_{\text{opt}} = S_{\text{opt}}^{\rho}$ for $\rho > \bar{\rho}$. The conclusion follows. □

Note that, in general, it is unrealistic to aim at finding conditions such that

$$\lim_{N\to\infty} (\arg\min \psi_N) = \arg\min \psi,$$

except in cases where $\arg\min \psi$ consists of a single point, see comments in Rockafellar and Wets [28, p. 263].

5.2. Stationary points. We now move on to analyze the convergence of statistical estimators of the stationary points, denoted by (x^N, y^N, z^N), obtained from solving the penalized SAA problem (6). Recall that a feasible point (x^N, y^N, z^N) is said to be an M-stationary point of problem (6) if it satisfies the following first-order optimality condition:

$$0 \in \partial_{(x, y, z)}\psi_N(x, y, z, \rho_N) + \mathcal{N}_D(x, y) \times \{0\} + \{0\} \times \mathcal{N}_{\mathcal{W}}(y, z). \tag{30}$$

From a numerical perspective, it might be difficult to obtain an exact stationary point. This motivates us to consider the following approximate first-order optimality condition:

$$0 \in \mathcal{A}_\epsilon^N(x, y, z, \rho_N) + \mathcal{N}_D(x, y) \times \{0\} + \{0\} \times \mathcal{N}_{\mathcal{W}}(y, z), \tag{31}$$

where $\mathcal{A}_\epsilon^N(x, y, z, \rho_N)$ is defined by (28) and ϵ is a small positive number. Observe that

$$\begin{aligned}\partial_{(x, y, z)}\psi_N(x, y, z, \rho_N) &= \frac{1}{N}\sum_{k=1}^N \nabla_{x, y, z} f(x, y, \xi^k) + \rho\nabla_{x, y, z}H_N(x, y, z)^T \partial\|H_N(x, y, z)\|_1 \\ &\subseteq \mathcal{A}_\epsilon^N(x, y, z, \rho_N)\end{aligned}$$

and

$$\lim_{\epsilon\downarrow 0} \mathcal{A}_\epsilon^N(x, y, z, \rho_N) = \partial_{(x, y, z)}\psi_N(x, y, z, \rho_N).$$

By virtue of a perturbation result of generalized equations (Xu [37, Lemma 4.2]), this means a stationary point defined by (31) converges to an M-stationary point of SAA problem (6) when ϵ is driven to zero and this gives theoretical justification of the "approximation." Likewise, since

$$\partial_{(x, y, z)}\psi(x, y, z, \rho) \subset \mathbb{E}[\nabla_{(x, y, z)}f(x, y, \xi)] + \rho\mathbb{E}[\nabla_{(x, y, z)}H(x, y, z, \xi)]^T \partial\|\mathbb{E}[H(x, y, z, \xi)]\|_1 \subset \mathcal{A}_\epsilon(x, y, z, \rho),$$

where $\mathcal{A}_\epsilon(x, y, z, \rho)$ is defined by (27), we may consider approximate first-order optimality condition for the penalized true problem (5):

$$0 \in \mathcal{A}_\epsilon(x, y, z, \rho) + \mathcal{N}_D(x, y) \times \{0\} + \{0\} \times \mathcal{N}_{\mathcal{W}}(y, z). \tag{32}$$

The theorem below states the convergence of an approximate stationary point satisfying (31) as N increases.

Theorem 5.2. *Let $\{(x_\epsilon^N, y_\epsilon^N, z_\epsilon^N)\}$ be a sequence of ϵ-stationary points defined by (31) and $(x_\epsilon^*, y_\epsilon^*, z_\epsilon^*)$ be a cluster point w.p.1, let Assumptions 3.3 and 4.4 hold and $\rho_N \to \rho$. Then w.p.1 $(x_\epsilon^*, y_\epsilon^*, z_\epsilon^*)$ satisfies (32). Moreover, the convergence rate is exponential.*

Proof. By taking a subsequence if necessarily, we assume for the simplicity of notation that $\{(x_\epsilon^N, y_\epsilon^N, z_\epsilon^N)\} \to (x_\epsilon^*, y_\epsilon^*, z_\epsilon^*)$ w.p.1. The first part of the claim follows from Theorem 4.5(ii) in that the uniform exponential convergence implies the uniform a.s. convergence, which further implies that

$$\limsup_{N\to\infty} \mathcal{A}_\epsilon^N(x_\epsilon^N, y_\epsilon^N, z_\epsilon^N, \rho_N) \subseteq \mathcal{A}_\epsilon \psi(x_\epsilon^*, y_\epsilon^*, z_\epsilon^*, \rho).$$

The second part of the claim follows from Theorem 4.5(ii) and the perturbation theorem of generalized equations (Xu [37, Lemma 4.2]). We omit the details. □

Note that we are short of claiming that a stationary point satisfying (32) is an ϵ-M-stationary point in that the ϵ-convex subdifferential is different from the ϵ-limiting subdifferential, see a discussion by Mordukhovich [21, p. 96]. However, when ϵ is driven to zero, we have

$$\mathcal{A}_\epsilon(x, y, z, \rho) \to \partial\psi(x, y, z, \rho),$$

which means the ϵ-stationary point approximates the M-stationary point of (22) and, through Propositions 3.6 and 3.7, approximates the M-stationary point of true problem (1).

