Stability Analysis of Two Stage Stochastic Mathematical Programs with Complementarity Constraints via NLP-Regularization

Yongchao Liu[1]
School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China
(googleliu0717@yahoo.com.cn)

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

Gui-Hua Lin[2]
School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China
(lin_g_h@yahoo.com.cn)

February 11, 2010

Abstract. This paper presents numerical approximation schemes for a two stage stochastic programming problem where the second stage problem has a general nonlinear complementarity constraint: first, the complementarity constraint is approximated by a parameterized system of inequalities with a well-known regularization approach [44] in deterministic mathematical programs with equilibrium constraints; the distribution of the random variables of the regularized two stage stochastic program is then approximated by a sequence of probability measures. By treating the approximation problems as a perturbation of the original (true) problem, we carry out a detailed stability analysis of the approximated problems including continuity and local Lipschitz continuity of optimal value functions, and outer semicontinuity and continuity of the set of optimal solutions and stationary points. A particular focus is given to the case when the probability distribution is approximated by the empirical probability measure which is known as sample average approximation.

Key words. SMPCC, NLP-regularization, MPEC-MFCQ, stability analysis, sample average approximation.

2000 Mathematics Subject Classification. 90C30, 90C33, 90C15.

1 Introduction

Consider the following two stage stochastic mathematical program with complementarity constraints (SMPCC):

$$\begin{array}{ll} \min\limits_{x,\, y(\cdot)} & \mathbb{E}[f(x, y(\omega), \xi(\omega))] \\ \text{subject to} & x \in X \text{ and for almost every } \omega \in \Omega : \\ & g(x, y(\omega), \xi(\omega)) \leq 0, \\ & h(x, y(\omega), \xi(\omega)) = 0, \\ & 0 \leq G(x, y(\omega), \xi(\omega)) \ \perp H(x, y(\omega), \xi(\omega)) \geq 0, \end{array} \tag{1.1}$$

where X is a nonempty closed convex subset of $\mathbb{R}^n$, $f : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^q \to \mathbb{R}$, $g : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^q \to \mathbb{R}^s$, $h : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^q \to \mathbb{R}^r$, $G : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^q \to \mathbb{R}^m$ and $H : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^q \to \mathbb{R}^m$ are continuously differentiable, $\xi : \Omega \to \Xi$ is a vector of random variables defined on probability $(\Omega, \mathcal{F}, P)$ with support set $\Xi \subset \mathbb{R}^q$, and $\mathbb{E}[\cdot]$ denotes the expected value with respect to the distribution of ξ and '$\perp$' denotes the perpendicularity of two vectors.

[1]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.

[2]The work of the first and third authors was supported in part by NSFC Grant #10771025 and a Royal Society international travel grant 2009.

The SMPCC model differs from the classical two stage stochastic program in that it contains a stochastic complementarity constraint. It also extends deterministic mathematical programs with complementarity constraints (MPCC) by including a random vector ξ. The extension is driven by the practical need as well as theoretical interest. For instance, in an investment model for a firm, one may use a random vector to represent uncertainties arising from future market and a complementarity problem to describe competition from its competitors, see [12, 46]. Similar SMPCC models can also be found in engineering design, see for instance [9].

Patriksson and Wynter [32] first proposed a two stage stochastic mathematical programs with equilibrium constraints (SMPEC) model where the equilibrium constraint is represented by a general stochastic variational inequality. They investigated a number of fundamental issues including existence and uniqueness of optimal solutions, differentiability of upper stage objective function and numerical method for solving the problem. Over the past few years since the first SMPEC paper, there have been increasing discussions on the SMPECs, most of which focus on numerical methods. Shapiro [42] first applied the well-known Monte Carlo method to a general two stage SMPECs where the expected of random functions are approximated by their sample averages and investigated asymptotic convergence of optimal solutions and optimal values as sample size increases. Shapiro and Xu [43] presented a detailed analysis of SMPEC structure and demonstrated the exponential rate of convergence of sharp local minimizers of sample average approximate problems. Lin, Chen and Fukushima [23] first investigated SMPCCs and propose an implicit smoothing method for solving a discrete SMPCC with P_0-linear complementarity constraint. Xu and Meng [48] reformulated the SMPCC as a two stage stochastic minimization problem with nonsmooth equality constraints and applied the sample average approximation method to solve it. They obtained exponential rate of convergence of global optimal solutions obtained from solving the sample average approximation problem. Moreover, they used a uniform law of large numbers for random set-valued mappings to analyze almost sure convergence of generalized KKT points of the sample average approximated SMPCC when the complementarity constraint is strongly monotone. Along this direction, Meng and Xu [27] investigate convergence of stationary points obtained from solving sample average approximated SMPECs where the complementarity constraints are not necessarily monotone. More recently, Xu and Ye [49] derived first order optimality conditions for a two stage SMPEC in terms of limiting subdifferentials. For more details about the development of SMPECs, see a survey paper [24] and references therein.

In this paper, we are concerned with numerical approximation of the two stage SMPCC (1.1). We ask ourselves two fundamental questions: (a) can we approximate SMPCC (1.1) by an ordinary two stage stochastic program with equality and/or inequality constraints? (b) can we approximate the stochastic program by a deterministic nonlinear programming problem (NLP)? The answer to question (a) has been partially answered. For example, one can use NCP functions such as min-function or Fischer-Burmeister function to reformulate a complementarity problem as a nonsmooth system of equations and consequently SMPCC (1.1) as a two stage stochastic program with nonsmooth equality constraints, see [48, 27]. Question (b) is classical in stochastic programming. A simple answer is to use the well-known Monte Carlo sampling method. In the literature of MPECs, however, the reformulation through NCP functions are not most popular. Similarly, in the literature of stochastic programming, there exist discretization/approximation schemes other than Monte Carlo sampling to deal with the random variables.

In this paper, we apply a well-known regularization method ([44, 41, 15]) to tackle the complementarity constraint and then consider a sequence of probability measures to approximate the distribution of ξ with a particular focus on the empirical probability measure (which is known as sample average approximation). The basic idea of the regularization method is to approximate the complementarity constraint $0 \le x \perp y \ge 0$ by a system of parameterized nonlinear inequalities $x \ge 0, y \ge 0$, where the components of x and y satisfy $x_i y_i \le t$ for some small positive parameter t. The regularization method has now been widely applied to solve deterministic MPCCs. The main advantage of the method is that the regularized MPCC is an NLP which can be solved by existing NLP solvers. Moreover, the regularized NLP satisfies Mangasarian-Fromowitz Constraint Qualification (MFCQ) under so-called MPEC-MFCQ of the original problem. It is well-known that MFCQ is closely related to the numerical stability of the problem. In the context of SMPCC, the regularization approach allows one to approximate SMPCC (1.1) by a parameterized ordinary two stage stochastic program which paves the way for the numerical solution of the problem. However, there are a number of theoretical issues to be resolved in order to justify such approximation and this is indeed one of the motivations of this paper.

We include a brief literature review of the NLP-regularization approach for two stage SMPCCs. Shapiro and Xu [43] seemed first to apply the approach to a two stage SMPCC and then used the sample average approximation method to solve it. They predicted the convergence of the regularized SAA method for a class of SMPCCs with strong monotone complementarity constraint but did not give details of the convergence analysis. In a conference paper, Ralph, Xu and Meng [34] carried out some convergence analysis of the NLP-regularized SAA method for solving a class of SMPCCs with monotone complementarity constraint with a particular focus on optimal values and Clarke stationary points.

Since the NLP-regularization is a very popular approach for solving deterministic MPECs, we revisit the topic (the application of the approach to two stage SMPCCs) but from a different perspective and on a wider class of problems: we consider a two stage SMPCC with a general complementarity constraint which is not necessarily monotone; under some moderate conditions (MPEC-MFCQ), we present a detailed analysis on the approximation of optimal values, optimal solutions and stationary points as the regularization parameter tends to zero. Our analysis is carried out from stability point of view, that is, treating the regularized problem as a perturbation of the true SMPCC (1.1). Under some standard conditions in MPECs and sensitivity analysis of parametric MPECs, we demonstrate that the SMPCC problem (1.1) can be effectively approximated by its NLP-regularization which is an ordinary two stage stochastic program. Moreover, we carry out stability analysis of the regularized two stage problem when the probability distribution of ξ is approximated by a sequence of probability measures including the sample average approximation as a special case (empirical probability measure). This broadens the scope in approximating the expected values of the random functions.

2 Preliminaries

In this section, we present some preliminary results in deterministic MPECs, set-valued analysis and random set-valued mapping.

Throughout this paper, we use the following notation. $x^T y$ denotes the scalar product of vectors x and y, $\|\cdot\|$ denotes the Euclidean norm of a vector and a compact set of vectors. $d(x, D)$ represents the distance from point x to set D, that is, $d(x, D) := \inf_{x' \in D} \|x - x'\|$. For two compact sets D_1 and D_2, $\mathbb{D}(D_1, D_2) := \sup_{x \in D_1} d(x, D_2)$ denotes the deviation of D_1 from D_2 and $\mathbb{H}(D_1, D_2) := \max(\mathbb{D}(D_1, D_2), \mathbb{D}(D_2, D_1))$ denotes the Hausdorff distance between D_1 and D_2; $D_1 + D_2$ to denote the Minkowski addition of D_1 and D_2, that is, $D_1 + D_2 = \{x + y : x \in D_1, y \in D_2\}$. For a set C, we use conv C, cl C to denote the convex hull and closure of set C respectively. For a real-valued function $f(x)$, we use $\nabla f(x)$ to denote the gradient of f at x which is a column vector. When f is a vector valued function, $\nabla f(x)$ represents the Jacobian of f at x where the gradient of the j-th component of f forms the j-th column of the Jacobian. Finally, for a set $\{(x, y) = z : z \in Z\}$, $\Pi_x Z = \{x : \exists\, y \text{ such that } (x, y) \in Z\}$.

2.1 Some basics in deterministic MPECs

Consider the following mathematical program with complementary constraints (MPCC for short)

$$\begin{array}{ll} \min\limits_{z} & f(z) \\ \text{subject to} & g(z) \leq 0, \\ & h(z) = 0, \\ & 0 \leq G(z) \perp H(z) \geq 0, \end{array} \tag{2.2}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^s$, $h : \mathbb{R}^n \to \mathbb{R}^r$, $G : \mathbb{R}^n \to \mathbb{R}^m$ and $H : \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable. For a feasible point z^*, we define the following index sets:

$$\begin{aligned} \mathcal{I}_g(z^*) &:= \{i \,|\, g_i(z^*) = 0,\ i = 1, \cdots, s\}, \\ \mathcal{I}_G(z^*) &:= \{i \,|\, G_i(z^*) = 0,\ i = 1, \cdots, m\}, \\ \mathcal{I}_H(z^*) &:= \{i \,|\, H_i(z^*) = 0,\ i = 1, \cdots, m\}. \end{aligned}$$

Moreover, we define a family of nonempty index sets $J \subseteq \{1, \cdots, m\}$,

$$\mathcal{J}(z^*) := \{J \ : \ J \subseteq \mathcal{I}_G(z^*),\ J^c \subseteq \mathcal{I}_H(z^*)\}, \tag{2.3}$$

where $J^c := \{1, \cdots, m\} \backslash J$. We consider the following nonlinear program corresponding to index set J

$$\begin{array}{llll} \mathrm{NLP_J}: & \min\limits_{z} & f(z) & \\ & \text{subject to} & g(z) \leq 0, & \\ & & h(z) = 0, & \\ & & G_i(z) = 0, \ \ H_i(z) \geq 0, & i \in J, \\ & & G_i(z) \geq 0, \ \ H_i(z) = 0, & i \in J^c. \end{array} \tag{2.4}$$

In the literature of MPECs, each of the nonlinear programs corresponding to index set J is called an *NLP branch* of (2.2) and its feasible set is called a branch of the feasible set of MPEC. It is obvious that the branches over $J \in \mathcal{J}(z^*)$ form a neighborhood of z^* in the feasible set of (2.2), see [18].

Definition 2.1 MPCC (2.2) is said to satisfy *MPEC-Mangasarian-Fromowitz Constraint Qualification* (MPEC-MFCQ for short) at a feasible point z^* if the gradient vectors

$$\{\nabla h_i(z) : i = 1, \cdots, r;\ \nabla G_i(z^*),\ i \in \mathcal{I}_G(z^*);\ \nabla H_i(z^*) : i \in \mathcal{I}_H(z^*)\}$$

are linearly independent and there exists a vector $d \in R^n$ perpendicular to the vectors such that

$$\nabla g_i(z^*)^T d < 0, \qquad \forall i \in \mathcal{I}_g(z^*).$$

It is said to satisfy *MPEC Linear Independent Constraint Qualification* (MPEC-LICQ for short) at z^* if the gradient vectors

$$\{\nabla g_i : i \in \mathcal{I}_g(z^*);\ \nabla h_i(z) : i = 1, \cdots, r; \nabla G_i(z^*),\ i \in \mathcal{I}_G(z^*);\ \nabla H_i(z^*) : i \in \mathcal{I}_H(z^*)\}$$

are linearly independent.

Definition 2.2 [41] A point z^* is said to be a *weak stationary point* of (2.2) if there exist vectors $\alpha^* \in \mathbb{R}^s$, $\beta^* \in \mathbb{R}^r$, and $u^*, v^* \in \mathbb{R}^m$ such that

$$\begin{array}{l} 0 = \nabla f(z^*) + \nabla g(z^*)\alpha^* + \nabla h(z^*)\beta^* - \nabla G(z^*)u^* - \nabla H(z^*)v^*, \\ 0 \ \leq \ \alpha^* \ \perp \ g(z^*) \ \leq 0, \\ 0 = u_i^*, \qquad i \notin \mathcal{I}_G(z^*), \\ 0 = v_i^*, \qquad i \notin \mathcal{I}_H(z^*). \end{array}$$

Here α^*, β^*, u^* and v^* are known as the corresponding *Lagrange multipliers*. Moreover,

- z^* is called *C-stationary* to (2.2) if $u_i^* v_i^* \geq 0$ holds for each $i \in \mathcal{I}_G(z^*) \cap \mathcal{I}_H(z^*)$.
- z^* is called *M-stationary* to (2.2) if $\min(u_i^*, v_i^*) > 0$ or $u_i^* v_i^* = 0$ holds for each $i \in \mathcal{I}_G(z^*) \cap \mathcal{I}_H(z^*)$.
- z^* is called *S-stationary* to (2.2) if $u_i^* \geq 0$ and $v_i^* \geq 0$ hold for each $i \in \mathcal{I}_G(z^*) \cap \mathcal{I}_H(z^*)$.

2.2 Set-valued mapping and subdifferentials

Let X be a closed subset of $\mathbb{R}^n$. A set-valued mapping $F : X \to 2^{\mathbb{R}^m}$ is said to be *closed* at $x \in X$ if $F(x)$ is a closed set. The *Painlevé-Kuratowski upper limit* of F at $\bar{x}$ is defined as

$$\overline{\lim_{x \to \bar{x}}} F(x) := \{v \in R^m : \exists \text{ seqences } x_k \to \bar{x}, v_k \to v \text{ with } v_k \in F(x_k)\}.$$

F is said to be *outer semicontinuous* (osc for brevity) at $\bar{x} \in X$ relative to $X \subset \mathbb{R}^n$ if $\overline{\lim}_{x \to \bar{x}} F(x) \subseteq F(\bar{x})$ or equivalently $\lim_{x \to \bar{x}} \mathbb{D}(F(x), F(\bar{x})) = 0$. F is said to be *locally bounded* at $\bar{x}$ if there exists a neighborhood U of $\bar{x}$ such that $\bigcup_{x\in U} F(x)$ is bounded. If F is locally bounded at $\bar{x}$, then the outer semicontinuity of F at $\bar{x}$ is equivalent to that $F(\bar{x})$ is closed and for every open set $O \supset F(\bar{x})$, there is a neighborhood U of $\bar{x}$ such that $\bigcup_{x\in U} F(x) \subset O$, see [38].

Consider now a random set-valued mapping $F(\cdot, \xi(\cdot)) : X \times \Omega \to 2^{\mathbb{R}^n}$ (we are slightly abusing the notation F) where X is a closed subset of $\mathbb{R}^n$ and ξ is a random vector defined on probability space $(\Omega, \mathcal{F}, P)$. Let $x \in X$ be fixed and consider the measurability of set-valued mapping $F(x, \xi(\cdot)) : \Omega \to 2^{\mathbb{R}^n}$. Let $\mathfrak{B}$ denote the space of nonempty, closed subsets of $\mathbb{R}^n$. Then $F(x, \xi(\cdot))$ can be viewed as a single valued mapping from Ω to $\mathfrak{B}$. Using [38, Theorem 14.4], we know that $F(x, \xi(\cdot))$ is measurable if and only if for every $B \in \mathfrak{B}$, $F(x, \xi(\cdot))^{-1}B$ is $\mathcal{F}$-measurable.

Recall that $a(x, \xi(\omega)) \in F(x, \xi(\omega))$ is said to be a *measurable selection* of the random set $F(x, \xi(\omega))$, if $a(x, \xi(\omega))$ is measurable. The *expectation of* $F(x, \xi(\omega))$, denoted by $\mathbb{E}[F(x, \xi(\omega))]$, is defined as the collection of $\mathbb{E}[a(x, \xi(\omega))]$, where $a(x, \xi(\omega))$ is an integrable measurable selection. The expected value is also known as Aumann's integral [3].

Definition 2.3 Let $f : \mathbb{R}^n \to \mathbb{R}$ be a lower semicontinuous function and finite at $x \in \mathbb{R}^n$. The *proximal subdifferential* ([38, Definition 8.45]) of f at x is defined as

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

the *limiting subdifferential* (Mordukhovich or basic [28]) of f at x is defined as

$$\partial^M f(x) := \overline{\lim_{x' \xrightarrow{f} x}} \, \partial^\pi f(x'),$$

and *singular limiting subdifferential*

$$\partial^\infty f(x) := \{v \in \mathbb{R}^n : v = \lim_{k\to\infty} a^k v^k \ \textit{with} \ v^k \in \partial^\pi f(x^k) \ \textit{and} \ a^k \downarrow 0, x^k \xrightarrow{f} x\},$$

where $x' \xrightarrow{f} x$ signifies that x' and $f(x')$ converge to x and $f(x)$ respectively.

It is well-known that a function $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz continuous near $\bar{x}$ if and only if $\partial^\infty f(\bar{x}) = \{0\}$, see for example [25, Proposition 2.4].

Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a locally Lipschitz continuous function. The *Clarke subdifferential* (also known as generalized gradient) of f at $x \in \mathbb{R}^n$ is defined as

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

where D denotes the set of points at which f is Fréchet differentiable, $\nabla f(y)$ denotes the usual gradient of f. It is well-known that the Clarke generalized gradient $\partial f(x)$ is a convex compact set and it is upper semi-continuous, see [10, Proposition 2.1.2 and 2.1.5]. When f is locally Lipschitz continuous near x, the Clarke subdifferential of f at x coincides with the convex hull of the limiting subdifferential, that is,

$$\partial f(x) = \text{conv}\, \partial^M f(x),$$

see [38, Theorem 9.61].

2.3 Sensitivity of generalized equations

Consider the following generalized equations

$$0 \in \Gamma(x) + K, \tag{2.5}$$

where $\Gamma(\cdot) : \mathbb{R}^n \to 2^{\mathbb{R}^m}$ is a closed set-valued mapping and $K \subseteq \mathbb{R}^m$ is a closed set. Let $\bar{\Gamma}(x)$ be a perturbation of $\Gamma(x)$ and consider the perturbed generalized equations

$$0 \in \bar{\Gamma}(x) + K. \tag{2.6}$$

The following lemma states that when $\mathbb{D}(\bar{\Gamma}(x), \Gamma(x))$ is sufficiently small uniformly with respect to x, the solution set of (2.6) is close to the solution set of (2.5).

Lemma 2.1 *Let X be a compact subset of $\mathbb{R}^n$. Let X^* denote the set of solutions to (2.5) within set X and Y^* the set of solutions to (2.6) within X. Assume that both X^* and Y^* are nonempty. Then*

(i) if Γ is outer semicontinuous in X, then for any $\epsilon > 0$, there exists a $\delta > 0$ such that $\mathbb{D}(Y^*, X^*) < \epsilon$, for $\sup_{x\in X} \mathbb{D}(\bar{\Gamma}(x), \Gamma(x)) < \delta$;

(ii) if, in addition, $\bar{\Gamma}(x)$ is also outer semicontinuous in X, then for any $\epsilon > 0$, there exists a $\delta > 0$ such that $\mathbb{H}(Y^*, X^*) < \epsilon$, for $\sup_{x\in X} \mathbb{H}(\bar{\Gamma}(x), \Gamma(x)) < \delta$.

This lemma is similar to [47, Lemma 4.2]. The only difference is that here we consider a general set K rather than a normal cone but this does not affect the conclusion.

Definition 2.4 [21] A set-valued mapping $F : X \subseteq \mathbb{R}^n \to 2^{\mathbb{R}^m}$ is said to be *Pseudo-Lipschitzian* at (z^*, x^*), where $x^* \in X$ and $z^* \in F(x^*)$, if there exist neighborhoods U of z^*, V of x^*, and a positive real number σ such that

$$F(x') \cap U \subset F(x'') + \sigma\|x' - x''\|\mathcal{B}, \qquad \forall x', x'' \in V,$$

where $\mathcal{B}$ is closed unit ball in $\mathbb{R}^m$.

The Pseudo-Lipschitz property is equivalent to F^{-1} having a linear rate of openness as well as to F^{-1} being metrically regular, that is, there exists a positive constant C such that

$$d(x, F^{-1}(z)) \le Cd(z, F(x)), \text{ for all } (x, z) \text{ close to } (x^*, z^*),$$

see [13].

3 NLP-regularization and stability analysis

In this section, we apply the NLP-regularization scheme [44] to SMPCC (1.1) and analyze the stability of the regularized SMPCC in the sense of continuity and local Lipschitz continuity of optimal value functions together with outer semicontinuity and continuity of set mappings of optimal solutions and stationary points. While our analysis follows general steps in the stability analysis of parametric programming [20, 21, 6], we need to tackle a number of new challenges and complications arising from: (a) a mix of parameters with entirely different roles including the first stage decision variable, the random vector and the regularization parameter in the second stage problem, and (b) the subtle relationship between the constraint qualification of the true problems and that of the regularized problems.

3.1 NLP-regularization

In order to apply the NLP-regularization scheme, we first need to reformulate the SMPCC (1.1). Under some moderate conditions, problem (1.1) can be written as

$$\begin{array}{lll} \mathrm{P}_\vartheta : & \min\limits_x & \vartheta(x) = \mathbb{E}[v(x, \xi(\omega))] \\ & \text{subject to} & x \in X, \end{array} \tag{3.7}$$

where $v(x, \xi)$ denotes the optimal value function of the following second stage problem:

$$\begin{array}{lll} \mathrm{MPCC}(x, \xi) : & \min\limits_y & f(x, y, \xi) \\ & \text{subject to} & g(x, y, \xi) \le 0, \\ & & h(x, y, \xi) = 0, \\ & & 0 \le G(x, y, \xi) \perp H(x, y, \xi) \ge 0. \end{array} \tag{3.8}$$

The reformulation is well-known, see for example a discussion in [43, Section 2]. We apply the NLP-regularization scheme ([44, 41, 15]) to the second stage problem MPCC(x, ξ) by replacing the complementarity constraint with a parameterized system of inequalities, that is,

$$G(x, y, \xi) \geq 0, \; H(x, y, \xi) \geq 0, \; G(x, y, \xi) \circ H(x, y, \xi) \leq te,$$

where $t \geq 0$ is a nonnegative parameter, $e \in \mathbb{R}^m$ is a vector with components 1 and "$\circ$" denotes the Hadamard product. Consequently we consider the following regularized second stage problem:

$$\begin{array}{lll} \text{REG}(x, \xi, t): & \min\limits_{y} & f(x, y, \xi) \\ & \text{subject to} & g(x, y, \xi) \leq 0, \\ & & h(x, y, \xi) = 0, \\ & & G(x, y, \xi) \geq 0, \\ & & H(x, y, \xi) \geq 0, \\ & & G(x, y, \xi) \circ H(x, y, \xi) \leq te. \end{array} \tag{3.9}$$

Following the terminology in deterministic MPECs, we call (3.9) a regularized NLP approximation of the second stage problem (3.8). Let $\hat{v}(x, \xi, t)$ denote the optimal value of the regularized problem. Then the corresponding first stage problem can be written as

$$\begin{array}{lll} \text{P}_{\hat{\vartheta}}: & \min\limits_{x} & \hat{\vartheta}(x, t) = \mathbb{E}[\hat{v}(x, \xi(\omega), t)] \\ & \text{subject to} & x \in X. \end{array} \tag{3.10}$$

Observe that when $t = 0$, REG(x, ξ, t) coincides with MPCC(x, ξ) and $P_{\hat{\vartheta}}$ coincides with P_{ϑ}. The underlying reason for us to consider the regularization scheme here is that the regularized problem is an ordinary stochastic NLP to which existing numerical methods in the literature of stochastic programming may be applied. From numerical perspective, t often takes a small positive value because REG(x, ξ, t) never satisfies MFCQ (which is equivalent to numerical stability) at $t = 0$. Our focus in this and the following section is to provide a theoretical justification of the NLP-regularization approximation as $t \to 0$. Specifically, we analyze continuity of optimal value functions and set of optimal solutions for both the first and the second stage problems particularly when t tends to 0. Note that this kind of stability analysis can be found to some extent in [44, 41, 15] where the NLP-regularization is applied to deterministic MPCCs with nonmonotonic complementarity constraints. Here the SMPCC involves two stages and at the second stage the first stage decision vector x, the random variable ξ are both treated as parameters together with the regularization parameter t. However the three parameters have to be treated in a different way, which means that we cannot directly apply the stability results established in [44, 41, 15] where t is the only parameter.

Some notation are in place. We use $\mathcal{F}(x, \xi)$ and $\hat{\mathcal{F}}(x, \xi, t)$ to denote respectively the feasible set of the second stage problem (3.8) and (3.9); $Y_{sol}(x, \xi)$ and $\hat{Y}_{sol}(x, \xi, t)$ the set of global optimal solutions; X_{sol} and $\hat{X}_{sol}(t)$ the optimal solution set of the first stage problem (3.7) and (3.10). We use $\phi(t)$ to denote the optimal value of $P_{\hat{\vartheta}}$. Observe that $\hat{\mathcal{F}}(x, \xi, 0) = \mathcal{F}(x, \xi)$, $\hat{Y}_{sol}(x, \xi, 0) = Y_{sol}(x, \xi)$ and $\hat{X}_{sol}(0) = X_{sol}$.

3.2 Continuity of optimal value functions and solution mappings

3.2.1 The second stage problem

We start by investigating the continuity of optimal value function $\hat{v}(x, \xi, t)$ and solution set mapping $\hat{Y}_{sol}(x, \xi, t)$ of the second stage regularized problem REG(x, ξ, t) with respect to x, ξ and t. We need the following inf-compactness condition.

Assumption 3.1 (Inf-compactness) *Let $x^* \in X$. There exist constants δ, $t^* > 0$, a compact set $Y \subset \mathbb{R}^m$ and a neighborhood U of x^* such that*

$$\emptyset \neq \{y : f(x, y, \xi) \leq \delta \text{ and } y \in \mathcal{F}(x, \xi, t)\} \subset Y,$$

for all $(x, \xi, t) \in U \times \Xi \times [0, t^*]$.

We make a few comments on the inf-compactness assumption.

1. Inf-compactness conditions are widely used in the stability analysis of parametric programming. The conditions here are slightly different from those in [6, Proposition 4.4] in that the parameters x, ξ and t are not treated equally. Specifically, x is the decision vector of the first stage problem and we need to discuss various topological properties of optimal values and solution mappings with respect to it, therefore we consider it in a neighborhood U of a considered point x^*; t is a regularization parameter and we are only interested in the case when it is close to 0, the fundamental reason that we are interested in a nonzero value of t is that the regularized problem satisfies MFCQ under the standard MPEC-MFCQ of the true problem when $t > 0$; finally ξ is a realization of the random vector $\xi(\omega)$, instead of requiring differentiability of optimal values of solution set mapping, we need measurability of these quantities with respect to ξ.

2. Both constants δ and t^* depend on x^*. The inf-compactness condition implies that the optimal solution set $\hat{Y}_{sol}(x, \xi, t)$ is nonempty and bounded by compact set Y for all $(x, \xi, t) \in U \times \Xi \times [0, t^*]$.

3. The inf-compactness condition holds when $f(x, \cdot, \xi)$ is uniformly coercive or strongly convex. Moreover, in the case when $G(x, y, \xi) = y$, the condition is implied by the monotonicity of $H(x, \cdot, \xi)$. For instance, if Ξ is bounded and $H(x, \cdot, \xi)$ is a R_0 function for every $(x, \xi) \in \mathcal{X} \times \Xi$, that is, if for any sequence $\{y^k\}$ with $\lim_{k\to\infty} \|y^k\| = +\infty$, $\liminf_{k\to\infty}(\min\{y_1^k, \cdots, y_m^k\})/\|y^k\| \geq 0$ and

$$\liminf_{k\to\infty} \min\{H_1(x, y^k, \xi), \cdots, H_m(x, y^k, \xi)\}/\|y^k\| \geq 0,$$

there exists an index j such that $\{y_j^k\} \to +\infty$ and $\{H_j(x, y^k, \xi)\} \to +\infty$. In such a case, the feasible set of problem (3.9) is uniformly bounded for $t \in [0, +\infty)$, see [8, Lemma 2.2] for more details.

Our first technical result is that under Assumption 3.1 the feasible set $\hat{\mathcal{F}}(x, \xi, t)$ of the second stage regularized problem is continuous with respect to (x, ξ, t) when it is restricted to set Y.

Proposition 3.1 *Let Assumption 3.1 hold at point $x^* \in X$ and $\mathcal{F}_Y(x, \xi, t) = Y \cap \hat{\mathcal{F}}(x, \xi, t)$. Then there exist a neighborhood U of x^* and a scalar $t^* > 0$ such that $\mathcal{F}_Y(x, \xi, t)$ is continuous on $U \times \Xi \times [0, t^*]$.*

Proof. Let U and t^* be given as in Assumption 3.1 and

$$R(x, y, \xi, t) = \begin{pmatrix} h(x, y, \xi) \\ g(x, y, \xi) \\ -G(x, y, \xi) \\ -H(x, y, \xi) \\ G(x, y, \xi) \circ H(x, y, \xi) - te \end{pmatrix}.$$

Then $\mathcal{F}_Y(x, \xi, t)$ is the set of solutions to the following generalized equations restricted to set Y:

$$0 \in R(x, y, \xi, t) + \mathcal{Q},$$

where $\mathcal{Q} = 0_r \times \mathbb{R}_+^{s+m+m+m}$. Under Assumption 3.1, $\mathcal{F}_Y(x, \xi, t)$ is nonempty for $(x, \xi, t) \in U \times \Xi \times [0, t^*]$. Moreover, $R(x, y, \xi, t)$ is single valued and continuous. By Lemma 2.1, $\mathcal{F}_Y(x, \xi, t)$ is Hausdorff continuous on $U \times \Xi \times [0, t^*]$. The proof is complete. ■

Using Proposition 3.1, we can establish the outer semicontinuity of the optimal solution set mapping and continuity of the optimal value function of the second stage regularized problem REG(x, ξ, t).

Theorem 3.1 (Stability of REG(x, ξ, t)) *Let Assumption 3.1 hold at point $x^* \in X$. Then there exist a neighborhood U of x^* and a scalar $t^* > 0$ such that*

(i) the optimal solution set $\hat{Y}_{sol}(x, \xi, t)$ of the second stage problem REG(x, ξ, t) is outer semi-continuous on $U \times \Xi \times [0, t^*]$;

(ii) the optimal value function $\hat{v}(x,\xi,t)$ of the second stage problem REG(x,ξ,t) is continuous on $U \times \Xi \times [0,t^*]$;

(iii) for any $x \in U$ and $t \in (0,t^*]$, $v(x,\cdot)$ and $\hat{v}(x,\cdot,t)$ are continuous on Ξ.

Proof. Let U and t^* be given as in Assumption 3.1. Observe first that $\hat{v}(x,\xi,t)$ is well-defined for all $(x,\xi,t) \in U \times \Xi \times [0,t^*]$, that is, $\hat{v}(x,\xi,t)$ takes a finite value. Moreover the optimal solution set $\hat{Y}_{sol}(x,\xi,t) \subset Y$.

Part (i). Let $\{(x^k,\xi^k,t_k)\}$ be any sequence in $U \times \Xi \times [0,t^*]$ such that $(x^k,\xi^k,t_k) \to (x,\xi,t)$. Let $\hat{y}^k \in \hat{Y}_{sol}(x^k,\xi^k,t_k)$ and $\hat{y}$ be an accumulation point of sequence $\{\hat{y}^k\}$. It suffices to show that $\hat{y} \in \hat{Y}_{sol}(x,\xi,t)$. Assume for a contradiction that $\hat{y} \notin \hat{Y}_{sol}(x,\xi,t)$, that is, $\hat{v}(x,\xi,t) < f(x,\hat{y},\xi)$. Let $y^* \in \hat{Y}_{sol}(x,\xi,t)$. Then

$$\hat{v}(x,\xi,t) = f(x,y^*,\xi) < f(x,\hat{y},\xi).$$

For the given y^*, it follows by Proposition 3.1, there exists a sequence $\{y^k\}$ such that $y^k \in \mathcal{F}_Y(x^k,\xi^k,t_k)$ and $y^k \to y^*$ as $k \to \infty$. Since f is continuous, there exists k_0 such that for $k \geq k_0$, $f(x^k,y^k,\xi^k) < f(x^k,\hat{y}^k,\xi^k)$, which contradicts the fact that $\hat{y}^k \in \hat{Y}_{sol}(x^k,\xi^k,t_k)$.

Part (ii). Given the outer semi-continuity of $\hat{Y}_{sol}(x,\xi,t)$ and the continuity of f, we can easily use [6, Proposition 4.4] to obtain the continuity of $\hat{v}(x,\xi,t)$ on $U \times \Xi \times [0,t^*]$. We omit the details.

Part (iii). The continuity of $v(x,\cdot)$ and $\hat{v}(x,\cdot,t)$ follows from Part (ii). ■

Recall that a set-valued mapping $\Gamma : \Omega \times X \to 2^{\mathbb{R}^m}$ is said to be *Carathédory* if for every x, $\Gamma(\cdot,x)$ is measurable and for every ω, $\Gamma(\omega,\cdot)$ is continuous, see [3]. By Theorem 3.1, we have the following.

Corollary 3.1 *Assume the settings and conditions of Proposition 3.1 and Theorem 3.1. Then $\mathcal{F}_Y(\cdot,\xi(\cdot),\cdot) : U \times \Omega \times [0,t^*] \to 2^{\mathbb{R}^m}$ is a Carathédory mapping and $\hat{v}(\cdot,\xi(\cdot),\cdot) : U \times \Omega \times [0,t^*] \to \mathbb{R}$ is a Carathéodory function.*

3.2.2 First stage problem

Next, we consider the first stage regularized problem $P_{\hat{\vartheta}}$. Under some moderate conditions, we establish the outer semi-continuity of the optimal solution set mapping and continuity of the optimal value function of the problem.

Theorem 3.2 (Stability of $P_{\hat{\vartheta}}$) *Let $\bar{X} \subseteq X$ be a compact set and Assumption 3.1 hold for every $x \in \bar{X}$. Suppose that there exists a positive constant $\bar{t}$ such that for all $t \in [0,\bar{t}]$, $\hat{X}_{sol}(t) \cap \bar{X} \neq \emptyset$. Then there exists a positive constant $t^* < \bar{t}$ such that*

(i) the optimal solution set mapping $\hat{X}_{sol}(\cdot) \cap \bar{X}$ is outer semi-continuous on $[0,t^*]$;

(ii) the optimal value function $\phi(t)$ of problem $P_{\hat{\vartheta}}$ is continuous on $[0,t^*]$.

Proof. Let $x \in \bar{X}$. Since Assumption 3.1 holds at x, by Theorem 3.1, there exist a neighborhood U_x of x and a scalar $t_x > 0$ (depending on x) such that $\hat{v}(x,\xi,t)$ is continuous on $U_x \times \Xi \times [0,t_x]$. What we need to prove here is to find a positive scalar t^* *independent* of x such that $\hat{v}(x,\xi,t)$ is continuous on $U_x \times \Xi \times [0,t^*]$ for all $x \in \bar{X}$. Our idea is to use the finite covering theorem: given the fact that we can find a neighborhood U_x for every point x and a positive number t_x such that $\hat{v}$ is continuous, we can find a finite number of such neighborhoods U_{x_i} and positive numbers t_{x_i}, $i = 1,\cdots,\hat{i}$ such that the union of the neighborhood $U = \bigcup_{i=1}^{\hat{i}} U_{x_i}$ covers the compact set $\bar{X}$, and $\hat{v}(\cdot,\cdot,\cdot)$ is continuous on $U \cap \bar{X} \times \Xi \times [0,t^*]$ where $t^* = \min_{i=1}^{\hat{i}} t_{x_i}$.

