SIAM J. OPTIM.
Vol. 17, No. 3, pp. 891–919

A REGULARIZED SAMPLE AVERAGE APPROXIMATION METHOD FOR STOCHASTIC MATHEMATICAL PROGRAMS WITH NONSMOOTH EQUALITY CONSTRAINTS*

FANWEN MENG† AND HUIFU XU‡

Abstract. We investigate a class of two stage stochastic programs where the second stage problem is subject to nonsmooth equality constraints parameterized by the first stage variant and a random vector. We consider the case when the parametric equality constraints have more than one solution. A regularization method is proposed to deal with the multiple solution problem, and a sample average approximation method is proposed to solve the regularized problem. We then investigate the convergence of stationary points of the regularized sample average approximation programs as the sample size increases. The established results are applied to stochastic mathematical programs with P_0-variational inequality constraints. Preliminary numerical results are reported.

Key words. sample average approximation, Karush–Kuhn–Tucker conditions, regularization methods, P_0-variational inequality, convergence of stationary points

AMS subject classifications. 90C15, 90C30, 90C31, 90C33

DOI. 10.1137/050638242

1. Introduction.

In this paper, we study the following stochastic mathematical program:

$$\begin{array}{ll} \min & \mathbb{E}\left[f(x, y(x, \xi(\omega)), \xi(\omega))\right] \\ \text{s.t.} & x \in \mathcal{X}, \end{array} \tag{1}$$

where $\mathcal{X}$ is a nonempty compact subset of $\mathbb{R}^m$, $f : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$ is continuously differentiable, $\xi : \Omega \to \Xi \subset \mathbb{R}^k$ is a vector of random variables defined on probability space $(\Omega, \mathcal{F}, P)$, $\mathbb{E}$ denotes the mathematical expectation, and $y(x, \xi(\omega))$ is *some measurable selection* (which will be reviewed in section 2.2) from the set of solutions of the following system of equations:

$$H(x, y, \xi(\omega)) = 0, \tag{2}$$

where $H : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$ is a piecewise smooth vector-valued function. Piecewise smooth function is a large class of locally Lipschitz continuous functions which cover most practical problems [29]. For the simplicity of notation, we will write $\xi(\omega)$ as ξ, and this should be distinguished from where ξ is a deterministic vector of Ξ in a context. Throughout this paper, we assume that the probability measure P of our considered space $(\Omega, \mathcal{F}, P)$ is nonatomic.

The model is slightly different from the standard two stage stochastic programming model where the second stage decision variate y is chosen to either minimize or maximize $f(x, y, \xi)$ for given x and every realization of $\xi(\omega)$. See an excellent survey

*Received by the editors August 16, 2005; accepted for publication (in revised form) May 15, 2006; published electronically October 24, 2006. This research was supported by United Kingdom Engineering and Physical Sciences Research Council grant GR/S90850/01.
http://www.siam.org/journals/siopt/17-3/63824.html
†School of Mathematics, University of Southampton, Southampton SO17 1BJ, United Kingdom. Current address: The Logistics Institute - Asia Pacific, National University of Singapore, 1 Law Link, 119260 Singapore (tlimf@nus.edu.sg).
‡School of Mathematics, University of Southampton, Southampton SO17 1BJ, United Kingdom (h.xu@soton.ac.uk).

by Ruszczyński and Shapiro [27, Chapters 1 and 2] for the latter. In some practical instances, finding an optimal solution for the second stage problem may be very difficult and/or expensive. For example, in a two stage stochastic Stackelberg–Nash–Cournot model [32], x is the leader's decision variable and y is the followers' Nash–Cournot equilibrium vector with each component representing a follower's decision variable. The followers' equilibrium problem can be reformulated as a system of nonsmooth equations like (2) which depends on the leader's decision variable x and realization of uncertainty ξ in market demand. In the case when the followers have multiple equilibria, the "selection" of an optimal $y(x, \xi)$ at the second stage can be interpreted as the leader's attitude towards the followers' multiple equilibria: an optimistic attitude leads to a selection in favor of his utility, whereas a pessimistic attitude goes to an opposite selection. See [32, section 2] for details. Alternatively such an optimal selection can be interpreted as that the leader puts in some resources so that the followers reach an equilibrium in his favor. In either interpretation, finding such an optimal $y(x, \xi)$ implies additional cost to the leader.

Our argument here is that the leader may not necessarily select an extreme equilibrium (which minimizes/maximizes $f(x, y, \xi)$); instead, he may select one of the feasible equilibria $y(x, \xi)$ under the following circumstances: (a) minimizing/maximizing $f(x, y, \xi)$ with respect to y may be difficult or even impossible numerically; for instance, one can obtain only a local optimal solution or a stationary point; (b) in the case when an optimal solution is obtainable, the cost for obtaining such a solution in the second stage overweighs the overall benefit; for instance, the leader is a dominant player, while the followers are small players and the range of possible followers' equilibria is very narrow; (c) the chance of followers reaching one equilibrium or another is equal, and the leader is unaware of which particular equilibrium may be actually reached in the future and has no intention of putting in any additional resources to influence it.

Of course, such a selection must be consistent for all x and ξ; in other words, $y(x, \xi)$ must be a single-valued function with some measurability or even continuity. Note that there may exist many such selections, and the leader considers only one of them. This means that the leader takes a neutral attitude towards the follower's every possible equilibrium. Note also that in this paper, the selection is not arbitrary and is guided by a regularization method to be discussed shortly.

The argument can be extended to two stage stochastic mathematical programs with equilibrium constraints (SMPECs)

$$
\begin{array}{ll} \min & \mathbb{E}[f(x, y(x, \xi), \xi)] \\ \text{s.t.} & x \in \mathcal{X}, \end{array} \tag{3}
$$

where $y(x, \xi)$ solves

$$
\begin{array}{ll} \min\limits_{y} & f(x, y, \xi) \\ \text{s.t.} & F(x, y, \xi)^T (v - y) \geq 0 \quad \forall v \in \mathcal{C}(x, \xi), \end{array} \tag{4}
$$

f and F are continuously differentiable function, ξ is a random variable, and $\mathcal{C}(x, \xi)$ is a random convex set. SMPECs were initially studied by Patriksson and Wynter [18]. Like deterministic MPECs, SMPECs have various potential applications in engineering and economics, etc. [9, 34]. Over the past few years, SMPECs have received increasing investigation from perspectives of both stochastic programming and MPEC; see [17, 30, 35, 36] and the references therein. Observe that the second stage problem

(4) is a deterministic parametric MPEC that is intrinsically nonconvex. Finding an optimal solution for MPECs is often difficult if not impossible. Consequently it may be a realistic approach to take *some* feasible measurable solution at the second stage (which is a solution of the variational inequality problem in the constraint) rather than trying to find an optimal one. Note that MPECs can be easily reformulated as a nonsmooth system of equations, and this is the very reason why we consider general nonsmooth equality constraints (2). We will discuss these in detail in section 5.

Having motivated our model, we next explain how to find the unspecified $y(x, \xi)$ in (1). Our idea can be outlined as follows. We approximate function H with some function R parameterized by a small positive number μ and then solve the following equation:

$$R(x, y, \xi, \mu) = 0. \tag{5}$$

Of course, R cannot be any function, and it must be constructed according to the structure of H. First, it must coincide with H when $\mu = 0$; second, it must have some nice topological properties such as Lipschitz continuity and directional differentiability. Finally and perhaps most importantly, (5) must have a unique solution for every $x \in \mathcal{X}$, $\xi \in \Xi$, and nonzero μ. We specify these needed properties in a definition of R (Definition 2.1) and regard R as a *regularization* in consistency with the terminology in the literature [19, 11]. Using the regularization method, we expect that an implicit function $\tilde{y}(x, \xi, \mu)$ defined by (5) approximates a measurable feasible solution $y(x, \xi)$ of (2), and consequently we can utilize the program

$$\begin{array}{ll} \min & \mathbb{E}\left[f(x, \tilde{y}(x, \xi, \mu), \xi)\right] \\ \text{s.t.} & x \in \mathcal{X} \end{array} \tag{6}$$

to approximate the true problem (1), where $y(x, \xi)$ is the limit of $\tilde{y}(x, \xi, \mu)$ as $\mu \to 0$.

We then propose a sample average approximation (SAA) method to solve (6). The SAA method and its variants, known under various names such as "stochastic counterpart method," "sample-path method," "simulated likelihood method," etc., were discussed in the stochastic programming and statistics literature over the years. See, for instance, [22, 25, 3, 31] for general stochastic problems and [17, 4, 30, 32, 36] for SMPECs.

We investigate the convergence of the SAA problem of (6) as $\mu \to 0$ and sample size tends to infinity. Since the underlying functions are piecewise smooth and nonconvex in general, our analysis focuses on stationary points rather than local or global optimal solutions. For this purpose, we study the optimality conditions for both the true and the regularized problems. We introduce a kind of generalized Karush–Kuhn–Tucker (KKT) condition for characterizing both true and regularized problems in terms of Clarke generalized Jacobians (subdifferentials). Rockafellar and Wets [23] investigated KKT conditions for a class of two stage convex stochastic programs and derived some "basic Kuhn–Tucker conditions" in terms of convex subdifferentials. More recent discussions on the optimality conditions can also be found in books by Birge and Louveaux [5] and Ruszczyński and Shapiro [27]. These conditions rely on the convexity of underlying functions and hence cannot be applied to our problems, which are nonconvex.