Note also that we are unable to establish the exponential rate of convergence for the M-stationary points of the penalized SAA problem (6) and this is indeed another underlying reason that we consider approximate stationary points in Theorem 5.2.

6. Preliminary numerical test results. In this paper, we proposed essentially two numerical schemes: a partially penalized SAA scheme (6) and a smoothed SAA scheme (7). For a given sample, the former is a deterministic MPEC with a nonsmooth objective function, whereas the latter is a specific smoothing of the former. We have carried out some numerical experiments on (7) and present a report of the test results in this section.

6.1. Convergence analysis of the smoothing scheme. The convergence analysis carried out in the preceding section is based on the assumption that an optimal solution or a stationary point is obtained from solving the partially penalized SAA problem (6). In doing so, we allow the SAA problem (6) to be solved by any deterministic MPEC solver, which can effectively deal with the nonsmoothness in the objective function. The convergence results, however, do not cover (7) as the smoothing parameter δ_N is positive. To fill out the gap, we start this section with a brief convergence analysis of (7).

Proposition 6.1. *Let $\{(x^N, y^N, z^N)\}$ be a sequence of optimal solutions of problem (7) and Assumption 3.3 hold. Let $\rho_N \to \rho$. Then,*

(i) *w.p.1 any accumulation point of the sequence $\{(x^N, y^N, z^N)\}$, denoted by (x^*, y^*, z^*), is an optimal solution of the true penalty problem (5) with the penalty parameter equal to ρ;*

(ii) *if, in addition, conditions (a)–(d) in Theorem 5.1 hold, then for any $\alpha > 0$, there exist positive constants $c(\alpha)$ and $k(\alpha)$, independent of N such that*

$$\mathrm{Prob}\{d((x^N, y^N, z^N), S_{\mathrm{opt}}) \geq \alpha\} \leq c(\alpha)e^{-Nk(\alpha)}$$

for N sufficiently large, where S_{opt} denotes the set of optimal solutions to (3).

The proof is essentially similar to Theorem 5.1 in that

$$\sum_{i=1}^{m} \sqrt{\left(\frac{1}{N}\sum_{k=1}^{N} F_i(x, y, \xi^k) - z_i\right)^2 + \delta_N} \longrightarrow \|\mathbb{E}[F(x, y, \xi)] - z\|_1$$

uniformly on any compact set of $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ at an exponential rate as the smoothing parameter $\delta_N \to 0$ as $N \to \infty$. Likewise, we can establish the convergence of stationary points generated by the scheme.

Proposition 6.2. *Let $\{(x^N, y^N, z^N; \beta_y^N, \beta_z^N)\}$ be a sequence of Karush-Kuhn-Tucker (KKT) pair of problem (7) and $(x^*, y^*, z^*; \beta_y^*, \beta_z^*)$ be an accumulation point. Suppose Assumption 3.3 holds. If $\rho_N \to \rho$, $\delta_N \to 0$ and $\rho > \|\beta_z^*\|_1$, then w.p.1 (x^*, y^*) is an M-stationary point of the true problem (1).*

Proof. Taking a subsequence if necessary, we may assume that

$$\lim_{N\to\infty} (x^N, y^N, z^N; \beta_y^N, \beta_z^N) = (x^*, y^*, z^*; \beta_y^*, \beta_z^*).$$

By definition, the M-stationary pair $(x^N, y^N, z^N; \beta_y^N, \beta_z^N)$ satisfies $(\beta_y^N, \beta_z^N) \in \mathcal{N}_{\mathcal{W}}(y^N, z^N)$ and

$$0 \in \nabla_{(x,y,z)} \hat{\psi}_N(x^N, y^N, z^N, \rho_N, \delta_N) + \mathcal{N}_D(x^N, y^N) \times \{0\} + \{(0, \beta_y^N, \beta_z^N)\}, \tag{33}$$

where

$$\nabla_{(x,y,z)} \hat{\psi}_N(x, y, z, \rho_N, \delta_N) = \frac{1}{N}\sum_{k=1}^N \nabla_{(x,y,z)} f(x, y, \xi^k) + \rho_N \sum_{i=1}^m \varpi_i^N(x, y, z) \varrho_i^N(x, y, z)$$

and

$$\varpi_i^N(x, y, z) = \frac{(1/N)\sum_{k=1}^N F_i(x, y, \xi^k) - z_i}{\sqrt{((1/N)\sum_{k=1}^N F_i(x, y, \xi^k) - z_i)^2 + \delta_N}} \tag{34}$$

$$\varrho_i^N(x, y, z) = \nabla_{(x,y,z)}\left(\frac{1}{N}\sum_{k=1}^N F_i(x, y, \xi^k) - z_i\right).$$

