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

Yongchao Liu,* Huifu Xu†and Jane J. Ye‡

November 11, 2010

Abstract

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.

Keywords: MPEC-metric regularity, NNAMCQ, error bound, partial exact penalization, M-stationary point.

AMS subject classification: 90C15, 65K15, 90C33, 65K10.

1 Introduction

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

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

where D is a nonempty closed subset of $\mathbb{R}^{n+m}$, $f : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}$, $F : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}^m$ are continuously differentiable, $\xi : \Omega \to \Xi \subset \mathbb{R}^d$ is a vector of random variables defined on the probability space $(\mathscr{F}, \Omega, P)$, $\mathbb{E}[\cdot]$ denotes the expected value with respect to the distribution of ξ and $a \perp b$ means that vector a is perpendicular to vector b. SMPCC is also known as stochastic mathematical program with equilibrium constraints (SMPEC) in that the

---

*School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China (googleliu0717@yahoo.com.cn). The work of this author is carried out while he is visiting the second author in the School of Mathematics, University of Southampton sponsored by China Scholarship Council.

†School of Mathematics,University of Southampton, Southampton, SO17 1BJ, UK (h.xu@soton.ac.uk).

‡Department of Mathematics and Statistics, University of Victoria, Victoria, B.C., Canada V8W 3R4 (janeye@uvic.ca). The work of this author was partly supported by NSERC.

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 [28]. Obviously our model is a one-stage SMPEC. Like deterministic MPECs, SMPEC models have many applications in economics, engineering, networks and management sciences, see for instances [4, 39–41] 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 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 $\xi^1, \cdots, \xi^N$ of random vector ξ, then we may use the sample average $\frac{1}{N} \sum_{k=1}^N f(x, y, \xi^k)$ and $\frac{1}{N} \sum_{k=1}^N F(x, y, \xi^k)$ 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:

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

We call SMPCC (1) the *true* problem and (2) the *sample average approximation* (SAA) problem. SAA method is a popular method in stochastic programming and it has been applied to solve SM-PECs over the past few years although most of the applications are focused on two-stage SMPECs, see for instance [37, 38, 43] and the references therein.

The SMPCC model (1) and its sample average approximation (2) are not new either. Indeed, Birbil et al [2] studied the model and applied the sample path optimization (SPO) method [31] to obtain some approximation results. SPO is essentially SAA although the former is slightly more general. More recently, Meng and Xu [21] 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 [2,21] are mainly for weak stationary points and they are obtained under very strong assumptions such as upper-level strict complementarity condition (ULSC), lower-level strict complementarity condition (LLSC) 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 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 LICQ; see [15] for 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 $\{i : \frac{1}{N} \sum_{k=1}^N F(x, y, \xi^k) = 0, \ y_i = 0\}$ 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, Ye and

Zhu [15] 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

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

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):

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

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

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

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

- Since the original problem (1) does not satisfy usual MFCQ, we cannot establish a full exact penalization of all constraints [1] under MFCQ. Partial exact penalization is, however, feasible under MPEC-GMFCQ [?, 46] or even weaker constraint qualification such as MPEC-metric regularity or equivalently MPEC-NNAMCQ to be defined in Section 2 as we keep the complementarity constraint in (5).

- 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 NLP-regularization method [35] depend on MPEC-LICQ. Indeed, this is a key motivation for Liu et al to consider the partial exact penalization in [15].

- 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

---

[1] Here full penalization means the whole complementarity constraint in (1) is penalized in 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 form of $|y^T z| + \|y_+\|_1 + \|z_+\|_1$, where $a_+ = \max(0,a)$ for a real number a and the maximum is taken componentwise when a is a vector.

of bounded penalty parameters in both (5) and (6) and this is indeed another departure from the existing research in the literature of SMPECs [13, 14]. Moreover, we consider a smoothing method proposed by Liu et al [15] to tackle the nonsmoothness. That is, we consider a smooth partial penalty problem of (4):

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

where

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

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

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 $\mathcal{M}$ is a compact set,

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

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

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

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

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

where $\mathcal{B}$ denotes the closed unit ball in the corresponding finite dimensional space here and throughout the paper. We use $\mathcal{C} + \mathcal{D}$ to denote the Minkowski addition of the two sets, that is, $\{C + D : C \in \mathcal{C}, D \in \mathcal{D}\}$; $B(x, \gamma)$ the closed ball with center x and radius γ; $a^T b$ the scalar product of vectors a and b, where a^T denotes the transpose of vector a. If A is a matrix, $A^T b$ 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 $\Gamma : X \rightrightarrows \mathbb{R}^n$ be a set-valued mapping. We say that Γ is *upper semi-continuous* at a point $x \in X$ if for any $\epsilon > 0$, there exists a number $\delta > 0$ (which may depends on x) such that

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

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

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

Note that many subdifferential mappings are upper semicontinuous but not uniformly upper semicontinuous. However, when $\mathcal{X}$ consists of a finite number of points, then the uniform upper semi-continuity is equivalent to pointwise upper semi-continuity. We need the concept in Lemma 4.1.

Definition 2.2 (Normal Cone [25]) Let $\mathcal{D}$ be a nonempty closed subset of $\mathbb{R}^n$. Given $z \in \mathcal{D}$, the convex cone

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

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

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

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

Note that the limiting normal cone is in general smaller than the Clarke normal cone which is defined as the polar cone of the Clarke tangent cone $\mathcal{T}_D^c(z)$, that is,

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

where $T_{\mathcal{D}}^c(z) = \liminf_{t \to 0, \ \mathcal{D} \ni z' \to z} \frac{1}{t}(\mathcal{D} - z')$. In the case when $\mathcal{D}$ is convex, the proximal normal cone, the limiting normal cone and the Clarke normal cone coincide, see [25, 32].

The following expressions for the limiting normal cone can be easily derived, see e.g. [45, Proposition 3.7].

Proposition 2.3 *Let* $\mathcal{W} = \{(y, z) \in \mathbb{R}^m \times \mathbb{R}^m : 0 \leq z \perp y \geq 0\}$. *The limiting normal cone of* $\mathcal{W}$ *at* $(y, z) \in \mathcal{W}$ *is*

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

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

$$\partial^{\pi} f(x) := \{\zeta \in \mathbb{R}^n :$$

$$\exists \sigma > 0, \delta > 0 \text{ such that } f(y) \geq f(x) + \zeta^T(y - x) - \sigma\|y - x\|^2 \quad \forall y \in B(x, \delta)\}$$

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

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

where $x' \xrightarrow{f} x$ signifies that x' and $f(x')$ converge to x and $f(x)$ respectively. When f is Lipschitz continuous near x, the convex hull of $\partial f(x)$ coincides with the *Clarke subdifferential [5]*, denoted by $\partial^c f(x)$, that is, $\text{conv}\partial f(x) = \partial^c f(x)$.