The main contributions of this paper can be summarized as follows: we show that under some conditions, the solution of (5) approximates a measurable solution of (2); we then show that with probability 1 (w.p.1 for short) an accumulation point of a sequence of generalized stationary points of the regularized problem is a generalized stationary point of the true problem. We propose an SAA method to solve the regularized problem (6) and show that w.p.1 an accumulation point of the sequence of the stationary points of the regularized SAA problem is a generalized stationary point of the true problem as sample size tends to infinity and parameter μ tends to zero. Finally, we apply the established results to a class of SMPECs where the underlying function is a P_0-function.

The rest of the paper is organized as follows. In section 2, we discuss the regularization scheme. In section 3, we investigate the generalized stationary points of both the regularized problem and the true problem. In section 4, we study the convergence of the SAA program of the regularized problem. We then apply the established results to a class of stochastic MPEC problems in section 5. Some preliminary numerical results are reported in section 6.

2. Preliminaries and a regularization scheme. In this section, we characterize the function R in (5) and investigate the approximation of (6) to (1) as $\mu \to 0$.

Throughout this paper, we use the following notation. We use $\|\cdot\|$ to denote the Euclidean norm of a vector, a matrix, and a compact set of matrices. Specifically, if $\mathcal{M}$ is a compact set of matrices, then $\|\mathcal{M}\| := \max_{M\in\mathcal{M}} \|M\|$. We use $\operatorname{dist}(x, \mathcal{D}) := \inf_{x'\in\mathcal{D}} \|x - x'\|$ to denote the distance between point x and set $\mathcal{D}$. Here $\mathcal{D}$ may be a subset of $\mathbb{R}^n$ or a subset of matrix space $\mathbb{R}^{n\times n}$. Given two compact sets $\mathcal{C}$ and $\mathcal{D}$, we use $\mathbb{D}(\mathcal{C}, \mathcal{D}) := \sup_{x\in\mathcal{C}} \operatorname{dist}(x, \mathcal{D})$ to denote the distance from set $\mathcal{C}$ to set $\mathcal{D}$. For two sets $\mathcal{C}$ and $\mathcal{D}$ in a metric space, $\mathcal{C} + \mathcal{D}$ denotes the usual Minkowski addition, and $\mathcal{C}\mathcal{D} := \{CD \mid \text{ for all } C \in \mathcal{C}, \text{ for all } D \in \mathcal{D}\}$ represents the multiplication. We use $\mathcal{B}(x, \delta)$ to denote the closed ball in $\mathbb{R}^n$ with radius δ and center x, that is, $\mathcal{B}(x, \delta) := \{x' \in \mathbb{R}^m : \|x'-x\| \le \delta\}$. For a vector-valued function $g : \mathbb{R}^m \to \mathbb{R}^l$, we use $\nabla g(x)$ to denote the classical Jacobian of g when it exists. In the case when $l = 1$, that is, g is a real-valued function, $\nabla g(x)$ denotes the gradient of g which is a row vector. We use $\overline{\lim}$ to denote the outer limit of a sequence of vectors and a set-valued mapping. We let $\mathbb{R}_{++} := \{x \mid x > 0, x \in \mathbb{R}\}$ and $\mathbb{R}^2_{++} := \{(x, y) \mid x > 0, y > 0, x, y \in \mathbb{R}\}$. For a set-valued mapping $\mathcal{A}(u, v) : \mathbb{R}^n \times \mathbb{R}^m \to 2^{\mathbb{R}^{n\times m}}$, $\pi_u\mathcal{A}(u, v)$ denotes the set of all $n\times n$ matrices M such that, for some $n\times m$ matrix N, the $n\times(n+m)$ matrix $[M\, N]$ belongs to $\mathcal{A}(u, v)$.

2.1. Preliminaries. We first present some preliminaries about the Clarke generalized Jacobian of random functions which will be used throughout the paper.

Let $G : \mathbb{R}^j \to \mathbb{R}^l$ be a locally Lipschitz continuous vector-valued function. Recall that the Clarke generalized Jacobian [10] of G at $x \in \mathbb{R}^j$ is defined as

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

where D_G denotes the set of points at which G is Fréchet differentiable, $\nabla G(y)$ denotes the usual Jacobian of G, and "conv" denotes the convex hull of a set. It is well known that the Clarke generalized Jacobian $\partial G(x)$ is a convex compact set [10]. In the case that $j = l$, $\partial G(x)$ consists of square matrices. We say $\partial G(x)$ is nonsingular if every matrix in set $\partial G(x)$ is nonsingular, and in this case we use $\partial G(x)^{-1}$ to denote the set of all inverse matrices of $\partial G(x)$.

In later discussions, particularly sections 2–4, we will have to deal with mathematical expectation of the Clarke generalized Jacobians of locally Lipschitz random functions. For this purpose, we recall some basics about measurability of a random set-valued mapping.

Let $V \subset \mathbb{R}^n$ be a compact set of $\mathbb{R}^n$ and $\xi : \Omega \to \Xi \subset \mathbb{R}^k$ be a random vector defined on the probability space $(\Omega, \mathcal{F}, P)$ (note that we use the same notation ξ and Ξ as in (1), although we do not have to in this general discussion). A random set-valued mapping $\mathcal{A}(\cdot, \xi) : V \to 2^{\mathbb{R}^{n\times m}}$ is said to be *closed-valued* if for every $v \in V$ and $\xi \in \Xi$ (a realization of $\xi(\omega)$), $\mathcal{A}(v, \xi)$ is a closed set. Let $\mathfrak{B}$ denote the space of nonempty, closed subsets of $\mathbb{R}^{n\times m}$ equipped with Hausdorff distance. Then $\mathcal{A}(v, \xi(\cdot))$ can also be viewed as a single-valued mapping from Ω to $\mathfrak{B}$. For a fixed $v \in V$, $\mathcal{A}(v, \xi(\cdot)) : \Omega \to 2^{\mathbb{R}^{n\times m}}$ is said to be *measurable* if for every closed set $B \subset \mathbb{R}^{n\times m}$, $\{\omega : \mathcal{A}(v, \xi(\omega)) \bigcap B \neq \emptyset\}$ belongs to the σ-algebra $\mathcal{F}$. Alternatively, by viewing $\mathcal{A}(v, \xi(\cdot))$ as a single-valued mapping, we can say that $\mathcal{A}(v, \xi(\cdot))$ is measurable if and only if for every $B \in \mathfrak{B}$, $\mathcal{A}(v, \xi(\cdot))^{-1}B$ is $\mathcal{F}$-measurable. See Theorem 14.4 of [24].

We now define the expectation of $\mathcal{A}(v, \xi(\omega))$. A *selection* of a random set $\mathcal{A}(v, \xi(\omega))$ is a random matrix $A(v, \xi) \in \mathcal{A}(v, \xi)$ (which means $A(v, \xi(\omega))$ is measurable). Selections exist. The *expectation of* $\mathcal{A}(v, \xi(\omega))$, denoted by $\mathbb{E}[\mathcal{A}(v, \xi(\omega))]$, is defined as the collection of $\mathbb{E}[A(v, \xi(\omega))]$, where $A(v, \xi(\omega))$ is a selection. For a detailed discussion in this regard, see [1, 2] and the references therein.

Finally, we need the following definitions concerning matrices and functions. A matrix $M \in \mathbb{R}^{l\times l}$ is called a P_0-matrix if for any $x \neq 0$, there exists $i \in \{1, \dots, l\}$ such that $x_i(Mx)_i \geq 0$ and $x_i \neq 0$. It is evident that a positive semidefinite matrix is a P_0-matrix. A function $G : \mathcal{D} \subset \mathbb{R}^l \to \mathbb{R}^l$ is said to be (over set $\mathcal{D}$) a P_0-function if for all $u, v \in \delta$ with $u \neq v$,

$$\max_{\substack{i\in\{1,\dots,l\}\\ u_i\neq v_i}} (u_i - v_i)[G_i(u) - G_i(v)] \geq 0.$$

For a continuously differentiable function G, if $\nabla G(x)$ is a P_0-matrix for all $x \in \mathcal{D}$, then $G(x)$ is a P_0-function on $\mathcal{D}$. For a comprehensive discussion of the properties of the above matrices and functions, we refer readers to the book [11].

2.2. A regularization scheme. We specify the regularized approximation outlined in section 1 and investigate the limiting behavior of the implicit function defined by the regularized approximation problem (6) as $\mu \to 0$.

Throughout this paper $\partial H(x, y, \xi)$ denotes the Clarke generalized Jacobian of H at (x, y, ξ), and $\partial R(x, y, \xi, \mu)$ denotes the Clarke generalized Jacobian of H at (x, y, ξ, μ).