Part (ii). Under Assumption 3.1, $\hat{v}(x,\xi,t) \leq \delta_x$ for some positive constant δ_x and from Part (i), $\hat{v}(\cdot,\cdot,\cdot)$ is continuous on $U_x \times \Xi \times [0,t_x]$. By [40, Proposition 1, Chapter 2], $\vartheta(x,t) = \mathbb{E}[\hat{v}(x,\xi,t)]$ is continuous on $U_x \times [0,t_x]$. Using the covering theorem as in the proof of Part (i), we can find $\delta = \max_{i=1}^{\hat{i}} \delta_{x_i}$ such that $\hat{v}(x,\xi,t)$ is bounded by δ and $\vartheta(x,t) = \mathbb{E}[\hat{v}(x,\xi,t)]$ is continuous on $\bar{X} \times [0,t^*]$, where t^* is given as in the proof of Part (i). Obviously the level set $\{x \in X : v(x,t) \leq \delta\}$ is nonempty and its interception with $\bar{X}$ is also nonempty. By applying [6, Proposition 4.4], we conclude that the optimal value function $\phi(t)$ of $P_{\hat{\vartheta}}$ is continuous on $[0,t^*]$. The proof is complete. ■

3.3 Lipschitz continuity of optimal value functions

We use the classical quantitative stability results in parametric programming to investigate the local Lipschitz continuity of the optimal value function $\hat{v}(x,\xi,t)$ of the second stage regularized problem REG(x,ξ,t) with respect to x, t and value function $v(x,\xi)$ of MPCC(x,ξ) with respect to x. A sufficient condition is the Pseudo-Lipschitz property of the feasible solution set mapping which is implied by the MFCQ of the problem, see a discussion by Klatte at page 3 in [20]. To this end, we discuss the MFCQ of the regularized problem REG(x,ξ,t) in Proposition 3.2 under the MPEC-MFCQ of MPCC(x,ξ).

Proposition 3.2 *Let $x^* \in X$, $\xi^* \in \Xi$ be fixed and $y^* \in \mathcal{F}(x^*,\xi^*)$. Assume that problem MPCC(x^*,ξ^*) satisfies MPEC-MFCQ at y^*. Then there exist neighborhoods of y^* and (x^*,ξ^*), denoted by U_{y^*} and $U_{(x^*,\xi^*)}$ respectively, and a scalar $t^* > 0$ such that for all $(x,\xi,t) \in U_{(x^*,\xi^*)} \times (0,t^*]$, the regularized second stage problem REG(x,ξ,t) satisfies the MFCQ at any point $y \in U_{y^*} \cap \hat{\mathcal{F}}(x,\xi,t)$.*

Proof. For the simplicity of notation, let $z = (x,y,\xi)$ and $z^* = (x^*,y^*,\xi^*)$ and throughout the proof, "∇" denote the gradient with respect to y. By the definition of MFCQ, it suffices to show that there exist a neighborhood U of z^* and a scalar $t^* > 0$ such that for any $t \in (0,t^*]$, $(x,\xi) \in X \times \Xi$ and feasible point y of REG(x,ξ,t) with $(x,y,\xi) = z \in U$, the gradient vectors

$$\nabla h_i(z) : i = 1, \cdots, r,$$

are linearly independent and there exists a vector $d(z)$ (depending on z) such that

$$\left\{ \begin{array}{ll} 0 = \nabla h_i(z)^T d(z), & i = 1, \cdots, r. \\ 0 > \nabla g_i(z)^T d(z), & i \in \mathcal{I}_g(z), \\ 0 > -\nabla G_i(z)^T d(z), & i \in \mathcal{I}_G(z), \\ 0 > -\nabla H_i(z)^T d(z), & i \in \mathcal{I}_H(z), \\ 0 > (H_i(z)\nabla G_i(z) + G_i(z)\nabla H_i(z))^T d(z), & i \in \mathcal{I}_{G\circ H}(z), \end{array} \right. \tag{3.11}$$

where $\mathcal{I}_{G\circ H}(z) := \{i \mid G_i(z)H_i(z) = t, \ \ i = 1, \cdots, m\}$. In what follows, we construct such a vector $d(z)$.

First, by assumption MPEC-MFCQ holds at y^* for problem MPCC(x^*,ξ^*). By the definition of MPEC-MFCQ, the gradient vectors

$$\{\nabla h_i(z^*), i = 1, \cdots r;\ \nabla G_i(z^*), i \in \mathcal{I}_G(z^*);\ \nabla H_i(z^*), i \in \mathcal{I}_H(z^*)\}$$

are linearly independent and there exists a vector $\bar{d} \in \mathbb{R}^n$ which is perpendicular to these gradient vectors and

$$\nabla g_i(z^*)^T \bar{d} < 0 \text{ for } i \in \mathcal{I}_g(z^*). \tag{3.12}$$

Second, it is not difficult to show that there exist a neighborhood U_1 of z^* and $t^* > 0$ such that for any $z \in U_1$ and $t \in (0,t^*]$, the following relations hold:

$$\left\{ \begin{array}{l} \mathcal{I}_g(z) \subseteq \mathcal{I}_g(z^*), \\ \mathcal{I}_G(z) \subseteq \mathcal{I}_G(z^*), \\ \mathcal{I}_H(z) \subseteq \mathcal{I}_H(z^*), \\ \mathcal{I}_G(z) \cap \mathcal{I}_{G\circ H}(z) = \emptyset, \\ \mathcal{I}_H(z) \cap \mathcal{I}_{G\circ H}(z) = \emptyset, \end{array} \right.$$

and the gradient vectors

$$\nabla h_i(z), i = 1, \cdots, r;\ \nabla G_i(z),\ i \in \mathcal{I}_G(z),\ \nabla H_i(z),\ i \in \mathcal{I}_H(z);\ H_i(z)\nabla G_i(z) + G_i(z)\nabla H_i(z),\ i \in \mathcal{I}_{G\circ H}(z)$$

are linearly independent.

Third, the linear independence of the gradient vectors in the second step implies that, for each fixed γ and any $z \in U_1$, there exists a nonzero vector $\hat{d}(z,\gamma)$ with bounded norm such that

$$\begin{array}{rcl} \gamma\nabla h_i(z)^T\bar{d} & = & \nabla h_i(z)^T\hat{d}(z,\gamma), \quad i = 1,\cdots,r, \\ 1 & = & \nabla G_i(z)^T\hat{d}(z,\gamma), \quad i \in \mathcal{I}_G(z), \\ 1 & = & \nabla H_i(z)^T\hat{d}(z,\gamma), \quad i \in \mathcal{I}_H(z), \\ -1 & = & (H_i(z)\nabla G_i(z) + G_i(z)\nabla H_i(z))^T\hat{d}(z,\gamma), \quad i \in \mathcal{I}_{G\circ H}(z). \end{array}$$

Indeed, if we use $A(z)^T$ to denote the coefficient matrix and $b(z,\gamma)$ the left hand side of the linear system of equations above, then we may choose

$$\hat{d}(z,\gamma) = A(z)[A(z)^TA(z)]^{-1}b(z,\gamma).$$

Denote $A^{\#}(z) := A(z)[A(z)^TA(z)]^{-1}$. The continuous differentiability of $h(z), G(z)$ and $H(z)$ implies that there exists a positive constant C such that $\|A^{\#}(z)\| \le C$ for all $z \in U_1$. Note that as z varies, the number of equations in the above system may change but our conclusion on the boundedness of $A^{\#}(z)$ holds.

Fourth, let $d(z,\gamma) = \gamma\bar{d} - \hat{d}(z,\gamma)$. Then

$$\nabla h_i(z)^Td(z,\gamma) = \nabla h_i(z)^T(\gamma\bar{d} - \hat{d}(z,\gamma)) = 0, \quad i = 1,\cdots,r. \tag{3.13}$$

Moreover, for any $i \in \mathcal{I}_g(z)$ and $z \in U_1$

$$\begin{array}{rcl} \nabla g_i(z)^Td(z,\gamma) & = & \gamma\nabla g_i(z)^T\bar{d} - \nabla g_i(z)^T\hat{d}(z,\gamma) \\ & = & \gamma\nabla g_i(z)^T\bar{d} - \nabla g_i(z)^T(A^{\#}(z)b(z,\gamma)) \\ & = & \gamma[\nabla g_i(z)^T\bar{d} - \nabla g_i(z)^T(A^{\#}_r(z)\nabla h(z)^T\bar{d})] - \nabla g_i(z)^T(A^{\#}_{r-}(z)(1,1,-1)^T), \end{array} \tag{3.14}$$

where $A^{\#}_r(z)$ denotes the matrix which takes the first r columns of $A^{\#}(z)$ and $A^{\#}_{r-}(z)$ denotes the other part of $A^{\#}(z)$. Note that $\nabla h(z)^T\bar{d}$ tends to zero, $\nabla g(z)^T\bar{d} \to \nabla g(z^*)^T\bar{d} < 0$ as $z \to z^*$ and $\nabla g_i(z)^T(A^{\#}_{r-}(z)(1,1,-1)^T)$ is independent of γ and bounded when z is close to z^*. Therefore there exist a positive scalar γ sufficiently large and a neighborhood $U_2 \subseteq U_1$ of z^* such that

$$\nabla g_i(z)^Td(z) < 0, \quad \forall z \in U_2.$$

Let γ be fixed. Since $\bar{d}$ is perpendicular to $\nabla G(z^*)$ and $\nabla H(z^*)$, we can choose a smaller neighborhood $U_3 \subseteq U_2$ of z^* such that for any $z \in U_3$

$$\left\{ \begin{array}{ll} -\nabla G_i(z)^Td(z,\gamma) = -\gamma\nabla G_i(z)^T\bar{d} - 1 < 0, & i \in \mathcal{I}_G(z), \\ -\nabla H_i(z)^Td(z,\gamma) = -\gamma\nabla H_i(z)^T\bar{d} - 1 < 0, & i \in \mathcal{I}_H(z), \end{array} \right. \tag{3.15}$$

and

$$\begin{array}{rcl} (H_i(z)\nabla G_i(z) + G_i(z)\nabla H_i(z))^Td(z,\gamma) & = & H_i(z)(-\gamma\nabla G_i(z)^T\bar{d} - 1) + G_i(z)(-\gamma\nabla H_i(z)^T\bar{d} - 1) \\ & < & 0, \quad i \in \mathcal{I}_{G\circ H}(z). \end{array} \tag{3.16}$$

Letting $U = U_3$, $U_{y^*} = \Pi_yU$, $U_{(x^*,\xi^*)} = \Pi_{(x,\xi)}U$ and combining (3.13)–(3.16), we obtain $d(z) = d(z,\gamma)$ satisfying (3.11) as desired and hence the conclusion. ■

Corollary 3.2 *Assume the conditions of Proposition 3.2. Then there exist a neighborhood $U_{(x^*,\xi^*)}$ of (x^*,ξ^*) and a neighborhood U_{y^*} of y^* such that for all $(x,\xi) \in U_{(x^*,\xi^*)}$, problem MPCC$(x,\xi)$ satisfies the MPEC-MFCQ at every feasible point $y \in U_{y^*}$.*

Proof. Let $z = (x,y,\xi)$ and $z^* = (x^*,y^*,\xi^*)$ and throughout the proof "∇" denote the gradient with respect to y. It is obvious that there exists a neighborhood U_1 of z^* such that

$$\mathcal{I}_g(z) \subseteq \mathcal{I}_g(z^*), \quad \mathcal{I}_G(z) \subseteq \mathcal{I}_G(z^*), \quad \mathcal{I}_H(z) \subseteq \mathcal{I}_H(z^*),$$

and the matrix $A(z)$ with columns

$$\nabla h_i(z),\ i=1,\cdots r;\quad \nabla G_i(z),\ i\in\mathcal{I}_G(z);\quad \nabla H_i(z),\ i\in\mathcal{I}_H(z),$$

has full column rank. Let $\bar{d}$ be a given vector which satisfies the MPEC-MFCQ at point y^* and let

$$d(z)=[I-A(z)(A(z)^TA(z))^{-1}A(z)^T]\bar{d}.$$

Since $d(z)\to\bar{d}$ as $z\to z^*$, there exists a neighborhood $U\subseteq U_1$ of z^* such that

$$\nabla g(z)^Td(z)<0,\qquad A(z)^Td(z)=0.$$

The claim holds for $U_{y^*}=\Pi_yU$ and $U_{(x^*,\xi^*)}=\Pi_{(x,\xi)}U$. ■

In what follows, we establish the local Lipschitz continuity of $\hat{v}(x,\xi,t)$ and $v(x,\xi)$ with respect to x and t for all $\xi\in\Xi$. We do so by exploiting the well-known stability results due to Klatte [20] and [21] for $\hat{v}(x,\xi,t)$ and a stability result on parametric MPEC by Hu and Ralph [18] for $v(x,\xi)$. The key argument we want to use from Klatte's stability results is that the local Lipschitz continuity of our objective function $f(x,y,\xi)$ and the Psudo-Lipschitzian of the feasible set $\hat{\mathcal{F}}(x,\xi,t)$ imply the local Lipschitz continuity of the optimal value function $\hat{v}(x,\xi,t)$. As for $v(x,\xi)$, Hu and Ralph observed that under the MPEC-LICQ, the quantitative stability of the optimal value function is essentially the same as that in the parametric nonlinear programming.

Theorem 3.3 *Let $x^*\in X$ and Assumption 3.1 hold at point x^*. Let $\xi\in\Xi$ be fixed and problem MPCC(x^*,ξ) satisfy MPEC-MFCQ at every point in the optimal solution set $Y_{sol}(x^*,\xi)$. Then*

(i) there exist a neighborhood U of x^* and a scalar $t^*>0$ such that $\hat{v}(\cdot,\xi,\cdot)$ is locally Lipschitz continuous on $U\times(0,t^*]$;

(ii) there exists a neighborhood U of x^* such that $v(\cdot,\xi)$ is locally Lipschitz continuous on U.

Proof. Part (i). Let U_1 and $t_1>0$ be given as in Assumption 3.1. We first claim that there exist a neighborhood $U\subseteq U_1$ of x^* and a scalar $0<t^*\le t_1$ such that REG(x,ξ,t) satisfies MFCQ at every point in the optimal solution set $\hat{Y}_{sol}(x,\xi,t)$ for $x\in U$ and $t\in(0,t^*]$.

Assume for a contradiction that there exist sequences $\{x^k\}\to x^*$, $\{t_k\}\to 0$ and $y^k\in\hat{Y}_{sol}(x^k,\xi,t_k)$ such that REG(x^k,ξ,t_k) fails to satisfy MFCQ at point y^k. Under Assumption 3.1, the optimal solution set $\hat{Y}_{sol}(x,\xi,t)$ is bounded for all $x\in U$ and $t\in(0,t^*]$. Moreover, it follows from Theorem 3.1 that the optimal solution set-mapping $\hat{Y}_{sol}(\cdot,\cdot,\cdot)$ is outer semi-continuous on $U\times\Xi\times[0,t^*]$ and contained in Y. Therefore the sequence $\{y^k\}$ must have an accumulation point $\bar{y}$ and any accumulation point must be in $Y_{sol}(x^*,\xi)$. Applying Proposition 3.2 at $\bar{y}$, there exist neighborhoods of U_{x^*} of x^*, $U_{\bar{y}}$ of $\bar{y}$ and $\bar{t}>0$ such that for $(x,t)\in U_{x^*}\times(0,\bar{t}]$, problem REG$(x,\xi,t)$ satisfies MFCQ at every feasible point $y\in U_{\bar{y}}$, this means that when x^k,t_k and y^k enter the neighborhood, the MFCQ holds at y^k, a contradiction!

Since functions g,h,G and H are continuously differentiable and MFCQ holds at every point in $\hat{Y}_{sol}(x,\xi,t)$ for $(x,t)\in U\times(0,t^*]$, by [20, Proposition 3], we have $\hat{\mathcal{F}}(x,\xi,t)$ is pseudo-Lipschitzian at $(y;x,t)$ where $(x,t)\in U\times(0,t^*]$ and $y\in\hat{Y}_{sol}(x,\xi,t)$. By [21, Theorem 1], $\hat{v}(x,\xi,t)$ is locally Lipschitz continuous at $(x,t)\in U\times(0,t^*]$.

Part (ii). Following Corollary 3.2 and a similar analysis of Part (i), there exists a neighborhood of U of x^* such that MPEC-MFCQ holds for every optimal solution of problem $MPCC(x,\xi)$, where $x\in U$. From [18, formula (8)], we have that for x near x^*,

$$v(x,\xi)=\min_{J\in\mathcal{J}(x^*,\xi)}v_J(x,\xi),\tag{3.17}$$

where $\mathcal{J}(x^*,\xi):=\{J\mid J\in\mathcal{J}(y),y\in Y_{sol}(x^*,\xi)\}$ and $\mathcal{J}(y)$ is defined by (2.3) with $z^*=(x^*,y,\xi)$. Denote the optimal solution set mapping of problem NLP$_\mathrm{J}(x,\xi)$ (see (2.4)) by $Y_J(x,\xi)$. For any $J\in\mathcal{J}(x^*,\xi)$, $Y_J(x^*,\xi)\cap Y_{sol}(x^*,\xi)$ is nonempty and thus $Y_J(x^*,\xi)\subseteq Y_{sol}(x^*,\xi)$. The MPEC-MFCQ assumption therefore gives the MFCQ for NLP$_\mathrm{J}(x,\xi)$ at each $y\in Y_J(x,\xi)$. By the proof of Part (i), $v_J(\cdot,\xi)$ is locally Lipschitz continuous and so is $v(\cdot,\xi)$ through (3.17). ■

It is important to note that we are short of claiming the locally Lipschitz continuity of $\hat{v}(x, \xi, t)$ at point $t = 0$ in Theorem 3.3. This is because the MFCQ established in Proposition 3.2 is satisfied only for $t > 0$. We will show the local Lipschitz continuity in Theorem 4.2 where we can use some estimates of Clarke subdifferentials of the optimal value function $\hat{v}$ for the proof.