For set-valued mappings, the definition for 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 $\Gamma : X \rightrightarrows Y$ be a set-valued mapping with a closed graph. Let $(\bar{x}, \bar{y}) \in \text{gph}\,\Gamma := \{(x, y) | y \in \Gamma(x)\}$. The set-valued mapping $D^*\Gamma(\bar{x}|\bar{y})$ from, Y to X defined by*

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

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 $\bar{y} = \Gamma(\bar{x})$.

In the 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 *([24, Proposition 2.11]) Let X, Y be finite dimensional spaces and $\Gamma : X \to Y$ be a single-valued and Lipschitz continuous near $\bar{x}$. Then $D^*\Gamma(\bar{x})(\eta) = \partial\langle\eta, \Gamma\rangle(\bar{x}) \quad \forall\eta \in Y$.*

The following sum rule will be useful.

Proposition 2.7 *(see [24, Corollary 4.4]) Let X, Y be finite dimensional spaces and $\phi : X \to Y$ is strictly differentiable near $\bar{x}$ and Ω is closed. Let $\Gamma(x) = -\phi(x) + \Omega$ and $0 \in \Gamma(\bar{x})$. Then*

$$D^*\Gamma(\bar{x}|0)(\eta) = -\nabla\phi(\bar{x})^T\eta \qquad -\eta \in N_\Omega(\phi(\bar{x})).$$

Definition 2.8 (Metric Regularity) Let X, Y be finite dimensional spaces and $\Gamma : X \rightrightarrows Y$ be a set-valued mapping. Let $(\bar{x}, \bar{y}) \in \text{gph}\,\Gamma$. We say Γ is *metrically regular* at $\bar{x}$ for $\bar{y}$ if there exist constants $\kappa > 0, \delta > 0$ such that

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

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

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

Proposition 2.9 (Estimate for Lipschitz perturbations) *[7, Theorem 3.3] Consider any set-valued mapping $\Gamma : X \rightrightarrows Y$ and any $(\bar{x}, \bar{y}) \in \text{gph}\Gamma$ at which $\text{gph}\,\Gamma$ is locally closed. Consider also a mapping $G : X \to Y$. If $\text{reg}\,\Gamma(\bar{x}|\bar{y}) < \kappa < \infty$ and $\text{lip}\,G(\bar{x}) < \lambda < \kappa^{-1}$, then*

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

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

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

Proposition 2.10 (Mordukhovich's Criteria for Metric Regularity) *[23, Corollary 5.4] For an arbitrary (closed graph) multifunction Γ and $(\bar{x}, \bar{y}) \in \text{gph}\Gamma$, Γ is metrically regular at $\bar{x}$ for $\bar{y}$ if and only if*

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

2.2 MPEC constraint qualification and stationarity

Consider now the following deterministic MPCC:

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

where X is a closed subset of $\mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, $G : \mathbb{R}^n \to \mathbb{R}^m$ and $H : \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable. When X is a system of smooth equalities and inequalities, it is well-known that the classical MFCQ fails at any feasible solution (see [47, Proposition 1.1]). Since the classical MFCQ

when X is a system of smooth equalities and inequalitiesis is equivalent to the metrical 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 never 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). We say that *MPEC-Metric Regularity* holds at x if the set-valued mapping defined by

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

is metrically regular at x for 0, where $\mathcal{W} := \{(y, z) : 0 \le z \perp y \ge 0\}$.

The metric regularity is, however, not easy to verify by definition. By using Mordukhovich's criteria for metric regularitiy (Proposition 2.10), the sum rule for coderivatives (Proposition 2.7) and the expression for the normal cone of $\mathcal{W}$ (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

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

$$\lambda_i = 0, \text{ if } G_i(x) > 0; \beta_i = 0, \text{ if } H_i(x) > 0,$$

$$\lambda_i < 0, \beta_i < 0 \text{ or } \lambda_i \beta_i = 0, \text{ if } G_i(x) = H_i(x) = 0,$$

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 [45, 46] 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 the following extended MPEC-NNAMCQ which was introduced in [15]. 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):

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & x \in X, \\ & z = G(x), \\ & y = H(x), \\ & 0 \le z \perp y \ge 0. \end{aligned} \tag{13}$$

A point (x, y, z) is said to be a *weak feasible point* of problem (13) 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 (13) and there exist no nonzero vectors $(\lambda, \beta) \in \mathbb{R}^m \times \mathbb{R}^m$ such that

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

$$\lambda_i = 0, \text{ if } y_i > 0; \beta_i = 0, \text{ if } z_i > 0,$$

$$\lambda_i < 0, \beta_i < 0 \text{ or } \lambda_i \beta_i = 0, \text{ if } y_i = z_i = 0.$$

For easy reference we review MPEC stationarities 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

$$0 \in \nabla f(x) + \nabla G(x)^T \lambda + \nabla H(x)^T \beta + \mathcal{N}_X(x), \tag{14}$$

$$\lambda_i = 0, \text{ if } G_i(x) > 0; \beta_i = 0, \text{ if } H_i(x) > 0. \tag{15}$$

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 (14)-(15) hold and

$$\lambda_i \beta_i \geq 0 \text{ if } G_i(x) = H_i(x) = 0,$$

$$\lambda_i < 0, \beta_i < 0 \text{ or } \lambda_i \beta_i = 0, \text{ if } G_i(x) = H_i(x) = 0,$$

$$\lambda_i \leq 0, \beta_i \leq 0 \text{ if } G_i(x) = H_i(x) = 0,$$

respectively.

The following relationship between MPEC stationary points is well known:

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

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 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 system of equalities and inequality and its perturbation.

Let $g^N : \mathbb{R}^n \to \mathbb{R}^l$ and $h^N : \mathbb{R}^n \to \mathbb{R}^m$, $N = 1, 2, 3, \cdots$, be two sequences of continuously differentiable mappings and $\mathcal{C}$ 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 $\mathcal{C}$ as $N \to \infty$. Denote by

$$\begin{aligned} S &:= \{x : g(x) \leq 0, h(x) = 0, x \in \mathcal{C}\}, \\ S^N &:= \{x : g^N(x) \leq 0, h^N(x) = 0, x \in \mathcal{C}\}. \end{aligned}$$

Let

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

and

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

The system $\{g(x) \leq 0, h(x) = 0, x \in \mathcal{C}\}$ 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 $\{g(x) \leq 0, h(x) = 0, x \in \mathcal{C}\}$ satisfies metric regularity at a feasible point $\bar{x} \in S$ with regularity modulus equal to κ. Then there exists a neighborhood of $\bar{x}$, denoted by $U_{\bar{x}}$, such that*

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

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

$$d(x, (\Gamma^N)^{-1}(y)) \leq 2\kappa d(y, \Gamma^N(x)), \forall (x, y) \in U_{\bar{x}} \times \delta \mathcal{B}; \tag{16}$$

(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*

$$\begin{cases} 0 & \in \nabla g(\bar{x})^T \lambda + \nabla h(\bar{x})^T \beta + \mathcal{N}_{\mathcal{C}}(\bar{x}), \\ 0 & \leq \lambda \perp -g(\bar{x}) \geq 0, \end{cases}$$

Proof. Part (i): By the definition of metrical regularity, there is an open neighborhood of $\bar{x}$ and $\delta > 0$ such that

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

Now let $\widetilde{x} \in U_{\bar{x}} \cap S$. Then there is a small enough neighborhood of $\widetilde{x}$, denoted by $U_{\widetilde{x}}$ such that

$$U_{\widetilde{x}} \times \delta \mathcal{B} \subset U_{\bar{x}} \times \delta \mathcal{B}.$$

Therefore

$$d(x, \Gamma^{-1}(y)) \leq \kappa d(y, \Gamma(x)), \quad \forall (x, y) \in U_{\widetilde{x}} \times \delta \mathcal{B}$$

which means that $\Gamma(x)$ is metrically regular at $\widetilde{x}$ for 0. Hence (i) holds.