DEFINITION 2.1. *Let $\mu \in [0, \mu_0]$, where μ_0 is a positive number. A continuous function $R : \mathcal{X} \times \mathbb{R}^n \times \Xi \times [0, \mu_0] \to \mathbb{R}^n$ is said to be a* regularization *of H if the following hold:*

(i) *for every $x \in \mathcal{X}$, $y \in \mathbb{R}^n$, $\xi \in \Xi$, $R(x, y, \xi, 0) = H(x, y, \xi)$;*

(ii) *$R(x, y, \xi, \mu)$ is locally Lipschitz continuous and piecewise smooth on $\mathcal{X} \times \mathbb{R}^n \times \Xi \times [0, \mu_0]$;*

(iii) *for every $x \in \mathcal{X}$, $y \in \mathbb{R}^n$, and $\xi \in \Xi$,*

$$\overline{\lim_{\mu\downarrow 0}}\, \pi_x \partial R(x, y, \xi, \mu) \subset \pi_x \partial H(x, y, \xi), \qquad \overline{\lim_{\mu\downarrow 0}}\, \pi_y \partial R(x, y, \xi, \mu) \subset \pi_y \partial H(x, y, \xi);$$

(iv) *equation $R(x, y, \xi, \mu) = 0$ defines a unique locally Lipschitz continuous function $\tilde{y} : \mathcal{X} \times \Xi \times (0, \mu_0) \to \mathbb{R}^n$ such that $R(x, \tilde{y}(x, \xi, \mu), \xi, \mu) = 0$ for every $x \in \mathcal{X}, \mu \in (0, \mu_0),\ \xi \in \Xi$.*

We call μ a regularization parameter (or variable).

The definition contains three elements. First, a regularization is a parameterized continuous approximation and is locally Lipschitz continuous with respect to the regularization parameter when it is viewed as an additional variable. Second, the regularization (part (iii)) satisfies some kind of Jacobian consistency [8] that was widely used in smoothing methods when R is a smoothing of H; see [21] and the references therein. A sufficient condition for this is that R is strictly differentiable at $\mu = 0$ (when μ is treated as a variable). Third, the regularization scheme defines a unique function $\tilde{y}$ that approximates a measurable solution $y(x, \xi)$ of (2). We shall investigate the existence of such $\tilde{y}$ in Proposition 2.3.

Remark 2.2. Part (iv) of Definition 2.1 is implied by the following *uniform nonsingularity condition*: for every $(x, \xi, \mu) \in \mathcal{X} \times \Xi \times (0, \mu_0)$, there exists y such that $R(x, y, \xi, \mu) = 0$; $\pi_y \partial R(x, y, \xi, \mu)$ is uniformly nonsingular; i.e., there exists a positive constant $C > 0$ such that for every $x \in \mathcal{X}, y \in \mathbb{R}^n, \xi \in \Xi, \mu \in (0, \mu_0)$, $\|[\pi_y \partial R(x, y, \xi, \mu)]^{-1}\| \leq C$. The uniform nonsingularity implies that the outer limit of $\pi_y \partial R(x, y, \xi, \mu)$ as $\mu \to 0$ is a strict subset of $\pi_y \partial H(x, y, \xi)$, which does not include singular matrices.

PROPOSITION 2.3. *Let R be a function satisfying conditions* (i)–(iii) *of Definition* 2.1 *and the uniform nonsingularity condition hold. Then R is a regularization of H.*

Proof. It suffices to verify (iv) in Definition 2.1; that is, (5) defines a unique locally Lipschitz continuous implicit function $\tilde{y} : \mathcal{X} \times \Xi \times (0, \mu_0) \to \mathbb{R}^n$ such that $R(x, \tilde{y}(x, \xi, \mu), \xi, \mu) = 0$ for all $x \in \mathcal{X}, \xi \in \Xi, \mu \in (0, \mu_0)$. With the uniform nonsingularity of $\pi_y \partial R$, the existence of such an implicit function on $\mathcal{X} \times \mathbb{R}^n \times (0, \mu_0)$ comes straightforwardly from [36, Lemma 2.3]. ∎

In the analysis of sections 3 and 4, we will not assume the uniform nonsingularity. Instead we will assume a regularization R with nonsingularity of $\pi_y \partial R$ and other conditions which are weaker than the uniform nonsingularity. Note that not every function has a regularized approximation. Our definition here is motivated by the functions reformulated from equilibrium constraints. See section 5, particularly Example 5.5, for a detailed explanation. In what follows we investigate the properties of $\tilde{y}(x, \xi, \mu)$, in particular, its limit as the regularization parameter μ tends to zero.

THEOREM 2.4. *Let R be a regularization of H and $\tilde{y}(x, \xi, \mu)$ be the implicit function defined as in Definition* 2.1. *Assume that $\lim_{\mu \downarrow 0} \tilde{y}(x, \xi, \mu)$ exists for every $x \in \mathcal{X}$ and $\xi \in \Xi$, that is,*

$$y(x, \xi) := \lim_{\mu \downarrow 0} \tilde{y}(x, \xi, \mu), \quad x \in \mathcal{X}, \xi \in \Xi. \tag{7}$$

Suppose that there exists a positive measurable function $\kappa_1(\xi) > 0$ such that $\|\tilde{y}(x, \xi, \mu)\| \leq \kappa_1(\xi)$ for all $(x, \mu) \in \mathcal{X} \times (0, \mu_0)$ and that $\mathbb{E}[\kappa_1(\xi)] < \infty$. Then the following statements hold:

(i) *$y(x, \xi)$ is a solution function of* (2) *on $\mathcal{X} \times \Xi$, and $y(x, \xi(\cdot)) : \Omega \to \mathbb{R}^n$ is measurable for every $x \in \mathcal{X}$;*

(ii) *if, in addition, there exists a measurable positive function $L(\xi) > 0$ such that*

$$\|\tilde{y}(x'', \xi, \mu) - \tilde{y}(x', \xi, \mu)\| \leq L(\xi)\|x'' - x'\| \quad \forall x', x'' \in \mathcal{X}, \tag{8}$$

then $y(\cdot, \xi)$ is Lipschitz continuous on $\mathcal{X}$ for every $\xi \in \Xi$.

Proof. Part (i). By Definition 2.1, we have $R(x, \tilde{y}(x, \xi, \mu), \xi, \mu) = 0$ for $x \in \mathcal{X}, \xi \in \Xi, \mu \in (0, \mu_0)$. By Definition 2.1 and the continuity of R in y, it follows that

$$\lim_{\mu \downarrow 0} R(x, \tilde{y}(x, \xi, \mu), \xi, \mu) = R(x, y(x, \xi), \xi, 0) = H(x, y(x, \xi), \xi),$$

which indicates that $y(x, \xi)$ is a solution of (2) for $(x, \xi) \in \mathcal{X} \times \Xi$.

To show the measurability of $y(x, \xi(\omega))$, observe that since $\tilde{y}(x, \xi, \mu)$ is continuous in ξ, then $\tilde{y}(x, \xi(\cdot), \mu) : \Omega \to \mathbb{R}^n$ is measurable. Moreover, since $\tilde{y}(x, \xi, \mu)$ is bounded by $\kappa_1(\xi)$ and $\mathbb{E}[\kappa_1(\xi)] < \infty$, by the Lebesgue dominated convergence theorem, $y(x, \xi(\cdot))$ is measurable.

Part (ii). For $x', x'' \in \mathcal{X}$, by (8), $\|\tilde{y}(x'', \xi, \mu) - \tilde{y}(x', \xi, \mu)\|$ is dominated by $L(\xi)\|x'' - x'\|$. The latter is integrable. By the Lebesgue dominated convergence theorem, we have from (8)

$$\|y(x'', \xi) - y(x', \xi)\| = \|\lim_{\mu\downarrow 0}(\tilde{y}(x'', \xi, \mu) - \tilde{y}(x', \xi, \mu))\|$$
$$= \lim_{\mu\downarrow 0}\|\tilde{y}(x'', \xi, \mu) - \tilde{y}(x', \xi, \mu)\| \le L(\xi)\|x'' - x'\| \quad \forall x', x'' \in \mathcal{X}.$$

This completes the proof. ∎

The theorem above shows that we may obtain a measurable solution of (2) through the process of regularization. Note that our assumption on the existence of limit (7) may be relaxed. Indeed if the sequence of functions $\tilde{y}(\cdot, \cdot, \mu)$ has multiple accumulation points, each of which is Lipschitz continuous, then $y(x, \xi)$ can be taken from any of them. Our assumption is to simplify the consequent discussion, and also we expect this to be satisfied in practical instances; see Example 5.5. The boundedness condition for $\tilde{y}(x, \xi, \mu)$ holds under the uniform nonsingularity condition (Remark 2.2). Throughout the rest of this paper, the $y(x, \xi)$ in the true problem (1) refers to the limit of $\tilde{y}(x, \xi, \mu)$ as $\mu \to 0$.

3. Generalized Karush–Kuhn–Tucker (GKKT) conditions.

In this section, we investigate the KKT conditions of both true problem (1) and regularized program (6). Our purpose is to show that w.p.1 the stationary points of the regularized problem converge to a stationary point of the true problem (1) as the regularization parameter is driven to zero; therefore the regularized problem (6) is a reasonable approximation of the true problem.