By Proposition 4.3(i), $(1/N)\sum_{k=1}^N F_i(x, y, \xi^k) - z_i$, $\sqrt{((1/N)\sum_{k=1}^N F_i(x, y, \xi^k) - z_i)^2 + \delta_N}$, and $\varrho_i^N(x, y, z)$ converge to $\mathbb{E}[F_i(x, y, \xi)] - z_i$, $|\mathbb{E}[F_i(x, y, \xi)] - z_i|$, and $\nabla_{(x,y,z)}(\mathbb{E}[F_i(x, y, \xi)] - z_i)$ uniformly on any compact set in $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$, respectively, for $i = 1, \ldots, m$. Let

$$A^N(x, y, z) := [\varpi_1^N(x, y, z), \ldots, \varpi_m^N(x, y, z)].$$

Since $|\varpi_i^N(x, y, z)| \le 1$ for any $(x, y, z) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ and w.p.1,

$$\lim_{N\to\infty} \varpi_i^N(x^N, y^N, z^N) \in \begin{cases} \{1\}, & \mathbb{E}[F_i(x^*, y^*, \xi)] - z_i^* > 0, \\ [-1, 1], & \mathbb{E}[F_i(x^*, y^*, \xi)] - z_i^* = 0, \\ \{-1\}, & \mathbb{E}[F_i(x^*, y^*, \xi)] - z_i^* < 0, \end{cases}$$

then the limit of the sequence $\{A^N(x^N, y^N, z^N)\}$ is contained in the set $\mathcal{G}(x^*, y^*, z^*)$ w.p.1, where $\mathcal{G}(x, y, z)$ is defined by (23). Consequently, we have

$$\lim_{N\to\infty} d(\nabla_{(x,y,z)} \hat{\psi}_N(x^N, y^N, z^N, \rho_N, \delta_N), \partial_{(x,y,z)} \psi(x^*, y^*, z^*, \rho)) = 0. \tag{35}$$

By (33) and the property of $\mathbb{D}$,

$$\begin{aligned} &d(0, \partial_{(x,y,z)} \psi(x^*, y^*, z^*, \rho) + \mathcal{N}_D(x^*, y^*) \times \{0\} + \{(0, \beta_y^*, \beta_z^*)\}) \\ &\quad \le \mathbb{D}(\nabla_{(x,y,z)} \hat{\psi}_N(x^N, y^N, z^N, \rho_N, \delta_N) + \mathcal{N}_D(x^N, y^N) \times \{0\} + \{(0, \beta_y^N, \beta_z^N)\}, \\ &\qquad\quad \partial_{(x,y,z)} \psi(x^*, y^*, z^*, \rho) + \mathcal{N}_D(x^*, y^*) \times \{0\} + \{(0, \beta_y^*, \beta_z^*)\}) \\ &\quad \le d(\nabla_{(x,y,z)} \psi_N(x^N, y^N, z^N, \rho_N, \delta_N), \partial_{(x,y,z)} \psi(x^*, y^*, z^*, \rho)) \\ &\qquad + \mathbb{D}(\mathcal{N}_D(x^N, y^N) \times \{0\} + \{(0, \beta_y^N, \beta_z^N)\}, \mathcal{N}_D(x^*, y^*) \times \{0\} + \{(0, \beta_y^*, \beta_z^*)\}), \end{aligned}$$

where the first inequality follows from the definition of $\mathbb{D}$ and the second inequality follows from (29). The first term at the right-hand side of the second inequality of the formula above tends to zero by (35); the second term tends to zero by the upper semicontinuity of the limiting normal cone mapping $\mathcal{N}_D(\cdot)$ and $(\beta_y^N, \beta_z^N) \to (\beta_y^*, \beta_z^*)$. This shows w.p.1 (x^*, y^*, z^*) is an M-stationary point of the true penalty problem (5). The rest follows straightforwardly from Propositions 3.6 and 3.7. $\square$

From a practical point of view, it might be interesting to estimate the penalty parameter ρ_N in problem (7). The proposition below provides some insights about how this could be possibly achieved through the Lagrange multipliers of the problem.

Proposition 6.3. *Let $\{(x^N, y^N, z^N; \beta_y^N, \beta_z^N)\}$ be a sequence of KKT pair of problem (7) and the penalty parameter ρ^N satisfies $\rho_N \ge \|\beta_z^N\|_1 + 1$. Let (x^*, y^*, z^*) be a limiting point of $\{(x^N, y^N, z^N)\}$. Suppose Assumption 3.3 holds. If the extended MPEC-NNAMCQ holds at (x^*, y^*, z^*), then w.p.1 (β_y^N, β_z^N) is bounded and (x^*, y^*) is an M-stationary point of the true problem (1).*

Proof. We first show the boundedness of (β_y^N, β_z^N). Assume for a contradiction that this is not the case. Let $t_N = \|(\beta_y^N, \beta_z^N)\|$. Then $t_N \to \infty$. Dividing (33) by t_N, we have

$$0 \in \nabla_{(x,y,z)} \hat{\psi}_N(x^N, y^N, z^N, \rho_N, \delta_N)/t_N + \mathcal{N}_D(x^N, y^N) \times \{0\} + \{(0, \beta_y^N/t_N, \beta_z^N/t_N)\}.$$

Under Assumption 3.3, $\mathbb{E}[\nabla_{(x,y,z)} f(x^N, y^N, \xi)]$ is bounded. Taking a limit on both sides of the formula above, we have

$$0 \in \sum_{i=1}^m \theta_i \nabla_{(x,y,z)}(\mathbb{E}[F_i(x^*, y^*, \xi)] - z_i^*) + \mathcal{N}_D(x^*, y^*) \times \{0\} + \{(0, \beta_y, \beta_z)\},$$

where

$$\theta_i = \lim_{N\to\infty} \rho_N \varpi_i^N(x^N, y^N, z^N)/t_N, \qquad \beta_y = \lim_{N\to\infty} \beta_y^N/t_N, \qquad \beta_z = \lim_{N\to\infty} \beta_z^N/t_N,$$

and $\varpi_i^N(\cdot,\cdot,\cdot)$ is defined by (34). Note that $\|(\beta_y, \beta_z)\| = 1$, which contradicts the extended MPEC-NNAMCQ holds at point (x^*, y^*, z^*). Then shows the boundedness of (β_y^N, β_z^N) as desired.