4 Stability analysis of stationary points

In this section, we investigate the stability of stationary points of the regularized first stage problem $P_{\hat{\vartheta}}$ with respect to parameter t. This complements our discussion on the stability analysis of the optimal values and optimal solution set mappings in the preceding subsection and the topic is particularly relevant given the nonconvex nature of the regularized problem. We start our discussion with the second stage problem REG(x, ξ, t), namely the outer semicontinuity of the set of the stationary points as x, ξ and t vary.

4.1 Second stage problems

Define the Lagrangian function of the second stage problem MPCC(x, ξ):

$$\mathcal{L}(x, y, \xi; \alpha, \beta, u, v) := f(x, y, \xi) + g(x, y, \xi)^T \alpha + h(x, y, \xi)^T \beta - G(x, y, \xi)^T u - H(x, y, \xi)^T v.$$

We consider the following KKT conditions of MPCC(x, ξ):

$$\begin{cases} 0 = \nabla_y \mathcal{L}(x, y, \xi; \alpha, \beta, u, v), \\ y \in \mathcal{F}(x, \xi), \\ 0 \le \alpha \perp -g(x, y, \xi) \ge 0, \\ 0 = u_i, & i \notin \mathcal{I}_G(x, y, \xi), \\ 0 = v_i, & i \notin \mathcal{I}_H(x, y, \xi), \\ 0 \le u_i v_i, & i \in \mathcal{I}_G(x, y, \xi) \cap \mathcal{I}_H(x, y, \xi). \end{cases} \tag{4.18}$$

Let $\mathcal{W}(x, \xi)$ denote the set of KKT pairs $(y; \alpha, \beta, u, v)$ satisfying the above conditions and $S(x, \xi)$ the corresponding set of stationary points, that is, $S(x, \xi) = \Pi_y \mathcal{W}(x, \xi)$. For each $(y; \alpha, \beta, u, v)$, y is a *C stationary point* of problem MPCC(x, ξ) and (α, β, u, v) the corresponding Lagrange multipliers. When the stationary points are restricted to global minimizers, we denote the set of KKT pairs by $\mathcal{W}^*(x, \xi)$, i.e., $\mathcal{W}^*(x, \xi) = \{(y; \alpha, \beta, u, v) \in \mathcal{W}(x, \xi), y \in Y_{sol}(x, \xi)\}$.

Analogously, we can define the Lagrangian function of REG(x, ξ, t):

$$\begin{aligned} \hat{\mathcal{L}}(x, y, \xi, t; \alpha, \beta, \gamma, \theta, \lambda) \quad := \quad & f(x, y, \xi) + g(x, y, \xi)^T \alpha + h(x, y, \xi)^T \beta - G(x, y, \xi)^T \gamma - H(x, y, \xi)^T \theta \\ & + (G(x, y, \xi) \circ H(x, y, \xi) - te)^T \lambda. \end{aligned}$$

The KKT conditions of REG(x, ξ, t) can be written as:

$$\begin{cases} 0 = \nabla_y \hat{\mathcal{L}}(x, y, \xi, t; \alpha, \beta, \gamma, \theta, \lambda), \\ 0 \le -g(x, y, \xi) \perp \alpha \ge 0, \\ 0 = h(x, y, \xi), \\ 0 \le G(x, y, \xi) \perp \gamma \ge 0, \\ 0 \le H(x, y, \xi) \perp \theta \ge 0, \\ 0 \le te - G(x, y, \xi) \circ H(x, y, \xi) \perp \lambda \ge 0. \end{cases} \tag{4.19}$$

Let $\hat{\mathcal{W}}(x, \xi, t)$ denote the set of KKT pairs $(y; \alpha, \beta, \gamma, \theta, \lambda)$ satisfying the above conditions and $\hat{S}(x, \xi, t)$ the corresponding set of stationary points, that is, $\hat{S}(x, \xi) = \Pi_y \hat{\mathcal{W}}(x, \xi, t)$. When the stationary points are restricted to global minimizers, we denote the set of KKT pairs by $\hat{\mathcal{W}}^*(x, \xi, t)$.

Remark 4.1 *Under Assumption 3.1 and MPEC-MFCQ, both $\mathcal{W}^*(x, \xi)$ and $\hat{\mathcal{W}}^*(x, \xi, t)$ are nonempty and bounded.*

Assumption 4.1 *Let $x^* \in X$. There exist constants δ, $t^* > 0$, a compact set $Y \subset \mathbb{R}^m$ and a neighborhood U of x^* such that*

$$\emptyset \neq \hat{\mathcal{F}}(x, \xi, t) \subset Y,$$

for all $(x, \xi, t) \in U \times \Xi \times [0, t^*]$.

Assumption 4.1 implies the inf-compactness condition (Assumption 3.1) in that the latter only ensures the boundedness of global optimal solutions to REG(x, ξ, t). In the stability analysis of the stationary points, we need the former which ensures the set of stationary points to be bounded. Under Assumption 4.1, we have the following proposition which describes a relationship between $S(x, \xi)$ and $\hat{S}(x, \xi, t)$.

Proposition 4.1 *Let $\{(x^k, \xi^k, t_k)\} \subset X \times \Xi \times (0, +\infty)$ be a sequence such that $x^k \to x^*$, $\xi^k \to \xi$ and $t_k \downarrow 0$. Consider the regularized second stage problem REG(x^k, ξ^k, t_k). Let $y^k \in \hat{S}(x^k, \xi^k, t_k)$ and y^* be an accumulation point of sequence $\{y^k\}$.*

(i) If problem MPCC(x^*, ξ) satisfies MPEC-MFCQ at y^*, then y^* is a C-stationary point of MPCC(x^*, ξ).

(ii) If, in addition, Assumption 4.1 holds at point x^* and MPEC-MFCQ holds at every $y \in \mathcal{F}(x^*, \xi)$, then

$$\lim_{x^k \to x^*, \xi^k \to \xi, t_k \downarrow 0} \mathbb{D}(\hat{S}(x^k, \xi^k, t_k), S(x^*, \xi)) = 0.$$

Proof. Part (i). For the simplicity of notation, we write $\mathcal{I}_g(x^k, y^k, \xi^k)$ and $\mathcal{I}_g(x^*, y^*, \xi)$ as $\mathcal{I}_g^k$ and $\mathcal{I}_g^*$. Similar simplification applies to $\mathcal{I}_G$, $\mathcal{I}_H$ and $\mathcal{I}_{G \circ H}$ where

$$\mathcal{I}_{G \circ H}^k = \{i : G_i(x^k, y^k, \xi^k) H_i(x^k, y^k, \xi^k) = t_k, i = 1, \cdots, m\}.$$

Since y^k is a stationary point of REG(x^k, ξ^k, t_k), there exist multipliers $\alpha^k \in \mathbb{R}^s, \beta^k \in \mathbb{R}^r, \gamma^k \in \mathbb{R}^m, \theta^k \in \mathbb{R}^m, \lambda^k \in \mathbb{R}^m$ such that

$$\begin{aligned} 0 = \nabla_y f(x^k, y^k, \xi^k) &+ \sum_{i \in \mathcal{I}_g^k} \alpha_i^k \nabla_y g_i(x^k, y^k, \xi^k) + \sum_{i=1}^{r} \beta_i^k \nabla_y h_i(x^k, y^k, \xi^k) - \sum_{i \in \mathcal{I}_G^k} \gamma_i^k \nabla_y G_i(x^k, y^k, \xi^k) \\ &- \sum_{i \in \mathcal{I}_H^k} \theta_i^k \nabla_y H_i(x^k, y^k, \xi^k) + \sum_{i \in \mathcal{I}_{G \circ H}^k} \lambda_i^k \nabla_y [H_i(x^k, y^k, \xi^k) G_i(x^k, y^k, \xi^k)], \end{aligned} \tag{4.20}$$

$$\left\{ \begin{array}{l} 0 \leq -g(x^k, y^k, \xi^k) \perp \alpha^k \geq 0, \\ 0 = h(x^k, y^k, \xi^k), \\ 0 \leq G(x^k, y^k, \xi^k) \perp \gamma^k \geq 0, \\ 0 \leq H(x^k, y^k, \xi^k) \perp \theta^k \geq 0, \\ 0 \leq t_k e - G(x^k, y^k, \xi^k) \circ H(x^k, y^k, \xi^k) \perp \lambda^k \geq 0. \end{array} \right. \tag{4.21}$$

Let

$$\bar{\alpha}_i^k := \left\{ \begin{array}{ll} \alpha_i^k, & i \in \mathcal{I}_g^* \cap \mathcal{I}_g^k, \\ 0, & \text{otherwise}, \end{array} \right.$$

$$u_i^k := \left\{ \begin{array}{ll} \gamma_i^k, & i \in \mathcal{I}_G^* \cap \mathcal{I}_G^k, \\ -\lambda_i^k H_i(x^k, y^k, \xi^k), & i \in \mathcal{I}_G^* \cap \mathcal{I}_{G \circ H}^k, \\ 0, & \text{otherwise}, \end{array} \right.$$

$$v_i^k := \left\{ \begin{array}{ll} \theta_i^k, & i \in \mathcal{I}_H^* \cap \mathcal{I}_H^k, \\ -\lambda_i^k G_i(x^k, y^k, \xi^k), & i \in \mathcal{I}_H^* \cap \mathcal{I}_{G \circ H}^k, \\ 0, & \text{otherwise}. \end{array} \right.$$

Note that for k sufficiently large, we have $\mathcal{I}_g^k \subseteq \mathcal{I}_g^*$, $\mathcal{I}_G^k \subseteq \mathcal{I}_G^*$, $\mathcal{I}_H^k \subseteq \mathcal{I}_H^*$, $\mathcal{I}_G^k \cap \mathcal{I}_{G \circ H}^k = \emptyset$ and $\mathcal{I}_H^k \cap \mathcal{I}_{G \circ H}^k = \emptyset$. Then (4.20) can be rewritten as

$$\begin{aligned} 0 &= \nabla_y f(x^k, y^k, \xi^k) + \nabla_y g(x^k, y^k, \xi^k) \bar{\alpha}^k \\ &+ \nabla_y h(x^k, y^k, \xi^k) \beta^k - \nabla_y G(x^k, y^k, \xi^k) u^k - \nabla_y H(x^k, y^k, \xi^k) v^k + R_k(x^k, y^k, \xi^k), \end{aligned} \tag{4.22}$$

where

$$R_k(x^k,y^k,\xi^k) = \sum_{i\in\mathcal{I}^k_{G\circ H}\cap(\mathcal{I}^*_G)^c} \lambda^k_i H_i(x^k,y^k,\xi^k)\nabla_y G_i(x^k,y^k,\xi^k) + \sum_{i\in\mathcal{I}^k_{G\circ H}\cap(\mathcal{I}^*_H)^c} \lambda^k_i G_i(x^k,y^k,\xi^k)\nabla_y H_i(x^k,y^k,\xi^k).$$

Since MPEC-MFCQ holds at y^*, by Proposition 3.2 there exists k_0 sufficiently large such that MFCQ holds at point y^k for $k \geq k_0$. Moreover, the MFCQ implies that α^k, β^k, γ^k, θ^k and λ^k are uniformly bounded (see the proof of [16, Theorem 3.4]). Taking a further subsequence if necessary, we may assume that the limits

$$\alpha^*_i = \lim_{k\to\infty}\bar{\alpha}^k_i,\quad \beta^*_i = \lim_{k\to\infty}\beta^k_i,\quad u^*_i = \lim_{k\to\infty}u^k_i,\quad v^*_i = \lim_{k\to\infty}v^k_i$$

exist. Moreover, $(\mathcal{I}^*_G)^c \subseteq \mathcal{I}^*_H$ and $(\mathcal{I}^*_H)^c \subseteq \mathcal{I}^*_G$. Consequently, the limit on (4.22) implies

$$\nabla_y f(x^*,y^*,\xi) + \nabla_y g(x^*,y^*,\xi)\alpha^* + \nabla_y h(x^*,y^*,\xi)\beta^* - \nabla_y G(x^*,y^*,\xi)u^* - \nabla_y H(x^*,y^*,\xi)v^* = 0.$$

By the definitions of u^* and v^*, for $i \in \mathcal{I}^*_G\cap\mathcal{I}^*_H$, if $i\in\mathcal{I}^k_{G\circ H}$

$$u^*_i v^*_i = \lim_{k\to\infty}\big(-\lambda^k_i H_i(x^k,y^k,\xi^k)\big)\big(-\lambda^k_i G_i(x^k,y^k,\xi^k)\big) \geq 0,$$

and if $i\notin\mathcal{I}^k_{G\circ H}$,

$$u^*_i v^*_i = \lim_{k\to\infty}(\gamma^k_i \text{ or } 0)(\theta^k_i \text{ or } 0) \geq 0,$$

which indicates that y^* is a C-stationary point of problem MPCC(x^*,ξ).

Part (ii). Under the additional condition, the set of stationary points $S(x,\hat{\xi})$ and $\hat{S}(x,\hat{\xi},t)$ are bounded for $(x,\hat{\xi})$ close to (x^*,ξ) and t sufficiently small. By Proposition 3.1, $\hat{\mathcal{F}}(x,\xi,t)$ is continuous on $U\times\Xi\times[0,t^*]$. Since MPEC-MFCQ holds at every $y\in\mathcal{F}(x^*,\xi)$, we obtain part (ii) from part (i). The proof is complete. ■

Note that Proposition 4.1 deals with the Clarke stationary points. Under some moderate conditions, the stationary points of REG(x,ξ,t) may converge to an M-stationary point or a S-stationary point of MPCC(x,ξ). For more details of these conditions, see [22] and [41]. In what follows, we investigate the stability of the optimal value functions $v(x,\xi)$ and/or $\hat{v}(x,\xi,t)$ in terms of Clarke subdifferentials. The result is crucial for establishing our main result Theorem 4.1 and it is also of independent interest.

Proposition 4.2 *Suppose that Assumption 3.1 holds at point x^* and problem MPCC(x^*,ξ) satisfies MPEC-MFCQ at every point y in set $Y_{sol}(x^*,\xi)$. Then there exist a neighborhood U of x^* and a scalar $t^*>0$ such that*

(i) for any $x\in U$ and $\xi\in\Xi$,

$$\partial_x v(x,\xi) \subseteq \Phi(x,\xi), \tag{4.23}$$

where

$$\Phi(x,\xi) = \text{conv}\Big\{\bigcup_{(y;\alpha,\beta,u,v)\in\mathcal{W}^*(x,\xi)} \nabla_x\mathcal{L}(x,y,\xi;\alpha,\beta,u,v)\Big\}; \tag{4.24}$$

(ii) for any $x\in U$, $\xi\in\Xi$ and $t\in(0,t^*]$,

$$\partial_x\hat{v}(x,\xi,t)\subseteq\hat{\Phi}(x,\xi,t),\qquad \partial_t\hat{v}(x,\xi,t)\subseteq\Lambda(x,\xi,t), \tag{4.25}$$

where

$$\hat{\Phi}(x,\xi,t) = \text{conv}\left\{\bigcup_{(y;\alpha,\beta,\gamma,\theta,\lambda)\in\hat{\mathcal{W}}^*(x,\xi,t)} \nabla_x\hat{\mathcal{L}}(x,y,\xi,t;\alpha,\beta,\gamma,\theta,\lambda)\right\} \tag{4.26}$$

and $\Lambda(x,\xi,t) = \Pi_\lambda\hat{\mathcal{W}}^*(x,\xi,t)$, equality in (4.25) holds if the MPEC-MFCQ is replaced by the MPEC-LICQ;

(iii) $\Phi(\cdot,\cdot)$ *is outer semi-continuous on* $U \times \Xi$ *and* $\hat{\Phi}(\cdot,\cdot,\cdot)$ *is outer semi-continuous on* $U \times \Xi \times (0, t^*]$*;*

(iv) for every $(x, \xi) \in U \times \Xi$*,*

$$\lim_{x^k \to x, \xi^k \to \xi, t_k \downarrow 0} \mathbb{D}(\hat{\Phi}(x^k, \xi^k, t_k), \Phi(x, \xi)) = 0.$$

Proof. By a similar analysis of the proof of Theorem 3.3, there exist a neighborhood U of x^* and a scalar $t^* > 0$ such that for $x \in U$, $\xi \in \Xi$ and $t \in (0, t^*]$, REG(x, ξ, t) satisfies MFCQ at every point in the optimal solution set $\hat{Y}_{sol}(x, \xi, t)$ and MPCC(x, ξ) satisfies MPEC-MFCQ at every point in the optimal solution set $Y_{sol}(x, \xi)$.

Part (i). Following a similar argument in the proof of [25, Theorem 4.8], we can show that, for any $x \in U$ and $\xi \in \Xi$

$$\partial_x^M v(x, \xi) \subseteq \left\{ \bigcup_{(y;\alpha,\beta,u,v) \in \mathcal{W}^*(x,\xi)} \nabla_x \mathcal{L}(x, y, \xi; \alpha, \beta, u, v) \right\}. \tag{4.27}$$

Taking the convex hull on both sides of the above inclusion and using the fact that v is locally Lipschitz continuous with respect to x and

$$\text{conv}\, \partial_x^M v(x, \xi) = \partial_x v(x, \xi),$$

we obtain (4.23).

Part (ii) follows from [16, Theorem 5.3 and Corollary 5.4].

Part (iii). We only prove the outer semicontinuity of $\hat{\Phi}$ as the proof for Φ is similar.

We first prove the outer semicontinuity of $\hat{\mathcal{W}}^*(\cdot,\cdot,\cdot)$. Let (x^k, ξ^k, t_k) be an arbitrary sequence in $U \times \Xi \times (0, t^*]$ such that $(x^k, \xi^k, t_k) \to (x, \xi, t)$, where $t > 0$ and $(y^k; \alpha^k, \beta^k, \gamma^k, \theta^k, \lambda^k) \in \hat{\mathcal{W}}^*(x^k, \xi^k, t_k)$. Since that MFCQ holds at every point of optimal solution set $\hat{Y}_{sol}(x^k, \xi^k, t_k)$ for k sufficiently large, by the proof of [16, Theorem 3.4], $(y^k; \alpha^k, \beta^k, \gamma^k, \theta^k, \lambda^k) \in \hat{\mathcal{W}}(x^k, \xi^k, t_k)$ are bounded. Taking a subsequence if necessary, we may assume for the simplicity of notation that

$$(y^k; \alpha^k, \beta^k, \gamma^k, \theta^k, \lambda^k) \to (y; \alpha, \beta, \gamma, \theta, \lambda).$$

Then $(y; \alpha, \beta, \gamma, \theta, \lambda) \in \hat{\mathcal{W}}(x, \xi, t)$ as the underlying functions defining the KKT system are continuous. Moreover, considering a smaller neighborhood U of x^* and a smaller number t^* if necessary, we have, through Theorem 3.1 (i), that $\hat{Y}_{sol}(\cdot,\cdot,\cdot)$ is outer semicontinuous on $U \times \Xi \times [0, t^*]$, which implies $y \in \hat{Y}_{sol}(x, \xi, t)$ and hence $(y; \alpha, \beta, \gamma, \theta, \lambda) \in \hat{\mathcal{W}}^*(x, \xi, t)$, the outer semicontinuity of $\hat{\mathcal{W}}^*(\cdot,\cdot,\cdot)$.