3.1. GKKT conditions of the true problem.

Let R be a regularization of H and $\tilde{y}(x, \xi, \mu)$ the solution of (5), let

$$y(x, \xi) = \lim_{\mu\downarrow 0}\tilde{y}(x, \xi, \mu) \quad \text{ for } x \in \mathcal{X}, \xi \in \Xi,$$

and let $y(\cdot, \xi)$ be Lipschitz continuous. In this subsection, we investigate the true problem (1) associated with $y(x, \xi)$. We first define a set which resembles the set of Lagrange multipliers in nonlinear programming.

DEFINITION 3.1. *For $(x, \xi) \in \mathcal{X} \times \Xi$, let*

$$\Lambda(x, \xi) := \mathrm{conv}\{\lambda(x, \xi) \in \mathbb{R}^n \mid 0 \in \nabla_y f(x, y(x, \xi), \xi) + \lambda(x, \xi)\pi_y \partial H(x, y(x, \xi), \xi)\}. \tag{9}$$

Note that $\lambda(x, \xi)$ is a row vector. Note also that when $y(x, \xi)$ is an optimal solution of

$$\begin{array}{ll} \min\limits_{y} & f(x, y, \xi) \\ \text{s.t.} & H(x, y, \xi) = 0, \end{array} \tag{10}$$

$\Lambda(x, \xi)$ contains the Lagrange multipliers of (10) in that the first-order necessary condition of (10) can be written as $0 \in \nabla_y f(x, y(x, \xi), \xi) + \lambda(x, \xi)\partial_y H(x, y(x, \xi), \xi)$, and by [10, Proposition 2.3.16], $\partial_y H(x, y(x,\xi), \xi) \subset \pi_y \partial H(x, y(x,\xi), \xi)$. Note further that $\Lambda(x,\xi)$ is nonempty if and only if the set $\{\lambda(x,\xi) \in \mathbb{R}^n \mid 0 \in \nabla_y f(x, y(x,\xi), \xi) + \lambda(x,\xi)\, \pi_y \partial H(x, y(x,\xi), \xi)\}$ is nonempty. For every $\lambda(x,\xi)$ of the latter set, there exists a matrix $M \in \pi_y \partial H(x, y(x,\xi), \xi)$ such that $0 = \nabla_y f(x, y(x,\xi), \xi) + \lambda(x,\xi) M$. A necessary and sufficient condition for $\Lambda(x,\xi)$ to be nonempty is that there exists $M \in \pi_y \partial H(x, y(x,\xi), \xi)$ such that $\mathrm{rank}([\nabla_y f^T, M^T]) = \mathrm{rank}(M^T)$. Moreover, the set Λ is bounded if and only if M is of full row rank. The following remark discusses the particular case when $\pi_y \partial H$ is nonsingular.

Remark 3.2. If $\pi_y \partial H(x, y(x,\xi), \xi)$ is nonsingular for $x \in \mathbb{R}^m$ and $\xi \in \Xi$, then we have

$$\Lambda(x,\xi) := -\nabla_y f(x, y(x,\xi), \xi) \mathrm{conv}([\pi_y \partial H(x, y(x,\xi), \xi)]^{-1}).$$

The set contains the Lagrange multipliers of the standard second stage minimization problem (10), since the nonsingularity of the Jacobian guarantees $y(x,\xi)$ to be the only feasible solution and hence trivially the optimal solution! If H is continuously differentiable in x, y, and ξ, then $\pi_y \partial H(x, y, \xi)$ reduces to $\nabla_y H(x, y, \xi)$, and $\Lambda(x,\xi)$ reduces to a singleton,

$$\Lambda(x,\xi) = -\nabla_y f(x, y(x,\xi), \xi) \nabla_y H(x, y(x,\xi), \xi)^{-1}.$$

This corresponds to the classical Lagrange multiplier of a standard second stage minimization problem (10).

In this paper, we consider the case when (2) has multiple solutions; therefore $\pi_y \partial H(x, y, \xi)$ cannot be nonsingular. For the simplicity of discussion, we make a blanket assumption that $\Lambda(x,\xi)$ is nonempty for $x \in \mathcal{X}$ and $\xi \in \Xi$, which implies that there exists at least one matrix $M \in \pi_y \partial H(x, y(x,\xi), \xi)$ such that $\mathrm{rank}([\nabla_y f^T, M^T]) = \mathrm{rank}(M^T)$. Using the notion of Λ, we can define the following optimality conditions for the true problem associated with $y(x,\xi)$.

DEFINITION 3.3. *Let $\Lambda(x,\xi)$ be defined as in Definition* 3.1. *A point $x \in \mathbb{R}^m$ is called a* generalized stationary point *of the true problem* (1) *if*

$$0 \in \mathbb{E}[\nabla_x f(x, y(x,\xi), \xi) + \Lambda(x,\xi) \pi_x \partial H(x, y(x,\xi), \xi)] + \mathcal{N}_{\mathcal{X}}(x), \tag{11}$$

where the expectation is taken over the integrable elements of the set-valued integrand, and $\mathcal{N}_{\mathcal{X}}(x)$ denotes the normal cone of $\mathcal{X}$ at $x \in \mathbb{R}^m$ [6]; *that is,*

$$\mathcal{N}_{\mathcal{X}}(x) := [T_{\mathcal{X}}(x)]^- = \{\varsigma \in \mathbb{R}^m \mid \langle \varsigma, d \rangle \le 0 \quad \forall d \in T_{\mathcal{X}}(x)\},$$

where $T_{\mathcal{X}}(x) := \limsup_{t \downarrow 0} (\mathcal{X} - x)/t$. We call (11) *a GKKT condition of the true problem* (1).

For the set of generalized stationary points to be well defined, it must contain all local minimizers of the true problem. In what follows we discuss this and the relationship between the GKKT conditions with other possible KKT conditions. For this purpose, we need to state the following implicit function theorem for piecewise smooth functions.

LEMMA 3.4. *Consider an underdetermined system of nonsmooth equations $P(y,z) = 0$, where $P : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^m$ is piecewise smooth. Let $(\bar{y}, \bar{z}) \in \mathbb{R}^m \times \mathbb{R}^n$ be such that $P(\bar{y}, \bar{z}) = 0$. Suppose that $\pi_y \partial P(\bar{y}, \bar{z})$ is nonsingular. Then*

(i) *there exist neighborhoods Z of $\bar{z}$ and Y of $\bar{y}$ and a piecewise smooth function $y : Z \to Y$ such that $y(\bar{z}) = \bar{y}$ and, for every $z \in Z$, $y = y(z)$ is the unique solution of the problem $P(y,z) = 0$, $y \in Y$;*

(ii) *for* $z \in Z$,

$$\partial y(z) \subset \text{conv}\{-V^{-1}U : [V,U] \in \partial P(y(z),z),\ V \in \mathbb{R}^{m\times m},\ U \in \mathbb{R}^{m\times n}\}. \tag{12}$$

Proof. The existence of an implicit function comes essentially from the Clarke implicit function theorem [10]. The piecewise smoothness and the differential inclusion (12) follow from [21, Proposition 4.8] straightforwardly. ∎

Note that the purpose of (12) is to give an estimate of the Clarke generalized Jacobian of the implicit function using the Clarke generalized Jacobian of P, which is relatively easier to obtain. The estimate may be improved under some index consistency conditions; see [21] for details. With Lemma 3.4, we are ready to discuss our GKKT condition. The proposition below establishes a relation between (local) minimizers of the true problem and the generalized stationary points defined in Definition 3.3 under some circumstances.

PROPOSITION 3.5. *Let x^* be a local minimizer of the true problem* (*associated with the limit function $y(x,\xi)$*) *and $[\pi_y \partial H(x^*, y(x^*,\xi),\xi)]^{-1}$ be nonsingular. Let $\nabla f(x, y(x,\xi),\xi)$, $[\pi_y \partial H(x, y(x,\xi),\xi)]^{-1}$, and $\pi_x \partial H(x, y(x,\xi),\xi)$ be bounded by a positive integrable function for all x in a neighborhood of x^*. Then x^* is a generalized stationary point of the true problem.*