The boundedness of (β_y^N, β_z^N) implies the boundedness of $\|\beta_z^N\|_1$, which means that we can choose a bounded sequence $\{\rho_N\}$ such that $\rho_N \ge \|\beta_z^N\|_1 + 1$. Let

$$(x^*, y^*, z^*; \beta_y^*, \beta_z^*; \rho^*) = \lim_{N\to\infty} (x^N, y^N, z^N; \beta_y^N, \beta_z^N; \rho_N),$$

and note that $\rho^* \ge \|\beta_z^*\|_1 + 1$. By Propositions 3.6 and 3.7, (x^*, y^*) is an M-stationary point of problem (1). The proof is complete. $\square$

6.2. Numerical implementation. We carried out a number of numerical experiments on (7) in Matlab R2009a installed in a PC with Windows XP operating system. In the tests, we employed the random number generator `rand` in Matlab R2009a to generate the samples and solver `fmincon` to solve problem (7). To deal with the complementary constraint $0 \le y \perp z \ge 0$, we use the well-known regularization method (Scholtes [30]) in the literature of MPEC to approximate it with a parameterized system of inequalities

$$y \ge 0, \qquad z \ge 0, \qquad y \circ z \le te,$$

where $t \downarrow 0$ is a small positive parameter, $e \in \mathbb{R}^m$ is a vector with components 1 and "$\circ$" denotes the Hadamard product. The approximation is theoretically guaranteed as the complementarity constraints satisfy the MPEC-LICQ at any feasible point.

We have constructed five academic problems for the tests. The first problem is a one-stage SMPCC with two decision variables and one random parameter:

$$\begin{aligned} \min\ & \mathbb{E}[(x^2 + y^2) - \xi] \\ \text{s.t.}\ & x \ge 1, \\ & 0 \le y \perp \mathbb{E}[-x + y + 2 + \xi^2] \ge 0, \end{aligned} \tag{36}$$

where ξ satisfies the uniform distribution on Burke [3] and Castaing and Valadier [4]. The example is varied from a deterministic MPEC example in Luo et al. [17, p. 12]. Through some elementary calculations, we can easily obtain a closed form; that is, the expected values of the underlying functions, and hence the SMPCC can be transformed into a deterministic MPCC. The problem has a unique optimal solution $(\sqrt{3.5}, 0)$ with optimal value 0. We consider this example purely for testing the performance of our proposed numerical scheme.

The second test problem is also a one-stage SMPCC with two decision variables and one random parameter:

$$\begin{aligned} \min\ & \mathbb{E}[\cos(y\xi)] + x^2 + y^2 \\ \text{s.t.}\ & x \ge 0, \\ & 0 \le y \perp \mathbb{E}[3\sin(x\xi) + y - 1] \ge 0, \end{aligned} \tag{37}$$

where the random variable ξ satisfies uniform distribution on $(0,1]$. Different from (36), it is difficult to obtain a closed form of the expected values of the underlying random functions and true optimal solution to the problem.

The third test problem is a combination of (36) and (37):

$$\begin{aligned}
\min\ & \mathbb{E}[\cos(y_1\xi)]+x_1+x_2+y_2\\
\text{s.t.}\ & x_1\geq 4, \quad x_2\geq 0,\\
& 0\leq y_1\perp \mathbb{E}[y_1\xi+x_1-y_2]\geq 0,\\
& 0\leq y_2\perp \mathbb{E}\left[\cos(x_2\xi)+y_1+y_2+\frac{\xi^2}{4}-\pi\right]\geq 0,
\end{aligned}\tag{38}$$

where the random variable ξ satisfies uniform distribution over $(0,1]$. Different from (37), we know the true optimal solution $(x^*,y^*)=((4,0),(0,\pi-13/12))$ with optimal value $\pi+47/12$.

The fourth test problem is:

$$\begin{aligned}
\min\ & \mathbb{E}[(2(x-1)^2+y_1^2+(y_2-1)^2+(y_3-1)^2+y_4x)\xi]\\
\text{s.t.}\ & x\geq 1,\\
& \mathbb{E}[0\leq y_1\xi+y_2^2+y_3^2+x^2]\perp y_1\geq 0,\\
& 0\leq \mathbb{E}[-y_2-\xi+2x\xi+y_4]\perp y_2\geq 0,\\
& 0\leq \mathbb{E}[x\xi+y_4-2y_2y_3\xi]\perp y_3\geq 0,\\
& 0\leq \mathbb{E}[xy_1+2y_4-4\xi+y_1]\perp y_4\geq 0,
\end{aligned}\tag{39}$$

where the random variable ξ satisfies uniform distribution over $(0,1]$. In what follows, we analyze the feasible solution of the problem. For any fixed $x\geq 1$, to ensure the first and fourth complementarity constraints hold, we must have $y_1=0$ and $y_4=1$. Substituting $y_1=0$ and $y_4=1$ into the second and third complementarity constraints, we obtain the following: (a) $y_2=0$, $y_3=0$; (b) $y_2=\mathbb{E}[2x\xi]+y_4-\mathbb{E}[\xi]$, $y_3=1$; (c) $y_2=\mathbb{E}[2x\xi]+y_4-\mathbb{E}[\xi]$, $y_3=0$. Through a simple analysis, we find the optimal solution is $(1,0,1.5,1,1)$ with the corresponding optimal value 1.25. Moreover, $(1,0,0,0,1)$ and $(1,0,1.5,0,1)$ are only a local minimizer.