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

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

Then the Lipschitz modulus of G^N at x is equal to

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

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

$$\text{reg}(\Gamma + G^N)(\bar{x}|G^N(\bar{x})) \leq 2\kappa$$

and hence (16) holds for N sufficiently large.

Part (iii) follows from Mordukhovich's criteria for metric regularitiy (Proposition 2.10) and the sum rule for coderivatives (Proposition 2.7). It is also covered by a recent result by Ioffe and Outrata [10, Proposition 3.5]. ∎

Using Proposition 3.1, we are able to establish local error bounds 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 $\{g(x) \leq 0, h(x) = 0, x \in \mathcal{C}\}$ satisfies metric regularity at point $\bar{x}$ with regularity modulus κ, then*

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

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

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

$$d(x, S^N) \leq 2\kappa(\|g^N(x)_+\|_1 + \|h^N(x)\|_1), \quad \forall x \in \mathcal{C} \cap B(\bar{x}, \delta); \tag{17}$$

(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 $\mathcal{C}$. 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

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

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

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

for all $x \in (\bar{x} + \delta\mathcal{B}) \cap \mathcal{C}$. This show Part (i).

Part (ii). In the same manner, we can derive from (16) that

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

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

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) to 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.

Usually it is easier to show that a solution to an original optimization problem is an solution of an exact penalized problem than the reverse. However the reverse statement is equally if not more important since one is hoping to solve the original problem by solving the penalized problem which is easier to solve. In the following theorem, extending the results of [29] we obtain the equivalence of the solutions under the assumption that D is compact and MPEC-metric regularity holds.

Assumption 3.3 *Let $f(x, y, \xi)$ and $F(x, y, \xi)$ be defined as in (1) and satisfy the following properties:*

(a) f and F are locally Lipschitz continuous w.r.t. (x, y), and their Lipschitz modulus are bounded by an integrable function $\kappa_1(\xi) > 0$.

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

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

(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). We use Proposition 3.2 (i) to prove the first claim. Let

$$\mathcal{C} := \{(x, y, z) \in D \times \mathbb{R}^m : 0 \le z \perp y \ge 0\},$$

$h(x, y, z) = \mathbb{E}[F(x, y, \xi)] - z$ and $S := \{(x, y, z) : h(x, y, z) = 0, (x, y, z) \in \mathcal{C}\}$. By the expression of the limiting normal cone in Proposition 2.3, it is easy to check that the MPEC-NNAMCQ for problem (1) at $(\bar{x}, \bar{y})$ is equivalent to the NNAMCQ of the system $\{h(x, y, z) = 0, (x, y, z) \in \mathcal{C}\}$ 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

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

In terms of the terminology of [15], $\|h(x, y, z)\|_1$ 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 [15, Theorem 3.3], $(\bar{x}, \bar{y}, \bar{z})$ is also a local minimizer of (5) for $\rho \ge \tilde{\rho}\kappa$ where κ is the Lipschitz modulus of function $\mathbb{E}[f(x, y, \xi)]$. Note that under Assumption 3.3, such an κ existed. This shows the existence of a positive constant $\rho^* = \tilde{\rho}\kappa$ such that for any $\rho \ge \rho^*$, $(\bar{x}, \bar{y}, \bar{z})$ 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, S_{opt} and S_{opt}^{ρ}, the sets of optimal solutions of problems (3) and (5), are nonempty. We first show there exists a constant $\bar{\rho} > 0$ such that for any $\rho \ge \bar{\rho}$, $S_{opt}^{\rho} \subseteq S_{opt}$. Assume for a contradiction that this is not true. Then for any $\rho_k > 0$, there exists $(x(\rho_k), y(\rho_k), z(\rho_k)) \in S_{opt}^{\rho_k}$ such that $(x(\rho_k), y(\rho_k), z(\rho_k)) \notin S_{opt}$. Let $\rho_k \to \infty$. The compactness of D implies that the sequence $\{(x(\rho_k), y(\rho_k), z(\rho_k))\}$ is bounded and therefore we can draw a subsequence if necessary such that $((x(\rho_k), y(\rho_k), z(\rho_k)) \to (x^*, y^*, z^*)$. Let $(\bar{x}, \bar{y}, \bar{z}) \in S_{opt}$. Since $(x(\rho_k), y(\rho_k), z(\rho_k)) \in S_{opt}^{\rho_k}$, we have

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

which implies that

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

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

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

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 $\rho \ge \hat{\rho}\kappa$, 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 $(x(\rho_k), y(\rho_k), z(\rho_k)) \in S_{opt}^{\rho_k}$, 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 $\rho_k \ge \hat{\rho}\kappa$. Consequently

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

which implies that

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

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

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

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

We are now ready to show that for any $\rho \geq \bar{\rho}$, $S_{opt} = S_{opt}^\rho$. Let $(\tilde{x}, \tilde{y}, \tilde{z}) \in S_{opt}$ and $(x(\rho), y(\rho), z(\rho)) \in S_{opt}^\rho$. Then for any $\rho > \bar{\rho}$, since $S_{opt}^\rho \subseteq S_{opt}$, we have $(x(\rho), y(\rho), z(\rho)) \in S_{opt}$. 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 ask conditions under which an optimal solution exists. To look at the issue, 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 [26] presented a detailed discussion of when a deterministic MPEC has an optimal solution; see section 1.3 and section 4.2 in [26] for details. In our context, if there exists a point $(x_0, y_0) \in D$ such that

$$(\mathbb{E}[F(x_0, y, \xi)] - \mathbb{E}[F(x_0, y_0, \xi)])^T (y - y_0)/\|y - y_0\| \to \infty, \text{ for } (x_0, y) \in D, \|y\| \to \infty, \qquad (18)$$

then (1) has a feasible solution. Moreove 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 (18): D is compact or there exists a nonnegative integrable function $\sigma(\xi)$ with $\mathbb{E}[\sigma(\xi)] > 0$ such that

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

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

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

Which implies the strong monotonicity of $\mathbb{E}[F(x_0, \cdot, \xi)]$. It is also possible to derive some weaker conditions using [26, 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 $\mathbb{E}[F(x, y, \xi)]$ implies 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 $S_{opt}^\rho \subset \mathcal{K}$. We can choose $\mathcal{K}$ sufficiently large such that $S_{opt}^\rho \subset \text{int}\mathcal{K}$, where "int" denotes the interior of a set. Theorem 3.4 (ii) states that $S_{opt}^\rho = S_{opt}$ for large enough ρ. Following the terminology of Pillo and Grippo [29, Definition 1], 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 any $\rho \geq \check{\rho}$ every local minimizer of problem (5) is a local optimal solution of problem (3) (in the terminology of [29], problem (5) is an *exact penalty* problem of (3)). Unfortunately we are unable to show the existence of such a parameter due to 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 counter example. We leave this to interested readers.