Proof. By Lemma 3.4,

$$\partial y(x,\xi) \subset \text{conv}([-\pi_y \partial H(x,y(x,\xi),\xi)]^{-1})\pi_x \partial H(x,y(x,\xi),\xi)$$

for x close to x^*. Let $v(x,\xi) := f(x,y(x,\xi),\xi)$. Under the boundedness conditions of $\nabla f(x,y(x,\xi),\xi)$, $[\pi_y \partial H(x,y(x,\xi),\xi)]^{-1}$, and $\pi_x \partial H(x,y(x,\xi),\xi)$, we have

$$\begin{aligned}
0 &\in \partial \mathbb{E}[v(x^*,\xi)] + \mathcal{N}_{\mathcal{X}}(x^*) \subset \mathbb{E}[\partial_x v(x^*,\xi)] + \mathcal{N}_{\mathcal{X}}(x^*)\\
&= \mathbb{E}[\nabla_x f(x^*,y(x^*,\xi),\xi) + \nabla_y f(x^*,y(x^*,\xi),\xi)\partial y(x^*,\xi)] + \mathcal{N}_{\mathcal{X}}(x^*)\\
&\subset \mathbb{E}[\nabla_x f(x^*,y(x^*,\xi),\xi) - \nabla_y f(x^*,y(x^*,\xi),\xi)\text{conv}([\pi_y \partial H(x^*,y(x^*,\xi),\xi)]^{-1})\\
&\quad \times \pi_x \partial H(x^*,y(x^*,\xi),\xi)] + \mathcal{N}_{\mathcal{X}}(x^*)\\
&= \mathbb{E}[\nabla_x f(x^*,y(x^*,\xi),\xi) + \Lambda(x^*,\xi)\pi_x \partial H(x^*,y(x^*,\xi),\xi)] + \mathcal{N}_{\mathcal{X}}(x^*). \quad (13)
\end{aligned}$$

The inclusion $\partial \mathbb{E}[v(x^*,\xi)] \subset \mathbb{E}[\partial_x v(x^*,\xi)]$ is deduced from the fact that the Clarke generalized directional derivative of $\mathbb{E}[v(x,\xi)]$ is bounded by the expected value of the Clarke generalized directional derivative of $v(x,\xi)$. See [33, Proposition 2.12]. The conclusion follows. ∎

Remark 3.6. In the case when ∂H is singular, if

$$\nabla_y f(x^*,y(x^*,\xi),\xi)\partial y(x^*,\xi) \subset \Lambda(x^*,\xi)\pi_x \partial H(x^*,y(x^*,\xi),\xi), \tag{14}$$

then we can draw a similar conclusion.

Note that if $v(x,\xi)$ is regular at x^* in the sense of Clarke [10, Definition 2.3.4] and $\|\partial_x v(x,\xi)\|$ is bounded by some integrable function $\eta(\xi)$, then by [15, Proposition 5.1], $\partial \mathbb{E}[v(x^*,\xi)] = \mathbb{E}[\partial_x v(x^*,\xi)]$; consequently, equality holds in the first inclusion of (13). This implies that the set of stationary points satisfying $0 \in \partial \mathbb{E}[v(x^*,\xi)] + \mathcal{N}_{\mathcal{X}}(x^*)$ coincides with the set of stationary points satisfying $0 \in \mathbb{E}[\partial_x v(x^*,\xi)] + \mathcal{N}_{\mathcal{X}}(x^*)$. Our discussions above show that all these stationary points are contained in the set of the generalized stationary points satisfying (11) under some appropriate conditions, which means the latter gives a bound or an estimate of the former. To see how precise the estimate is, we need to look at the second inclusion in (13) or the inclusion in (14). The former relies on the index consistency of the piecewise smooth function H

in y at the considered point [21]. The latter depends on the structure of Λ and $\pi_x \partial H$. In general the inclusions are strict but perhaps not very loose. See [21, section 5] for the comparisons of various GKKT conditions for deterministic nonsmooth equality constrained minimization problems.

3.2. GKKT conditions of the regularized problem. We now consider the GKKT conditions of the regularized program (6). Throughout this subsection and section 4, we make the following assumption.

Assumption 3.7. Let R be a regularization of H. $\pi_y \partial R(x, \tilde{y}(x, \xi, \mu), \xi, \mu)$ is nonsingular for $x \in \mathcal{X}$, $\xi \in \Xi$, $\mu \in (0, \mu_0)$.

This assumption is rather moderate and is satisfied by many regularizations. See section 5 for a detailed discussion. Let us define the mapping of multipliers of the regularized problem.

DEFINITION 3.8. *For* $(x, \xi, \mu) \in \mathcal{X} \times \Xi \times (0, \mu_0)$*, let*

$$\begin{aligned} \Lambda^{\text{reg}}(x, \xi, \mu) := &\text{conv}\{\lambda(x, \xi) \in \mathbb{R}^n \mid 0 \in \nabla_y f(x, \tilde{y}(x, \xi, \mu), \xi) \\ &+ \lambda(x, \xi) \pi_y \partial R(x, \tilde{y}(x, \xi, \mu), \xi, \mu)\}. \end{aligned} \tag{15}$$

Since $\pi_y \partial R(x, \tilde{y}(x, \xi, \mu), \xi, \mu)$ is nonsingular, then Λ^{reg} can be rewritten as

$$\Lambda^{\text{reg}}(x, \xi, \mu) = -\nabla_y f(x, \tilde{y}(x, \xi, \mu), \xi) \text{conv}([\pi_y \partial R(x, \tilde{y}(x, \xi, \mu), \xi, \mu)]^{-1}). \tag{16}$$

Obviously, Λ^{reg} contains the set of Lagrange multipliers of the trivial second stage regularized problem:

$$\min_y f(x, y, \xi) \ \text{ s.t. } \ R(x, y, \xi, \mu) = 0,$$

since $\tilde{y}(x, \xi, \mu)$ is the unique feasible solution. Using the notion of Λ^{reg}, we define the stationary point of the regularized problem.

DEFINITION 3.9. *Let* $\Lambda^{\text{reg}}(x, \xi, \mu)$ *be defined as in Definition* 3.8. *A point* $x \in \mathbb{R}^m$ *is called a* generalized stationary point *of the regularized problem* (6) *if*

$$0 \in \mathbb{E}[\nabla_x f(x, \tilde{y}(x, \xi, \mu), \xi) + \Lambda^{\text{reg}}(x, \xi, \mu) \pi_x \partial R(x, \tilde{y}(x, \xi, \mu), \xi, \mu)] + \mathcal{N}_{\mathcal{X}}(x). \tag{17}$$

We call condition (17) *a GKKT condition for the regularized problem* (6). *Note that this definition depends on the function* $\tilde{y}(x, \xi, \mu)$.

Let $\tilde{v}(x, \xi, \mu) := f(x, \tilde{y}(x, \xi, \mu), \xi)$. Obviously, $\tilde{v}(\cdot, \xi, \mu)$ is locally Lipschitz continuous, since $\tilde{y}(\cdot, \xi, \mu)$ is locally Lipschitz continuous by assumption. Note that by Lemma 3.4

$$\partial \tilde{y}(x, \xi, \mu) \subset -\text{conv}([\pi_y \partial R(x, \tilde{y}(x, \xi, \mu), \xi, \mu)]^{-1}) \pi_x \partial R(x, \tilde{y}(x, \xi, \mu), \xi, \mu).$$

If $x^* \in \mathcal{X}$ be a local minimizer of the regularized problem, then under some appropriate measurable conditions (of $\partial_x \tilde{v}$, etc.) we have

$$\begin{aligned} 0 &\in \partial \mathbb{E}[\tilde{v}(x^*, \xi, \mu)] + \mathcal{N}_{\mathcal{X}}(x^*) \subset \mathbb{E}[\partial_x \tilde{v}(x^*, \xi, \mu)] + \mathcal{N}_{\mathcal{X}}(x^*) \\ &= \mathbb{E}[\nabla_x f(x^*, \tilde{y}(x^*, \xi, \mu), \xi) + \nabla_y f(x^*, \tilde{y}(x^*, \xi, \mu), \xi) \partial_x \tilde{y}(x^*, \xi, \mu)] + \mathcal{N}_{\mathcal{X}}(x^*) \\ &\subset \mathbb{E}[\nabla_x f(x^*, \tilde{y}(x^*, \xi, \mu), \xi) - \nabla_y f(x^*, \tilde{y}(x^*, \xi, \mu), \xi) \\ &\quad \times \text{conv}([\pi_y \partial R(x^*, \tilde{y}(x^*, \xi, \mu), \xi, \mu)]^{-1}) \pi_x \partial R(x^*, \tilde{y}(x^*, \xi, \mu), \xi, \mu)] + \mathcal{N}_{\mathcal{X}}(x^*) \\ &= \mathbb{E}[\nabla_x f(x^*, \tilde{y}(x^*, \xi, \mu), \xi) + \Lambda^{\text{reg}}(x^*, \xi) \pi_x \partial R(x^*, \tilde{y}(x^*, \xi, \mu), \xi, \mu)] \\ &\quad + \mathcal{N}_{\mathcal{X}}(x^*), \end{aligned} \tag{18}$$

which implies that an optimal solution of the regularized problem (6) is a generalized stationary point.

3.3. Convergence analysis of the regularized problem. In this subsection, we investigate the convergence of the stationary points of regularized problem as $\mu \to 0$. We first state the following intermediate result.

