SIAM J. OPTIM.
Vol. 21, No. 3, pp. 669–705

STABILITY ANALYSIS OF TWO-STAGE STOCHASTIC MATHEMATICAL PROGRAMS WITH COMPLEMENTARITY CONSTRAINTS VIA NLP REGULARIZATION*

YONGCHAO LIU†, HUIFU XU‡, AND GUI-HUA LIN†

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 by Scholtes [*SIAM J. Optim.*, 11 (2001), pp. 918–936] 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 where the probability distribution is approximated by the empirical probability measure which is also known as sample average approximation.

Key words. stochastic mathematical program with complementarity constraints, nonlinear programming regularization, mathematical program with equilibrium constraints, Mangasarian–Fromowitz constraint qualification, stability analysis, sample average approximation

AMS subject classifications. 90C15, 90C30, 90C33

DOI. 10.1137/100785685

1. Introduction. Consider the following two-stage stochastic mathematical program with complementarity constraints (SMPCC):

$$
\begin{aligned}
&\min_{x,y(\cdot)\in\mathcal{Y}} \mathbb{E}[f(x,y(\omega),\xi(\omega))]\\
&\quad \text{s.t. } x \in X \text{ and for almost every } \omega \in \Omega,\\
&\qquad g(x,y(\omega),\xi(\omega)) \le 0,\\
&\qquad h(x,y(\omega),\xi(\omega)) = 0,\\
&\qquad 0 \le G(x,y(\omega),\xi(\omega)) \perp H(x,y(\omega),\xi(\omega)) \ge 0,
\end{aligned}
\tag{1.1}
$$

where X is a nonempty closed convex subset of $\mathbb{R}^n$, f, g, h, G, H are continuously differentiable functions from $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^q$ to $\mathbb{R}$, $\mathbb{R}^s$, $\mathbb{R}^r$, $\mathbb{R}^m$, $\mathbb{R}^m$, respectively, $\xi:\Omega \to \Xi$ 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 probability measure P. $\perp$ denotes the perpendicularity of two vectors, and $\mathcal{Y}$ is a space of functions $y(\cdot):\Omega \to \mathbb{R}^m$ such that $\mathbb{E}[f(x,y(\omega),\xi(\omega))]$ is well defined.

*Received by the editors February 11, 2010; accepted for publication (in revised form) April 15, 2011; published electronically July 21, 2011.
http://www.siam.org/journals/siopt/21-3/78568.html

†School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China (googleliu0717@yahoo.com.cn, lin_g_h@yahoo.com.cn). The work of the first author was carried out while he was visiting the second author at the School of Mathematics, University of Southampton, sponsored by the China Scholarship Council. The work of the first and third authors was supported in part by NSFC grants 10771025 and 11071028 and a Royal Society International Travel grant 2009.

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

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 practical need as well as by theoretical interest. For instance, in an investment model for a firm, one may use a random vector to represent market uncertainties and a complementarity problem to describe competition from its competitors; see [15], [45]. Similar SMPCC models can also be found in engineering design; see, for instance, [12].

Patriksson and Wynter [30] first proposed a two-stage stochastic mathematical program 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 a numerical method for solving the problem. In the years since the first SMPEC paper, there have been increasing discussions on SMPECs, most of which have focused on numerical methods. Shapiro [40] first applied the well-known sample average approximation (SAA) method (also known under different names such as Monte Carlo method, sample path optimization, and stochastic counterpart [31], [34]) to general two-stage SMPECs where the expected value of random functions are approximated by their sample averages, and he investigated asymptotic convergence of optimal solutions and optimal values as sample size increases. Shapiro and Xu [41] presented a detailed analysis of SMPEC structure and demonstrated the exponential rate of convergence of sharp local minimizers of sample average approximated problems. Lin, Chen, and Fukushima [25] first investigated SMPCCs and proposed an implicit smoothing method for solving a discrete SMPCC with a P_0-linear complementarity constraint. Xu and Meng [47] reformulated the SMPCC as a two-stage stochastic minimization problem with nonsmooth equality constraints and applied the SAA method to solve it. They obtained an exponential rate of convergence of global optimal solutions obtained from solving the SAA 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] investigated convergence of stationary points obtained from solving sample average approximated SMPECs where the second stage problem may have multiple solutions. Specifically they used a nonlinear complementarity problem (NCP) function which combines Tikhonov regularization and some smoothing technique to approximate the complementarity constraints with a smooth nonlinear system of equality constraints. The latter define a unique feasible solution. The NCP-regularization scheme is restricted to P_0 functions, and as the regularization parameter is driven to zero, the unique feasible solution of the regularized second stage problem converges to a measurable feasible solution of the true second stage problem which is not necessarily optimal. The method is applicable to the case when a decision maker is not able or keen to find an optimal solution of the second stage at each scenario; see detailed discussions in [27, p 892]. Putting this another way, an optimal solution or a stationary point obtained under the NCP-regularized SAA scheme does not necessarily converge to its true counterpart.

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 (NLP) problem? 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 [47], [27]. Question (b) is classical in stochastic programming. A simple answer is to use the well-known Monte Carlo sampling method. In the literature on MPECs, however, the reformulation through NCP functions is not the most popular. Likewise, in the literature on stochastic programming, there exist discretization/approximation schemes other than Monte Carlo sampling to deal with the random variables. This motivates us to consider different schemes to approximate the complementarity constraints and the probability measure P.

Here we apply a well-known regularization method [42], [38], [17] to tackle the complementarity constraint, and then we 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 SAA. The basic idea of the regularization method is to approximate the complementarity constraint $0 \leq x \perp y \geq 0$ by a system of parameterized nonlinear inequalities $x \geq 0$, $y \geq 0$, where the components of x and y satisfy $x_i y_i \leq t$ for some small positive parameter t. The regularization method has 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 such as the sequential quadratic programming methods [1], [17]. Moreover, the regularized NLP satisfies the 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 an 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 [41] appear to be the first to apply the approach to a two-stage SMPCC and then use the SAA method to solve it. They predicted the convergence of the regularized SAA method for a class of SMPCCs with strongly monotone complementarity constraints but did not give details of the convergence analysis. In a conference paper, Ralph, Xu, and Meng [32] carried out a convergence analysis of the NLP regularized SAA method for solving a class of SMPCCs with monotone complementarity constraints with a particular focus on optimal values and Clarke stationary points.

Since 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; we present a detailed stability analysis of the NLP regularized problem as the regularization parameter tends to zero. Moreover, for a fixed regularization parameter, we investigate stability of the NLP regularized two-stage SMPEC when the probability distribution of ξ is approximated by a sequence of probability measures. Finally, we combine the two stability analyses under an empirical probability measure.

As far as we are concerned, the main contributions of this paper can be summarized as follows:

(a) Different from [41], [47], [48], we consider a two-stage SMPCC where the second stage problem may have multiple solutions, and we apply the popular NLP regularization method to deal with the complementarity constraint. The regularization scheme is significantly different from NCP regularization in [27] which is applicable to a specific class of SMPECs. Under MPEC-MFCQ, instead of MPEC linear independent constraint qualification (MPEC-LICQ) as in [42, Lemma 2.1], we demonstrate that the regularized second stage problem satisfies MPEC-MFCQ. Consequently, we present a comprehensive stability analysis of the NLP regularized SMPEC including Lipschitz continuity of optimal value functions of both first stage and second stage problems as well as the outer semicontinuity of the set of optimal solutions and stationary points. This type of analysis is new in the research of SMPECs, and it addresses a fundamental problem: under some moderate conditions, two-stage SMPECs can be effectively approximated by ordinary two-stage stochastic NLPs. This paves the way for the application of existing numerical schemes developed for classical two-stage stochastic NLPs (e.g., [6], 39], [22], [50]) to the NLP regularized two-stage SMPECs.

(b) We carry out a stability analysis of the NLP regularized two-stage SMPECs. Differing from the existing research on SMPECs, our analysis is performed under general probability measure approximation including empirical probability measure, optimal scenarios generation, and many others. Our analysis covers optimal values and optimal solutions as well as stationary points. In particular, we establish, under general perturbation of the probability measure, uniform approximation of the Clarke subdifferential of the expected value of a nonsmooth random function and the expected value of the Clarke subdifferential of a nonsmooth random function. The result strengthens the earlier results on subdifferential approximation by Birge and Qi [9] and has potential applications in the research of general nonsmooth stochastic programming and stochastic equilibrium problems.

(c) We present a combined stability analysis due to NLP regularization and empirical probability measure and establish exponential convergence of optimal solutions and almost sure convergence of stationary points. This demonstrates how our stability analysis could generate concrete asymptotic convergence results when the probability approximation is restricted to empirical measure.

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^Ty 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; $D_1 + D_2$ denotes 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 jth component of f forms the jth column of the Jacobian. Finally, for a set $\{(x, y) = z : z \in Z\}$, $\Pi_x Z = \{x : \exists y \quad \text{such that } (x, y) \in Z\}$.

2.1. Some basics in deterministic MPECs. Consider the following MPCC:

$$\min_z f(z) \quad \text{s.t. } g(z) \le 0, \quad h(z) = 0, \quad 0 \le G(z) \perp H(z) \ge 0, \tag{2.1}$$

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

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

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

$$\mathcal{J}(z^*) := \{J\colon J \subseteq \mathcal{I}_G(z^*), J^c \subseteq \mathcal{I}_H(z^*)\}, \tag{2.2}$$

where $J^c := \{1, \dots, m\} \setminus J$. We consider the following NLP corresponding to index set J:

$$\begin{aligned}
\text{NLP}_J\colon \min_z\ & f(z)\\
\text{s.t. }\ & g(z) \le 0,\\
& h(z) = 0,\\
& G_i(z) = 0, \quad H_i(z) \ge 0, \quad i \in J,\\
& G_i(z) \ge 0, \quad H_i(z) = 0, \quad i \in J^c. \qquad (2.3)
\end{aligned}$$

In the literature of MPECs, each of the NLPs corresponding to index set J is called an *NLP branch* of (2.1), 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 (2.1); see [19].

DEFINITION 2.1. *MPCC* (2.1) *is said to satisfy the* MPEC-MFCQ *at a feasible point* z^* *if the gradient vectors*

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

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

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

It is said to satisfy the MPEC-LICQ *at* z^* *if the gradient vectors*

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

are linearly independent.