Let us now use the part (ii) of Proposition 3.2 and [15, Theorem 3.3] to establish 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 $\{(x^N, y^N, z^N)\}$. Let Assumption 3.3 hold. Then (x^*, y^*) is a feasible point of problem (1). If 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 sketch the proof although it is similar to that of Theorem 3.4 (i). The feasibility of (x^*, y^*) comes from Assumption 3.3 which ensures the uniform convergence of the underlying

functions. The MPEC-NNAMCQ at (x^*, y^*) is equivalent to the NNAMCQ at (x^*, y^*, z^*) with $z^* = \mathbb{E}[F(x^*, y^*, \xi)]$. Let $h(x, y, z) := \mathbb{E}[F(x, y, \xi)] - z$ and $\mathcal{C}$ be defined as in the proof of Theorem 3.4 (i). Let $h^N(x, y, z) := \frac{1}{N} \sum_{k=1}^{N} F(x, y, \xi^k) - z$ and $S^N := \{(x, y, z) : h^N(x, y, z) = 0, (x, y, z) \in \mathcal{C}\}$. 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

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

Applying the principle of partial exact penalization [15, Theorem 3.3], the inequality above implies (x^N, y^N, z^N) is also a local minimizer of (6) for $\rho \geq \rho_N \kappa_N$, where κ_N converges to the Lipschitz modulus of function $\mathbb{E}[f(x, y, \xi)]$ under Assumption 3.3. $\blacksquare$

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) but 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 due to 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 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

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

where $\mathcal{W} = \{(y, z) : 0 \leq z \perp y \geq 0\}$ and the limiting normal cone $\mathcal{N}_{\mathcal{W}}(y, z)$ 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

$$\begin{cases} 0 \in \nabla \mathbb{E}[f(x, y, \xi)] + \nabla \mathbb{E}[F(x, y, \xi)]^T \lambda + \mathcal{N}_D(x, y) + \{(0, \beta_y)\}, \\ 0 = -\lambda + \beta_z, \\ (\beta_y, \beta_z) \in \mathcal{N}_{\mathcal{W}}(y, z). \end{cases} \tag{19}$$

The equivalence of the two set of stationary points is straightforward. $\blacksquare$

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

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

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

$$0 \in \partial_{(x,y,z)} \psi(x, y, z, \rho) + \mathcal{N}_D(x, y) \times \{0\} + \{(0, \beta_y, \beta_z)\}. \tag{20}$$

Observe first that norm $\|\cdot\|_1$ is a convex function, $\mathbb{E}[f(x,y,\xi)]$ and $\mathbb{E}[F(x,y,\xi)] - z$ are continuously differentiable functions. By the sum rule ([25, Proposition 1.107]) and the chain rule ([32, Theorem 10.6] or [5, 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 $b = (b_1, \cdots, b_i, \cdots, b_m)^T$ with

$$b_i = \begin{cases} 1, & \text{if } \mathbb{E}[F_i(x,y,\xi)] - z_i > 0, \\ [-1,1], & \text{if } \mathbb{E}[F_i(x,y,\xi)] - z_i = 0, \\ -1, & \text{if } \mathbb{E}[F_i(x,y,\xi)] - z_i < 0. \end{cases} \tag{21}$$

Consequently (20) can be equivalently written as

$$\begin{cases} 0 \in \nabla\mathbb{E}[f(x,y,\xi)] + \rho\nabla\mathbb{E}[F(x,y,\xi)]^T \mathcal{G}(x,y,z) + \mathcal{N}_D(x,y) + \{(0,\beta_y)\}, \\ 0 \in -\rho\mathcal{G}(x,y,z) + \beta_z, \\ (\beta_y, \beta_z) \in \mathcal{N}_{\mathcal{W}}(y,z). \end{cases} \tag{22}$$

In what follows, we show that an M-stationary point (x,y,z) satisfying (22) is an M-stationary point of (3) defined by (19). Let $b^* \in \mathcal{G}(x,y,z)$ be such that

$$\begin{cases} 0 \in \nabla\mathbb{E}[f(x,y,\xi)] + \rho\nabla\mathbb{E}[F(x,y,\xi)]^T b^* + \mathcal{N}_D(x,y) + \{(0,\beta_y)\}, \\ 0 = -\rho b^* + \beta_z, \\ (\beta_y, \beta_z) \in \mathcal{N}_{\mathcal{W}}(y,z). \end{cases} \tag{23}$$

Then $(x,y,z;\rho b^*, \beta_y, \beta_z)$ satisfies (19). To show that it is a 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 $1 \leq i_0 \leq m$ such that $\mathbb{E}[F_{i_0}(x,y,\xi)] - z_{i_0} \neq 0$. Then we must have $|b^*_{i_0}| = 1$. By (23), $0 = -\rho b^*_{i_0} + [\beta_z]_{i_0}$. Then $\rho = |\rho b^*_{i_0}| = |[\beta_z]_{i_0}| \leq \|\beta_z\|_1$, which contradicts the fact that $\rho > \|\beta_z\|_1$.

Problem (3) $\Rightarrow$ Problem (5). Let $(x,y,z;\lambda, \beta_y, \beta_z)$ be a 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 $\rho \geq \|\lambda\|_1$ and $b^* = \lambda/\rho$. Then $b^*_i \in [-1,1]$ for each $i = 1, \cdots, m$ 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 (22). The proof is complete. ∎

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 sample average approximation 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 $\xi^1, \cdots, \xi^N$ be an iid sampling of ξ. Denote by

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

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 [42, Section 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*

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

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 integrably bounded. Then

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

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 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:

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

where ϵ is a fixed positive number, see [9]. It is well-known that $\partial_\epsilon Q(\cdot)$ is convex, compact set-valued and Hausdorff continuous, see [9, 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,

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

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

(ii) *if $\rho_N \to \rho$ as $N \to \infty$, then w.p.1 $\psi_N(x,y,z,\rho_N)$ converges to $\psi(x,y,z,\rho)$ 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*

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

Proof. Part (i) can be easily proved by virtue of [34, Section 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

$$H(v, \xi) := F(x, y, \xi) - z, \qquad H_N(v) := \frac{1}{N} \sum_{k=1}^{N} H(v, \xi^k), \qquad (24)$$

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

Since Q is a convex function, it is Clarke regular. By the chain rule [32, Theorem 10.6] or [5, 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 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 semi-continuity 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