The fifth example is varied from a deterministic MPEC problem in Luo et al. [16, p. 357]:

$$\begin{aligned}
\min\ & \mathbb{E}[2(-8x_1-4x_2+4y_1-40y_2-4y_3)\xi]\\
\text{s.t.}\ & x_i\geq 0, \quad i=1,2,\\
& x_1+2x_2-y_3\leq 1.3,\\
& 0\leq \mathbb{E}[(4-2y_4-4y_5+8y_6)\xi]\perp y_1\geq 0,\\
& 0\leq \mathbb{E}[1+2y_4\xi+4y_5-2y_6]\perp y_2\geq 0,\\
& 0\leq 2+y_4-y_5-y_6\perp y_3\geq 0,\\
& 0\leq \mathbb{E}[(2+2y_1-2y_2-2y_3)\xi]\perp y_4\geq 0,\\
& 0\leq \mathbb{E}[(4-8x_1+4y_1-8y_2+2y_3)\xi]\perp y_5\geq 0,\\
& 0\leq \mathbb{E}[2-8x_2\xi-8y_1\xi+2y_2+y_3]\perp y_6\geq 0,
\end{aligned}\tag{40}$$

where the random variable ξ satisfies uniform distribution over $(0,1]$. The MPEC problem is obtained from a primal-dual formulation for a bilevel optimization problem with (y_4,y_5,y_6) being the dual variables. As discussed in Luo et al. [16], the optimal solution $x=(0.5,0.8)$, $y=(0,0.2,0.8)$ and the optimal value is -18.4. In our test, we use the same initial point as in Luo et al. [16]; that is,

$$x^0=(0.5,1), \qquad y^0=(0.5,0.5,1,1,1,1), \qquad z^0=(1,0.1,0.1,0.1,0.1,0.1).$$

The numerical results are displayed in Tables 1–5. A few words about the notation. *Iter* denotes the number of iterations returned by `fmincon` at the end of each test, *Appr.Sol* denotes the approximate optimal solution, and *Appr.Val* denotes the optimal value $(1/N)\sum_{k=1}^{N}f(x^N,y^N,\xi^k)$. To check the feasibility of the approximate optimal solution, we also recorded the residual value of the constraints denoted by *Res*, which is defined as

Table 1. Numerical results for problem (36).

N	δ_N	*iter*	*Appr.Sol* x^N	*Appr.Sol* y^N	*Res*	*Appr.Val*
50	10^{-3}	36	1.858700	0.000000	3.1×10^{-7}	6.6×10^{-8}
100	10^{-4}	31	1.868530	0.000000	8.4×10^{-8}	1.1×10^{-6}
200	10^{-5}	36	1.865917	0.000000	4.1×10^{-8}	1.2×10^{-8}
400	10^{-6}	40	1.866691	0.000000	3.1×10^{-6}	3.8×10^{-7}

Table 2. Numerical results for problem (37).

N	δ_N	*iter*	*Appr.Sol* x^N	*Appr.Sol* y^N	*Res*	*Appr.Val*
50	10^{-3}	23	0.428301	0.356502	0.000020	1.288359
100	10^{-4}	32	0.417516	0.326517	0.000006	1.261283
200	10^{-5}	61	0.426770	0.349296	0.000002	1.282946
400	10^{-6}	42	0.436715	0.341144	0.000001	1.279425

Table 3. Numerical results for problem (38).

N	δ_N	*iter*	*Appr.Sol* x^N	*Appr.Sol* y^N	*Res*	*Appr.Val*
50	10^{-3}	27	(4.000000, 0.000000)	$(1.3 \times 10^{-5}, 2.049814)$	0.001143	7.049814
100	10^{-4}	36	(4.000000, 0.000000)	$(5.2 \times 10^{-5}, 2.058361)$	0.000564	7.058361
200	10^{-5}	41	(4.000000, 0.000000)	$(5.1 \times 10^{-6}, 2.055591)$	0.000347	7.055591
400	10^{-6}	30	(4.000000, 0.000000)	$(2.5 \times 10^{-7}, 2.058639)$	0.000022	7.058639

Table 4. Numerical results for problem (39).

N	δ_N	*iter*	*Appr.Sol* x^N	*Appr.Sol* y^N	*Res*	*Appr.Val*
100	10^{-3}	40	1.000000	(0.000160, 0.000656, 0.000656, 1.015981)	0.000134	3.078464
400	10^{-4}	52	1.000000	(0.000030, 1.502982, 0.000067, 1.002028)	2.10×10^{-10}	2.259537
800	10^{-5}	84	1.046554	(0.000001, 1.533056, 0.000007, 0.991277)	2.11×10^{-8}	2.303452
1,600	10^{-6}	83	1.004118	(0.000000, 1.504954, 0.998079, 1.000559)	1.27×10^{-9}	1.260364

Table 5. Numerical results for problem (40).