LEMMA 3.10. *Let R be a regularization of H and $\Lambda^{\rm reg}$ be defined as in Definition* 3.8. *Suppose that there exists a function $\nu_1(\xi) > 0$ such that*

$$\|\Lambda^{\rm reg}(x,\xi,\mu)\| \le \nu_1(\xi) \quad \forall (x,\mu) \in \mathcal{X} \times (0,\mu_0) \tag{19}$$

and that $\mathbb{E}[\nu_1(\xi)] < \infty$. Then $\Lambda^{\rm reg}(\cdot,\xi,\cdot)$ is upper semicontinuous on $\mathcal{X} \times (0,\mu_0)$, and

$$\overline{\lim_{\mu \downarrow 0}} \Lambda^{\rm reg}(x,\xi,\mu) \subset \Lambda(x,\xi), \quad x \in \mathcal{X}. \tag{20}$$

Proof. The upper semicontinuity of $\Lambda^{\rm reg}$ on $\mathcal{X} \times (0,\mu_0)$ follows from (16), the upper semicontinuity of $\pi_y \partial R(x,y,\xi,\mu)$, and $\nabla_y f(x,y,\xi)$ with respect to x, y, μ. In what follows we show (20). By the definition of $\Lambda^{\rm reg}$,

$$-\nabla_y f(x,\tilde{y}(x,\xi,\mu),\xi) \in \Lambda^{\rm reg}(x,\xi,\mu)\pi_y \partial R(x,\tilde{y}(x,\xi,\mu),\xi,\mu). \tag{21}$$

By Definition 2.1(iii),

$$\overline{\lim_{\mu \downarrow 0}} \pi_y \partial R(x,\tilde{y}(x,\xi,\mu),\xi,\mu) \subset \pi_y \partial H(x,y(x,\xi),\xi).$$

Therefore $\pi_y \partial R(x,\tilde{y}(x,\xi,\mu),\xi,\mu)$ is bounded for μ close to 0. This and condition (19) allow us to take an outer limit on both sides of (21)

$$\begin{aligned} -\nabla_y f(x,y(x,\xi),\xi) &\in \overline{\lim_{\mu \downarrow 0}} [\Lambda^{\rm reg}(x,\xi,\mu)\pi_y \partial R(x,\tilde{y}(x,\xi,\mu),\xi,\mu)] \\ &\subset \overline{\lim_{\mu \downarrow 0}} \Lambda^{\rm reg}(x,\xi,\mu)\, \overline{\lim_{\mu \downarrow 0}} \pi_y \partial R(x,\tilde{y}(x,\xi,\mu),\xi,\mu). \end{aligned}$$

Then we arrive at

$$-\nabla_y f(x,y(x,\xi),\xi) \in \left[\overline{\lim_{\mu \downarrow 0}} \Lambda^{\rm reg}(x,\xi,\mu)\right] \pi_y \partial H(x,y(x,\xi),\xi)).$$

The conclusion follows immediately from the definition of Λ. ☐

Note that the boundedness condition (19) is satisfied if $[\pi_y \partial R(x,\tilde{y}(x,\xi,\mu),\xi,\mu)]^{-1}$ is uniformly bounded (see Remark 2.2) and f is uniformly Lipschitz continuous in y. We are now ready to present the main result of this section concerning the convergence of the stationary points of the regularized problem.

THEOREM 3.11. *Suppose that assumptions in Theorem* 2.4 *are satisfied. Suppose also that there exists a function $\kappa_2(\xi)$, where $\mathbb{E}[\kappa_2(\xi)] < \infty$, such that for all $(x,\xi,\mu) \in \mathcal{X} \times \Xi \times (0,\mu_0)$,*

$$\max\{\|\nabla_x f(x,\tilde{y}(x,\xi,\mu),\xi)\|, \|\pi_x \partial R(x,\tilde{y}(x,\xi,\mu),\xi,\mu)\|\} \le \kappa_2(\xi). \tag{22}$$

Let $\{x(\mu)\}$ be a sequence of generalized stationary points of the regularized problem (6). *Assume that x^* is an accumulation point of the sequence as $\mu \to 0$. Suppose that condition* (19) *holds and that $\mathbb{E}[\kappa_2(\xi)(1+\nu_1(\xi))] < \infty$. Then w.p.1 x^* is a generalized stationary point of the true problem* (1), *that is,*

$$0 \in \mathbb{E}[\nabla_x f(x^*,y(x^*,\xi),\xi) + \Lambda(x^*,\xi)\pi_x \partial H(x^*,y(x^*,\xi),\xi)] + \mathcal{N}_{\mathcal{X}}(x^*).$$

Proof. We use the Lebesgue dominated convergence theorem to prove the result. Let

$$\mathcal{K}(x,\xi,\mu) := \nabla_x f(x,\tilde{y}(x,\xi,\mu),\xi) + \Lambda^{\mathrm{reg}}(x,\xi,\mu)\pi_x\partial R(x,\tilde{y}(x,\xi,\mu),\xi,\mu).$$

First note that $x(\mu)$ is a generalized stationary point of (6), that is,

$$0 \in \mathbb{E}[\mathcal{K}(x(\mu),\xi,\mu)] + \mathcal{N}_{\mathcal{X}}(x(\mu)). \tag{23}$$

Note that, by Lemma 3.10, $\overline{\lim}_{\mu\downarrow 0}\Lambda^{\mathrm{reg}}(x,\xi,\mu) \subset \Lambda(x,\xi)$. By (19), $\Lambda^{\mathrm{reg}}(x,\xi,\mu)$ is uniformly dominated by $\nu_1(\xi)$ for μ sufficiently small and x close to x^*. On the other hand, by (22), $\nabla_x f(x,y,\xi)$ and $\pi_x\partial R(x,\tilde{y}(x,\xi,\mu),\xi,\mu)$ are uniformly dominated by $\kappa_2(\xi)$ for μ sufficiently small. Hence, $\mathcal{K}(x,\xi,\mu)$ is uniformly dominated by $\kappa_2(\xi)(1+\nu_1(\xi))$ for μ small enough and x sufficiently close to x^*. Note that $\mathbb{E}[\kappa_2(\xi)(1+\nu_1(\xi))] < \infty$; by the Lebesgue dominated convergence theorem, we then have

$$\begin{aligned}&\overline{\lim_{\mu\downarrow 0}}\,\mathbb{E}[\mathcal{K}(x(\mu),\xi,\mu)] = \mathbb{E}\left[\overline{\lim_{\mu\downarrow 0}}\,\mathcal{K}(x(\mu),\xi,\mu)\right]\\ &= \mathbb{E}\left[\overline{\lim_{\mu\downarrow 0}}\,[\nabla_x f(x(\mu),\tilde{y}(x(\mu),\xi,\mu),\xi) + \Lambda^{\mathrm{reg}}(x(\mu),\xi,\mu)\pi_x\partial R(x(\mu),\tilde{y}(x(\mu),\xi,\mu),\xi,\mu)]\right].\end{aligned}$$

By Theorem 2.4 and Definition 2.1, we have by taking a subsequence if necessary on $\{x(\mu)\}$

$$\overline{\lim_{\mu\downarrow 0}}\,\nabla_x f(x(\mu),\tilde{y}(x(\mu),\xi,\mu),\xi) = \nabla_x f(x^*,y(x^*,\xi),\xi)$$

and

$$\overline{\lim_{\mu\downarrow 0}}\,\pi_x\partial R(x(\mu),\tilde{y}(x(\mu),\xi,\mu),\xi,\mu) \subset \pi_x\partial H(x^*,y(x^*,\xi),\xi).$$

In addition, notice that $\overline{\lim}_{\mu\downarrow 0}\Lambda^{\mathrm{reg}}(x(\mu),\xi,\mu) \subset \Lambda(x^*,\xi)$. Thus, with (23), it yields that

$$0 \in \mathbb{E}[\nabla_x f(x^*,y(x^*,\xi),\xi) + \Lambda(x^*,\xi)\pi_x\partial H(x^*,y(x^*,\xi),\xi)] + \mathcal{N}_{\mathcal{X}}(x^*).$$

This completes the proof. ∎

Note that when $f(x,y,\xi)$ and $H(x,y,\xi)$ are uniformly Lipschitz in x, condition (22) is satisfied, since $\pi_x\partial R(x,\tilde{y}(x,\xi,\mu),\xi,\mu)$ approximates $\pi_x\partial H(x,y(x,\xi),\xi)$ following Definition 2.1(iii).

4. Sample average approximations. In this section, we propose a sample average approximation (SAA) method for solving the regularized program (6). SAA methods have been extensively investigated in SMPECs recently. See, for instance, [17, 30, 32, 35, 36]. Our convergence analysis is similar to that in [36]. However, there are two main differences: (a) $\tilde{y}(x,\xi,\mu)$ is a solution of a regularized equation (2), which may be nonsmooth, while in [36] $\tilde{y}(x,\xi,\mu)$ is an implicit smoothing of $y(x,\xi)$ and is smooth in x; (b) $y(x,\xi)$ is the limit of $\{\tilde{y}(x,\xi,\mu)\}_{\mu\to 0}$ which satisfies (2) but it not necessarily a unique implicit function of (2).