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

The rest follows straightforwardly from the fact that

$$\mathbb{D}(\partial_{(x,y,z)}\psi_N(x, y, z, \rho_N), \partial_{(x,y,z)}\psi(x, y, z, \rho))$$

$$\leq \left\| \frac{1}{N} \sum_{k=1}^{N} \nabla_{(x,y,z)} f(x, y, \xi^k) - \mathbb{E}[\nabla_{(x,y,z)} f(x, y, \xi)] \right\| + \mathbb{D}(\rho_N \partial G_N(v), \rho \partial G(v))$$

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

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 $\frac{1}{N} \sum_{k=1}^{N} \nabla_{(x,y,z)} f(x, y, \xi^k)$ and $\frac{1}{N} \sum_{k=1}^{N} F(x, y, \xi^k)$, and the feasibility to secure the uniform convergence of $\partial_{(x,y,z)} \left(\left\| \frac{1}{N} \sum_{k=1}^{N} F(x, y, \xi^k) - z \right\|_1 \right)$ because $\frac{1}{N} \sum_{k=1}^{N} F(x, y, \xi^k) - z$ 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 $\mathcal{W}$ be a compact set of $\mathbb{R}^n \times \mathbb{R}^m$ and $\vartheta : \mathcal{W} \times \Xi \to \mathbb{R}$ 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 $w \in \mathcal{W}$, the moment generating function

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

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 $\kappa : \mathbb{R}^n \to \mathbb{R}_+$ and constant $\gamma > 0$ such that

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

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

(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*

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

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*

$$\mathcal{A}_\epsilon(x, y, z, \rho) = \mathbb{E}[\nabla_{(x,y,z)}f(x, y, \xi)] + \mathscr{A}_\epsilon(x, y, z, \rho), \tag{25}$$

and

$$\mathcal{A}_\epsilon^N(x, y, z, \rho_N) = \frac{1}{N}\sum_{k=1}^{N}\nabla_{(x,y,z)}f(x, y, \xi^k) + \mathscr{A}_\epsilon^N(x, y, z, \rho_N), \tag{26}$$

where

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

and

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

H and H_N are defined by (24). Then for any $\alpha > 0$, there exist positive constants $c_3(\alpha)$ and $k_3(\alpha)$ independent of N, such that

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

for N sufficiently large.

Proof. Part (i). The claim follows straightforwardly from [38, 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 $A, B, C, D \subseteq \mathbb{R}^m$,

$$\mathbb{D}(A + C, B + D) \leq \mathbb{D}(A, B) + \mathbb{D}(C, D), \tag{27}$$

Using the inequality, we have

$$\text{Prob}\left\{\sup_{(x,y,z)\in\mathcal{K}} \mathbb{D}\left(\mathcal{A}_\epsilon^N(x, y, z, \rho_N), \mathcal{A}_\epsilon(x, y, z, \rho)\right) \geq \alpha\right\}$$

$$\leq \text{Prob}\left\{\sup_{(x,y,z)\in\mathcal{K}} \mathbb{D}\left(\frac{1}{N}\sum_{k=1}^{N}\nabla_{(x,y,z)}f(x, y, \xi^k), \mathbb{E}[\nabla_{(x,y,z)}f(x, y, \xi)]\right) \geq \frac{\alpha}{2}\right\}$$

$$+ \text{Prob}\left\{\sup_{(x,y,z)\in\mathcal{K}} \mathbb{D}\left(\mathscr{A}_\epsilon^N(x, y, z, \rho_N), \mathscr{A}_\epsilon(x, y, z, \rho)\right) \geq \frac{\alpha}{2}\right\}.$$

By [38, 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 (42),

$$\text{Prob}\left\{\sup_{(x,y,z)\in\mathcal{K}}\mathbb{D}\left(\mathscr{A}_\epsilon^N(x,y,z,\rho_N),\mathscr{A}_\epsilon(x,y,z,\rho)\right)\geq\frac{\alpha}{2}\right\}$$

$$\leq\text{Prob}\left\{\sup_{(x,y,z)\in\mathcal{K}}\|\rho_N\nabla H_N(x,y,z)-\rho\nabla\mathbb{E}[H(x,y,z,\xi)]\|\geq\frac{\alpha}{\displaystyle\sup_{(x,y,z)\in\mathcal{K}}4\|\partial_\epsilon\|H_N(x,y,z)\|_1\|}\right\}$$

$$+\text{Prob}\left\{\sup_{(x,y,z)\in\mathcal{K}}\mathbb{D}\left(\partial_\epsilon\|H_N(x,y,z)\|_1,\partial_\epsilon\|\mathbb{E}[H(x,y,z,\xi)]\|_1\right)\geq\frac{\alpha}{\displaystyle\sup_{(x,y,z)\in\mathcal{K}}4\rho\|\nabla\mathbb{E}[H(x,y,z,\xi)]\|}\right\}$$

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 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 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 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 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 approximate first order optimality condition in section 5.2. See also the comments at the end of section 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*

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

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 [42, Lemma 4.1].

Part (ii). The exponential rate of convergence of (x^N, y^N, z^N) to S^ρ_{opt} follows from Theorem 4.5 (i) and [42, Lemma 4.1]. Moreover, by Theorem 3.4 (ii), $S_{opt} = S^\rho_{opt}$ 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 [32, page 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:

$$0 \in \partial_{(x,y,z)} \psi_N(x, y, z, \rho_N) + \mathcal{N}_D(x, y) \times \{0\} + \{0\} \times \mathcal{N}_\mathcal{W}(y, z). \tag{28}$$

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:

$$0 \in \mathcal{A}^N_\epsilon(x, y, z, \rho_N) + \mathcal{N}_D(x, y) \times \{0\} + \{0\} \times \mathcal{N}_\mathcal{W}(y, z). \tag{29}$$

where $\mathcal{A}^N_\epsilon(x, y, z, \rho_N)$ is defined by (26) and ϵ is a small positive number. Observe that

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

and

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

Using a perturbation result of generalized equations [42, Lemma 4.2], this means a stationary point defined by (29) 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

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

where $\mathcal{A}_\epsilon(x, y, z, \rho)$ is defined by (25), we may consider approximate first order optimality condition for the penalized true problem (5):

$$0 \in \mathcal{A}_\epsilon(x, y, z, \rho) + \mathcal{N}_D(x, y) \times \{0\} + \{0\} \times \mathcal{N}_\mathcal{W}(y, z). \tag{30}$$

The theorem below states the convergence of an approximate stationary point satisfying (29) as N increases.