The outer semicontinuity of $\hat{\Phi}$ follows from the fact that it is essentially a composite mapping of $\nabla_x \hat{\mathcal{L}}$ and $\hat{\mathcal{W}}^*$ while $\nabla_x \hat{\mathcal{L}}$ is continuous.

Part (iv). The proof is similar to that of Part (iii) except $t = 0$. Mimicking the proof of Proposition 4.1 (replacing $\hat{S}(x^k, \xi^k, t_k)$ with $\hat{Y}_{sol}(x^k, \xi^k, t_k)$), we can prove that

$$\nabla_x \hat{\mathcal{L}}(x^k, y^k, \xi^k, t_k; \alpha^k, \beta^k, \gamma^k, \theta^k, \lambda^k) \xrightarrow{k \to \infty} \nabla_x \mathcal{L}(x, y^*, \xi; \alpha, \beta, u, v),$$

where $(y^k; \alpha^k, \beta^k, \gamma^k, \theta^k, \lambda^k) \in \hat{\mathcal{W}}^*(x^k, \xi^k, t_k)$ and $(y^*; \alpha, \beta, u, v) \in \mathcal{W}^*(x, \xi)$. The conclusion follows. ■

It might be helpful to note that the equality in (4.25) under MPEC-LICQ implies that the outer bound of the Clarke subdifferentials cannot be improved. Indeed, this is a key result for establishing the subdifferential consistency in Theorem 4.1. In the literature of MPECs, Lucent and Ye established a number of estimates for the limiting subdifferentials of optimal value functions of parametric mathematical programs with variational inequality constraints without MFCQ. When the variational inequality constraint reduces to a system of equalities, their results recover Gauvin and Dubeau's result [16, Theorem 5.3] under MFCQ. However, it seems an open question as to whether the upper estimates of the limiting subdifferentials of the optimal value functions can be improved. In our context, it is unclear under which conditions equality in (4.27) holds.

Remark 4.2 *In the definition of Φ and $\hat{\Phi}$, we use the KKT pairs at the global optimal solutions of the second stage problems. It is possible to all KKT pairs in the definitions, that is, replace $\mathcal{W}^*$ and $\hat{\mathcal{W}}^*$ with $\mathcal{W}$ and $\hat{\mathcal{W}}$. Consequently we may obtain larger outer bounds Ψ and $\hat{\Psi}$ defined as follows for the Clarke subdifferentials of the optimal value functions:*

$$\Psi(x, \xi) := \text{conv} \left\{ \bigcup_{(y;\alpha,\beta,u,v)\in\mathcal{W}(x,\xi)} \nabla_x \mathcal{L}(x, y, \xi; \alpha, \beta, u, v) \right\} \tag{4.28}$$

and

$$\hat{\Psi}(x, \xi, t) := \text{conv}\left\{ \bigcup_{(y;\alpha,\beta,\gamma,\theta,\lambda)\in\hat{\mathcal{W}}(x,\xi,t)} \nabla_x \hat{\mathcal{L}}(x, y, \xi, t; \alpha, \beta, \gamma, \theta, \lambda) \right\}. \tag{4.29}$$

4.2 First stage problems

We now move on to investigate stability of stationary points of the regularized first stage problem $P_{\hat{\vartheta}}$ at $t = 0$. Our focus is on the Clarke stationary points. There are two underlying reasons: (a) the optimal value function $\hat{v}(x, \xi, t)$ is locally Lipschitz continuous in x and t for all $t > 0$ and under mild conditions $\mathbb{E}[\hat{v}(x, \xi, t)]$ is also locally Lipschitz continuous which means that the Clarke generalized gradient of both functions are well-defined; (b) we need some consistency property of the subdifferentials of $\hat{v}(x, \xi, t)$ (see equation (4.33) in Theorem 4.1) and it turns out that the Clarke subdifferentials can fulfil this under MPEC-LICQ through Proposition 4.2 (ii) while it is an open question as to whether or not the limiting subdifferential can do the job.

Let us start with the KKT conditions of problem P_{ϑ}:

$$0 \in \partial\mathbb{E}[v(x, \xi)] + \mathcal{N}_X(x),$$

where $\partial\mathbb{E}[v(x, \xi)]$ denotes the Clarke generalized gradient of $\mathbb{E}[v(x, \xi)]$ and $\mathcal{N}_X(x)$ is the normal cone to X at point x. In Theorem 3.3, $v(x, \xi)$ is proved to be locally Lipschitz continuous under MPEC-MFCQ. If the Lipschitz modulus is integrably bounded, then $\mathbb{E}[v(x, \xi)]$ is also globally Lipschitz continuous and hence $\partial\mathbb{E}[v(x, \xi)]$ is well-defined.

From computational point of view, it might be easier to calculate the subdifferential $\partial_x v(x, \xi)$ and its expectation. Consequently we may consider the following KKT conditions:

$$0 \in \mathbb{E}[\partial_x v(x, \xi)] + \mathcal{N}_X(x). \tag{4.30}$$

It is well-known that $\partial\mathbb{E}[v(x, \xi)] \subseteq \mathbb{E}[\partial_x v(x, \xi)]$, see for instance [45]. We call (4.30) the *weak KKT condition* of the first stage problem (3.7). Likewise, we may consider weak KKT conditions of $P_{\hat{\vartheta}}$:

$$0 \in \mathbb{E}[\partial_x \hat{v}(x, \xi, t)] + \mathcal{N}_X(x). \tag{4.31}$$

Let X_{sta} and $\hat{X}_{sta}(t)$ denote respectively the set of stationary points satisfying (4.30) and (4.31). In what follows, we establish a relationship between the two sets as $t \to 0$.

Theorem 4.1 *Let Assumption 3.1 hold at point x^* and $\xi \in \Xi$.*

(i) If problem MPCC(x^*, ξ) satisfies MPEC-MFCQ at every point in $Y_{sol}(x^*, \xi)$, then

$$\lim_{x\to x^*, t\downarrow 0} \mathbb{D}(\hat{\Phi}(x, \xi, t), \Phi(x^*, \xi)) = 0. \tag{4.32}$$

(ii) If the MPEC-MFCQ is replaced by the MPEC-LICQ, then

$$\lim_{x\to x^*, t\downarrow 0} \mathbb{D}(\partial_x \hat{v}(x, \xi, t), \partial_x v(x^*, \xi)) = 0, \tag{4.33}$$

moreover if: (a) X is a compact set, (b) Assumption 3.1 holds at every point x in X and MPEC-LICQ holds at any point in $Y_{sol}(x,\xi)$ for every $x \in X$ and $\xi \in \Xi$, and (c) $\partial_x \hat{v}(x,\xi,t)$ is integrably bounded[3]*, i.e., there exists $\kappa(\xi)$ such that $\|\partial_x \hat{v}(x,\xi,t)\| \le \kappa(\xi)$, then*

$$\lim_{t \downarrow 0} \mathbb{D}(\hat{X}_{sta}(t), X_{sta}) = 0. \tag{4.34}$$

Proof. Part (i). By Theorem 3.3, there exist a neighborhood U of x^* and a positive scalar t^* such that $\hat{v}(\cdot,\xi,\cdot)$ is locally Lipschitz continuous on $U \times (0,t^*]$ and $v(\cdot,\xi)$ is locally Lipschitz continuous on U. By Theorem 3.1, any accumulation point y^* of $\{y^k\}$ with $y^k \in \hat{Y}_{sol}(x^k,\xi,t_k)$ is contained in $Y_{sol}(x^*,\xi)$. Mimicking the proof of Proposition 4.1 (replacing $\hat{S}(x^k,\xi,t_k)$ with $\hat{Y}_{sol}(x^k,\xi,t_k)$), we can prove that

$$\nabla_x \hat{\mathcal{L}}(x^k, y^k, \xi, t_k; \alpha^k, \beta^k, \gamma^k, \theta^k, \lambda^k) \xrightarrow{k \to \infty} \nabla_x \mathcal{L}(x^*, y^*, \xi; \alpha^*, \beta^*, u^*, v^*),$$

where $(y^k; \alpha^k, \beta^k, \gamma^k, \theta^k, \lambda^k) \in \mathcal{W}^*(x^k,\xi,t_k)$ and $(y^*; \alpha^*, \beta^*, u^*, v^*) \in \mathcal{W}^*(x^*,\xi)$.

Part (ii). Let us first prove the subdifferential consistency (4.33). Under MPEC-LICQ, the application of [16, Corollary 5.4] to the regularized second stage problem MPEC(x^*,ξ,t) gives

$$\partial_x \hat{v}(x^*,\xi,t) = \hat{\Phi}(x^*,\xi,t). \tag{4.35}$$

On the other hand, it follows from (4.23) that $\partial_x v(x^*,\xi) \subseteq \Phi(x^*,\xi)$. In what follows, we show

$$\partial_x v(x^*,\xi) \supseteq \Phi(x^*,\xi) = \bigcup_{(y;\alpha,\beta,u,v) \in \mathcal{W}^*(x^*,\xi)} \Big\{ \nabla_x \mathcal{L}(x^*, y, \xi; \alpha, \beta, u, v) \Big\}. \tag{4.36}$$

Under the assumption that MPEC-LICQ holds at every point in optimal solution set $Y_{sol}(x^*,\xi)$, it follows by virtue of [18, Theorem 2, formula (7)] that

$$v'(x^*,\xi;q) = \min_{(y;\alpha,\beta,u,v) \in \mathcal{W}^*(x^*,\xi)} \Big\{ \nabla_x \mathcal{L}(x^*, y, \xi; \alpha, \beta, u, v)^T q \Big\},$$

where the directional derivative "v'" is with respect to x. Therefore

$$(-v)'(x^*,\xi;q) = \max_{(y;\alpha,\beta,u,v) \in \mathcal{W}^*(x^*,\xi)} \Big\{ -\nabla_x \mathcal{L}(x^*, y, \xi; \alpha, \beta, u, v)^T q \Big\}.$$

Let

$$\eta \in \bigcup_{(y;\alpha,\beta,u,v) \in \mathcal{W}^*(x^*,\xi)} \Big\{ -\nabla_x \mathcal{L}(x^*, y, \xi; \alpha, \beta, u, v) \Big\}.$$

Then there exists a KKT pair $(y;\alpha,\beta,u,v) \in \mathcal{W}^*(x^*,\xi)$ such that $\eta = -\nabla_x \mathcal{L}(x^*, y, \xi; \alpha, \beta, u, v)$ and for any $q \in \mathbb{R}^n$

$$\eta^T q = -\nabla_x \mathcal{L}(x^*, y, \xi; \alpha, \beta, u, v)^T q \le (-v)'(x^*,\xi;q) \le (-v)^o(x^*,\xi;q),$$

where $(-v)^o(x^*,\xi;q)$ denotes Clarke generalized derivative ([10]) of $-v$ in x. By the definition of Clarke generalized gradient [10, page 27], $\eta \in \partial_x(-v)(x^*,\xi)$ and by [10, Proposition 2.3.1], $\partial_x(-v)(x^*,\xi) = -\partial_x v(x^*,\xi)$. This shows $\eta \in -\partial_x v(x^*,\xi)$ and hence

$$\bigcup_{(y;\alpha,\beta,u,v) \in \mathcal{W}^*(x^*,\xi)} \Big\{ -\nabla_x \mathcal{L}(x^*, y, \xi; \alpha, \beta, u, v) \Big\} \subseteq \partial_x(-v)(x^*,\xi) = -\partial_x v(x^*,\xi),$$

which implies (4.36). This shows $\partial_x v(x^*,\xi) = \Phi(x^*,\xi)$, and through (4.32) and (4.35), the subdifferential consistency (4.33).

Let us now prove (4.34). Since Assumption 3.1 holds at every point x in X and MPEC-LICQ holds at any point in $Y_{sol}(x,\xi)$ for every $x \in X$ and $\xi \in \Xi$, we have from the subdifferential consistency (4.33) that

$$\lim_{x' \to x, t \downarrow 0} \mathbb{D}(\partial_x \hat{v}(x',\xi,t), \partial_x v(x,\xi)) = 0 \tag{4.37}$$

[3]The condition is satisfied under Assumption 4.2.

for every $(x, \xi) \in X \times \Xi$. Let $x(t) \in \hat{X}_{sta}(t)$, that is,

$$0 \in \mathbb{E}[\partial_x \hat{v}(x(t), \xi, t)] + \mathcal{N}_X(x(t)). \tag{4.38}$$

The compactness of X implies the boundedness of $\hat{X}_{sta}(t)$. Therefore we may assume without loss of generality that $x(t) \to \hat{x}$, where $\hat{x} \in X$. From (4.38), we have

$$\begin{array}{rcl} 0 & \in & \overline{\lim_{t\to 0}}(\mathbb{E}[\partial_x \hat{v}(x(t), \xi, t)] + \mathcal{N}_X(x(t))) \\ & \subseteq & \mathbb{E}[\overline{\lim_{t\to 0}}\, \partial_x \hat{v}(x(t), \xi, t)] + \mathcal{N}_X(\hat{x}) \\ & \subseteq & \mathbb{E}[\partial_x v(\hat{x}, \xi)] + \mathcal{N}_X(\hat{x}), \end{array}$$

where the first inclusion follows from [4, Proposition 4.1] under the integrable boundedness of $\partial_x \hat{v}(x(t), \xi, t)$ and the outer semicontinuity of the normal cone $\mathcal{N}(\cdot)$, and the second inclusion follows from (4.37). This implies that $\hat{x}$ is a weak KKT point satisfying (4.30). The proof is complete. ■

The first order optimality conditions (4.30)–(4.31) require the derivative information of the optimal value function $v(x, \xi)$ which may be difficult to calculate. Motivated by the outer bounds of $\partial_x v(x, \xi)$ and $\partial_x \hat{v}(x, \xi, t)$ established in Proposition 4.2, we may consider optimality conditions by replacing $\partial_x v(x, \xi)$ with $\Phi(x, \xi)$ in the weak KKT conditions (4.30) and $\partial_x \hat{v}(x, \xi, t)$ with $\hat{\Phi}(x, \xi, t)$ in the weak KKT conditions (4.31). This kind of optimality conditions are considered by Outrata and Römisch [29, Theorem 3.5] and more recently by Ralph and Xu [35] for classical two stage stochastic programs. We will not go to details in this direction as this is not the main interest of this paper. Likewise, we can consider the KKT condition by replacing the subgradients with Ψ and $\hat{\Psi}$ as defined in Remark 4.2. We give a formal definition for the latter as we need them in Section 6.

Definition 4.1 We call the following stochastic generalized equations

$$0 \in \mathbb{E}[\Psi(x, \xi)] + \mathcal{N}_X(x) \tag{4.39}$$

the *relaxed KKT conditions* of the first stage true problem (3.7), and

$$0 \in \mathbb{E}[\hat{\Psi}(x, \xi, t)] + \mathcal{N}_X(x), \tag{4.40}$$

the *relaxed KKT conditions* of the first stage regularized problem (3.10). A point $x^* \in X$ satisfying the equation (4.39) is called a *relaxed stationary point* of the true problem if for almost every $\xi \in \Xi$, MPEC-MFCQ holds at any point in the set of stationary points $S(x^*, \xi)$. A point $x^* \in X$ satisfying equation (4.40) is called a *relaxed stationary point* of the regularized problem if for almost every $\xi \in \Xi$, MFCQ holds at any point in the set of stationary points $\hat{S}(x^*, \xi, t)$.

Note that the MPEC-MFCQ and MFCQ are needed in Definition 4.1 in order to guarantee that the generalized equations are relevant to the first order optimality conditions in that under the constraint qualifications and Assumption 3.1, the two optimal value functions v and $\hat{v}$ are locally Lipschitz continuous with respect to x on a neighborhood of x^* and the estimates for the Clarke subdifferentials in Proposition 4.2 are valid.

Note also that in the literature stochastic programming, this type of relaxed KKT conditions were considered by Ralph and Xu [35] for an ordinary two stage stochastic program with equality and inequality constraints and by Xu and Ye in deriving first order optimality conditions for a two stage SMPEC with variational inequality constraints [49].

Assumption 4.2 *For every $x \in X$, there exist an integrable function $\kappa(\xi)$, a neighborhood $\bar{U}$ of x and a scalar $\bar{t} > 0$ such that $\mathbb{E}[\kappa(\xi)^3] < \infty$ and*

$$\max \Big\{ \|\nabla_x f(x, y, \xi)\|, \|\nabla_x g(x, y, \xi)\|, \|\nabla_x h(x, y, \xi)\|, \|G(x, y, \xi)\|, \|H(x, y, \xi)\|, \|\nabla_x G(x, y, \xi)\|, \\ \|\nabla_x H(x, y, \xi)\|, \|\Pi_{(\alpha,\beta,u,v)} \mathcal{W}(x, \xi)\|, \|\Pi_{(\alpha,\beta,\gamma,\theta,\lambda)} \hat{\mathcal{W}}(x, \xi, t)\| \Big\} \le \kappa(\xi),$$

for all $x \in \bar{U}$, $\xi \in \Xi$, $t \in [0, \bar{t}]$ and $y \in \hat{S}(x, \xi, t)$.

Note that Assumption 4.2 holds when the support set Ξ of $\xi(\omega)$ is bounded. The boundedness of $\|G(x, y, \xi)\|$ and $\|H(x, y, \xi)\|$ can be weakened to the boundedness of the two quantities at a fixed point $x_0 \in U$ because the latter together with the boundedness of $\|\nabla_x G(x, y, \xi)\|$ and $\|\nabla_x H(x, y, \xi)\|$ imply the former. Moreover, under Assumption 4.2, we can easily verify that $\nabla_x \mathcal{L}$ and $\nabla_x \hat{\mathcal{L}}$ are bounded respectively by $\kappa(\xi)^2$ and $\kappa(\xi)^3$ for all $x \in \bar{U}$, $\xi \in \Xi$, $t \in [0, \bar{t}]$ and $y \in \hat{S}(x, \xi, t)$.