Let $\xi^1,\ldots,\xi^N$ be an independent, identically distributed sample of ξ. We consider the following SAA program:

$$\begin{aligned}\min_{x\in\mathcal{X},y^1,\ldots,y^N}\quad & f_N(x,y^1,\ldots,y^N) := \frac{1}{N}\sum_{i=1}^N f(x,y^i,\xi^i)\\ \text{s.t.}\quad & R(x,y^i,\xi^i,\mu) = 0,\quad i=1,\ldots,N.\end{aligned} \tag{24}$$

Here $\mu > 0$ is a small positive number which may depend on sample size N in practical computation. Problem (24) is essentially a deterministic continuous minimization problem with variables x and $y^1, \ldots, y^N$. It can also be regarded as a two stage stochastic program with finite discrete distribution. Choosing which numerical method for solving (24) depends on the structure and size of the problem. If the problem is of relatively small size, and R is smooth, then many existing nonlinear programming methods may be readily applied to solving the problem. When R is not continuously differentiable, we need to employ those which can deal with nonsmoothness. Bundle methods and aggregate subgradient methods are effective ones.

In the case when the problem size is large, decomposition methods which are popular in dealing with large scale stochastic programs seem to be the choice. Of course, choosing which particular decomposition method also depends on the structure of the problem such as linearity, convexity, separability, and sparsity of the underlying functions. Higle and Sen [13] and Ruszczyński [26] presented a comprehensive discussion and review of various decomposition methods for solving two stage stochastic programs. We refer readers to them and the references therein for the methods.

Note that our model (1) is motivated by SMPECs; hence it might be helpful to explain how (24) is possibly solved when applied to SMPECs. For many practical SMPEC problems such as the stochastic leader-followers problem and capital expansion problem, f is convex in y, whereas $f(x, y(x, \xi), \xi)$ is usually nonconvex in x. Moreover, the feasible set of variable y is governed by a complementarity constraint and is often nonconvex. This means that the feasible set of variable y^i defined by an equality constraint in (24) is nonconvex when $\mu = 0$. However, since we assume $\pi_y R$ is nonsingular for $\mu > 0$, the equation has a unique solution y^i for given x and ξ^i; that is, the feasible set of y^i is a singleton. This implies the minimization with respect to variable y^i is trivial theoretically, albeit not numerically, and this can be achieved by solving an N system of equations simultaneously. Based on these observations, if we can solve (24), then we are likely to obtain a point $(x_N(\mu), y_N^1(\mu), \ldots, y_N^N(\mu))$ with $x_N(\mu)$ being a stationary point, while $y_N^i(\mu)$ is the unique global minimizer which depends on $x_N(\mu)$. Alternatively, we can say that $x_N(\mu)$ is a stationary point of (6).

In what follows, we focus on the Clarke stationary points of (24) given the nonsmooth and nonconvex nature of the problem. Following Hiriart-Urruty [14], we can write down the GKKT conditions of (24) as follows:

$$(25)\qquad \begin{cases} 0 \in \dfrac{1}{N} \displaystyle\sum_{i=1}^{N} \nabla_x f(x, y^i, \xi^i) + \sum_{i=1}^{N} \lambda^i \partial_x R(x, y^i, \xi^i, \mu) + \mathcal{N}_{\mathcal{X}}(x), \\[2ex] 0 \in \dfrac{1}{N} \begin{pmatrix} \nabla_y f(x, y^1, \xi^1) \\ \vdots \\ \nabla_y f(x, y^N, \xi^N) \end{pmatrix} + \begin{pmatrix} \lambda^1 \partial_y R(x, y^1, \xi^1, \mu) \\ \vdots \\ \lambda^N \partial_y R(x, y^N, \xi^N, \mu) \end{pmatrix}, \\[2ex] 0 = R(x, y^i, \xi^i, \mu), \quad i = 1, \ldots, N. \end{cases}$$

Since by assumption for every $(x, \xi) \in \mathcal{X} \times \Xi$, equation $R(x, y, \xi, \mu) = 0$ has a unique solution $\tilde{y}(x, \xi, \mu)$, then y^i in (25) can be expressed as $\tilde{y}(x, \xi^i, \mu)$, $i = 1, \ldots, N$.

Consequently, the above GKKT conditions can be rewritten as

(26)
$$\begin{cases} 0 \in \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \nabla_x f(x, \tilde{y}(x, \xi^i, \mu), \xi^i) + \sum_{i=1}^{N} \lambda^i \partial_x R(x, \tilde{y}(x, \xi^i, \mu), \xi^i, \mu) + \mathcal{N}_{\mathcal{X}}(x), \\ 0 \in \dfrac{1}{N} \begin{pmatrix} \nabla_y f(x, \tilde{y}(x, \xi^1, \mu), \xi^1) \\ \vdots \\ \nabla_y f(x, \tilde{y}(x, \xi^N, \mu), \xi^N) \end{pmatrix} + \begin{pmatrix} \lambda^1 \partial_y R(x, \tilde{y}(x, \xi^1, \mu), \xi^1, \mu) \\ \vdots \\ \lambda^N \partial_y R(x, \tilde{y}(x, \xi^N, \mu), \xi^N, \mu) \end{pmatrix}. \end{cases}$$

Note that by [10, Proposition 2.3.16],

$$\partial_x R(x, y, \xi, \mu) \subset \pi_x \partial R(x, y, \xi, \mu) \quad \text{and} \quad \partial_y R(x, y, \xi, \mu) \subset \pi_y \partial R(x, y, \xi, \mu).$$

In addition, under Assumption 3.7, $\pi_y \partial R(x, y^i, \xi^i, \mu)$ is nonsingular; then we replace λ^i, $i = 1, \ldots, N$, in (25) with

$$-\frac{1}{N} \nabla_y f(x, y^i, \xi^i) \mathrm{conv}([\pi_y \partial R(x, y^i, \xi^i, \mu)]^{-1}), \quad i = 1, \ldots, N.$$

By writing y^i as $\tilde{y}(x, \xi^i, \mu)$, we may consider a weaker GKKT condition than (26) as

$$\begin{aligned} 0 \in & \frac{1}{N} \sum_{i=1}^{N} [\nabla_x f(x, \tilde{y}(x, \xi^i, \mu), \xi^i) \\ & - \nabla_y f(x, \tilde{y}(x, \xi^i, \mu), \xi^i) \mathrm{conv}([\pi_y \partial R(x, \tilde{y}(x, \xi^i, \mu), \xi^i, \mu)]^{-1}) \pi_x \partial R(x, \tilde{y}(x, \xi^i, \mu), \xi^i, \mu)] \\ & + \mathcal{N}_{\mathcal{X}}(x). \end{aligned}$$

The "weaker" is in the sense that a point x satisfying (26) must satisfy the above equation but not vice versa. Let $\Lambda^{\mathrm{reg}}(x, \xi, \mu)$ be defined as in (16). Then the above equation can be written as

$$\begin{aligned} 0 \in & \frac{1}{N} \sum_{i=1}^{N} \left[\nabla_x f(x, \tilde{y}(x, \xi^i, \mu), \xi^i) + \Lambda^{\mathrm{reg}}(x, \xi^i, \mu) \pi_x \partial R(x, \tilde{y}(x, \xi^i, \mu), \xi^i, \mu)\right] \\ & + \mathcal{N}_{\mathcal{X}}(x). \end{aligned} \tag{27}$$

We say that a point $x \in \mathcal{X}$ is a *generalized stationary point* of the reduced regularized SAA problem (24) if it satisfies (27). In what follows, we investigate the convergence of the generalized stationary points as the sample size tends to infinity. We consider two cases: (a) μ is set small and fixed, and the sample size N tends to infinity; (b) μ depends on the sample size and is reduced to zero as N increases to infinity.

We establish the following theorem, which states the convergence results of generalized stationary points.

THEOREM 4.1. *Let assumptions in Theorem* 3.11 *hold,* $\kappa_3(\xi) := \max(\nu_1(\xi), \kappa_2(\xi))$, *and* $\mathbb{E}[\kappa_3(\xi)(1 + \kappa_3(\xi))] < \infty$. *Then the following statements hold:*

(i) *Let* $\mu > 0$ *be fixed. If* $\{x_N(\mu)\}$ *is a sequence of generalized stationary points which satisfy* (27), *then w.p.*1 *an accumulation point of the sequence is a generalized stationary point of the regularized problem* (6); *that is,*

$$0 \in \mathbb{E}\left[\mathcal{G}(x, \xi, \mu)\right] + \mathcal{N}_{\mathcal{X}}(x),$$

where $\mathcal{G}(x, \xi, \mu) := \nabla_x f(x, \tilde{y}(x, \xi, \mu), \xi) + \Lambda^{\mathrm{reg}}(x, \xi, \mu) \pi_x \partial R(x, \tilde{y}(x, \xi, \mu), \xi, \mu)$.

(ii) *Let* $\mu = \mu_N$, *where* $\mu_N \to 0$ *as* $N \to \infty$, *and* $\{x(\mu_N)\}$ *be a sequence of generalized stationary points which satisfy* (27). *Suppose that* $\|\pi_x \partial H(x, y(x,\xi), \xi)\|$ *is also bounded by* $\kappa_2(\xi)$ *in* (22). *If* x^* *is an accumulation point of* $\{x(\mu_N)\}$, *then w.p.1* x^* *is a generalized stationary point of the true problem* (1); *that is,* x^* *satisfies*

$$0 \in \mathbb{E}[\mathcal{L}(x,\xi)] + \mathcal{N}_{\mathcal{X}}(x), \tag{28}$$

where $\mathcal{L}(x,\xi) := \nabla_x f(x, y(x,\xi), \xi) + \Lambda(x,\xi)\pi_x \partial H(x, y(x,\xi), \xi)$, Λ, *and* $y(x,\xi)$ *are as given in section* 3.