2.2. Set-valued mapping and subdifferentials. Let X be a closed subset of $\mathbb{R}^n$. A set-valued mapping $F\colon X \to 2^{\mathbb{R}^m}$ 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

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

F is said to be *outer semicontinuous* 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 $O^{F(\bar{x})}$, there is a neighborhood U of $\bar{x}$ such that $\bigcup_{x\in U} F(x) \subset O$; see [35].

DEFINITION 2.2 (see [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*

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

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

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 [35, 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 selection. The expected value is also known as Aumann's integral [4].

DEFINITION 2.3. *Let $f: \mathbb{R}^n \to \mathbb{R}$ be a lower semicontinuous function, and let it be finite at $x \in \mathbb{R}^n$. The* proximal subdifferential [35, Definition 8.45] *of f at x is defined as*

$$\partial^\pi f(x) := \{\zeta \in \mathbb{R}^n : \exists \sigma > 0, \delta > 0 \ \ s.t. \ \ f(y) \geq f(x) + \zeta^T(y - x) - \sigma\|y - x\|^2 \\ \forall y \in B(x, \delta)\},$$

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 the singular limiting subdifferential *is defined as*

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

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, [26, 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

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

where D denotes the set of points at which f is Fréchet differentiable, and $\nabla f(y)$ 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 semicontinuous; see [13, Propositions 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,

$$\partial f(x) = \mathrm{conv}\partial^M f(x);$$

see [35, Theorem 9.61].

3. NLP regularization and stability analysis. In this section, we apply the NLP regularization scheme [42] to SMPCC (1.1), and we 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-valued mappings of optimal solutions and stationary points. While our analysis follows general steps in the stability analysis of parametric programming [20], [21], [10], 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). Problem (1.1) can be written as

$$\begin{aligned} P_\vartheta: \qquad & \min_x \vartheta(x) = \mathbb{E}[v(x,\xi(w))] \\ & \text{s.t. } x \in X, \end{aligned} \tag{3.1}$$

as long as $\mathbb{E}[(v(x,\xi))_+] < \infty$ and $\mathbb{E}[(-v(x,\xi))_+] < \infty$, where $(a)_+ = \max(0,a)$ and $v(x,\xi)$ denotes the optimal value function of the following second stage problem:

$$\begin{aligned} \mathrm{MPCC}(x,\xi)\colon \min_y\ & f(x,y,\xi) \\ \text{s.t. } & g(x,y,\xi) \le 0, \\ & h(x,y,\xi) = 0, \\ & 0 \le G(x,y,\xi) \perp H(x,y,\xi) \ge 0. \end{aligned} \tag{3.2}$$

The reformulation is well known in stochastic programming; see, for example, [37, Chapter 1, Proposition 5] and a discussion in [41, section 1] in the context of two-stage SMPECs. We apply the NLP regularization scheme [42], [38], [17] to the second stage problem MPCC(x,ξ) by replacing the complementarity constraint with a parameterized system of inequalities, that is,

$$G(x,y,\xi) \ge 0, \quad H(x,y,\xi) \ge 0, \quad G(x,y,\xi) \circ H(x,y,\xi) \le te,$$

where $t \ge 0$ 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:

$$\begin{aligned}\text{REG}(x,\xi,t)\colon \min_y\ & f(x,y,\xi)\\ \text{s.t.}\ \ & g(x,y,\xi)\le 0,\\ & h(x,y,\xi)=0,\\ & G(x,y,\xi)\ge 0,\\ & H(x,y,\xi)\ge 0,\\ & G(x,y,\xi)\circ H(x,y,\xi)\le te.\end{aligned} \tag{3.3}$$

Following the terminology in deterministic MPECs, we call (3.3) a regularized NLP approximation of the second stage problem (3.2). Let $\hat{v}(x,\xi,t)$ denote the optimal value of the regularized problem. Then the corresponding first stage problem can be written as

$$\begin{aligned}P_{\hat{\vartheta}}\colon \qquad & \min_x \hat{\vartheta}(x,t)=\mathbb{E}[\hat{v}(x,\xi(\omega),t)]\\ & \text{s.t. } x\in X.\end{aligned} \tag{3.4}$$

Observe that when $t=0$, $\text{REG}(x,\xi,t)$ coincides with $\text{MPCC}(x,\xi)$, 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 a numerical perspective, t often takes a small positive value because $\text{REG}(x,\xi,t)$ never satisfies the 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 the 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 [42], [38], [17] where 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 and 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 [42], [38], [17] where t is the only parameter.

Throughout this section, we use the following notation. $\mathcal{F}(x,\xi)$ and $\hat{\mathcal{F}}(x,\xi,t)$ denote, respectively, the feasible sets of the second stage problems (3.2) and (3.3); $Y_{\text{sol}}(x,\xi)$ and $\hat{Y}_{\text{sol}}(x,\xi,t)$ denote the sets of global optimal solutions; X_{sol} and $\hat{X}_{\text{sol}}(t)$ denote the optimal solution sets of the first stage problems (3.1) and (3.4). We use $\phi(t)$ to denote the optimal value of $P_{\hat{\vartheta}}$. Observe that $\hat{\mathcal{F}}(x,\xi,0)=\mathcal{F}(x,\xi)$, $\hat{Y}_{\text{sol}}(x,\xi,0)=Y_{\text{sol}}(x,\xi)$, and $\hat{X}_{\text{sol}}(0)=X_{\text{sol}}$.

In order to avoid technical difficulties in our analysis, we assume throughout this paper that $\mathcal{F}(x,\xi)$ is nonempty for all x,ξ, which implies relatively complete recourse of the second stage problem. A direct consequence of this assumption is that $\hat{\mathcal{F}}(x,\xi,t)\neq\emptyset$, as the former is a subset of the latter.

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 $\hat{Y}_{\text{sol}}(x,\xi,t)$ of the second stage regularized problem $\text{REG}(x,\xi,t)$ 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 $\delta\in(-\infty,+\infty)$, $t^*>0$, a compact set $Y\subset\mathbb{R}^m$, and a neighborhood U of x^* such that

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

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 [10, Proposition 4.4] in that the parameters x, ξ, and t are not treated in a similar fashion. 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 interested in the case only 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 $\hat{Y}_{\text{sol}}(x, \xi, t)$ 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 some kinds of monotonicity of $H(x, \cdot, \xi)$. For instance, if Ξ is bounded and $H(x, \cdot, \xi)$ is an R_0 function for every $(x, \xi) \in \mathcal{X} \times \Xi$; that is, if, for any sequence $\{y^k\}$ with $\lim_{k\to\infty} \|y^k\| = +\infty$, $\liminf_{k\to\infty} \min\{y_1^k, \ldots, y_m^k\} / \|y^k\| \geq 0$, and

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

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 (3.3) is uniformly bounded for $t \in [0, +\infty)$; see [23] 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) as long as it is restricted to set Y.

PROPOSITION 3.2. *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 exists 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 $\mathcal{Q} = 0_r \times \mathbb{R}_+^{s+m+m+m}$ and 0_r is the r-dimensional 0 vector. 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 4.2 in [46],[1] $\mathcal{F}_Y(x, \xi, t)$ is Hausdorff continuous on $U \times \Xi \times [0, t^*]$. The proof is complete. ∎

Using Proposition 3.2, 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 3.3 (stability of REG(x, ξ, t)). *Let Assumption* 3.1 *hold at point* $x^* \in X$. *Then there exists a neighborhood* U *of* x^* *and a scalar* $t^* > 0$ *such that*

(i) *the optimal solution set* $\hat{Y}_{\text{sol}}(x, \xi, t)$ *of the second stage problem* REG(x, ξ, t) *is outer semicontinuous 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 $\hat{Y}_{\text{sol}}(x, \xi, t) \subset Y$.

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

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

For the given y^*, it follows by Proposition 3.2 that there exists a sequence $\{y^k\}$ such that $\mathbf{y}^k \in \mathcal{F}_Y(x^k, \xi^k, t_k)$ 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 $\hat{y}^k \in \hat{Y}_{\text{sol}}(x^k, \xi^k, t_k)$.

Part (ii). Given the outer semicontinuity of $\hat{Y}_{\text{sol}}(x, \xi, t)$ and the continuity of f, we can easily use [10, 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). ∎

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 semicontinuity of the optimal solution set mapping and continuity of the optimal value function of the problem.

THEOREM 3.4 (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* $t \in [0, \bar{t}]$, $\hat{X}_{\text{sol}}(t) \cap \bar{X} \neq \emptyset$. *Then there exists a positive constant* $t^* < \bar{t}$ *such that*

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

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

Proof. Part (i). Let $x \in \bar{X}$. Since Assumption 3.1 holds at x, by Theorem 3.3 there exists 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 that we can 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

[1]It is obvious that conclusions of the lemma hold when the normal cone is replaced by any closed set-valued mapping.

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}, $i = 1, \ldots, \hat{i}$, 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 [37, Chapter 2, Proposition 1], $\vartheta(x, t) = \mathbb{E}[\hat{v}(x, \xi, t)]$ 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 [10, 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 in [20, p. 3]. To this end, we discuss the MFCQ of the regularized problem REG(x, ξ, t) in Proposition 3.5 under the MPEC-MFCQ of MPCC(x, ξ).