Proposition 4.3 *Suppose that Assumption 4.1 holds at point x and MPEC-MFCQ holds for MPCC(x, ξ) at every $y \in \mathcal{F}(x, \xi)$ and $\xi \in \Xi$. Then there exist a neighborhood of U of x and a scalar $t^* > 0$ such that*

(i) Both $\hat{\mathcal{W}}(x, \xi, t)$ and $\mathcal{W}(x, t)$ are nonempty for $(x, \xi, t) \in U \times \Xi \times (0, t^*]$, $\hat{\mathcal{W}}(\cdot, \cdot, \cdot)$ is outer semicontinuous on $U \times \Xi \times (0, t]$ and $\mathcal{W}(\cdot, \cdot)$ is outer semicontinuous on $U \times \Xi$;

(ii) for every $(x^*, \xi^*) \in U \times \Xi$,

$$\lim_{(x,\xi,t)\to(x^*,\xi^*,0)} \mathbb{D}(\hat{\Psi}(x, \xi, t), \Psi(x^*, \xi^*)) = 0; \tag{4.41}$$

(iii) under Assumption 4.2, $\mathbb{E}[\hat{\Psi}(x, \xi, t)]$ and $\mathbb{E}[\Psi(x, \xi)]$ are well-defined for any $x \in U$ and $t \in (0, t^*]$ and

$$\lim_{x\to x^*, t\downarrow 0} \mathbb{D}(\mathbb{E}[\hat{\Psi}(x, \xi, t)], \mathbb{E}[\Psi(x^*, \xi)]) = 0 \quad \forall x \in U. \tag{4.42}$$

Proof. Part (i). By Assumption 4.1, there exist a neighborhood U of x and a scalar $t^* > 0$ such that the feasible set $\mathcal{F}(x, \xi)$ and $\hat{\mathcal{F}}(x, \xi, t)$ are bounded for $x \in U$ and $t \in (0, t^*]$. Then the sets of stationary points of both MPCC(x, ξ) and REG(x, ξ, t) are nonempty. Following a similar proof to that in Proposition 4.2 (iii), we can show that $\hat{\mathcal{W}}(\cdot, \cdot, \cdot)$ is outer semicontinuous on $U \times \Xi \times (0, t^*]$ and $\mathcal{W}(\cdot, \cdot)$ is outer semicontinuous on $U \times \Xi$.

Part (ii). The proof is similar to that of Proposition 4.2 (iv). We omit the details.

Part (iii). Viewing $\hat{\Psi}$ as a composition of $\nabla\hat{\mathcal{L}}$ and $\hat{\mathcal{W}}$, we claim that $\hat{\Psi}$ is outer semicontinuous and through [38, Theorem 14.13] the measurability. The well-definedness then follows from the boundedness of $\hat{\Psi}$ under Assumption 4.2 and the definition of Aumann's integral. Finally, we prove equation (4.42). Notice that $\hat{\Psi}$ is a closed set-valued mapping on $U \times \Xi \times (0, t^*]$ and it is integrable bounded under Assumption 4.2. Note that the above analysis also holds for Ψ. The conclusion follows via application of [14, Theorems 2.5] (or [14, Theorem 2.8] and the following remark). The proof is complete. ■

Note that Proposition 4.3 (iii) implies that any stationary point satisfying (4.40) converges to the set of stationary points satisfying (4.39). We will use this in Section 6.

4.3 Lipschitz continuity at $t = 0$

In this subsection, we study the Lipschitz continuity of $\hat{v}(x, \xi, t)$ at $t = 0$. We are unable to do this in Theorem 4.2 as it requires some complex arguments related to singular subdifferentials, limiting subdifferentials, Clarke subdifferentials of $\hat{v}(x, \xi, t)$ and their approximations.

Theorem 4.2 *Suppose that Assumption 3.1 holds at point x^* and problem MPCC(x^*, ξ) satisfies MPEC-MFCQ at every point in the optimal solution set $Y_{sol}(x^*, \xi)$ for every $\xi \in \Xi$. Then*

(i) there exist a neighborhood U of x^* and a scalar $t^* > 0$ such that $\hat{v}(\cdot, \xi, \cdot)$ is locally Lipschitz continuous on $U \times [0, t^*]$ for each fixed $\xi \in \Xi$;

(ii) if Assumption 4.2 holds at point x^*, then there exist a neighborhood U of x^* and a scalar t^* such that $\mathbb{E}[\hat{v}(\cdot, \xi, \cdot)]$ is locally Lipschitz continuous on $U \times [0, t^*]$;

(iii) if, in addition, the conditions of Theorem 3.2 are satisfied and Assumption 4.2 holds for all $x \in \bar{X}$ ($\bar{X}$ is given in Theorem 3.2), then there exists a scalar $t^* > 0$ such that $\phi(t)$ is globally Lipschitz continuous on $[0, t^*]$.

Proof. Part (i). By Theorem 3.3, there exist a close neighborhood U of x^* and a scalar $t^* > 0$ such that $\hat{v}(\cdot, \xi, \cdot)$ is locally Lipschitz continuous on $U \times (0, t^*]$ and $v(\cdot, \xi)$ is locally Lipschitz continuous on U. To complete the proof, we only need to show that $\hat{v}(x, \xi, t)$ is Lipschitz continuous at point $(x, 0)$ for every $x \in U$. By [25, Proposition 2.4], it suffices to show that $\partial^{\infty}_{(x,t)} v(x, \xi, 0) = \{0\}$. From Proposition 4.2 (see (4.25)) and [10, Proposition 2.3.15], we have

$$\partial_{(x,t)} \hat{v}(x, \xi, t) \subseteq \hat{\Phi}(x, \xi, t) \times \Pi_\lambda \hat{\mathcal{W}}^*(x, \xi, t).$$

If we can show the boundedness of $\hat{\Phi}(x, \xi, t)$ and $\Pi_\lambda \hat{\mathcal{W}}^*(x, \xi, t)$ for all $x \in U$ and $t \in (0, t^*)$, then $\partial_{(x,t)} \hat{v}(x, \xi, t)$ is bounded and so is $\partial^{\pi}_{(x,t)} \hat{v}(x, \xi, t)$, subsequently we have $\partial^{\infty}_{(x,t)} \hat{v}(x, \xi, 0) = \{0\}$ (see the definition of the singular subdifferential). Note that the boundedness of $\hat{\Phi}(x, \xi, t)$ and $\Pi_\lambda \hat{\mathcal{W}}^*(x, \xi, t)$ is implied by the boundedness of $\hat{\mathcal{W}}^*(x, \xi, t)$. Under Assumption 3.1, $\hat{Y}_{sol}(x, \xi, t)$ is bounded. Since MPEC-MFCQ holds at every point in the optimal solution set $Y_{sol}(x^*, \xi)$, by the proof of Theorem 3.3, there exist a neighborhood U of x^* and a scalar $t^* > 0$ such that for $x \in U$ and $t \in (0, t^*]$, REG(x, ξ, t) satisfies MFCQ at every point in the optimal solution set $\hat{Y}_{sol}(x, \xi, t)$. Under the MFCQ, the boundedness of $\hat{\mathcal{W}}^*(x, \xi, t)$ follows from the proof of [16, Theorem 3.4].

Part (ii). The Lipschitz modulus of $\hat{v}(\cdot, \xi, \cdot)$ at point (x, t) is bounded by $\|\partial_{(x,t)} \hat{v}(x, \xi, t)\|$. By Propositions 4.2 and Assumption 4.2, the Lipschitz modulus is bounded by integrable function $\kappa(\xi)^3$ for $x \in U_1 \cap U_2$ and $t \in [0, \min\{t_1, t_2\}]$, where U_1, t_1 are given as in Part (i) and U_2, t_2 are given as in Assumption 4.2. From Proposition 2 of [40, Chaper 2] and $\hat{v}(x, \cdot, t)$ is continuous on Ξ, $\mathbb{E}[\hat{v}(x, \xi, t)]$ is locally Lipschitz continuous on $U \times [0, t^*]$, where $U = U_1 \cap U_2$ and $t^* = \min\{t_1, t_2\}$.

Part (iii). Applying the conclusion in part (i) to every point x in $\bar{X}$, we can show through the finite covering theorem (due to the compactness of $\bar{X}$) that there exists a scalar t_1 such that $\hat{v}(x, \xi, t)$ is locally Lipschitz continuous on $\bar{X} \times [0, t_1]$. Moreover, since Assumption 4.2 holds for every $x \in \bar{X}$, then $\hat{v}(x, \xi, t)$ is integrably bounded and $\hat{\vartheta}(x, t) = \mathbb{E}[\hat{v}(x, \xi, t)]$ is globally Lipschitz continuous on $\bar{X} \times [0, t_1]$. On the other hand, there exists a scalar $t_2 > 0$ such that for all $t \in [0, t_2]$, $X_{sol}(t) \cap \bar{X} \neq \emptyset$. Let $t^* = \min\{t_1, t_2\}$ and $t', t'' \in [0, t^*]$ with $t' < t''$. It is easy to verify that

$$|\phi(t') - \phi(t'')| \quad \leq \quad \sup_{x \in \bar{X}} |\hat{\vartheta}(x, t') - \hat{\vartheta}(x, t'')|.$$

By Lebourg's mean value theorem [10, Theorem 2.3.7] and Proposition 4.2 (ii),

$$\begin{aligned}
|\hat{\vartheta}(x, t') - \hat{\vartheta}(x, t'')| &\leq \sup_{t \in [t' t'']} \|\partial_t \hat{\vartheta}(x, t)\| |t' - t''| \\
&\leq \sup_{t \in [t', t'']} \mathbb{E}[\|\partial_t \hat{v}(x, \xi, t)\|] |t' - t''| \\
&\leq \sup_{t \in [t', t'']} \mathbb{E}[\|\Pi_\lambda \hat{\mathcal{W}}(x, \xi, t)\|] |t' - t''| \\
&\leq \mathbb{E}[\kappa(\xi)] |t' - t''|.
\end{aligned}$$

The last inequality is due to Assumption 4.2. The conclusion follows. ■

Note that Theorem 4.2 plays an essential role in the proof of Theorem 6.1.

5 Stability analysis with respect to the probability measure

The regularization scheme discussed in the preceding section is proposed to deal with complementarity constraints. In this section, we discuss another main challenge in SMPCC (1.1), that is, the mathematical expectation operation in the objective. If we can obtain a closed form of the expected values of $\mathbb{E}[v(x, \xi(\omega))]$ and $\mathbb{E}[\hat{v}(x, \xi(\omega), t)]$, then the resulting first stage problems are deterministic minimization problems. However, in many practical instances, it is often difficult to obtain an explicit expression of the optimal value function of the second stage problems and hence its mathematical expectation. Consequently, we need some kind of approximation of the expected value.

In this section, we discuss a general probability approximation scheme. Specifically, we write $\mathbb{E}[\hat{v}(x, \xi, t)]$ as $\int_\Xi \hat{v}(x, \xi, t) dP(\xi)$ and then consider a sequence of probability measures $\{P_\nu\}$ approximating P. Here

P_ν is assumed to be numerically more tractable than P. A specific example of such probability approximation is the empirical probability measure. To simplify the discussion, we fix the regularization parameter t throughout this section.

Consider the first stage regularized problem (3.10). Let Ξ be the support set of $\xi(\omega)$ and P be a Borel probability measure on Ξ. Problem (3.10) can be equivalently written as

$$\begin{array}{ll} \min\limits_x & \hat{\vartheta}_P(x,t) = \int_\Xi \hat{v}(x,\xi,t)dP(\xi) \\ \text{subject to} & x \in X. \end{array} \tag{5.43}$$

Let P_ν be a sequence of probability measures $\{P_\nu\}$ approximating P in distribution as $\nu \to \infty$. Instead of solving (5.43) directly, we solve the following approximation problem:

$$\begin{array}{ll} \min\limits_x & \hat{\vartheta}_{P_\nu}(x,t) = \int_\Xi \hat{v}(x,\xi,t)dP_\nu(\xi) \\ \text{subject to} & x \in X. \end{array} \tag{5.44}$$

We study the perturbation of the optimal value, and the set of optimal solutions and stationary points of (5.44) as $P_\nu \to P$. In the literature of stochastic programming, this kind of perturbation analysis is known as stability and/or sensitivity analysis, see a comprehensive review by Römisch [39] and references therein.

Let $\phi_P(t)$, $\phi_{P_\nu}(t)$, $X^*_P(t)$ and $X^*_{P_\nu}(t)$ denote the optimal values and solutions of (5.43) and (5.44) respectively.

Theorem 5.1 *Let $\bar{X}$ be a compact subset of X and Assumption 3.1 hold at every $x \in \bar{X}$. Suppose that there exist a positive constant $\bar{t}$ and a positive integer $\bar{\nu}$ such that $X^*_P(t) \cap \bar{X} \neq \emptyset$ and $X^*_{P_\nu}(t) \cap \bar{X} \neq \emptyset$ for any $t \in [0,\bar{t}]$ and $\nu \geq \bar{\nu}$. Then there exists a positive scalar $\hat{t} < \bar{t}$ such that for every fixed $t \in [0,\hat{t}]$*

(i) $\lim\limits_{\nu\to\infty} \mathbb{D}(X^*_{P_\nu}(t) \cap \bar{X}, X^*_P(t) \cap \bar{X}) = 0$,

(ii) $\lim\limits_{\nu\to\infty} \phi_{P_\nu}(t) = \phi_P(t)$.

Proof. By the covering theorem and Theorem 3.1, there exist positive constants $\hat{t} < \bar{t}$ and $\hat{\delta}$ such that $\hat{v}(x,\xi,t)$ is continuous on $\bar{X} \times \Xi \times [0,\hat{t}]$ and $\hat{v}(x,\xi,t) \leq \hat{\delta}$. By [40, Chapter 2, Proposition 1], $\hat{\vartheta}_P(x,t)$ and $\hat{\vartheta}_{P_\nu}(x,t)$, $\nu = 1,2,\cdots$, are continuous on $\bar{X} \times [0,\hat{t}]$ and hence they are bounded on the set. Since $P_\nu(\xi)$ converges to $P(\xi)$ in distribution by assumption, then

$$\lim_{\nu\to\infty} \sup_{(x,t)\in\bar{X}\times[0,\hat{t}]} (\hat{\vartheta}_{P_\nu}(x,t) - \hat{\vartheta}_P(x,t)) = \lim_{\nu\to\infty} \sup_{(x,t)\in\bar{X}\times[0,\hat{t}]} \int_\Xi \hat{v}(x,\xi,t)d(P_\nu(\xi) - P(\xi)) = 0$$

It is well known that the uniform convergence of $\hat{\vartheta}_{P_\nu}(\cdot,t)$ to $\hat{\vartheta}(\cdot,t)$ over compact set $\bar{X}$ implies the convergence of its optimal value and optimal solutions, see for instance [47, Lemma 4.1]. ■

In what follows, we investigate the stability of the set of stationary points. It is easy to verify that if $\hat{v}(x,\xi,t)$ is Lipschitz continuous w.r.t x for almost every ξ and t and its Lipschitz constant is integrably bounded under the probability measure P and P_ν, then $\hat{\vartheta}_P(x,t)$ and $\hat{\vartheta}_{P_\nu}(x,t)$ are Lipschitz continuous with respect to x. The KKT conditions of (5.43) and (5.44) can be written respectively as

$$0 \in \partial_x \hat{\vartheta}_P(x,t) + \mathcal{N}_X(x) \tag{5.45}$$

and

$$0 \in \partial_x \hat{\vartheta}_{P_\nu}(x,t) + \mathcal{N}_X(x), \tag{5.46}$$

where ∂ denotes the Clarke subdifferential. Let $S^*_P(t)$ and $S^*_{P_\nu}(t)$ denote the set of stationary points satisfying (5.45) and (5.46) respectively. Following a similar argument to that in section 3.2, we may consider weaker KKT conditions of (5.43) and (5.44) defined respectively as

$$0 \in \int_\Xi \partial_x \hat{v}(x,\xi,t)dP(\xi) + \mathcal{N}_X(x) \tag{5.47}$$

and

$$0 \in \int_{\Xi} \partial_x \hat{v}(x,\xi,t) dP_\nu(\xi) + \mathcal{N}_X(x), \tag{5.48}$$

where $\partial_x \hat{\vartheta}(x,t) \subset \int_{\Xi} \partial_x \hat{v}(x,\xi,t)dP(\xi)$ and $\partial_x \hat{\vartheta}_{P_\nu}(x,t) \subset \int_{\Xi} \partial_x \hat{v}(x,\xi,t)dP_\nu(\xi)$. Let $S_P^w(t)$ and $S_{P_\nu}^w(t)$ denote the set of stationary points satisfying (5.47) and (5.48) respectively. We investigate the approximation of $S_P^w(t)$ and $S_P^*(t)$ by $S_{P_\nu}^w(t)$ and $S_{P_\nu}^*(t)$ respectively as $\nu \to \infty$. To this end, we need to show, under some moderate conditions, that $\partial_x \hat{\vartheta}_{P_\nu}(x,t)$ approximates $\partial_x \hat{\vartheta}_P(x,t)$ and $\int_{\Xi} \partial_x \hat{v}(x,\xi,t)dP_\nu(\xi)$ approximates $\int_{\Xi} \partial_x \hat{v}(x,\xi,t)dP(\xi)$ uniformly as $\nu \to \infty$.