N	δ_N	*iter*	*Appr.Sol* x^N	*Appr.Sol* y^N	*Res*	*Appr.Val*
100	10^{-3}	87	(0.684957, 0.772538)	(0.000000, 0.052389, 0.949784, 0.799280, 0.633791, 2.162078)	0.003198	−14.203291
400	10^{-4}	61	(0.593800, 0.790798)	(0.000012, 0.124951, 0.875395, 0.657016, 0.610835, 2.046179)	0.038421	−15.852703
800	10^{-5}	85	(0.541006, 0.795950)	(0.000001, 0.167201, 0.832907, 0.000000, 0.500003, 1.499985)	0.045778	−16.952109
1,600	10^{-6}	86	(0.507108, 0.799307)	(0.000000, 0.194311, 0.805721, 0.005924, 0.500988, 1.504936)	0.132091	−19.093353

$\|(1/N)\sum_{k=1}^{N} F(x^N, y^N, \xi^k) - z^N\|_1$. The regularization parameter $t = \delta_N$ and the exact penalty is fixed with $\rho = 1{,}000$. For fixed-sample size N and parameter δ_N, we run the algorithm three times. The results depend on sampling in each run and we record the best result. Note that `fmincon` requires an initial point. We set the initial point to be a zero vector for problems (36)–(38), $(1, 1, 1, 1, 1, 1, 1, 1, 1)$ for problem (39), and (x^0, y^0, z^0) for problem (40).

The results show that the numerical scheme performed reasonably well, but more tests might be needed to confirm the claim. Note that the results rely heavily on the Matlab built-in NLP solver fmincon. It would be possible to display stronger results if one uses a more robust NLP solver. Moreover, it might be interesting to carry out numerical tests on (6). This may require to develop a numerical method, which incorporates the existing MPEC solvers with well-known techniques in nonsmooth optimization such as the bundle method and aggregate subgradient method (Lemarchal [11], Kiewiel [10]). We leave this for our future work.

6.3. Concluding remarks. The results established in this paper are presented in terms of M-stationary points in that from a theoretical point of view, M-stationarity is stronger than C-stationarity. However, from numerical perspective, it is often easier to obtain a C-stationary point than an M-stationary point, as the latter usually requires more conditions, see comments in Scholtes [30]. It is therefore interesting to know whether our results in this paper can be extended to C-stationary point. The answer is "yes." Let us sketch how this works: if we reformulate the complementarity constraint $0 \le y \perp z \ge 0$ as a nonsmooth system of equations $\Phi(y, z) := \min(y, z) = 0$, then all of the optimality conditions and convergence results will be in the sense of Clarke's [6]. We omit the details.

Note also that it is possible to include ordinary stochastic equality and inequality constraints into SMPCC model (1). Under some appropriate metric regularity conditions as we discussed in §3, we can move these constraints to the objective through exact partial penalization. In other words, the partial penalization scheme and the SAA in this paper apply to classical stochastic programs with stochastic equality and inequality constraints (by dropping the complementarity constraints). This complements the existing asymptotic and/or stability analysis by Shapiro [31] and Rachev and Römisch [26] for this type of problems.

Acknowledgments. The work of the first author was carried out while he was visiting the second author at the School of Mathematics, University of Southampton, sponsored by China Scholarship Council. The authors thank the two anonymous referees for valuable comments, and the associate editor for effective handling of the review.

Appendix.

Proof of Lemma 4.1. Observe first that the uniform convergence of $H_N(v)$ to $\mathbb{E}[H(v, \xi)]$ is equivalent to

$$H_N(v) \in \mathbb{E}[H(v, \xi)] + \delta\mathcal{B} \quad \text{w.p.1} \tag{41}$$

for N sufficiently large. Let us estimate $\sup_{v\in\mathcal{V}} \mathbb{D}(\mathcal{A}G_N(v), \mathcal{A}G(v))$. To this end, we need to review some elementary properties of $\mathbb{D}$. Let $\mathcal{D}_1$ and $\mathcal{D}_2$ be two compact subsets in $\mathbb{R}^m$ and M_1 and M_2 be two matrices in $\mathbb{R}^{n\times m}$. It is easy to verify that

$$\mathbb{D}(M_1\mathcal{D}_1, M_1\mathcal{D}_2) \le \|M_1\|\mathbb{D}(\mathcal{D}_1, D_2) \tag{42}$$

and

$$\mathbb{D}(M_1\mathcal{D}_1, M_2\mathcal{D}_1) \le \|M_1 - M_2\|\|\mathcal{D}_1\|. \tag{43}$$