Theorem 5.2 *Let $\{(x^N_\epsilon, y^N_\epsilon, z^N_\epsilon)\}$ be a sequence of ϵ-stationary points defined by (29) and $(x^*_\epsilon, y^*_\epsilon, z^*_\epsilon)$ is 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 (30). Moreover the convergence rate is exponential.*

Proof. By taking a subsequence if necessarily we assume for the simplicity of notation that $\{(x^N_\epsilon, y^N_\epsilon, z^N_\epsilon)\} \to (x^*_\epsilon, y^*_\epsilon, z^*_\epsilon)$ 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

$$\limsup_{N \to \infty} \mathcal{A}^N_\epsilon(x^N_\epsilon, y^N_\epsilon, z^N_\epsilon, \rho_N) \subseteq \mathcal{A}_\epsilon \psi(x^*_\epsilon, y^*_\epsilon, z^*_\epsilon, \rho).$$

The second part of the claim follows from Theorem 4.5 (ii) and the perturbation theorem of generalized equations [42, Lemma 4.2]. We omit the details. ∎

Note that we are short of claiming a stationary point satisfying (30) is an ϵ-M-stationary point in that the ϵ-convex subdifferential is different from the ϵ-limiting subdifferential, see a discussion by Mordukhovich in [25, page 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 (20) 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*

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

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

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

uniformly on any compact set of $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ w.p.1 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 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

$$0 \in \nabla_{(x,y,z)}\hat{\psi}_N(x^N, y^N, z^N, \rho_N, \delta_N) + \mathcal{N}_D(x^N, y^N) \times \{0\} + \{(0, \beta_y^N, \beta_z^N)\}, \qquad (31)$$

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

$$\varpi_i^N(x, y, z) = \frac{\frac{1}{N}\sum_{k=1}^N F_i(x, y, \xi^k) - z_i}{\sqrt{(\frac{1}{N}\sum_{k=1}^N F_i(x, y, \xi^k) - z_i)^2 + \delta_N}}, \qquad (32)$$

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

By Proposition 4.3 (i), $\frac{1}{N}\sum_{k=1}^N F_i(x, y, \xi^k) - z_i$, $\sqrt{(\frac{1}{N}\sum_{k=1}^N F_i(x, y, \xi^k) - z_i)^2 + \delta_N}$ 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 $i = 1, \cdots, m$. Let

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

Since $|\varpi_i^N(x, y, z)| \leq 1$ 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 (21). Consequently, we have

$$\lim_{N\to\infty} d\left(\nabla_{(x,y,z)}\hat{\psi}_N(x^N, y^N, z^N, \rho_N, \delta_N), \partial_{(x,y,z)}\psi(x^*, y^*, z^*, \rho)\right) = 0. \qquad (33)$$

By (31) and the property of $\mathbb{D}$,

$$d\left(0, \partial_{(x,y,z)}\psi(x^*, y^*, z^*, \rho) + \mathcal{N}_D(x^*, y^*) \times \{0\} + \{(0, \beta_y^*, \beta_z^*)\}\right)$$

$$\leq \mathbb{D}\left(\nabla_{(x,y,z)}\hat{\psi}_N(x^N, y^N, z^N, \rho_N, \delta_N) + \mathcal{N}_D(x^N, y^N) \times \{0\} + \{(0, \beta_y^N, \beta_z^N)\},\right.$$

$$\left.\partial_{(x,y,z)}\psi(x^*, y^*, z^*, \rho) + \mathcal{N}_D(x^*, y^*) \times \{0\} + \{(0, \beta_y^*, \beta_z^*)\}\right)$$

$$\leq d\left(\nabla_{(x,y,z)}\psi_N(x^N, y^N, z^N, \rho_N, \delta_N), \partial_{(x,y,z)}\psi(x^*, y^*, z^*, \rho)\right)$$

$$+\mathbb{D}\left(\mathcal{N}_D(x^N, y^N) \times \{0\} + \{(0, \beta_y^N, \beta_z^N)\}, \mathcal{N}_D(x^*, y^*) \times \{0\} + \{(0, \beta_y^*, \beta_z^*)\}\right),$$

where the first inequality follows from the definition of $\mathbb{D}$ and the second inequality follows from (27). The first term at the right hand of the second inequality of the formula above tends to zero by (33); the second term tends to zero by the upper semi-continuity 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. ∎

From 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* $\rho_N \geq \|\beta_z^N\|_1 + 1$. *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 (31) 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

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

and $\varpi_i^N(\cdot, \cdot, \cdot)$ is defined by (32). 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 $\rho_N \geq \|\beta_z^N\|_1 + 1$. 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 $\rho^* \geq \|\beta_z^*\|_1 + 1$. By Propositions 3.6 and 3.7, (x^*, y^*) is an M-stationary point of problem (1). The proof is complete. ∎

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 $0 \leq y \perp z \geq 0$, we use the well-known regularization method [35] in the literature of MPEC to approximate it with a parameterized system of inequalities

$$y \geq 0, \quad z \geq 0, \quad y \circ z \leq te,$$

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:

$$
\begin{aligned}
\min \quad & \mathbb{E}[(x^2 + y^2) - \xi] \\
\text{s.t.} \quad & x \geq 1, \\
& 0 \leq y \perp \mathbb{E}[-x + y + 2 + \xi^2] \geq 0,
\end{aligned}
\tag{34}
$$

where ξ satisfies the uniform distribution on $[3, 4]$. The example is varied from a deterministic MPEC example in [20, page 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:

$$\begin{array}{ll} \min & \mathbb{E}[\cos(y\xi)] + x^2 + y^2 \\ \text{s.t.} & x \geq 0, \\ & 0 \leq y \perp \mathbb{E}[3\sin(x\xi) + y - 1] \geq 0, \end{array} \tag{35}$$

where the random variable ξ satisfies uniform distribution on $(0, 1]$. Different from (34), 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 (34) and (35):

$$\begin{array}{ll} \min & \mathbb{E}[\cos(y_1\xi)] + x_1 + x_2 + y_2 \\ \text{s.t.} & x_1 \geq 4, \ x_2 \geq 0, \\ & 0 \leq y_1 \perp \mathbb{E}[y_1\xi + x_1 - y_2] \geq 0, \\ & 0 \leq y_2 \perp \mathbb{E}[\cos(x_2\xi) + y_1 + y_2 + \frac{\xi^2}{4} - \pi] \geq 0, \end{array} \tag{36}$$

where the random variable ξ satisfies uniform distribution over $(0, 1]$. Different from (35), 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:

$$\begin{array}{ll} \min & \mathbb{E}[(2(x-1)^2 + y_1^2 + (y_2-1)^2 + (y_3-1)^2 + y_4x)\xi] \\ \text{s.t.} & x \geq 1, \\ & \mathbb{E}[0 \leq y_1\xi + y_2^2 + y_3^2 + x^2] \perp y_1 \geq 0, \\ & 0 \leq \mathbb{E}[-y_2 - \xi + 2x\xi + y_4] \perp y_2 \geq 0, \\ & 0 \leq \mathbb{E}[x\xi + y_4 - 2y_2y_3\xi] \perp y_3 \geq 0, \\ & 0 \leq \mathbb{E}[x*y_1 + 2y_4 - 4\xi + y_1] \perp y_4 \geq 0, \end{array} \tag{37}$$