Lemma 5.1 (Approximation of subdifferentials) *Let $F(x,\xi) : \mathbb{R}^n \times \Xi \to \mathbb{R}^m$ be a continuous function, $\{P_\nu\}$ be a sequence of probability measures and $\mathcal{X}$ be a compact subset. Assume: (a) $F(x,\xi)$ is locally Lipschitz continuous with respect to x for almost every ξ with modulus $L(x,\xi)$ which is bounded by a positive constant C; (b) $\{P_\nu\}$ converges to P in distribution. Then*

(i) for every fixed x, $\partial \mathbb{E}_{P_\nu}[F(x,\xi)]$ and $\partial \mathbb{E}_P[F(x,\xi)]$ are well-defined and

$$\lim_{\nu\to\infty} \sup_{x\in\mathcal{X}} \mathbb{H}(\partial \mathbb{E}_{P_\nu}[F(x,\xi)], \partial \mathbb{E}_P[F(x,\xi)]) = 0; \tag{5.49}$$

(ii) if $\partial_x F(x,\xi)$ is osc in ξ, then

$$\lim_{\nu\to\infty} \sup_{x\in\mathcal{X}} \mathbb{D}(\mathbb{E}_{P_\nu}[\partial_x F(x,\xi)], \mathbb{E}_P[\partial_x F(x,\xi)]) = 0; \tag{5.50}$$

if, in addition, $\partial_x F(x,\xi)$ is Hausdorff continuous in ξ, then

$$\lim_{\nu\to\infty} \sup_{x\in\mathcal{X}} \mathbb{H}(\mathbb{E}_{P_\nu}[\partial_x F(x,\xi)], \mathbb{E}_P[\partial_x F(x,\xi)]) = 0. \tag{5.51}$$

Proof. Part (i). For the simplicity of notation, let $f_{P_\nu}(x) = \mathbb{E}_{P_\nu}[F(x,\xi)]$ and $f_P(x) = \mathbb{E}_P[F(x,\xi)]$. Under condition (a), both $f_{P_\nu}(x)$ and $f_P(x)$ are globally Lipschitz continuous, therefore Clarke's generalized derivatives of $f_{P_\nu}(x)$ and $f_P(x)$, denoted by $f^o_{P_\nu}(x;h)$ and $f^o_P(x;h)$ respectively, are well-defined for any fixed nonzero vector $h \in \mathbb{R}^n$, where

$$f^o_{P_\nu}(x;h) = \limsup_{x'\to x, \tau\downarrow 0} \frac{1}{\tau}(f_{P_\nu}(x'+\tau h) - f_{P_\nu}(x'))$$

and

$$f^o_P(x;h) = \limsup_{x'\to x, \tau\downarrow 0} \frac{1}{\tau}(f_P(x'+\tau h) - f_P(x')).$$

Our idea is to study the Hausdorff distance $\mathbb{H}(\partial f_{P_\nu}(x), \partial f_P(x))$ through certain "distance" of the Clarke generalized derivatives $f^o_{P_\nu}(x;h)$ and $f^o_P(x;h)$. Let D_1, D_2 be two convex and compact subsets of $\mathbb{R}^m$. Let $\sigma(D_1,u)$ and $\sigma(D_2,u)$ denote the support functions of D_1 and D_2 respectively. Then

$$\mathbb{D}(D_1,D_2) = \max_{\|u\|\le 1} (\sigma(D_1,u) - \sigma(D_2,u))$$

and

$$\mathbb{H}(D_1,D_2) = \max_{\|u\|\le 1} |\sigma(D_1,u) - \sigma(D_2,u)|.$$

The above relationships are known as Hömander's formulae, see [7, Theorem II-18]. Applying the second formula to our setting, we have

$$\mathbb{H}(\partial f_{P_\nu}(x), \partial f_P(x)) \quad = \quad \sup_{\|h\|\le 1} \left|\sigma(\partial f_{P_\nu}(x),h) - \sigma(\partial f_{P_\nu}(x),h)\right|.$$

Using the relationship between Clarke's subdifferential and Clarke's generalized derivative, we have that $f^o_{P_\nu}(x;h) = \sigma(\partial f_{P_\nu}(x),h)$ and $f^o_P(x;h) = \sigma(\partial f_P(x),h)$. Consequently,

$$\begin{aligned}\mathbb{H}(\partial f_{P_\nu}(x), \partial f_P(x)) &= \sup_{\|h\|\le 1} \left|f^o_P(x;h) - f^o_{P_\nu}(x;h)\right| \\ &= \sup_{\|h\|\le 1} \left| \limsup_{x'\to x,\tau\downarrow 0} \frac{1}{\tau}(f_P(x'+\tau h) - f_P(x')) - \limsup_{x'\to x,\tau\downarrow 0} \frac{1}{\tau}(f_{P_\nu}(x'+\tau h) - f_{P_\nu}(x'))\right|.\end{aligned}$$

Note that for any bounded sequence $\{a_k\}$ and $\{b_k\}$, we have

$$|\limsup_{k\to\infty} a_k - \limsup_{k\to\infty} b_k| \leq \limsup_{k\to\infty} |a_k - b_k|. \tag{5.52}$$

To see this, let $\{a_{k_j}\}$ be a subsequence such that $\limsup_{k\to\infty} a_k = \lim_{k_j\to\infty} a_{k_j}$. Then

$$\begin{array}{rcl}\limsup_{k\to\infty} |a_k - b_k| & \geq & \limsup_{k_j\to\infty} |a_{k_j} - b_{k_j}| \\ & \geq & \limsup_{k_j\to\infty} (a_{k_j} - b_{k_j}) \\ & = & \limsup_{k\to\infty} a_k + \limsup_{k_j\to\infty}(-b_{k_j}) \\ & \geq & \limsup_{k\to\infty} a_k + \liminf_{k_j\to\infty}(-b_{k_j}) \\ & \geq & \limsup_{k\to\infty} a_k + \liminf_{k\to\infty}(-b_k) \\ & = & \limsup_{k\to\infty} a_k - \limsup_{k\to\infty} b_k.\end{array}$$

Since a_k and b_k are in a symmetric position, we have that

$$\limsup_{k\to\infty} |a_k - b_k| \geq \limsup_{k\to\infty} b_k - \limsup_{k\to\infty} a_k.$$

This verifies (5.52). Using (5.52), we have

$$\begin{array}{rcl}\mathbb{H}(\partial f_{P_\nu}(x), \partial f_P(x)) & \leq & \sup_{\|h\|\leq 1} \limsup_{x'\to x, \tau\downarrow 0} \left| \frac{1}{\tau}(f_P(x'+\tau h) - f_P(x')) - \frac{1}{\tau}(f_{P_\nu}(x'+\tau h) - f_{P_\nu}(x')) \right| \\ & & = \sup_{\|h\|\leq 1} \limsup_{x'\to x, \tau\downarrow 0} \left| \int_\Xi \frac{1}{\tau}(F(x'+\tau h, \xi) - F(x', \xi)) d(P - P_\nu)(\xi) \right|.\end{array}$$

Since P_ν converges to P in distribution, and the integrand $\frac{1}{\tau}(F(x'+\tau h,\xi) - F(x',\xi))$ is continuous w.r.t ξ and it is bounded by L, then

$$\lim_{\nu\to\infty} \sup_{x\in\mathcal{X}} \sup_{\|h\|\leq 1} \limsup_{x'\to x, \tau\downarrow 0} \left| \int_\Xi \frac{1}{\tau}(F(x'+\tau h, \xi) - F(x', \xi)) d(P - P_\nu)(\xi) \right| = 0.$$

Part (ii). We first show that $\mathbb{E}_{P_\nu}[\partial_x F(x,\xi)]$ and $\mathbb{E}_P[\partial_x F(x,\xi)]$ are well-defined. The continuity of $F(x,\xi)$ in ξ implies the measurability of $F(x,\xi(\cdot))$ and $F^o(x,\xi(\cdot);h)$ by virtue of [3, Theorem 8.2.5]. Since $F^o(x,\xi;h)$ is the support function of $\partial_x F(x,\xi)$, by [3, Theorem 8.2.14], $\partial_x F(x,\xi(\cdot))$ is also measurable. Moreover, the Clarke subdifferential $\partial_x F(x,\xi)$ is compact set-valued and bounded by C (under condition (a)), which implies that $\mathbb{E}_P[\partial_x F(x,\xi)]$ is nonempty, compact set-valued, and $\mathbb{E}_P[\|\partial_x F(x,\xi)\|] \leq C$. By [1], $\mathbb{E}_P[\partial_x F(x,\xi)]$ is well-defined. Using the same argument, we can show the well definedness of $\mathbb{E}_{P_\nu}[\partial_x F(x,\xi)]$. Note that $\partial_x F(x,\xi)$ is convex set-valued, we obtain (5.50) through [2, Theorem 4.2], and (5.51) by virtue of [2, Theorem 3.1]. The proof is complete. ■

We make a few comments about Lemma 5.1 because it is not only prepared for establishing our main result, Theorem 5.2, but also of general interest. First, Birge and Qi [5] investigated pointwise approximation of $\partial\mathbb{E}_{P_\nu}[F(x,\xi)]$ to $\partial\mathbb{E}_P[F(x,\xi)]$ (i.e. for fixed x) under the condition that P_ν is a particular class of continuous probability measures whose distribution function has a piecewise continuous density function, see [5, Theorem 4.1] for details. Our result (5.49) is stronger than the convergence result in [5, equation (4.1)] in the sense that the convergence here is unform and there is no restriction on the distribution of P_ν. Second, Artstein and Wets [2] established a number of convergence results for the integral of random set-valued mappings when the probability measures P_ν converges weakly to P. Lemma 5.1 (ii) is a direct application of their results to Clarke subdifferentials. Third, consider a popular special case that P_ν is an empirical probability measure. That is,

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

where $\xi^1, \cdots, \xi^\nu$ is an independent and identically distributed sampling of ξ and

$$\mathbb{1}_{\xi^k}(\omega) := \begin{cases} 1, & \text{if } \xi(\omega) = \xi^k, \\ 0, & \text{if } \xi(\omega) \neq \xi^k. \end{cases}$$

In this case

$$\partial \mathbb{E}_{P_\nu}[F(x,\xi)] = \partial \left(\frac{1}{\nu} \sum_{k=1}^{\nu} F(x, \xi^k) \right)$$

and

$$\mathbb{E}_{P_\nu}[\partial_x F(x,\xi)] = \frac{1}{\nu} \sum_{k=1}^{\nu} \partial_x F(x, \xi^k).$$

From the calculus of Clarke subdifferential, we know that

$$\partial \mathbb{E}_{P_\nu}[F(x,\xi)] \subseteq \mathbb{E}_{P_\nu}[\partial_x F(x,\xi)]$$

and equality holds when $F(\cdot, \xi^k)$, $k = 1, \cdots, \nu$, is Clarke regular at x. Moreover, Lemma 5.1 indicates that

$$\lim_{\nu \to \infty} \sup_{x \in \mathcal{X}} \mathbb{H} \left(\partial \left(\frac{1}{\nu} \sum_{k=1}^{\nu} F(x, \xi^k) \right), \partial \mathbb{E}_P[\partial_x(x,\xi)] \right) = 0$$

and

$$\lim_{\nu \to \infty} \sup_{x \in \mathcal{X}} \mathbb{D} \left(\frac{1}{\nu} \sum_{k=1}^{\nu} \partial_x F(x, \xi^k), \mathbb{E}_P[\partial_x F(x,\xi)] \right) = 0.$$

If, in addition, $\partial_x F(x,\xi)$ is Hausdorff continuous in ξ, then

$$\lim_{\nu \to \infty} \sup_{x \in \mathcal{X}} \mathbb{H} \left(\frac{1}{\nu} \sum_{k=1}^{\nu} \partial_x F(x, \xi^k), \mathbb{E}_P[\partial_x F(x,\xi)] \right) = 0.$$

It seems open whether these results hold when $\partial_x F(x,\xi)$ is integrably bounded and/or $\partial_x F(x,\xi)$ is merely outer semicontinuous in ξ.

Theorem 5.2 [Stability of stationary points] *Let X be a compact set and Assumptions 3.1 and 4.2 hold for all $x \in X$. Let $\{P_\nu\}$ be a sequence of probability measures converging to P in distribution. Then there exists a constant $t^* > 0$ such that*

(i) $\hat{v}(x,\xi,t)$ is continuous on $X \times \Xi \times [0,t^*]$ and for any fixed $\xi \in \Xi$, $\hat{v}(\cdot,\xi,\cdot)$ is Lipschitz continuous on $X \times [0,t^*]$;

(ii) if the Lipschitz modulus of $\hat{v}(x,\xi,t)$, denoted by $\hat{L}(x,\xi,t)$, is bounded by a constant C for any $(x,\xi,t) \in X \times \Xi \times [0,t^*]$, then

$$\lim_{\nu \to \infty} \mathbb{H}(S^*_{P_\nu}(t), S^*_P(t)) = 0$$

and

$$\lim_{\nu \to \infty} \mathbb{D}(S^w_{P_\nu}(t), S^w_P(t)) = 0.$$

Proof. Part (i) follows from Theorem 3.1 and Theorem 4.2. Part (ii) follows from [47, Lemma 4.2] and Lemma 5.1. ■

6 Sample average approximation

In this section, we discuss sample average approximation of the regularized two stage problem. This is a combination of the stability analysis in sections 3-5 but has an independent interest: we investigate the behavior of optimal solutions and stationary points when the regularization parameter t is driven to 0 and the probability measure P is approximated by the empirical probability measure (sample average). By focusing on sample average approximation, we are able to obtain some stronger results which we cannot do under general probability measures in Section 5.

We start by writing the regularized two stage problem (3.9) and (3.10) in a compact form:

$$\begin{array}{ll} \min\limits_{x,y(\cdot)} & \mathbb{E}[f(x,y(\omega),\xi(\omega))] \\ \text{subject to} & x \in X, \text{and for a.e. } \omega \in \Omega: \\ & g(x,y(\omega),\xi(\omega)) \leq 0, \\ & h(x,y(\omega),\xi(\omega)) = 0, \\ & -G(x,y(\omega),\xi(\omega)) \leq 0, \\ & -H(x,y(\omega),\xi(\omega)) \leq 0, \\ & G(x,y(\omega),\xi(\omega)) \circ H(x,y(\omega),\xi(\omega)) \leq te. \end{array} \tag{6.53}$$

The equivalence between (6.53) and (3.9)–(3.10) is well documented in the stochastic programming literature (e.g. [40, Chapter 1, section 2.4]). Let $\xi^1, \cdots, \xi^N$ be an independent identically distributed (i.i.d. for short) sample. We consider the following sample average approximation of the regularized problem (6.53):

$$\begin{array}{ll} \min\limits_{x;y^1,\cdots,y^N} & \frac{1}{N}\sum_{i=1}^N f(x,y^i,\xi^i) \\ \text{subject to} & x \in X, \text{and for } i = 1,\cdots,N: \\ & g(x,y^i,\xi^i) \leq 0, \\ & h(x,y^i,\xi^i) = 0, \\ & -G(x,y^i,\xi^i) \leq 0, \\ & -H(x,y^i,\xi^i) \leq 0, \\ & G(x,y^i,\xi^i) \circ H(x,y^i,\xi^i) \leq t_N e, \end{array} \tag{6.54}$$

where $t_N \downarrow 0$ as $N \to \infty$. Note that the dependence of the regularization parameter on sample size is numerically important as it allows one to change the parameter value as the sampling changes.

If we use $\hat{v}(x,\xi^i,t)$, $i = 1,\cdots,N$, to denote the optimal value of the regularized second stage problem (3.9) with $\xi = \xi^i$ and assume that $(x;y^1,\cdots,y^N)$ is a global optimal solution, then problem (6.54) can be written in an implicit form, that is,

$$\begin{array}{ll} \min\limits_{x} & \frac{1}{N}\sum_{i=1}^N \hat{v}(x,\xi^i,t_N) \\ \text{subject to} & x \in X, \end{array} \tag{6.55}$$

which is the sample average approximation of the first stage (3.10). Here "implicit" is in the sense the (6.55) does not explicitly involve the underlying functions of the second stage problem. The terminology is used by Ralph and Xu in [35] where SAA is applied to a classical two stage stochastic program.

Sample average approximation is a very popular method in stochastic programming, it is known under various names such as Monte Carlo sampling, sample path optimization and stochastic counterpart, see [33, 36, 40] for SAA in general stochastic programming and [42, 48, 27] for recent application of the method to SMPECs.

The regularized SAA scheme for a two stage SMPEC problem was first considered in [43] and with some detailed convergence analysis in a conference paper [34] where $G(x,y,\xi) = y$ and $H(x,y,\xi)$ is uniformly strongly monotone with respect to y. In this section, we carry out convergence analysis under weaker conditions, that is, the second stage problem MPEC(x,ξ) satisfies MPEC-MFCQ.

We start with a convergence analysis of first stage optimal solutions. Specifically, by assuming that $\{x^N;y^1,\cdots,y^N\}$ is a global optimal solution to SAA problem (6.54), we investigate an accumulation point of $\{x^N\}$ as the sample size N increases. From numerical perspective, if we obtain an approximate global optimal solution from solving (6.54) and observe a tendency of convergence of x^N as N increases, then we want to know how the convergent sequence is related to the optimal solution of true problem (1.1).

Theorem 6.1 *Let $\{(x^N; y^1, \cdots, y^N)\}$ be a sequence of global optimal solutions of problem (6.54) and $\hat{x}$ be an accumulation point of $\{x^N\}$. Let $\bar{X}$ be a closed subset of X such that w.p.1. $x^N \in \bar{X}$ for N sufficiently large and $\bar{X}$ contains a global optimal solution x^* of the true first stage problem (3.7). Suppose: (a) Assumptions 3.1 and 4.2 are satisfied at every point x in $\bar{X}$, (b) problem MPCC(x,ξ) satisfies MPEC-MFCQ at every point in the optimal solution set $Y_{sol}(x,\xi)$ for $(x,\xi) \in \bar{X} \times \Xi$. Then*

- *(i) w.p.1 $\hat{x}$ is an optimal solution to the true problem (3.7).*
- *(ii) Suppose, in addition, that: (c) there exists a positive constant $\hat{t}$ such that for every $x \in \bar{X}$ and $t \in [0, \hat{t}]$, the moment generating function $\mathbb{E}[e^{(\hat{v}(x,\xi,t)-\mathbb{E}[\hat{v}(x,\xi,t)])\tau}]$ of the random variable $\hat{v}(x,\xi,t) - \mathbb{E}[\hat{v}(x,\xi,t)]$ is finite valued for τ close to 0; (d) the moment generating function $\mathbb{E}[e^{\kappa(\xi)^2\tau}]$ of the random variable $\kappa(\xi)^2$ is finite valued for τ close to 0, where $\kappa(\xi)$ is defined in Assumption 4.2. Then $\{x^N\}$ converges to $\hat{x}$ with probability approaching one exponentially fast with the increase of sample size N, that is, for every $\epsilon > 0$, there exist positive constants $C(\epsilon)$ and $\beta(\epsilon)$ such that*

$$\text{Prob}(d(x^N, X_{sol}) \geq \epsilon) \leq C(\epsilon)e^{-\beta(\epsilon)N} \tag{6.56}$$

for N sufficiently large.