Using (42)–(43) and the triangle inequality of $\mathbb{D}$, we have

$$\begin{aligned}
&\mathbb{D}(\nabla H_N(v)^T \mathcal{A}Q(H_N(v)), \nabla\mathbb{E}[H(v, \xi)]^T \mathcal{A}Q(\mathbb{E}[H(v, \xi)])) \\
&\quad \le \mathbb{D}\big(\nabla H_N(v)^T \mathcal{A}Q(H_N(v)), \nabla\mathbb{E}[H(v, \xi)]^T \mathcal{A}Q(H_N(v))\big) \\
&\qquad + \mathbb{D}\left(\nabla\mathbb{E}[H(v, \xi)]^T \mathcal{A}Q(H_N(v)), \nabla\mathbb{E}[H(v, \xi)]^T \mathcal{A}Q(\mathbb{E}[H(v, \xi)])\right) \\
&\quad \le \|\mathcal{A}Q(H_N(v))\|\|\nabla H_N(v) - \nabla\mathbb{E}[H(v, \xi)]\| \\
&\qquad + \|\nabla\mathbb{E}[H(v, \xi)]\|\mathbb{D}(\mathcal{A}Q(H_N(v)), \mathcal{A}Q(\mathbb{E}[H(v, \xi)])).
\end{aligned} \tag{44}$$

We estimate the last two terms in the above equation. By (41),

$$\|\mathcal{A}Q(H_N(v))\| \le \|\mathcal{A}Q(\mathbb{E}[H(v, \xi)]) + \delta\mathcal{B}\|.$$

The right-hand side in the inequality is bounded for all $v \in \mathcal{V}$ since $\mathcal{A}Q$ is compact set valued and uniformly upper semicontinuous on W. On the other hand, $\nabla H_N(v)$ converges to $\nabla\mathbb{E}[H(v, \xi)]$ uniformly on $\mathcal{V}$, this shows $\|\mathcal{A}Q(H_N(v))\|\|\nabla H_N(v) - \nabla\mathbb{E}[H(v, \xi)]\| \to 0$ uniformly on $\mathcal{V}$. Note that under integrable boundedness of $\nabla H(v, \xi)$, $\nabla\mathbb{E}[H(v, \xi)] = \mathbb{E}[\nabla_v H(v, \xi)]$. To complete the proof, we estimate the second term. By assumption, $\mathcal{A}Q(w)$ is uniformly upper semicontinuous on W, which means that for any ϵ, there exists a δ_1 such that

$$\mathcal{A}Q(w') \subseteq \mathcal{A}Q(w) + \epsilon\mathcal{B}, \quad \forall\, w' \in w + \delta_1\mathcal{B} \text{ and } w \in W.$$

Let $v \in \mathcal{V}$ and $w = \mathbb{E}[H(v, \xi)]$. Then $w \in W$. By (41), we have from the inclusion above by setting $\delta \leq \delta_1$

$$\mathcal{A}Q(H_N(v)) \subseteq \mathcal{A}Q(\mathbb{E}[H(v, \xi)]) + \epsilon\mathcal{B}, \quad \forall\, v \in \mathcal{V},$$

which is equivalent to

$$\sup_{v \in \mathcal{V}} \mathbb{D}(\mathcal{A}Q(H_N(v)), \mathcal{A}Q(\mathbb{E}[H(v, \xi)])) \leq \epsilon.$$

The conclusion follows as ϵ can be arbitrarily small and $\|\nabla \mathbb{E}[H(v, \xi)]\|$ is bounded. The proof is complete. □