PROPOSITION 3.5. *Let $x^* \in X$, $\xi^* \in \Xi$ be fixed, and $y^* \in \mathcal{F}(x^*, \xi^*)$. Assume that problem* MPCC(x^*, ξ^*) *satisfies the 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 simplicity of notation, let $z = (x, y, \xi)$ and $z^* = (x^*, y^*, \xi^*)$, and throughout the proof, ∇ denotes the gradient with respect to y. By the definition of MFCQ, it suffices to show that there exists 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 $\nabla h_i(z) : i = 1, \ldots, r$, are linearly independent, and there exists a vector $d(z)$ (depending on z) such that

$$
(3.5) \qquad \begin{cases} 0 = \nabla h_i(z)^T d(z), & i = 1, \ldots, r, \\ 0 > \nabla g_i(z)^T d(z), & i \in \mathcal{I}_g(z), \\ 0 > -\nabla G_i(z)^T d(z), & i \in \mathcal{I}_G(z), \\ 0 > -\nabla H_i(z)^T d(z), & i \in \mathcal{I}_H(z), \\ 0 > (H_i(z) \nabla G_i(z) + G_i(z) \nabla H_i(z))^T d(z), & i \in \mathcal{I}_{G \circ H}(z), \end{cases}
$$

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

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

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

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

$$\nabla g_i(z^*)^T \bar{d} < 0 \quad \text{for } i \in \mathcal{I}_g(z^*). \tag{3.6}$$

Second, it is not difficult to show that there exists 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:

$$\begin{cases} \mathcal{I}_g(z) \subseteq \mathcal{I}_g(z^*), \\ \mathcal{I}_G(z) \subseteq \mathcal{I}_G(z^*), \\ \mathcal{I}_H(z) \subseteq \mathcal{I}_H(z^*), \\ \mathcal{I}_G(z) \cap \mathcal{I}_{G\circ H}(z) = \varnothing, \\ \mathcal{I}_H(z) \cap \mathcal{I}_{G\circ H}(z) = \varnothing, \end{cases}$$

and the gradient vectors

$$\begin{aligned} &\nabla h_i(z), \quad i = 1, \ldots, r; \quad \nabla G_i(z), \quad i \in \mathcal{I}_G(z), \quad \nabla H_i(z), \quad i \in \mathcal{I}_H(z); \\ &\quad H_i(z)\nabla G_i(z) + G_i(z)\nabla H_i(z), \quad i \in \mathcal{I}_{G\circ H}(z), \end{aligned}$$

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

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

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

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

Denote $A^{\#}(z) := A(z)[A(z)^T A(z)]^{-1}$. The well-definedness of $A^{\#}(z)$ (hence of $\hat{d}(z, \gamma)$) follows from the linear independence of the column vectors of $A(z)$ as discussed in the second step. 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

$$\nabla h_i(z)^T d(z, \gamma) = \nabla h_i(z)^T (\gamma \bar{d} - \hat{d}(z, \gamma)) = 0, \quad i = 1, \ldots, r. \tag{3.7}$$

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

$$\begin{aligned} \nabla g_i(z)^T d(z, \gamma) &= \gamma \nabla g_i(z)^T \bar{d} - \nabla g_i(z)^T \hat{d}(z, \gamma) \\ &= \gamma \nabla g_i(z)^T \bar{d} - \nabla g_i(z)^T (A^{\#}(z) b(z, \gamma)) \\ &= \gamma [\nabla g_i(z)^T \bar{d} - \nabla g_i(z)^T (A_r(z) \nabla h(z)^T \bar{d})] \\ &\quad - \nabla g_i(z)^T (A_{r-}(z)(1, 1, -1)^T), \end{aligned} \tag{3.8}$$

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 exists a positive scalar γ sufficiently large and a neighborhood $U_2 \subseteq U_1$ of z^* such that $\nabla g_i(z)^T d(z) < 0$ for all $z \in U_2$. 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 $z \in U_3$,

$$\begin{cases} -\nabla G_i(z)^T d(z,\gamma) = -\gamma \nabla G_i(z)^T \bar{d} - 1 < 0, & i \in \mathcal{I}_G(z), \\ -\nabla H_i(z)^T d(z,\gamma) = -\gamma \nabla H_i(z)^T \bar{d} - 1 < 0, & i \in \mathcal{I}_H(z), \end{cases} \tag{3.9}$$

and

$$\begin{aligned} (H_i(z)\nabla G_i(z) + G_i(z)\nabla H_i(z))^T d(z,\gamma) &= H_i(z)(-\gamma \nabla G_i(z)^T \bar{d} - 1) \\ &\quad + G_i(z)(-\gamma \nabla H_i(z)^T \bar{d} - 1) \\ &< 0, \quad i \in \mathcal{I}_{G\circ H}(z). \end{aligned} \tag{3.10}$$

Letting $U = U_3$, $U_{y^*} = \Pi_y U$, $U_{(x^*,\xi^*)} = \Pi_{(x,\xi)} U$, and combining (3.7)–(3.10), we obtain $d(z) = d(z,\gamma)$, satisfying (3.5) as desired and hence the conclusion.

COROLLARY 3.6. *Assume the conditions of Proposition* 3.5. *Then there exists 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, ∇ denotes 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

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

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

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

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

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

The claim holds for $U_{y^*} = \Pi_y U$ 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], [21] for $\hat{v}(x,\xi,t)$ and a stability result on parametric MPEC by Hu and Ralph [19] 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 pseudo-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 NLP.

THEOREM 3.7. *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* $Y_{\rm sol}(x^*, \xi)$. *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 $\text{REG}(x, \xi, t)$ satisfies MFCQ at every point in the optimal solution set $\hat{Y}_{\rm sol}(x, \xi, t)$ 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 $y^k \in \hat{Y}_{\rm sol}(x^k, \xi, t_k)$ such that $\text{REG}(x^k, \xi, t_k)$ fails to satisfy MFCQ at point y^k. Under Assumption 3.1, the optimal solution set $\hat{Y}_{\rm sol}(x, \xi, t)$ is bounded for all $x \in U$ and $t \in (0, t^*]$. Moreover, it follows from Theorem 3.3 that the optimal solution set mapping $\hat{Y}_{\rm sol}(\cdot, \cdot, \cdot)$ is outer semicontinuous on $U \times \Xi \times [0, t^*]$ and is contained in Y. Therefore, the sequence $\{y^k\}$ must have an accumulation point $\bar{y}$, and any accumulation point must be in $Y_{\rm sol}(x^*, \xi)$. Applying Proposition 3.5 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 $\text{REG}(x, \xi, t)$ 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 $\hat{Y}_{\rm sol}(x, \xi, t)$ for $(x, t) \in U \times (0, t^*]$, by [20, Proposition 3], we have that $\hat{\mathcal{F}}(x, \xi, t)$ is pseudo-Lipschitzian at $(y; x, t)$, where $(x, t) \in U \times (0, t^*]$ and $y \in \hat{Y}_{\rm sol}(x, \xi, t)$. 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 3.6 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, ξ), where $x \in U$. From [19, Formula (8)], we have that for x near x^*,

$$v(x, \xi) = \min_{J \in \mathcal{J}(x^*, \xi)} v_J(x, \xi), \tag{3.11}$$

where $\mathcal{J}(x^*, \xi) := \{J | J \in \mathcal{J}(y), y \in Y_{\rm sol}(x^*, \xi)\}$ and $\mathcal{J}(y)$ is defined by (2.2) with $z^* = (x^*, y, \xi)$. Denote the optimal solution set mapping of problem $\text{NLP}_J(x, \xi)$ (see (2.3)) by $Y_J(x, \xi)$. For any $J \in \mathcal{J}(x^*, \xi)$, $Y_J(x^*, \xi) \cap Y_{\rm sol}(x^*, \xi)$ is nonempty, and thus $Y_J(x^*, \xi) \subseteq Y_{\rm sol}(x^*, \xi)$. The MPEC-MFCQ assumption therefore gives the MFCQ for $\text{NLP}_J(x, \xi)$ at each $y \in Y_J(x, \xi)$. By the proof of part (i), $v_J(\cdot, \xi)$ is local Lipschitz continuous and so is $v(\cdot, \xi)$ through (3.11). ∎

It is important to note that we are short of claiming the local Lipschitz continuity of $\hat{v}(x, \xi, t)$ at point $t = 0$ in Theorem 3.7. This is because the MFCQ established in Proposition 3.5 is satisfied only for $t > 0$. We will show the local Lipschitz continuity in Theorem 4.10, 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 as MPCC(x, ξ):

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

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

$$\text{(4.1)} \qquad \begin{cases} 0 = \nabla_y \mathcal{L}(x, y, \xi; \alpha, \beta, u, v), \\ y \in \mathcal{F}(x, \xi), \\ 0 \le \alpha \perp -g(x, y, \xi) \ge 0, \\ 0 = u_i, & i \notin \mathcal{I}_G(x, y, \xi), \\ 0 = v_i, & i \notin \mathcal{I}_H(x, y, \xi), \\ 0 \le u_i v_i, & i \in \mathcal{I}_G(x, y, \xi) \cap \mathcal{I}_H(x, y, \xi). \end{cases}$$

Let $\mathcal{W}(x, \xi)$ denote the set of KKT pairs $(y; \alpha, \beta, u, v)$ satisfying the above conditions for given (x, ξ), and denote by $S(x, \xi)$ 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) are the corresponding Lagrange multipliers. When the stationary points are restricted to global minimizers, we denote the set of KKT pairs by $\mathcal{W}^*(x, \xi)$, that is, $\mathcal{W}^*(x, \xi) = \{(y; \alpha, \beta, u, v) \in \mathcal{W}(x, \xi); y \in Y_{\text{sol}}(x, \xi)\}$.

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

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

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

$$\text{(4.2)} \qquad \begin{cases} 0 = \nabla_y \hat{\mathcal{L}}(x, y, \xi, t; \alpha, \beta, \gamma, \theta, \lambda), \\ 0 \le -g(x, y, \xi) \perp \alpha \ge 0, \\ 0 = h(x, y, \xi), \\ 0 \le G(x, y, \xi) \perp \gamma \ge 0, \\ 0 \le H(x, y, \xi) \perp \theta \ge 0, \\ 0 \le te - G(x, y, \xi) \circ H(x, y, \xi) \perp \lambda \ge 0. \end{cases}$$

Let $\hat{\mathcal{W}}(x, \xi, t)$ denote the set of KKT pairs $(y; \alpha, \beta, \gamma, \theta, \lambda)$ satisfying the above conditions and $\hat{S}(x, \xi, t)$ denote 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 4.2. 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 $\emptyset \ne \hat{\mathcal{F}}(x, \xi, t) \subset Y$ for all $(x, \xi, t) \in U \times \Xi \times [0, t^*]$.

Assumption 4.2 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 4.2, we have the following proposition which describes a relationship between $S(x,\xi)$ and $\hat{S}(x,\xi,t)$.

PROPOSITION 4.3. *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* $\mathrm{REG}(x^k,\xi^k,t_k)$. *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* $\mathrm{MPCC}(x^*,\xi)$ *satisfies the MPEC-MFCQ at* y^*, *then* y^* *is a C-stationary point of* $\mathrm{MPCC}(x^*,\xi)$.