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 $x \geq 1$, in order 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 [20, page 357]:

$$\begin{array}{ll} \min & \mathbb{E}[2(-8x_1 - 4x_2 + 4y_1 - 40y_2 - 4y_3)\xi] \\ \text{s.t.} & x_i \geq 0, \ i = 1, 2, \\ & x_1 + 2x_2 - y_3 \leq 1.3, \\ & 0 \leq \mathbb{E}[(4 - 2y_4 - 4y_5 + 8y_6)\xi] \perp y_1 \geq 0, \\ & 0 \leq \mathbb{E}[1 + 2y_4\xi + 4y_5 - 2y_6] \perp y_2 \geq 0, \\ & 0 \leq 2 + y_4 - y_5 - y_6 \perp y_3 \geq 0, \\ & 0 \leq \mathbb{E}[(2 + 2y_1 - 2y_2 - 2y_3)\xi] \perp y_4 \geq 0, \\ & 0 \leq \mathbb{E}[(4 - 8x_1 + 4y_1 - 8y_2 + 2y_3)\xi] \perp y_5 \geq 0, \\ & 0 \leq \mathbb{E}[2 - 8x_2\xi - 8y_1\xi + 2y_2 + y_3] \perp y_6 \geq 0, \end{array} \tag{38}$$

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 [20], 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 [20], that is,

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

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 $\frac{1}{N}\sum_{k=1}^{N} f(x^N, y^N, \xi^k)$. 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 $\|\frac{1}{N}\sum_{k=1}^{N} F(x^N, y^N, \xi^k) - z^N\|_1$. The regularization parameter $t = \delta_N$ and the exact penalty is fixed with $\rho = 1000$. 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 (34)-(36), $(1,1,1,1,1,1,1,1,1)$ for problem (37) and (x^0, y^0, z^0) for problem (38).

Table 1: Numerical results for problem (34)

N	δ_N	*iter*	*Appr.Sol*		*Res*	*Appr.Val*
			x^N	y^N		
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 (35)

N	δ_N	*iter*	*Appr.Sol*		*Res*	*Appr.Val*
			x^N	y^N		
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 (36)

N	δ_N	*iter*	*Appr.Sol*		*Res*	*Appr.Val*
			x^N	y^N		
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 (37)

N	δ_N	*iter*	*Appr.Sol*		*Res*	*Appr.Val*
			x^N	y^N		
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
1600	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 (38)

N	δ_N	iter	Appr.Sol		Res	Appr.Val
			x^N	y^N		
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
1600	10^{-6}	86	(0.507108, 0.799307)	(0.000000, 0.194311, 0.805721, 0.005924, 0.500988, 1.504936)	0.132091	-19.093353

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 [11, 12]. 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 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 [35]. 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 $0 \leq y \perp z \geq 0$ 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. 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 Section 3, we can move these constraints to the objective through exact partial penalization. In other words, the partial penalization scheme and the subsequent sample average approximation 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 [36] and Römisch and Rachev [33] for this type of problems.

References

[1] R. J. Aumann, Integrals of set-valued functions, *J. Math. Anal. Appl.*, Vol. 12, pp. 1-12, 1965.

[2] S. I. Birbil, G. Gürkan, and O. Listes, Solving stochastic mathematical programs with complementarity constraints using simulation, *Math. Oper. Res.*, Vol. 31, pp. 739-760, 2006.

[3] C. Castaing and M. Valadier, *Convex analysis and measurable multifunctions, Lecture notes in mathematics*, Springer, Berlin, 1977.

[4] S. Christiansen, M. Patriksson and L. Wynter, Stochastic bilevel programming in structural optimization, *Stru. mult. optim.,* Vol. 21, pp. 361-371, 2001.

[5] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.

[6] L. Dai, C. H. Chen and J.R. Birge, Convergence properties of two stage stochastic programming, *J. Optim. Theo. Appl.*, Vol. 106, pp. 489-509, 2000.

[7] A.L. Dontchev, A.S. Lewis and R.T. Rockafellar, The radius of metric regularity, *Trans. AMS*, Vol. 355, pp. 493-517.

[8] R. Fletcher, *Practical Methods of Optimization*, Wiley, New York, 1987.

[9] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex Analysis and Minimization Algorithms: II*, Springer-Verlag, Berlin, 1993.

[10] A. D. Ioffe and J. V. Outrata, On metric and calmness qualification conditions in subdifferential calculus, *Set-Valued Anal.*, Vol. 16, pp. 199227, 2008.

[11] K. C. Kiewiel, *Methods of Descent for Nondifferentiable Optimization*, Lecture Notes in Mathematics, 1133, Springer-Verlag, New York, 1985.

[12] C. Lemarchal, Bundle methods in nonsmooth optimization, *Nonsmooth optimization*, C. Lemarechal and R. Mifflin eds., Pergamon Press, Oxford, pp. 79-102, 1978.

[13] G.-H. Lin, X. Chen and M. Fukushima, Solving stochastic mathematical programs with equilibrium constraints via approximation and smoothing implicit programming with penalization, *Math. Program*, Vol. 116, pp. 343-368, 2009.

[14] G.-H. Lin, H. Xu and M. Fukushima, Monte Carlo and quasi-Monte Carlo sampling methods for a class of stochastic mathematical programs with equilibrium constraints, *Math. Meth. Oper. Res.*, Vol. 67, pp. 423-441, 2008.

[15] G. Liu, J. J. Ye and J. Zhu, Partial exact penalty for mathematical program with equilibrium constraints, *Set-Valued Anal.*, Vol 16, pp. 785-804, 2008.

[16] Y. Liu, H. Xu and G.-H. Lin, Stability analysis of two stage stochastic mathematical programs with complementarity constraints via NLP-regularization, Optimization online, March 2010.

[17] Y. Lucet and J. J. Ye, Sensitivity analysis of the value function for optimization problems with variational inequality constraints, *SIAM J. Contr. Optim.*, Vol 40, pp. 699-723, 2001.

[18] Y. Lucet and J. J. Ye, Erratum: Sensitivity analysis of the value function for optimization problems with variational inequality constraints, *SIAM J. Contr. Optim.*, Vol 41, pp. 1315-1319, 2002.

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

[20] Z. Q. Luo, J-S. Pang and D. Ralph, *Piece-wise sequential quadratic programming for mathematical programs with nonlinear complementarity constraints*, in: M.C. Ferris and J.S. Pang, eds., Complementarity and Variational Problems: State of the Art, SIAM Publications, 1997.

[21] F. Meng and H. Xu, Exponential convergence of sample average approximation methods for a class of stochastic mathematical programs with complementarity constraints, *J. Comput. Math.*, Vol. 24, pp. 733-748, 2006.

[22] F. Meng and H. Xu, A regularized sample average approximation method for stochastic mathematical programs with nonsmooth equality constraints, *SIAM J. Optim.*, Vol. 17, pp. 891-919, 2006.

[23] B. S. Mordukhovich, Complete characterization of openness, metric regularity, and Lipschitzian properties of multifunctions, *Trans. Amer. Math. Soc.*, Vol. 340, pp. 1-35, 1993.