Proof. Part (i). It suffices to show that $\frac{1}{N}\sum_{i=1}^N \hat{v}(x,\xi^i,t_N)$ converges uniformly to $\mathbb{E}[v(x,\xi)]$ over the compact set $\bar{X}$, that is,

$$\lim_{N\to\infty} \sup_{x\in\bar{X}} \left| \frac{1}{N}\sum_{i=1}^N \hat{v}(x,\xi^i,t_N) - \mathbb{E}[v(x,\xi)] \right| = 0 \quad \text{w.p.1.} \tag{6.57}$$

Indeed, if (6.57) holds, then we can claim, by virtue of [47, Lemma 4.1] or [38, Theorem 7.33] (as uniform convergence implies epi-convergence), that the set of global minimizers of the sample average function $\frac{1}{N}\sum_{i=1}^N \hat{v}(x,\xi^i,t_N)$ within $\bar{X}$ converges to that of $\mathbb{E}[v(x,\xi)]$ within $\bar{X}$ w.p.1. This implies that w.p.1 $\hat{x}$ is a global minimizer of $\mathbb{E}[v(x,\xi)]$ in $\bar{X}$ and hence $\mathbb{E}[v(\hat{x},\xi)] = \mathbb{E}[v(x^*,\xi)]$. In what follows, we prove (6.57).

Since Assumption 3.1 holds at every point $x \in \bar{X}$, it follows from Theorem 3.3 (ii) that $v(\cdot,\cdot)$ is continuous on $\bar{X} \times \Xi$ and $v(\cdot,\xi)$ is locally Lipschitz continuous on $\bar{X}$ for every fixed $\xi \in \Xi$. Moreover, by Proposition 4.2 (i),

$$\|\partial_x v(x,\xi)\| \leq \|\Phi(x,\xi)\|.$$

Under Assumption 4.2,

$$\|\Phi(x,\xi)\| \leq \kappa(\xi)^2, \forall (x,\xi) \in \bar{X} \times \Xi,$$

where $\kappa(\xi)$ is given in Assumption 4.2 and $\mathbb{E}[\kappa(\xi)^2] < \infty$. Further, the condition $x^* \in \bar{X}$ implies $\mathbb{E}[v(x^*,\xi)] < \infty$. Therefore for every $x \in \bar{X}$,

$$|v(x,\xi)| \leq |v(x^*,\xi)| + \kappa(\xi)^2\|x - x^*\|$$

and hence $\mathbb{E}[v(x,\xi)]$ is well-defined and

$$\mathbb{E}[v(x,\xi)] \leq \mathbb{E}[|v(x^*,\xi)|] + \mathbb{E}[\kappa(\xi)^2]\|x - x^*\| < \infty.$$

This implies, through the classical uniform law of large numbers [40, Lemma A1], that

$$\lim_{N\to\infty} \sup_{x\in\bar{X}} \left| \frac{1}{N}\sum_{i=1}^N v(x,\xi^i) - \mathbb{E}[v(x,\xi)] \right| = 0 \quad \text{w.p.1.} \tag{6.58}$$

On the other hand, under Assumption 3.1, we know through Theorem 4.2 that $\hat{v}(\cdot,\xi,\cdot)$ is locally Lipschitz continuous at $(x,0)$ for $x \in \bar{X}$. Moreover, by Proposition 4.2 (ii) and Assumption 4.2

$$\|\partial_t \hat{v}(x,\xi,t)\| \leq \|\Pi_\lambda \hat{\mathcal{W}}(x,\xi,t)\| \leq \kappa(\xi), \text{ for all } (x,\xi) \in \bar{X} \times \Xi,$$

where $\kappa(\xi)$ is given in Assumption 4.2. Consequently, we have

$$\begin{aligned}\left|\frac{1}{N}\sum_{i=1}^N \hat{v}(x,\xi^i,t_N) - \mathbb{E}[v(x,\xi)]\right| &\leq \frac{1}{N}\sum_{i=1}^N |\hat{v}(x,\xi^i,t_N) - v(x,\xi^i)| + \left|\frac{1}{N}\sum_{i=1}^N v(x,\xi^i) - \mathbb{E}[v(x,\xi)]\right| \\ &\leq \frac{1}{N}\sum_{i=1}^N \kappa(\xi^i)t_N + \left|\frac{1}{N}\sum_{i=1}^N v(x,\xi^i) - \mathbb{E}[v(x,\xi)]\right|. \end{aligned} \tag{6.59}$$

Combining (6.58) and (6.59) together with the fact that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \kappa(\xi^i) = \mathbb{E}[\kappa(\xi)],$$

we obtain (6.57).

Part (ii). Let $\epsilon > 0$ be given. By [11, Lemma 3.2] (or [47, Lemma 4.1]), there exists a $\delta(\epsilon) > 0$ such that if

$$\lim_{N\to\infty}\sup_{x\in\bar{X}}\left|\frac{1}{N}\sum_{i=1}^N \hat{v}(x,\xi^i,t_N) - \mathbb{E}[v(x,\xi)]\right| \leq \delta(\epsilon),$$

then $d(x^N, X_{sol}) \leq \|x^N - \hat{x}\| \leq \epsilon$. Under condition (d), there exist positive constants $C_1(\epsilon)$, $\beta_1(\epsilon)$ and N_0 sufficiently large such that for $N \geq N_0$

$$\text{Prob}\left(\frac{1}{N}\sum_{i=1}^N \kappa(\xi^i)t_N \geq \frac{1}{2}\delta(\epsilon)\right) \leq C_1(\epsilon)e^{-\beta_1(\epsilon)N}.$$

On the other hand, by virtue of the Lipschitz continuity of $v(\cdot,\xi)$ together with conditions (c) and (d) of this theorem, we can apply [43, Theorem 5.1] to the sample average $\frac{1}{N}\sum_{i=1}^N v(x,\xi^i)$, that is, for given $\delta(\epsilon) > 0$, there exist positive constants $C_2(\epsilon)$, $\beta_2(\epsilon)$ and $N_1 \geq N_0$ such that for $N \geq N_1$

$$\text{Prob}\left(\sup_{x\in\bar{X}}\frac{1}{N}\left|\sum_{i=1}^N v(x,\xi^i) - \mathbb{E}[v(x,\xi)]\right| \geq \frac{1}{2}\delta(\epsilon)\right) \leq C_2(\epsilon)e^{-\beta_2(\epsilon)N}.$$

Combining the above two inequalities with (6.59), we have

$$\text{Prob}\left(\lim_{N\to\infty}\sup_{x\in\bar{X}}\left|\frac{1}{N}\sum_{i=1}^N \hat{v}(x,\xi^i,t_N) - \mathbb{E}[v(x,\xi)]\right| \geq \delta(\epsilon)\right) \leq C_1(\epsilon)e^{-\beta_1(\epsilon)N} + C_2(\epsilon)e^{-\beta_2(\epsilon)N}.$$

The conclusion follows by setting $C(\epsilon) = C_1(\epsilon) + C_2(\epsilon)$ and $\beta(\epsilon) = \min(\beta_1(\epsilon), \beta_2(\epsilon))$. ■

We now move on to discuss the case when a solution $\{x^N; y^1, \cdots, y^N\}$ obtained from solving the SAA problem (6.54) is a stationary point but not a global optimal solution. This happens in numerical solution in that MPECs are generically nonconvex and so are their counterparts via NLP-regularization. This motivates us a separate discussion on the convergence of x^N.

Consider the KKT conditions of the regularized SAA program (6.54):

$$0 \in \frac{1}{N}\sum_{i=1}^N \nabla_x\hat{\mathcal{L}}(x,y^i,\xi^i,t;\alpha^i,\beta^i,\gamma^i,\theta^i,\lambda^i) + \mathcal{N}_X(x) \tag{6.60}$$

and for $i = 1, \cdots, N$

$$\begin{cases} 0 = \nabla_y\hat{\mathcal{L}}(x,y^i,\xi^i,t;\alpha^i,\beta^i,\gamma^i,\theta^i,\lambda^i), \\ 0 \leq -g(x,y^i,\xi^i) \perp \alpha^i \geq 0, \\ 0 = h(x,y^i,\xi^i), \\ 0 \leq G(x,y^i,\xi^i) \perp \gamma^i \geq 0, \\ 0 \leq H(x,y^i,\xi^i) \perp \theta^i \geq 0, \\ 0 \leq t_N e - G(x,y^i,\xi^i) \circ H(x,y^i,\xi^i) \perp \lambda^i \geq 0. \end{cases} . \tag{6.61}$$

We note that $(y^1;\alpha^1,\beta^1,\gamma^1,\theta^1,\lambda^1),\cdots(y^N;\alpha^N,\beta^N,\gamma^N,\theta^N,\lambda^N)$ change as N changes. So it would be more accurate to denote each y^i by $y^{i,N}$ and similarly with the other vectors. To keep the notation simple we will take this point as understood.

The KKT conditions (6.61) imply that $(y^i;\alpha^i,\beta^i,\gamma^i,\theta^i,\lambda^i)$ is a KKT pair of REG(x,ξ^i,t_N), i.e., $(y^i;\alpha^i,\beta^i,\gamma^i,\theta^i,\lambda^i)\in\hat{\mathcal{W}}(x,\xi^i,t_N)$. By the definition of $\hat{\Psi}(x,\xi,t)$ (see (4.28)),

$$\nabla_x\hat{\mathcal{L}}(x,y^i,\xi^i,t;\alpha^i,\beta^i,\gamma^i,\theta^i,\lambda^i)\in\hat{\Psi}(x,\xi^i,t).$$

Combining this with (6.60), we arrive at

$$0\in\frac{1}{N}\sum_{i=1}^{N}\hat{\Psi}(x,\xi^i,t_N)+\mathcal{N}_X(x), \tag{6.62}$$

which implies that (6.62) is a sample average approximation of the relaxed KKT condition (4.40).

Theorem 6.2 *Let $\{x^N;y^1,\cdots,y^N\}$ be a stationary point of problem (6.54) and $\hat{x}$ be an accumulation point of $\{x^N\}$. Suppose that Assumptions 4.1 and 4.2 hold at $\hat{x}$, and problem MPCC(x,ξ) satisfies MPEC-MFCQ at every point in the feasible set $\mathcal{F}(\hat{x},\xi)$ for every $\xi\in\Xi$. Then w.p.1 $\hat{x}$ is relaxed stationary point of true problem (3.7), that is, $\hat{x}$ satisfies (4.39).*

Proof. Let

$$\mathcal{A}(x,\xi,t):=\begin{cases}\hat{\Psi}(x,\xi,t), & t\neq 0,\\ \Psi(x,\xi), & t=0.\end{cases}$$

By Proposition 4.3, there exist a neighborhood U of $\hat{x}$ and a scalar $t^*>0$ such that $\mathcal{A}(\cdot,\cdot,\cdot)$ is outer semi-continuous on $U\times\Xi\times[0,t^*]$. Under Assumption 4.2, $\mathcal{A}(x,\xi(\cdot),t)$ is measurable, integrably bounded. The conclusion follows by application of [50, Theorem 4.3]. The proof is complete. ■

References

[1] Z. Artstein and R.A. Vitale, A strong law of large numbers for random compact sets, *The Annals of Probability*, 3 (1975), 879–882.

[2] Z. Artstein and R.J-B. Wets, Approximating the integral of a multifunction, *Multivariate analysis*, 24 (1988), 285–308.

[3] R.J. Aubin and H. Frankowska, Set-Valued Analysis, Birkhauser, Boston, 1990.

[4] R.J. Aumann, Integrals of set-valued functions, *Journal of Mathematical Analysis and Applications*, 12 (1965), 1-12.

[5] J.R. Birge and L. Qi, Subdifferential convergence in stochastic programs, *SIAM Journal on Optimization*, 5 (1995), 436–453.

[6] J.F. Bonnans and A. Shapiro, Perturbation analysis of optimization problems, Springer, New York, 2000.

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

[8] X. Chen and M. Fukushima, Expected residual minimization method for stochastic linear complementarity problems, *Mathematics of Operations Research*, 30 (2005), 1022–1038.

[9] S. Christiansen, M. Patriksson and L. Wynter, Stochastic bilevel programming in structural optimization, *Structural and multidisciplinary optimization*, 21 (2001), 361–371.

[10] F.H. Clarke, Optimization and nonsmooth analysis, Wiley, New York, 1983.

[11] L. Dai, C.H. Chen and J.R. Birge, Convergence properties of two stage stochastic programming, *Journal of Optimization Theory and Applications*, 106 (2000), 489–509.

[12] D. De Wolf and Y. Smeers, A stochastic version of a Stackelverg-Nash-Cournot equilibrium model, *Management Sciences*, 43 (1997), 190–197.

[13] A.L. Dontchev and R.T. Rockafellar, Characterizations of strong regularity for variational inequalities over polyhedral convex sets, *SIAM Journal on Optimization*, 6 (1996), 1087–1105.

[14] F. Hiai, Convergence of conditional expectations and strong law of large numbers for multivalued random variables, *Transaction of The American Mathematical Society*, 291 (1985), 613–627.

[15] R. Fletcher, S. Leyffer, D. Ralph and S. Scholtes, Local convergence of SQP methods for mathematical programs with equilibrium constraints, *SIAM Journal on Optimization*, 17 (2006), 259–286.

[16] J. Gauvin and F. Dubeau, Differential properties of the marginal function in mathematical programming, *Mathematical Programming Studies*, 19 (1982), 101–119.

[17] W.W. Hongan, Point-to-set maps in mathematical programming, *SIAM Review*, 15 (1973), 591–603.

[18] X. Hu and D. Ralph, A note on sensetivity of value functions of mathematical programs with complementarity constraints, *Mathematical Programming*, 93 (2002), 265–279.

[19] P. Kall, Approximation to optimization problems: An elementary review, *Mathematics of Operations Research*, 11 (1986), 9–18.

[20] D. Klatte, On the stability of local and global optimal solutions in parametric problems of nonlinear programming, Part I: Basic results, *Seminarbericht Nr. 75*, Sektion Mathematik, Humboldt-Universität zu Berlin, Berlin (1985), 1–21.

[21] D. Klatte, A note on quantitative stability results in nonlinear optimization, *Seminarbericht Nr. 90*, Sektion Mathematik, Humboldt-Universität zu Berlin, Berlin (1987), 77–86.

[22] G.H. Lin and M. Fukushima, New relaxation method for mathematical programs with complementarity constraints, *Journal of Optimization Theory and Applications*, 118 (2003), 81–116.

[23] G.H. Lin, X. Chen and M. Fukushima, Solving stochastic mathematical programs with equilibrium constraints via approximation and smoothing implicit programming with penalization, *Mathematical Programming*, 116 (2009), 343–368.

[24] G.H. Lin and M. Fukushima, Stochastic equilibrium problems and stochastic mathematical programs with equilibrium constraints: A survey, *Pacific Journal of Optimization*, to appear.

[25] Y. Lucet and J.J. Ye, Sensitivity analysis of the value function for optimization problems with variational inequality constraints, *SIAM Journal onControl and Optimization*, 40 (2001), 699–723.

[26] Z.Q. Luo, J.S. Pang and D. Ralph, Mathematical programs with equilibrium constraints, Cambridge University Press, Cambridge, UK, 1996.

[27] F. Meng and H. Xu, A regularized sample average approximation method for stochastic mathematical programs with nonsmooth equality constraints, *SIAM Journal on Optimization*, 17 (2006), 891–919.

[28] B.S. Mordukhovich, Variational analysis and generalized differentiation, I: basic theory. Grundlehren Series (Fundamental Principles of Mathematical Sciences), Vol. 330, Springer, 2006.

[29] J. Outrata and W. Römisch, On optimization conditions for some nonsmooth optimization problems over L^p spaces, *Journal of Optimization Theory and Applcations*, 126 (2005), 411-438.

[30] J. Outrata, M. Kocvara and J. Zowe, Nonsmooth approach to optimization problems with equilibrium constraints: theory, applications, and numerical results, Kluwer Academic Publishers, Dordrect, The Netherlands, 1998.

[31] N.S. Papageorgiou, On the theory of Banach space valued multifunctions. I: Integration and conditional expectation, *Journal of multivariate analysis*, 17 (1985), 185–206.

[32] M. Patriksson and L. Wynter, Stochastic mathematical programs with equilibrium constraints, *Operations Research Letters*, 25 (1999), 159–167.

[33] E.L. Plambeck, B.R. Fu, S.M. Robinson and R. Suri, Sample-path optimization of convex stochastic performances functions, *Mathematical Programming*, 75 (1996), 137–176.

[34] D. Ralph, H. Xu and F. Meng, Regularization methods for stochastic mathematical programs with complementarity constraints, European conference on operational research, 21st (EURO XXI), 2–5 July (2006), Rejkjavik, Iceland.

[35] D. Ralph and H. Xu, Asympototic analysis of stationary points of sample average two stage stochastic programs: A generalized equation approach, Manuscript, 2008. Submitted to *Mathematics of Operational Research*, under revision.

[36] S.M. Robinson, Analysis of sample-path optimization, *Mathematics of Operations Research*, 21 (1996), 513–528.

[37] R.T. Rockafellar, Convex analysis, Princeton University press, Princeton, 1970.

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

[39] W. Römisch, Stability of Stochastic Programming Problems, in: Stochastic Programming (A. Ruszczynski and A. Shapiro eds.), Handbooks in Operations Research and Management Science Vol. 10, Elsevier, Amsterdam 2003, 483–554.

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

[41] H. Scheel and S. Scholtes, Mathematical programs with complementarity constraints: Stationarity, optimality, and sensivity, *Mathematics of Operations Research*, 25 (2000), 1–22.

[42] A. Shapiro, Stochastic mathematical programs with equilibrium constraints, *Journal of Optimization Theory and Applications*, 128 (2006), 223–243.

[43] A. Shapiro and H. Xu, Stochastic mathematical programs with equilibrium constraints, modeling and sample average approximation, *Optimization*, 57 (2008), 395–418.

[44] S. Scholtes, Convergence properties of a regularization sheme for mathematical programs with complementarity constraints, *SIAM Journal on Optimization*, 11 (2001), 918–936.

[45] R. J-B. Wets, Stochastic Programming, Handbooks in OR & MS, Vol. 1, G. L. Nemhauser et al eds, 1989, 573–629.

[46] H. Xu, An MPCC approach for stochastic Stackelberg-Nash-Cournot Equilibrium, *Optimization*, 54 (2005), 27–57.

[47] H. Xu, Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming, School of Mathematics, University of Southampton, November 2008.

[48] H. Xu and F. Meng, Convergence analysis of sample average approximation methods for a class of stochastic mathematical programs with equality constraints, *Mathematics of Operations Research*, 32 (2007), 648–668.

[49] H. Xu and J.J. Ye, Necessary Optimality Conditions for two-stage Stochastic Programs with Equilibrium Constraints, *SIAM Journal on Optimization*, 20 (2010), 1685–1715.

[50] H. Xu and D. Zhang, Smooth sample average approximation of stationary points in nonsmooth stochastic optimization and applications, *Mathematical Programming*, 119 (2009), 371–401.