(ii) *If, in addition, Assumption* 4.2 *holds at point* x^* *and the 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

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

Since y^k is a stationary point of $\mathrm{REG}(x^k,\xi^k,t_k)$, 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

$$\begin{aligned} 0 = {} & \nabla_y f(x^k,y^k,\xi^k) + \sum_{i\in\mathcal{I}_g^k} \alpha_i^k \nabla_y g_i(x^k,y^k,\xi^k) + \sum_{i=1}^r \beta_i^k \nabla_y h_i(x^k,y^k,\xi^k) \\ & - \sum_{i\in\mathcal{I}_G^k} \gamma_i^k \nabla_y G_i(x^k,y^k,\xi^k) - \sum_{i\in\mathcal{I}_H^k} \theta_i^k \nabla_y H_i(x^k,y^k,\xi^k) \\ & + \sum_{i\in\mathcal{I}_{G\circ H}^k} \lambda_i^k \nabla_y [H_i(x^k,y^k,\xi^k) G_i(x^k,y^k,\xi^k)], \end{aligned} \tag{4.3}$$

$$\begin{cases} 0 \le -g(x^k,y^k,\xi^k) \perp \alpha^k \ge 0, \\ 0 = h(x^k,y^k,\xi^k), \\ 0 \le G(x^k,y^k,\xi^k) \perp \gamma^k \ge 0, \\ 0 \le H(x^k,y^k,\xi^k) \perp \theta^k \ge 0, \\ 0 \le t_k e - G(x^k,y^k,\xi^k) \circ H(x^k,y^k,\xi^k) \perp \lambda^k \ge 0. \end{cases} \tag{4.4}$$

Let

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

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

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

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 (4.3) can be rewritten as

$$\begin{aligned} 0 &= \nabla_y f(x^k, y^k, \xi^k) + \nabla_y g(x^k, y^k, \xi^k)\bar{\alpha}^k \\ &\quad + \nabla_y h(x^k, y^k, \xi^k)\beta^k - \nabla_y G(x^k, y^k, \xi^k)u^k - \nabla_y H(x^k, y^k, \xi^k)v^k + R_k(x^k, y^k, \xi^k), \end{aligned} \tag{4.5}$$

where

$$\begin{aligned} R_k(x^k, y^k, \xi^k) = &\sum_{i\in\mathcal{I}_{G\circ H}^k \cap (\mathcal{I}_G^*)^c} \lambda_i^k H_i(x^k, y^k, \xi^k)\nabla_y G_i(x^k, y^k, \xi^k) \\ &+ \sum_{i\in\mathcal{I}_{G\circ H}^k \cap (\mathcal{I}_H^*)^c} \lambda_i^k G_i(x^k, y^k, \xi^k)\nabla_y H_i(x^k, y^k, \xi^k). \end{aligned}$$

Since MPEC-MFCQ holds at y^*, by Proposition 3.5 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 [18, Theorem 3.4]). Taking a further subsequence if necessary, we may assume that the limits

$$\alpha_i^* = \lim_{k\to\infty} \bar{\alpha}_i^k, \quad \beta_i^* = \lim_{k\to\infty} \beta_i^k, \quad u_i^* = \lim_{k\to\infty} u_i^k, \quad v_i^* = \lim_{k\to\infty} v_i^k$$

exist. Moreover, $(\mathcal{I}_G^*)^c \subseteq \mathcal{I}_H^*$ and $(\mathcal{I}_H^*)^c \subseteq \mathcal{I}_G^*$. Consequently, the limit on (4.5) implies

$$\begin{aligned} &\nabla_y f(x^*, y^*, \xi) + \nabla_y g(x^*, y^*, \xi)\alpha^* + \nabla_y h(x^*, y^*, \xi)\beta^* - \nabla_y G(x^*, y^*, \xi)u^* \\ &\quad - \nabla_y H(x^*, y^*, \xi)v^* = 0. \end{aligned}$$

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

$$u_i^* v_i^* = \lim_{k\to\infty} (-\lambda_i^k H_i(x^k, y^k, \xi^k))(-\lambda_i^k G_i(x^k, y^k, \xi^k)) \geq 0,$$

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

$$u_i^* v_i^* = \lim_{k\to\infty} (\gamma_i^k \,\mathrm{or}\, 0)(\theta_i^k \,\mathrm{or}\, 0) \geq 0,$$

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 $\hat{S}(x, \hat{\xi}, t)$ is bounded for $(x, \hat{\xi})$ close to (x^*, ξ) and t sufficiently small. By Proposition 3.2, $\hat{\mathcal{F}}(x, \xi, t)$ 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 4.3 deals with the Clarke stationary points. If the smallest eigenvalue of $\nabla_y^2 \mathcal{L}(x^k, y^k, \xi^k; \alpha^k, \beta^k, u^k, v^k)$ is lower bounded (the Hessian could be singular) independent of k, then we can show similar to the proof of [24, Theorem 3.5] that the stationary points of REG(x^k, ξ^k, t_k) converge to an S-stationary point of MPCC(x^*, ξ) (note that in [24, Theorem 3.5] Lin and Fukushima derived B-stationarity which is equivalent to S-stationarity under MPEC-LICQ, but under MPEC-MFCQ, one can easily derive the S-stationarity under the eigenvalue condition). We leave the details for interested readers as they are beyond the scope of this paper. 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 4.6), and it is also of independent interest.

PROPOSITION 4.4. *Suppose that Assumption* 3.1 *holds at point* x^* *and problem* MPCC(x^*,ξ) *satisfies MPEC-MFCQ at every point* y *in set* $Y_{\rm sol}(x^*,\xi)$*. Then there exists a neighborhood* U *of* x^* *and a scalar* $t^*>0$ *such that*

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

$$\partial_x v(x,\xi)\subseteq \Phi(x,\xi), \tag{4.6}$$

where

$$\Phi(x,\xi)=\operatorname{conv}\left\{\bigcup_{(y;\alpha,\beta,u,v)\in\mathcal{W}^*(x,\xi)}\nabla_x\mathcal{L}(x,y,\xi;\alpha,\beta,u,v)\right\}; \tag{4.7}$$

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

$$\partial_x\hat{v}(x,\xi,t)\subseteq\hat{\Phi}(x,\xi,t),\ \ \partial_t\hat{v}(x,\xi,t)\subseteq\Lambda(x,\xi,t), \tag{4.8}$$

where

$$\hat{\Phi}(x,\xi,t)=\operatorname{conv}\left\{\bigcup_{(y;\alpha,\beta,\gamma,\theta,\lambda)\in\hat{\mathcal{W}}^*(x,\xi,t)}\nabla_x\hat{\mathcal{L}}(x,y,\xi,t;\alpha,\beta,\gamma,\theta,\lambda)\right\} \tag{4.9}$$

and $\Lambda(x,\xi,t)=\Pi_\lambda\hat{\mathcal{W}}^*(x,\xi,t)$*; the equality in* (4.8) *holds if the MPEC-MFCQ is replaced by the MPEC-LICQ;*

(iii) $\Phi(\cdot,\cdot)$ *is outer semicontinuous on* $U\times\Xi$ *and* $\hat{\Phi}(\cdot,\cdot,\cdot)$ *is outer semicontinuous 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 an analysis similar to the proof of Theorem 3.7, there exists 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 $\hat{Y}_{\rm sol}(x,\xi,t)$, and MPCC$(x,\xi)$ satisfies MPEC-MFCQ at every point in the optimal solution set $Y_{\rm sol}(x,\xi)$.

Part (i). Following an argument similar to the proof of [26, Theorem 4.8], we can show that for any $x\in U$ and $\xi\in\Xi$,

$$\partial_x^M v(x,\xi)\subseteq\left\{\bigcup_{(y;\alpha,\beta,u,v)\in\mathcal{W}^*(x,\xi)}\nabla_x\mathcal{L}(x,y,\xi;\alpha,\beta,u,v)\right\}. \tag{4.10}$$

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 $\operatorname{conv}\partial_x^M v(x,\xi)=\partial_x v(x,\xi)$, we obtain (4.6).

Part (ii) follows from [18, Theorem 5.3 and Corollary 5.4].

Part (iii). We prove only 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 MFCQ holds at every point of optimal solution set $\hat{Y}_{\rm sol}(x^k,\xi^k,t_k)$ for k sufficiently large, by the proof of [18, 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 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 3.3 (i), that $\hat{Y}_{\rm sol}(\cdot,\cdot,\cdot)$ is outer semicontinuous on $U \times \Xi \times [0,t^*]$, which implies $y \in \hat{Y}_{\rm sol}(x,\xi,t)$, 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 $\hat{\mathbf{\Phi}}$ 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 4.3 (replacing $\hat{S}(x^k,\xi^k,t_k)$ with $\hat{Y}_{\rm sol}(x^k,\xi^k,t_k)$), we can prove that

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

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. $\square$

It might be helpful to note that the equality in (4.8) 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 4.6. In the literature on MPECs, Lucet and Ye [26] 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 [18, 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 the equality in (4.10) holds.

Remark 4.5. In the definition of $\mathbf{\Phi}$ and $\hat{\mathbf{\Phi}}$, we use the KKT pairs at the global optimal solutions of the second stage problems. It is possible to cover all KKT pairs in the definitions, that is, to replace $\mathcal{W}^*$ and $\hat{\mathcal{W}}^*$ with $\mathcal{W}$ and $\hat{\mathcal{W}}$. Consequently, we may obtain larger outer bounds $\mathbf{\Psi}$ and $\hat{\mathbf{\Psi}}$, defined as follows for the Clarke subdifferentials of the optimal value functions:

$$\mathbf{\Psi}(x,\xi) := \text{conv}\left\{\bigcup_{(y;\alpha,\beta,u,v)\in\mathcal{W}(x,\xi)} \nabla_x\mathcal{L}(x,y,\xi;\alpha,\beta,u,v)\right\} \tag{4.11}$$

and

$$\hat{\mathbf{\Psi}}(x,\xi,t) := \text{conv}\left\{\bigcup_{(y;\alpha,\beta,\gamma,\theta,\lambda)\in\hat{\mathcal{W}}(x,\xi,t)} \nabla_x\hat{\mathcal{L}}(x,y,\xi,t;\alpha,\beta,\gamma,\theta,\lambda)\right\}. \tag{4.12}$$

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 gradients of both functions are well defined; (b) we need some consistency property of the subdifferentials of $\hat{v}(x, \xi, t)$ (see (4.16) in Theorem 4.6), and it turns out that the Clarke subdifferentials can fulfill this under MPEC-LICQ through Proposition 4.4 (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 3.7, $v(x, \xi)$ 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 the 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:

$$0 \in \mathbb{E}[\partial_x v(x, \xi)] + \mathcal{N}_X(x). \tag{4.13}$$

It is well known that $\partial \mathbb{E}[v(x, \xi)] \subseteq \mathbb{E}[\partial_x v(x, \xi)]$, and the equality holds when v is Clarke regular; see, for instance, [13, Theorem 2.8.2], [44], and [46, section 4.2] for recent discussions related to limiting subdifferentials.