References

[1] Aumann, R. J. 1965. Integrals of set-valued functions. *J. Math. Anal. Appl.* **12**(1) 1–12.
[2] Birbil, S. I., G. Gürkan, O. Listes. 2006. Solving stochastic mathematical programs with complementarity constraints using simulation. *Math. Oper. Res.* **31**(4) 739–760.
[3] Burke, J. V. 1991. An exact penalization viewpoint of constrained optimization. *SIAM J. Control Optim.* **29**(4) 968–998.
[4] Castaing, C., M. Valadier. 1977. *Convex Analysis and Measurable Multifunctions, Lecture Notes in Mathematics*, Vol. 580. Springer-Verlag, Berlin/Heidelberg.
[5] Christiansen, S., M. Patriksson, L. Wynter. 2001. Stochastic bilevel programming in structural optimization. *Structual Multidisciplinary Optim.* **21**(5) 361–371.
[6] Clarke, F. H. 1983. *Optimization and Nonsmooth Analysis*. John Wiley & Sons, New York.
[7] Dontchev, A. L., A. S. Lewis, R. T. Rockafellar. 2003. The radius of metric regularity. *Trans. AMS* **355**(2) 493–517.
[8] Hiriart-Urruty, J.-B., C. Lemaréchal. 1993. *Convex Analysis and Minimization Algorithms: II*. Springer-Verlag, Berlin/Heidelberg.
[9] Ioffe, A. D., J. V. Outrata. 2008. On metric and calmness qualification conditions in subdifferential calculus. *Set-Valued Anal.* **16**(2–3) 199–227.
[10] Kiewiel, K. C. 1985. *Methods of Descent for Nondifferentiable Optimization, Lecture Notes in Mathematics*, Vol. 1133. Springer-Verlag, New York.
[11] Lemarchal, C. 1978. Bundle methods in nonsmooth optimization. C. Lemarechal, R. Mifflin, eds. *Nonsmooth Optimization*. Pergamon Press, Oxford, UK, 79–102.
[12] Lin, G.-H., X. Chen, M. Fukushima. 2009. Solving stochastic mathematical programs with equilibrium constraints via approximation and smoothing implicit programming with penalization. *Math. Programming* **116**(1) 343–368.
[13] Lin, G.-H., H. Xu, M. Fukushima. 2008. Monte Carlo and quasi-Monte Carlo sampling methods for a class of stochastic mathematical programs with equilibrium constraints. *Math. Methods Oper. Res.* **67**(3) 423–441.
[14] Liu, G., J. J. Ye, J. Zhu. 2008. Partial exact penalty for mathematical program with equilibrium constraints. *Set-Valued Anal.* **16**(5–6) 785–804.
[15] Lucet, Y., J. J. Ye. 2002. Erratum: Sensitivity analysis of the value function for optimization problems with variational inequality constraints. *SIAM J. Control Optim.* **41**(4) 1315–1319.
[16] Luo, Z. Q., J.-S. Pang, D. Ralph. 1996. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, Cambridge, UK.
[17] Luo, Z. Q., J.-S. Pang, D. Ralph. 1998. Piece-wise sequential quadratic programming for mathematical programs with nonlinear complementarity constraints. A. Migdalas, P. Pardas, P. Värbrand, eds. *Multilevel Optimization: Algorithms Complexity and Application*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 209–229.
[18] Meng, F., H. Xu. 2006. Exponential convergence of sample average approximation methods for a class of stochastic mathematical programs with complementarity constraints. *J. Comput. Math.* **24**(6) 733–748.
[19] Mordukhovich, B. S. 1993. Complete characterization of openness, metric regularity, and Lipschitzian properties of multifunctions. *Trans. Amer. Math. Soc.* **340**(1) 1–35.
[20] Mordukhovich, B. S. 1994. Generalized differential calculus for nonsmooth and set-valued mappings. *J. Math. Anal. Appl.* **183** 250–288.
[21] Mordukhovich, B. S. 2006. *Variational Analysis and Generalized Differentiation I: Basic Theory.* A Series of Comprehensive Studies in Mathematics, Vol. 330. Springer-Verlag, Berlin/Heidelberg.
[22] Outrata, J. V. 1999. Optimality conditions for a class of mathematical programs with equilibrium constraints. *Math. Oper. Res.* **24**(3) 627–644.
[23] Outrata, J., M. Kočvara, J. Zowe. 1998. *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints: Theory, Applications and Numerical Constraints*. Kluwer Publishers, Dordrecht, The Netherlands.
[24] Patriksson, M., L. Wynter. 1999. Stochastic mathematical programs with equilibrium constraints. *Oper. Res. Lett.* **25**(4) 159–167.
[25] Pillo, G. D., L. Grippo. 1989. Exact penalty functions in constrained optimization. *SIAM J. Optim.* **27**(6) 1333–1360.
[26] Rachev, S. T., W. Römisch. 2002. Quantitative stability in stochastic programming: The method of probability metrics. *Math. Oper. Res.* **27**(4) 792–818.
[27] Robinson, S. M. 1996. Analysis of sample-path optimization. *Math. Oper. Res.* **21**(3) 513–528.
[28] Rockafellar, R. T., R. J.-B. Wets. 1998. *Variational Analysis*. Springer-Verlag, Berlin/Heidelberg.
[29] Ruszczyński, A., A. Shapiro. 2003. Stochastic programming. A. Ruszczyński, A. Shapiro, eds. *Handbook in Operations Research and Management Science*, Vol. 10. North-Holland Publishing Company, Amsterdam.
[30] Scholtes, S. 2001. Convergence properties of a regularization sheme for mathematical programs with complementarity constraints. *SIAM J. Optim.* **11**(4) 918–936.
[31] Shapiro, A. 1991. Asymptotic analysis of stochastic programs. *Ann. Oper. Res.* **30**(1) 169–186.
[32] Shapiro, A. 2006. Stochastic mathematical programs with equilibrium constraints. *J. Optim. Theory Appl.* **128**(1) 223–243.
[33] Shapiro, A., H. Xu. 2008. Stochastic mathematical programs with equilibrium constraints, modeling and sample average approximation. *Optim.* **57**(3) 395–418.

[34] Tomasgard, A., Y. Smeers, K. Midthun. 2009. Capacity booking in a transportation network with stochastic demand. *Proc. 20th Internat. Sympos. Math. Programming, Chicago*.

[35] Werner, A. S. 2004. Bilevel stochastic programming problems: Analysis and application to telecommunications. Doctoral dissertation, Norwegian University of Science and Technology, Trondheim, Norway.

[36] Werner, A. S., Q. Wang. 2009. Resale in vertically separated markets: Profit and consumer surplus implications. *Proc. 20th Internat. Sympos. Math. Programming, Chicago*.

[37] Xu, H. 2010. Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming. *J. Math. Anal. Appl.* **368**(2) 692–710.

[38] Xu, H., F. Meng. 2007. Convergence analysis of sample average approximation methods for a class of stochastic mathematical programs with equality constraints. *Math. Oper. Res.* **32**(3) 648–668.

[39] Ye, J. J. 2000. Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints. *SIAM J. Optim.* **10**(4) 943–962.

[40] Ye, J. J. 2005. Necessary and sufficient optimality conditions for mathemtical programs with equilibrium constraints. *J. Math. Anal. Appl.* **307**(1) 305–369.

[41] Ye, J. J., D. L. Zhu, Q. J. Zhu. 1997. Exact penalization and necessary optimality conditions for generalized bilevel programming problems. *SIAM J. Optim.* **7**(2) 481–507.