[24] B. S. Mordukhovich, Generalized differential calculus for nonsmooth and set-valued mappings, *J. Math. Anal. Appl.*, Vol. 183, pp. 250-288.

[25] B. S. Mordukhovich, *Variational analysis and generalized differentiation, I: Basic theory.* Grundlehren series (Fundamental principles of mathematical sciences), Vol. 330, Springer, 2006.

[26] , J.V. Outrata, Optimality conditions for a class of mathematical programs with equilibrium constraints, *Math. Oper. Res.*, Vol. 24, pp. 627-644, 1999.

[27] J. Outrata, M. Kočvara and J. Zowe, *Nonsmooth approach to optimization problems with equilibrium constraints–Theory, applications and numerical constraints*, Kluwer Publishers, 1998.

[28] J. Outrata and W. Römisch, On optimization conditions for some nonsmooth optimization problems over L^p spaces, *J. Optim. Theo. Appl.*, Vol. 126, pp. 411-438, 2005.

[29] M. Patriksson and L. Wynter, Stochastic mathematical programs with equilibrium constraints, *Oper. Res. lett.*, Vol. 25, 159-167, 1999.

[30] G. D. Pillo and L. Grippo, *Exact penalty functions in constrained optimization, SIAM J. Optim.*, Vol 27, pp. 1333-1360, 1989.

[31] R. A. Poliquin and R. T. Rockafellar, Tilt stability of a local minimum, *SIAM J. Optim.*, Vol 8, pp. 287-299, 1998.

[32] S. M. Robinson, Analysis of sample-path optimization, *Math. Oper. Res.*, Vol. 21, pp. 513-528, 1996.

[33] R. T. Rockafellar and R. J.-B. Wets, *Variational analysis*, Springer-Verlag, Berlin, 1998.

[34] S.T. Rachev and W. Römisch, Quantitative stability in stochastic programming: The method of probability metrics, *Math. Oper. Res.*, Vol. 27, pp. 792-818, 2002.

[35] A. Ruszczynski and A. Shapiro, *Stochastic programming, handbook in operations research and management science*, Elsevier, 2003.

[36] S. Scholtes, Convergence properties of a regularization sheme for mathematical programs with complementarity constraints, *SIAM J. Optim.*, Vol. 11 , pp. 918-936, 2001.

[37] A. Shapiro, Asymptotic analysis of stochastic programs, *Ann. Oper. Res.*, Vol. 30, pp. 169-186, 1991.

[38] A. Shapiro, Stochastic mathematical programs with equilibrium constraints, *J. Optim. Theo. Appl.*, Vol. 128, pp. 223-243, 2006.

[39] A. Shapiro and H. Xu, Stochastic mathematical programs with equilibrium constraints, modeling and sample average approximation, *Optimization*, Vol. 57, pp. 395-418, 2008.

[40] A. Tomasgard, Y. Smeers and K. Midthun, Capacity booking in a transportation network with stochastic demand, The Proceedings of 20th International Symposium on Mathematical Programming, Chicago, USA, August 2009.

[41] A. S. Werner, Bilevel stochastic programming problems: analysis and application to telecommunications, PhD dissertation, Norwegian University of Science and Technology, 2004.

[42] A. S. Werner and Q. Wang, Resale in vertically separated markets: profit and consumer surplus implications, The Proceedings of 20th International Symposium on Mathematical Programming, Chicago, USA, August 2009.

[43] H. Xu, Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming, *J. Math. Anal. Appl.*, Vol. 368, pp. 692-710, 2010.

[44] H. Xu and F. Meng, Convergence analysis of sample average approximation methods for a class of stochastic mathematical programs with equality constraints, *Math. Oper. Res.*, Vol. 32, pp. 648-668, 2007.

[45] J. J. Ye, Optimality conditions for optimization problems with complementarity constraints, *SIAM J. Optim.*, Vol. 9, pp. 374-387, 1999.

[46] J. J. Ye, Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints, *SIAM J. Optim.*, Vol. 10, pp. 943-962, 2000.

[47] J. J. Ye, Necessary and sufficient optimality conditions for mathemtical programs with equilibrium constraints, *J. Math. Anal. Appl.*, Vol. 307, pp. 305-369, 2005.

[48] J. J. Ye, D. L. Zhu and Q. J. Zhu, Exact penalization and necessary optimality conditions for generalized bilevel programming problems, *SIAM J. Optim.*, Vol. 2, pp. 481-507, 1997.

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

$$H_N(v) \in \mathbb{E}[H(v,\xi)] + \delta\mathcal{B} \qquad \text{w.p.1,} \tag{39}$$

for N sufficiently large. Let us estimate $\sup_{v\in\mathcal{V}} \mathbb{D}\big(\mathcal{A}G_N(v), \mathcal{A}G(v)\big)$. 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 $\mathrm{I\!R}^m$ and M_1 and M_2 be two matrices in $\mathrm{I\!R}^{n\times m}$. It is easy to verify that

$$\mathbb{D}(M_1\mathcal{D}_1, M_1\mathcal{D}_2) \le \|M_1\|\mathbb{D}(\mathcal{D}_1, D_2), \tag{40}$$

and

$$\mathbb{D}(M_1\mathcal{D}_1, M_2\mathcal{D}_1) \le \|M_1 - M_2\|\|\mathcal{D}_1\|. \tag{41}$$

Using (40)-(41) and the triangle inequality of $\mathbb{D}$, we have

$$\mathbb{D}\left(\nabla H_N(v)^T \mathcal{A}Q(H_N(v)), \nabla\mathbb{E}[H(v,\xi)]^T \mathcal{A}Q(\mathbb{E}[H(v,\xi)])\right)$$
$$\le \mathbb{D}\big(\nabla H_N(v)^T \mathcal{A}Q(H_N(v)), \nabla\mathbb{E}[H(v,\xi)]^T \mathcal{A}Q(H_N(v))\big)$$
$$+\mathbb{D}\left(\nabla\mathbb{E}[H(v,\xi)]^T \mathcal{A}Q(H_N(v)), \nabla\mathbb{E}[H(v,\xi)]^T \mathcal{A}Q(\mathbb{E}[H(v,\xi)])\right)$$
$$\le \|\mathcal{A}Q(H_N(v))\|\|\nabla H_N(v) - \nabla\mathbb{E}[H(v,\xi)]\|$$
$$+\|\nabla\mathbb{E}[H(v,\xi)]\|\mathbb{D}\left(\mathcal{A}Q(H_N(v)), \mathcal{A}Q(\mathbb{E}[H(v,\xi)])\right). \tag{42}$$

We estimate the last two terms in the above equation. By (39),

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

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 semi-continuous on W, which means that for any ϵ, there exists a δ_1 such that

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

Let $v \in \mathcal{V}$ and $w = \mathbb{E}[H(v,\xi)]$. Then $w \in W$. By (39), we have from the inclusion above by setting $\delta \le \delta_1$

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

which is equivalent to

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

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