We call (4.13) the *weak KKT condition* of the first stage problem (3.1). Likewise, we may consider weak KKT conditions of $P_{\hat{\vartheta}}$:

$$0 \in \mathbb{E}[\partial_x \hat{v}(x, \xi, t)] + \mathcal{N}_X(x). \tag{4.14}$$

Let $X_{\rm sta}$ and $\hat{X}_{\rm sta}(t)$ denote, respectively, the set of stationary points satisfying (4.13) and (4.14). In what follows we establish a relationship between the two sets as $t \to 0$.

THEOREM 4.6. *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 $Y_{\rm sol}(x^*, \xi)$, then*

$$\lim_{x \to x^*, t \downarrow 0} \mathbb{D}(\hat{\Phi}(x, \xi, t), \Phi(x^*, \xi)) = 0. \tag{4.15}$$

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

$$\lim_{x \to x^*, t \downarrow 0} \mathbb{D}(\partial_x \hat{v}(x, \xi, t), \partial_x v(x^*, \xi)) = 0. \tag{4.16}$$

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 $Y_{\rm sol}(x, \xi)$ for every $x \in X$ and $\xi \in \Xi$, and* (c) *$\partial_x \hat{v}(x, \xi, t)$ is integrably bounded*[2] *(that is, there exists $\kappa(\xi)$ such that $\|\partial_x \hat{v}(x, \xi, t)\| \le \kappa(\xi)$) and the probability measure is nonatomic, then*

$$\lim_{t \downarrow 0} \mathbb{D}(\hat{X}_{\rm sta}(t), X_{\rm sta}) = 0. \tag{4.17}$$

Proof. Part (i). By Theorem 3.7, there exists 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

[2]The condition is satisfied under Assumption 4.8.

locally Lipschitz continuous on U. By Theorem 3.3, any accumulation point y^* of $\{y^k\}$ with $y^k \in \hat{Y}_{\rm sol}(x^k, \xi, t_k)$ is contained in $Y_{\rm sol}(x^*, \xi)$. Mimicking the proof of Proposition 4.3 (replacing $\hat{S}(x^k, \xi, t_k)$ with $\hat{Y}_{\rm sol}(x^k, \xi, t_k)$), we can prove that

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

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 (4.16). Under MPEC-LICQ, the application of [18, Corollary 5.4] to the regularized second stage problem MPEC(x^*, ξ, t) gives

$$\partial_x \hat{v}(x^*, \xi, t) = \hat{\Phi}(x^*, \xi, t). \tag{4.18}$$

On the other hand, it follows from (4.6) that $\partial_x v(x^*, \xi) \subseteq \Phi(x^*, \xi)$. In what follows we show

$$\partial_x v(x^*, \xi) \supseteq \Phi(x^*, \xi) = \bigcup_{(y;\alpha,\beta,u,v)\in\mathcal{W}^*(x^*,\xi)} \{\nabla_x \mathcal{L}(x^*, y, \xi; \alpha, \beta, u, v)\}. \tag{4.19}$$

Under the assumption that MPEC-LICQ holds at every point in optimal solution set $Y_{\rm sol}(x^*, \xi)$, it follows by virtue of [19, Theorem 2, Formula (7)] that

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

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

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

Let

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

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 the Clarke generalized derivative [13] of $-v$ in x. By the definition of Clarke generalized gradient [13, p. 27], $\eta \in \partial_x(-v)(x^*, \xi)$, and by [13, 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

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

which implies (4.19). This shows $\partial_x v(x^*, \xi) = \Phi(x^*, \xi)$, and through (4.15) and (4.18), the subdifferential consistency (4.16) is shown.

Let us now prove (4.17). Since Assumption 3.1 holds at every point x in X and MPEC-LICQ holds at any point in $Y_{\rm sol}(x,\xi)$ for every $x \in X$ and $\xi \in \Xi$, we have from the subdifferential consistency (4.16) that

$$\lim_{x' \to x, t \downarrow 0} \mathbb{D}(\partial_x \hat{v}(x', \xi, t), \partial_x v(x, \xi)) = 0 \tag{4.20}$$

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

$$0 \in \mathbb{E}[\partial_x \hat{v}(x(t), \xi, t)] + \mathcal{N}_X(x(t)). \tag{4.21}$$

The compactness of X implies the boundedness of $\hat{X}_{\rm sta}(t)$. Therefore, we may assume without loss of generality that $x(t) \to \hat{x}$, where $\hat{x} \in X$. From (4.21), we have

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

where the first inclusion follows from [5, 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 (4.20). This implies that $\hat{x}$ is a weak KKT point satisfying (4.13). The proof is complete. ∎

The first-order optimality conditions (4.13)–(4.14) 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 4.4, we may consider optimality conditions by replacing $\partial_x v(x,\xi)$ with $\mathbf{\Phi}(x,\xi)$ in the weak KKT conditions (4.13) and by replacing $\partial_x \hat{v}(x,\xi,t)$ with $\hat{\mathbf{\Phi}}(x,\xi,t)$ in the weak KKT conditions (4.14). These kinds of optimality conditions are considered by Outrata and Römisch [29, Theorem 3.5] and more recently by Ralph and Xu [33] for classical two-stage stochastic programs. We will not go into 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 $\mathbf{\Psi}$ and $\hat{\mathbf{\Psi}}$ as defined in Remark 4.5. We give a formal definition for the latter as we need them in section 6.

Definition 4.7. *We call the stochastic generalized equation*

$$0 \in \mathbb{E}[\mathbf{\Psi}(x,\xi)] + \mathcal{N}_X(x) \tag{4.22}$$

the relaxed KKT conditions *of the first stage true problem* (3.1)*, and we call*

$$0 \in \mathbb{E}[\hat{\mathbf{\Psi}}(x,\xi,t)] + \mathcal{N}_X(x) \tag{4.23}$$

the relaxed KKT conditions *of the first stage regularized problem* (3.4)*. A point $x^* \in X$ satisfying* (4.22) *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* (4.23) *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 the MFCQ are needed in Definition 4.7 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 4.4 are valid.

Note also that in the stochastic programming literature, these types of relaxed KKT conditions were considered by Ralph and Xu [33] 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 [48].

Assumption 4.8. For every $x \in X$, there exists 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

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

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 4.8 holds when the support set Ξ of $\xi(\omega)$ is bounded and $\nabla_x f(x,y,\xi)$, $\nabla_x g(x,y,\xi)$, $\nabla_x h(x,y,\xi)$, $\nabla_x G(x,y,\xi)$, $\nabla_x H(x,y,\xi)$ are continuous in ξ. To see this, we note that both $\mathcal{W}$ and $\hat{\mathcal{W}}$ are outer semicontinuous, and this property is retained under orthogonal projections Π. It is easy to prove using the finite covering theorem that a compact outer semicontinuous set-valued mapping is bounded over a compact set. All other quantities in the curly brackets are continuous functions and are hence bounded over a compact set. 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 4.8, 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 4.9. *Suppose that Assumption* 4.2 *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 exists a neighborhood 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 $(x^*,\xi^*) \in U \times \Xi$,*

$$\lim_{(x,\xi,t)\to(x^*,\xi^*,0)} \mathbb{D}(\hat{\mathbf{\Psi}}(x,\xi,t), \mathbf{\Psi}(x^*,\xi^*)) = 0; \tag{4.24}$$

(iii) *under Assumption 4.8, $\mathbb{E}[\hat{\mathbf{\Psi}}(x,\xi,t)]$ and $\mathbb{E}[\mathbf{\Psi}(x,\xi)]$ are well defined for any $x \in U$ and $t \in (0,t^*]$ and if the probability measure is nonatomic, then*

$$\lim_{x\to x^*, t\downarrow 0} \mathbb{D}(\mathbb{E}[\hat{\mathbf{\Psi}}(x,\xi,t)], \mathbb{E}[\mathbf{\Psi}(x^*,\xi)]) = 0 \quad \forall x \in U. \tag{4.25}$$

Proof. Part (i). By Assumption 4.2, there exists a neighborhood U of x and a scalar $t^* > 0$ such that the feasible sets $\mathcal{F}(x,\xi)$ 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 proof similar to that in Proposition 4.4 (iii), we can show that $\hat{\mathcal{W}}(\cdot,\cdot,\cdot)$ is outer semicontinuous on $U \times \Xi \times (0,t^*]$ and that $\mathcal{W}(\cdot,\cdot)$ is outer semicontinuous on $U \times \Xi$.

Part (ii). The proof is similar to that of Proposition 4.4 (iv). We omit the details.

Part (iii). Viewing $\hat{\mathbf{\Psi}}$ as a composition of $\nabla\hat{\mathcal{L}}$ and $\hat{\mathcal{W}}$, we claim that $\hat{\mathbf{\Psi}}$ is outer semicontinuous and, through [35, Theorem 14.13], the measurability. The well-definedness then follows from the boundedness of $\hat{\Psi}$ under Assumption 4.8 and the definition of Aumann's integral. Finally, we prove (4.25). Notice that $\hat{\Psi}$ is a closed set-valued mapping on $U \times \Xi \times (0, t^*]$, and it is integrably bounded under Assumption 4.8. Note that the above analysis also holds for Ψ. The conclusion follows via application of [16, Theorem 2.5] (or [16, Theorem 2.8] and the following remark). The proof is complete. ∎

Note that Proposition 4.9 (iii) implies that any stationary point satisfying (4.23) converges to the set of stationary points satisfying (4.22). 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 4.4 as it requires some complex arguments related to singular subdifferentials, limiting subdifferentials, Clarke subdifferentials of $\hat{v}(x, \xi, t)$, and their approximations.

THEOREM 4.10. *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 $Y_{\rm sol}(x^*, \xi)$ for every $\xi \in \Xi$. Then*

(i) *there exists 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* 4.8 *holds at point x^*, then there exists 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* 3.4 *are satisfied and Assumption* 4.8 *holds for all $x \in \bar{X}$ ($\bar{X}$ is given in Theorem* 3.4*), then there exists a scalar $t^* > 0$ such that $\phi(t)$ is globally Lipschitz continuous on $[0, t^*]$.*

Proof. Part (i). By Theorem 3.7, there exists 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 [26, Proposition 2.4], it suffices to show that $\partial^{\infty}_{(x,t)} \hat{v}(x, \xi, 0) = \{0\}$. From Proposition 4.4 (see (4.8)) and [13, 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, $\hat{Y}_{\rm sol}(x, \xi, t)$ is bounded. Since MPEC-MFCQ holds at every point in the optimal solution set $Y_{\rm sol}(x^*, \xi)$, by the proof of Theorem 3.7, there exists 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 $\hat{Y}_{\rm sol}(x, \xi, t)$. Under the MFCQ, the boundedness of $\hat{\mathcal{W}}^*(x, \xi, t)$ follows from the proof of [18, 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 Proposition 4.4 and Assumption 4.8, 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 4.8. From Proposition 2 of [37, Chapter 2] and $\hat{v}(x, \cdot, t)$ being 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 4.8 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 $t \in [0, t_2]$, $X_{\rm sol}(t) \cap \bar{X} \neq \varnothing$. Let $t^* = \min\{t_1, t_2\}$ and $t', t'' \in [0, t^*]$ with $t' < t''$. It is easy to verify that

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

By Lebourg's mean value theorem [13, Theorem 2.3.7] and Proposition 4.4 (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 4.8. The conclusion follows. ∎

Note that Theorem 4.10 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, this turns out to be very difficult or even impossible.

In this section, we discuss a scheme for approximating the probability measure P. Specifically, we write $\mathbb{E}[\hat{v}(x, \xi, t)]$ as $\int_\Xi \hat{v}(x, \xi, t) \mathrm{d}P(\xi)$ and then consider a sequence of probability measures $\{P_\nu\}$ approximating P. Here P_ν is assumed to be numerically more tractable than P. In practice, there are many schemes to approximate P or $\mathbb{E}[\hat{v}(x, \xi, t)]$. The most well known examples are empirical probability measure approximation (which is also known as SAA, to be discussed specifically in section 6) and optimal scenario generation technique; see sections 4.1–4.2 in the excellent review paper [36] by Römisch and the references therein. To simplify the discussion, we fix the regularization parameter t and the probability measure P is nonatomic throughout this section.

Consider the first stage regularized problem (3.4). Let Ξ be the support set of $\xi(\omega)$ and P be a Borel probability measure on Ξ. Problem (3.4) can be equivalently written as

$$\min_x \hat{\vartheta}_P(x, t) = \int_\Xi \hat{v}(x, \xi, t) \mathrm{d}P(\xi) \quad \text{s.t. } x \in X. \tag{5.1}$$

Let P_ν be a sequence of probability measures $\{P_\nu\}$ approximating P in distribution as $\nu \to \infty$. Instead of solving (5.1) directly, we solve the approximation problem

$$\min_x \hat{\vartheta}_{P_\nu}(x, t) = \int_\Xi \hat{v}(x, \xi, t) \mathrm{d}P_\nu(\xi) \quad \text{s.t. } x \in X. \tag{5.2}$$

We study the perturbation of the optimal value and the set of optimal solutions and stationary points of (5.2) 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 [36] and the 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 (5.1) and (5.2), 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 exists 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) $\lim_{\nu\to\infty} \mathbb{D}(X^*_{P_\nu}(t) \cap \bar{X}, X^*_P(t) \cap \bar{X}) = 0$,

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

Proof. By the covering theorem and Theorem 3.3, 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 [37, Chapter 2, Proposition 1], $\hat{\vartheta}_P(x, t)$ and $\hat{\vartheta}_{P_\nu}(x, t)$, $\nu = 1, 2, \ldots,$ 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

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

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, [46, 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 with respect to 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 (5.1) and (5.2) can be written, respectively, as

$$0 \in \partial_x \hat{\vartheta}_P(x, t) + \mathcal{N}_X(x) \tag{5.3}$$

and

$$0 \in \partial_x \hat{\vartheta}_{P_\nu}(x, t) + \mathcal{N}_X(x), \tag{5.4}$$

where ∂ denotes the Clarke subdifferential. Let $S^*_P(t)$ and $S^*_{P_\nu}(t)$ denote the set of stationary points satisfying (5.3) and (5.4), respectively. Following an argument similar to that in section 3.2, we may consider weaker KKT conditions of (5.1) and (5.2) defined, respectively, as

$$0 \in \int_\Xi \partial_x \hat{v}(x, \xi, t)\mathrm{d}P(\xi) + \mathcal{N}_X(x) \tag{5.5}$$

and

$$0 \in \int_\Xi \partial_x \hat{v}(x, \xi, t)\mathrm{d}P_\nu(\xi) + \mathcal{N}_X(x), \tag{5.6}$$

where $\partial_x \hat{\vartheta}(x, t) \subset \int_\Xi \partial_x \hat{v}(x, \xi, t)\mathrm{d}P(\xi)$ and $\partial_x \hat{\vartheta}_{P_\nu}(x, t) \subset \int_\Xi \partial_x \hat{v}(x, \xi, t)\mathrm{d}P_\nu(\xi)$. The equality holds when v is Clarke regular; see, for instance, [13, Theorem 2.8.2], [44], and [46, section 4.2] for recent discussions related to limiting subdifferentials. Let $S^w_P(t)$ and $S^w_{P_\nu}(t)$ denote the set of stationary points satisfying (5.5) and (5.6), 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 that $\int_\Xi \partial_x \hat{v}(x,\xi,t)\mathrm{d}P_\nu(\xi)$ approximates $\int_\Xi \partial_x \hat{v}(x,\xi,t)\mathrm{d}P(\xi)$ uniformly as $\nu \to \infty$.

LEMMA 5.2 (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 of $\mathbb{R}^n$. Assume the following:* (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 and $\frac{1}{\tau}(F(x+\tau h,\xi) - F(x,\xi))$ is uniformly continuous with respect to ξ for $x \in \chi$, $\|h\| \le 1$, and τ sufficiently small;* (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*

$$\lim_{\nu\to\infty} \sup_{x\in\mathcal{X}} \mathbb{H}(\partial \mathbb{E}_{P_\nu}[F(x,\xi)], \partial \mathbb{E}_P[F(x,\xi)]) = 0; \tag{5.7}$$

(ii) *if $\partial_x F(x,\xi)$ is outer semicontinuous in ξ, then*

$$\lim_{\nu\to\infty} \sup_{x\in\mathcal{X}} \mathbb{D}(\mathbb{E}_{P_\nu}[\partial_x F(x,\xi)], \mathbb{E}_P[\partial_x F(x,\xi)]) = 0; \tag{5.8}$$

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

$$\lim_{\nu\to\infty} \sup_{x\in\mathcal{X}} \mathbb{H}(\mathbb{E}_{P_\nu}[\partial_x F(x,\xi)], \mathbb{E}_P[\partial_x F(x,\xi)]) = 0. \tag{5.9}$$

Proof. Part (i). For 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 $f_{P_\nu}^o(x;h)$ and $f_P^o(x;h)$, respectively, are well defined for any fixed nonzero vector $h \in \mathbb{R}^n$, where

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

and

$$f_P^o(x;h) = \lim_{x'\to x, \tau\downarrow 0} \sup \frac{1}{\tau}(f_P(x'+\tau h) - f_P(x')).$$

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 $f_{P_\nu}^o(x;h)$ and $f_P^o(x;h)$. 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 formulas; see [11, Theorem II-18]. Applying the second formula to our setting, we have

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

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,

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

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

$$\left| \limsup_{k\to\infty} a_k - \limsup_{k\to\infty} b_k \right| \le \limsup_{k\to\infty} |a_k - b_k|. \tag{5.10}$$

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

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

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

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

This verifies (5.10). Using (5.10), we have

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

Since P_ν converges to P in distribution, and the integrand $\frac{1}{\tau}(F(x' + \tau h, \xi) - F(x', \xi))$ is uniformly continuous with respect to ξ and is bounded by L, then by virtue of [7, Theorem 2.1],

$$\limsup_{\nu\to\infty} \sup_{x\in\mathcal{X}} \sup_{\|h\|\le 1} \limsup_{x'\to x, \tau\downarrow 0} \left| \int_\Xi \frac{1}{\tau}(F(x' + \tau h, \xi) - F(x', \xi)) \mathrm{d}(P - P_\nu)(\xi) \right| = 0. \tag{5.11}$$

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 $F^o(x, \xi(\cdot); h)$ through [4, Theorem 8.2.5]. Since $F^o(x, \xi; h)$ is the support function of $\partial_x F(x, \xi)$, by [4, 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 and compact set-valued and that $\mathbb{E}_P[\|\partial_x F(x, \xi)\|] \le C$.

In view of a discussion by Artstein and Vitale [2], $\mathbb{E}_P[\partial_x F(x,\xi)]$ 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 (5.8) through [3, Theorem 4.2], and, (5.9) by virtue of [3, Theorem 3.1]. The proof is complete. $\square$

We make a few comments about Lemma 5.2 because it is prepared not only for establishing our main result, Theorem 5.3, but also for general interest such as the stability analysis of stationary points in general nonsmooth stochastic programming. First, Birge and Qi [9] 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 [9, Theorem 4.1] for details. Our result (5.7) is stronger than the convergence result in [9, equation (4.1)] in the sense that the convergence here is uniform and there is no restriction on the distribution of P_ν. Second, Artstein and Wets [3] established a number of convergence results for the integral of random set-valued mappings when the probability measure P_ν converges weakly to P. Lemma 5.2 (ii) is a direct application of their results to Clarke subdifferentials.

THEOREM 5.3 (stability of stationary points). *Let X be a compact set and Assumptions* 3.1 *and* 4.8 *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 and $\frac{1}{\tau}(\hat{v}(x+\tau h,\xi,t) - \hat{v}(x,\xi,t))$ is uniformly continuous with respect to ξ for any $(x,\xi,t) \in X \times \Xi \times [0,t^*]$, $\|h\| \le 1$, and τ sufficiently small, then*

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

and if $\partial_x \hat{v}(x,\xi,t)$ is outer semicontinuous in ξ, then

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

Proof. Part (i) follows from Theorems 3.3 and 4.10. Part (ii) follows from [46, Lemma 4.2] and Lemma 5.2. $\square$

Before concluding this section, we point out a popular special case when P_ν is an empirical probability measure. That is,

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

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

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

It is well known that P_ν converges weakly to P with probability one (w.p.1); see, for instance, [43]. 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 $F(\cdot,\xi^k)$, $k=1,\ldots,\nu$, is Clarke regular at x. Putting this into the context of Lemma 5.2, we have, from the law of large numbers, that (5.11) holds w.p.1 as long as $\frac{1}{\tau}(F(x'+\tau h,\xi)-F(x',\xi))$ is bounded by an integrable function $L(\xi)$ (independent of x and τ it is uniformly continuous with respect to ξ. Consequently, we have

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

w.p.1 and

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

w.p.1. If, in addition, $\partial_x F(x,\xi)$ is Hausdorff continuous in ξ, then

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

w.p.1. Let us point out a subtle difference between the convergence results here and those in Lemma 5.2. The Lipschitz modulus here is integrably bounded which is obviously weaker than condition (a) in the lemma and our convergence results are also weaker in that the limits hold w.p.1. In view of Theorem 5.3 (ii), the boundedness of $\hat{L}(x,\xi,t)$ by C may be weakened to $C(\xi)$ w.p.1 and the consequent limits hold w.p.1 rather than deterministically.

6. Sample average approximation. In this section, we discuss SAA of the regularized two-stage problem. This is a combination of the stability analyses in sections 3–5, but we have independent interest: we investigate the behavior of optimal solutions and stationary points when the regularization parameter t is driven to zero, and the probability measure P is approximated by the empirical probability measure (sample average). By focusing on SAA, 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 (3.3)–(3.4) in a compact form:

$$
\begin{aligned}
&\min_{x,y(\cdot)} \mathbb{E}[f(x, y(\omega), \xi(\omega))] \\
&\text{s.t. } x \in X, \text{ and for a.e. } \omega \in \boldsymbol{\Omega}\text{:} \\
&\quad g(x, y(\omega), \xi(\omega)) \leq 0, \\
&\quad h(x, y(\omega), \xi(\omega)) = 0, \\
&\quad -G(x, y(\omega), \xi(\omega)) \leq 0, \\
&\quad -H(x, y(\omega), \xi(\omega)) \leq 0, \\
&\quad G(x, y(\omega), \xi(\omega)) \circ H(x, y(\omega), \xi(\omega)) \leq te.
\end{aligned} \tag{6.1}
$$

The equivalence between (6.1) and (3.3)–(3.4) is well documented in the stochastic programming literature (see, e.g., [37, Chapter 1, section 2.4]). Let $\xi^1, \ldots, \xi^N$ be an independent identically distributed sample. We consider the following SAA of the regularized problem (6.1):

$$
\begin{aligned}
&\min_{x;y^1,\ldots,y^N} \frac{1}{N}\sum_{i=1}^{N} f(x, y^i, \xi^i) \\
&\text{s.t. } x \in X, \text{ and for } i = 1, \ldots, N\text{:} \\
&\quad g(x, y^i, \xi^i) \leq 0, \\
&\quad h(x, y^i, \xi^i) = 0, \\
&\quad -G(x, y^i, \xi^i) \leq 0, \\
&\quad -H(x, y^i, \xi^i) \leq 0, \\
&\quad G(x, y^i, \xi^i) \circ H(x, y^i, \xi^i) \leq t_N e,
\end{aligned} \tag{6.2}
$$

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 $\hat{v}(x, \xi^i, t)$, $i = 1, \ldots, N$, to denote the optimal value of the regularized second stage problem (3.3) with $\xi = \xi^i$ and assume that $(x; y^1, \ldots, y^N)$ is a global optimal solution, then problem (6.2) can be written in an implicit form, that is,

$$
\begin{aligned}
&\min_{x} \frac{1}{N}\sum_{i=1}^{N} \hat{v}(x, \xi^i, t_N) \\
&\text{s.t. } x \in X,
\end{aligned} \tag{6.3}
$$

which is the SAA of the first stage (3.4). Here "implicit" is in the sense that (6.3) does not explicitly involve the underlying functions of the second stage problem. The terminology is used by Ralph and Xu in [33], where SAA is applied to a classical two-stage stochastic program.

SAA 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 [31], [34], [37] for SAA in general stochastic programming and [8], [40], [47], [27] for recent application of the method to SMPECs.

The regularized SAA scheme for a two-stage SMPEC problem was first considered in [41] and with some detailed convergence analysis in a conference paper [32], 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 $\{x^N; y^1, \ldots, y^N\}$ is a global optimal solution to SAA problem (6.2), we investigate an accumulation point of $\{x^N\}$ as the sample size N increases. From the numerical perspective, if we obtain an approximate global optimal solution from solving (6.2) 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 the true problem (1.1).

THEOREM 6.1. *Let $\{(x^N; y^1, \ldots, y^N)\}$ be a sequence of global optimal solutions of problem* (6.2) *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* (3.1)*. Suppose the following:* (a) *Assumptions* 3.1 *and* 4.8 *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 $Y_{\text{sol}}(x, \xi)$ for $(x, \xi) \in \bar{X} \times \Xi$. Then*

(i) *w.p.1 $\hat{x}$ is an optimal solution to the true problem* (3.1)*.*

(ii) *suppose, in addition, the following:* (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 as in Assumption* 4.8*. 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*

$$\text{Prob}(d(x^N, X_{\text{sol}}) \geq \epsilon) \leq C(\epsilon) e^{-\beta(\epsilon) N} \tag{6.4}$$

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,

$$\lim_{N \to \infty} \sup_{x \in \bar{X}} \left| \frac{1}{N} \sum_{i=1}^{N} \hat{v}(x, \xi^i, t_N) - \mathbb{E}[v(x, \xi)] \right| = 0 \text{ w.p.1.} \tag{6.5}$$

Indeed, if (6.5) holds, then we can claim, by virtue of [46, Lemma 4.1] or [35, 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 (6.5).

Since Assumption 3.1 holds at every point $x \in \bar{X}$, it follows from Theorem 3.7 (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 4.4 (i),

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

Under Assumption 4.8,

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

where $\kappa(\xi)$ is as given in Assumption 4.8 and $\mathbb{E}[\kappa(\xi)^2] < \infty$. Further, the condition $x^* \in \bar{X}$ implies $\mathbb{E}[v(x^*, \xi)] < \infty$. Therefore, for every $x \in \bar{X}$,

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

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

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

This implies, through the classical uniform law of large numbers [37, Lemma A1], that

$$\lim_{N\to\infty} \sup_{x\in\bar{X}} \left| \frac{1}{N}\sum_{i=1}^{N} v(x,\xi^i) - \mathbb{E}[v(x,\xi)] \right| = 0 \quad \text{w.p.1.} \tag{6.6}$$

On the other hand, under Assumption 3.1, we know through Theorem 4.10 that $\hat{v}(\cdot,\xi,\cdot)$ is locally Lipschitz continuous at $(x,0)$ for $x \in \bar{X}$. Moreover, by Proposition 4.4 (ii) and Assumption 4.8,

$$\|\partial_t \hat{v}(x,\xi,t)\| \le \|\Pi_\lambda \hat{\mathcal{W}}(x,\xi,t)\| \le \kappa(\xi) \quad \forall\ (x,\xi) \in \bar{X} \times \Xi,$$

where $\kappa(\xi)$ is as given in Assumption 4.8. Consequently, we have

$$\begin{aligned}
\left| \frac{1}{N}\sum_{i=1}^{N} \hat{v}(x,\xi^i,t_N) - \mathbb{E}[v(x,\xi)] \right| &\le \frac{1}{N}\sum_{i=1}^{N} |\hat{v}(x,\xi^i,t_N) - v(x,\xi^i)| \\
&\quad + \left| \frac{1}{N}\sum_{i=1}^{N} v(x,\xi^i) - \mathbb{E}[v(x,\xi)] \right| \\
&\le \frac{1}{N}\sum_{i=1}^{N} \kappa(\xi^i) t_N + \left| \frac{1}{N}\sum_{i=1}^{N} v(x,\xi^i) - \mathbb{E}[v(x,\xi)] \right|.
\end{aligned} \tag{6.7}$$

Combining (6.6) and (6.7) 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 (6.5).

Part (ii). Let $\epsilon > 0$ be given. By [14, Lemma 3.2] (or [46, Lemma 4.1]), there exists a $\delta(\epsilon) > 0$ such that if

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

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

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

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 [41, 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 $N_1 \ge N_0$ such that for $N \ge N_1$,

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

Combining the above two inequalities with (6.7), we have

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

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

We now move on to discuss the case when a solution $\{x^N;y^1,\ldots,y^N\}$ obtained from solving the SAA problem (6.2) 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 to have a separate discussion on the convergence of x^N.

Consider the KKT conditions of the regularized SAA program (6.2):

$$0\in\frac{1}{N}\sum_{i=1}^{N}\nabla_x\hat{\mathcal{L}}(x,y^i,\xi^i,t;\alpha^i,\beta^i,\gamma^i,\theta^i,\lambda^i)+\mathcal{N}_X(x), \tag{6.8}$$

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

$$\begin{cases}0=\nabla_y\hat{\mathcal{L}}(x,y^i,\xi^i,t;\alpha^i,\beta^i,\gamma^i,\theta^i,\lambda^i),\\ 0\leq -g(x,y^i,\xi^i)\perp\alpha^i\geq 0,\\ 0=h(x,y^i,\xi^i),\\ 0\leq G(x,y^i,\xi^i)\perp\gamma^i\geq 0,\\ 0\leq H(x,y^i,\xi^i)\perp\theta^i\geq 0,\\ 0\leq t_Ne-G(x,y^i,\xi^i)\circ H(x,y^i,\xi^i)\perp\lambda^i\geq 0.\end{cases} \tag{6.9}$$

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

The KKT conditions (6.9) imply that $(y^i;\alpha^i,\beta^i,\gamma^i,\theta^i,\lambda^i)$ is a KKT pair of $\mathrm{REG}(x,\xi^i,t_N)$; that is,

$$(y^i;\alpha^i,\beta^i,\gamma^i,\theta^i,\lambda^i)\in\hat{\mathcal{W}}(x,\xi^i,t_N).$$

By the definition of $\hat{\boldsymbol{\Psi}}(x,\xi,t)$ (see (4.11)),

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

Combining this with (6.8), we arrive at

$$0\in\frac{1}{N}\sum_{i=1}^{N}\hat{\boldsymbol{\Psi}}(x,\xi^i,t_N)+\mathcal{N}_X(x), \tag{6.10}$$

which implies that (6.10) is an SAA of the relaxed KKT condition (4.23).

THEOREM 6.2. *Let $\{x^N;y^1,\ldots,y^N\}$ be a stationary point of problem (6.2) and $\hat{x}$ be an accumulation point of $\{x^N\}$. Suppose that Assumptions 4.2 and 4.8 hold at $\hat{x}$, and suppose* *that problem* MPCC(x, ξ) *satisfies MPEC-MFCQ at every point in the feasible set* $\mathcal{F}(\hat{x}, \xi)$ *for every* $\xi \in \Xi$ *and that the probability measure is nonatomic. Then w.p.*1 $\hat{x}$ *is a relaxed stationary point of the true problem* (3.1)*;that is,* $\hat{x}$ *satisfies* (4.22).

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 4.9, there exists a neighborhood U of $\hat{x}$ and a scalar $t^* > 0$ such that $\mathcal{A}(\cdot, \cdot, \cdot)$ is outer semicontinuous on $U \times \Xi \times [0, t^*]$. Under Assumption 4.8, $\mathcal{A}(x, \xi(\cdot), t)$ is measurable and integrably bounded. The conclusion follows by application of [49, Theorem 4.3]. The proof is complete. ◻

Theorem 6.2 addresses almost sure convergence of stationary points. It might be both theoretically and practically interesting to discuss exponential convergence so that one can estimate the sample size for a prescribed precision. However, this would require a lot of complicated technical analysis, and such analysis would have significantly increased the length of this paper. We leave this for future research.

Let us now make a few comments on the numerical resolution of SAA problem (6.2). For a given sample, this is a deterministic NLP. Therefore, theoretically speaking, any existing NLP solver may be applied to solve it. However, when sample size increases, the problem size could be very large. Consequently, one may wish to apply some techniques which exploit the special structure of the problem before plugging it into an NLP solver. Research like this is perhaps well known in the literature of stochastic programming. Let us point out some recent development in this direction. Bastin [6] considered a trust-region-based Schur-complement scheme for solving stochastic NLPs; Shanbhag [39] proposed a line-search-based decomposition method for solving stochastic MPCCs. Kulkarni and Shanbhag [22] discussed a novel hybrid algorithm which combines the SQP-based method and Benders decomposition. Moreover, if (6.1) is a convex program, the dual method proposed by Zhao [50] may be applicable.

Acknowledgments. The authors are grateful to the two anonymous referees and Werner Römisch for their valuable suggestions and comments that have significantly helped to improve the presentation of the paper and to the associate editor Mihai Anitescu for effective handling of the article.

REFERENCE

[1] M. Anitescu, *On using the elastic mode in nonlinear programming approaches to mathematical programs with complementarity constraints*, SIAM J. Optim., 15 (2005), pp. 1203–1236.

[2] Z. Artstein and R. A. Vitale, *A strong law of large numbers for random compact sets*, Ann. Probab., 3 (1975), pp. 879–882.

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

[4] R. J. Aubin and H. Frankowska, *Set-Valued Analysis*, Birkhäuser Boston, Cambridge, MA, 1990.

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

[6] F. Bastin, *Trust-Region Algorithms for Nonlinear Stochastic Programming and Mixed Logit Models*, Ph.D. thesis, University of Namur, Namur, Belgium, 2004.

[7] P. Billingsley, *Convergence of Probability Measures*, John Wiley, New York, 1968.

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

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

[10] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer, New York, 2000.
[11] C. Castaing and M. Valadier, *Convex Analysis and Measurable Multifunctions*, Lecture Notes in Math. 580, Springer, Berlin, 1977.
[12] S. Christiansen, M. Patriksson, and L. Wynter, *Stochastic bilevel programming in structural optimization*, Struct. Multidiscip. Optim., 21 (2001), pp. 361–371.
[13] F. H. Clarke, *Optimization and Nonsmooth Analysis*, John Wiley, New York, 1983.
[14] L. Dai, C. H. Chen, and J. R. Birge, *Convergence properties of two-stage stochastic programming*, J. Optim. Theory Appl., 106 (2000), pp. 489–509.
[15] D. De Wolf and Y. Smeers, *A stochastic version of a Stackelberg–Nash–Cournot equilibrium model*, Manage. Sci., 43 (1997), pp. 190–197.
[16] F. Hiai, *Convergence of conditional expectations and strong law of large numbers for multivalued random variables*, Trans. Amer. Math. Soc., 291 (1985), pp. 613–627.
[17] R. Fletcher, S. Leyffer, D. Ralph, and S. Scholtes, *Local convergence of SQP methods for mathematical programs with equilibrium constraints*, SIAM J. Optim., 17 (2006), pp. 259–286.
[18] J. Gauvin and F. Dubeau, *Differential properties of the marginal function in mathematical programming*, Math. Programming Stud., 19 (1982), pp. 101–119.
[19] X. Hu and D. Ralph, *A note on sensitivity of value functions of mathematical programs with complementarity constraints*, Math. Program., 93 (2002), pp. 265–279.
[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, pp. 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, pp. 77–86.
[22] A. Kulkarni and U. V. Shanbhag, *Recourse-based stochastic nonlinear programming: Properties and Benders-SQP algorithms*, Comput. Optim. Appl., to appear, 2010.
[23] G. H. Lin, X. Chen, and M. Fukushima, *New restricted NCP functions and their applications to stochastic NCP and stochastic MPEC*, Optimization, 56 (2007), pp. 641–653.
[24] G. H. Lin and M. Fukushima, *A modified relaxation scheme for mathematical programs with complementarity constraints*, Ann. Oper. Res., 133 (2005), pp. 63–84.
[25] G. H. Lin, X. Chen, and M. Fukushima, *Solving stochastic mathematical programs with equilibrium constraints via approximation and smoothing implicit programming with penalization*, Math. Program., 116 (2009), pp. 343–368.
[26] Y. Lucet and J. J. Ye, *Sensitivity analysis of the value function for optimization problems with variational inequality constraints*, SIAM J. Control Optim., 40 (2001), pp. 699–723.
[27] F. Meng and H. Xu, *A regularized sample average approximation method for stochastic mathematical programs with nonsmooth equality constraints*, SIAM J. Optim., 17 (2006), pp. 891–919.
[28] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation.* I*: Basic Theory.* Grundlehren Math. Wiss. (Fundamental Principles of Mathematical Sciences) 330, Springer, Berlin, 2006.
[29] J. Outrata and W. Römisch, *On optimization conditions for some nonsmooth optimization problems over* L^p *spaces*, J. Optim. Theory Appl., 126 (2005), pp. 411–438.
[30] M. Patriksson and L. Wynter, *Stochastic mathematical programs with equilibrium constraints*, Oper. Res. Lett., 25 (1999), pp. 159–167.
[31] E. L. Plambeck, B. R. Fu, S. M. Robinson, and R. Suri, *Sample-path optimization of convex stochastic performances functions*, Math. Program., 75 (1996), pp. 137–176.
[32] D. Ralph, H. Xu, and F. Meng, *Regularization methods for stochastic mathematical programs with complementarity constraints*, Program Book of the 21st European Conference on Operational Research, Rejkjavik, Iceland, 2006.
[33] D. Ralph and H. Xu, *Asympototic analysis of stationary points of sample average two-stage stochastic programs: A generalized equation approach*, Math. Oper. Res., submitted, 2008.
[34] S. M. Robinson, *Analysis of sample-path optimization*, Math. Oper. Res., 21 (1996), pp. 513–528.
[35] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer, Berlin, 1998.
[36] W. Römisch, *Stability of stochastic programming problems*, in Stochastic Programming, Handbooks in Operations Research and Management Science 10, A. Ruszczynski and A. Shapiro eds., Elsevier, Amsterdam, 2003, pp. 483–554.
[37] A. Ruszczynski and A. Shapiro, *Stochastic Programming*, Handbook in Operations Research and Management Science 10, Elsevier, Amsterdam, 2003.

[38] H. Scheel and S. Scholtes, *Mathematical programs with complementarity constraints: Stationarity, optimality, and sensivity*, Math. Oper. Res., 25 (2000), pp. 1–22.

[39] U. V. Shanbhag, *Decomposition and Sampling Methods for Stochastic Equilibrium Problems*, Ph.D. thesis, Stanford University, Palo Alto, CA, 2006.

[40] A. Shapiro, *Stochastic mathematical programs with equilibrium constraints*, J. Optim. Theory Appl., 128 (2006), pp. 223–243.

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

[42] S. Scholtes, *Convergence properties of a regularization scheme for mathematical programs with complementarity constraints*, SIAM J. Optim., 11 (2001), pp. 918–936.

[43] G. R. Shorack and J. A. Wellner, *Empirical Processes with Applications to Statistics*, John Wiley, New York, 1986.

[44] R. J.-B. Wets, *Stochastic Programming,* , Handbooks in Operations Research and Management Science 1, G. L. Nemhauser ed., Elsevier, New York, 1989.

[45] H. Xu, *An MPCC approach for stochastic Stackelberg–Nash–Cournot equilibrium*, Optimization, 54 (2005), pp. 27–57.

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

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

[48] H. Xu and J. J. Ye, *Necessary optimality conditions for two-stage stochastic programs with equilibrium constraints*, SIAM J. Optim., 20 (2010), pp. 1685–1715.

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

[50] G. Zhao, *A Lagrangian dual method with self-concordant barriers for multi-stage stochastic convex programming*, Math. Program., 102 (2005), pp. 1–24.