Proof. Part (i). By assumption, there exists a unique $\tilde{y}(x,\xi,\mu)$ such that

$$R(x, \tilde{y}(x,\xi,\mu), \xi, \mu) = 0$$

for every $(x,\xi,\mu) \in \mathcal{X} \times \Xi \times (0, \mu_0)$. Since $\partial R(\cdot, \tilde{y}(\cdot,\xi,\mu), \xi, \mu)$ is an upper semicontinuous, compact set-valued mapping, then $\mathcal{G}(\cdot,\cdot,\mu)$ is also an upper semicontinuous and compact set-valued mapping on $\mathcal{X}$ for every $\xi \in \Xi$. It follows from (19) and (22) that $\mathcal{G}(x,\xi,\mu)$ is uniformly dominated by $\kappa_3(\xi)(1+\kappa_3(\xi))$, which is integrable by assumption. Assume without loss of generality that $\{x_N(\mu)\} \to \{x^*\}$. Since $x_N(\mu)$ is a generalized stationary point of problem (24), we have by definition

$$0 \in \frac{1}{N}\sum_{i=1}^{N} \mathcal{G}(x_N(\mu), \xi^i, \mu) + \mathcal{N}_{\mathcal{X}}(x_N(\mu)). \tag{29}$$

For any sufficiently small $\delta > 0, \gamma > 0$, we estimate

$$\mathbb{D}\left(\frac{1}{N}\sum_{i=1}^{N} \mathcal{G}(x_N(\mu), \xi^i, \mu), \quad \mathbb{E}[\mathcal{G}_\delta(x^*, \xi, \mu)] + \gamma\mathcal{B}\right),$$

where $\mathcal{G}_\delta(x^*,\xi,\mu) := \bigcup_{x \in \mathcal{B}(x^*,\delta)} \mathcal{G}(x,\xi,\mu)$ and $\mathbb{E}[\mathcal{G}_\delta(x^*,\xi,\mu)] = \bigcup_{G \in \mathcal{G}_\delta(x^*,\xi,\mu)} \mathbb{E}[G]$. Note that

$$\begin{aligned}
&\mathbb{D}\left(\frac{1}{N}\sum_{i=1}^{N} \mathcal{G}(x_N(\mu), \xi^i, \mu), \quad \mathbb{E}[\mathcal{G}_\delta(x^*, \xi, \mu)] + \gamma\mathcal{B}\right)\\
&\le \mathbb{D}\left(\frac{1}{N}\sum_{i=1}^{N} \mathcal{G}(x_N(\mu), \xi^i, \mu), \quad \frac{1}{N}\sum_{i=1}^{N} \mathcal{G}_{\delta/2}(x^*, \xi^i, \mu) + \gamma/2\mathcal{B}\right)\\
&+ \mathbb{D}\left(\frac{1}{N}\sum_{i=1}^{N} \mathcal{G}_{\delta/2}(x^*, \xi^i, \mu) + \gamma/2\mathcal{B}, \quad \mathbb{E}[\mathcal{G}_\delta(x^*, \xi, \mu)] + \gamma\mathcal{B}\right).
\end{aligned}$$

By Lemma 3.2 of [36], the second term on the right-hand side of the equation tends to zero w.p.1 as $N \to \infty$. On the other hand, for N large enough such that $x_N(\mu) \in \mathcal{B}(x^*, \delta)$, by definition $\mathcal{G}(x_N(\mu), \xi^i, \mu) \subset \mathcal{G}_\delta(x^*, \xi^i, \mu)$, which leads to

$$\mathbb{D}\left(\frac{1}{N}\sum_{i=1}^{N} \mathcal{G}(x_N(\mu), \xi^i, \mu), \quad \frac{1}{N}\sum_{i=1}^{N} \mathcal{G}_\delta(x^*, \xi^i, \mu) + \gamma\mathcal{B}\right) = 0$$

for N sufficiently large. Hence, by (29), it follows that

$$0 \in \mathbb{E}[\mathcal{G}_\delta(x^*, \xi, \mu)] + \mathcal{N}_{\mathcal{X}}(x^*) + \gamma\mathcal{B} \text{ w.p.1.}$$

By the Lebesgue dominated convergence theorem and noticing the arbitrariness of δ and γ, we get the desired conclusion.

Part (ii). We now treat μ in $\mathcal{G}(x,\xi,\mu)$ as a variable and define

$$\hat{\mathcal{G}}(x,\xi,\mu) := \begin{cases} \mathcal{G}(x,\xi,\mu), & \mu > 0, \\ \mathcal{A}(x,\xi), & \mu = 0, \end{cases}$$

where $\mathcal{A}(x,\xi) := \nabla_x f(x, y(x,\xi), \xi) + \overline{\lim}_{\mu\downarrow 0} \Lambda^{\mathrm{reg}}(x,\xi,\mu)\pi_x \partial H(x, y(x,\xi),\xi)$. By assumption, it follows that $\hat{\mathcal{G}}(\cdot,\xi,\cdot) : \mathcal{X} \times [0,\mu_0) \to 2^{\mathbb{R}^n}$ is an upper semicontinuous and compact set-valued mapping for every $\xi \in \Xi$. Conditions (19) and (22) and the bound $\kappa_2(\xi)$ on $\pi_x \partial H(x, y(x,\xi),\xi)$ imply that $\hat{\mathcal{G}}(x,\xi,\mu)$ is bounded by $\kappa_3(\xi)(1+\kappa_3(\xi))$, which is integrable by assumption. Since $x(\mu_N)$ is a generalized stationary point of problem (24) with $\mu = \mu_N$, it follows that

$$0 \in \frac{1}{N}\sum_{i=1}^{N} \hat{\mathcal{G}}(x(\mu_N), \xi^i, \mu_N) + \mathcal{N}_{\mathcal{X}}(x(\mu_N)). \tag{30}$$

Assume without loss of generality that $x(\mu_N) \to x^*$ as $N \to \infty$. For any small $\delta > 0$ and $\gamma > 0$, we will show that

$$\mathbb{D}\left(\frac{1}{N}\sum_{i=1}^{N} \hat{\mathcal{G}}(x(\mu_N), \xi^i, \mu_N), \quad \mathbb{E}\left[\mathcal{L}_\delta(x^*,\xi)\right] + \gamma\mathcal{B}\right) \to 0, \quad \text{w.p.1 as } N \to \infty,$$

where $\mathcal{L}_\delta(x^*,\xi) = \hat{\mathcal{G}}_\delta(x^*,\xi,0), \hat{\mathcal{G}}_\delta(x,\xi,\mu) = \bigcup_{(x',\mu')\in\mathcal{B}(x,\delta)\times[0,\delta]} \mathcal{G}(x',\xi,\mu')$. Note that

$$\begin{aligned}
&\mathbb{D}\left(\frac{1}{N}\sum_{i=1}^{N} \hat{\mathcal{G}}(x(\mu_N), \xi^i, \mu_N), \quad \mathbb{E}\left[\mathcal{L}_\delta(x^*,\xi)\right] + \gamma\mathcal{B}\right) \\
&\leq \mathbb{D}\left(\frac{1}{N}\sum_{i=1}^{N} \hat{\mathcal{G}}(x(\mu_N), \xi^i, \mu_N), \quad \frac{1}{N}\sum_{i=1}^{N} \hat{\mathcal{G}}_{\delta/2}(x^*,\xi^i,0) + \gamma/2\mathcal{B}\right) \\
&+ \mathbb{D}\left(\frac{1}{N}\sum_{i=1}^{N} \hat{\mathcal{G}}_{\delta/2}(x^*,\xi^i,0) + \gamma/2\mathcal{B}, \quad \mathbb{E}\left[\mathcal{L}_\delta(x^*,\xi)\right] + \gamma\mathcal{B}\right).
\end{aligned}$$

By Lemma 3.2 of [36], the second term on the right-hand side of the equation above tends to zero as $N \to \infty$. On the other hand, since $\hat{\mathcal{G}}(x(\mu_N), \xi^i, \mu_N) \subset \hat{\mathcal{G}}_\delta(x^*, \xi^i, 0)$ for any $(x(\mu_N), \mu_N) \in \mathcal{B}(x^*,\delta) \times [0,\delta]$, hence the first term on the right-hand side equals zero for N sufficiently large. Since γ and δ are arbitrarily small, thereby, the conclusion follows immediately by virtue of the Lebesgue dominated convergence theorem and (30). The proof is completed. ∎

Theorem 4.1 states that if μ is fixed, then w.p.1. an accumulation point of a sequence of the generalized stationary points of the regularized SAA problem (24) is a generalized stationary point of the regularized problem (6). In the case that μ depends on sample size N and is reduced to zero as $N \to \infty$, then w.p.1. an accumulation point of a sequence of the generalized stationary points of the regularized SAA problem (24) is a generalized stationary point of the true problem.

5. Applications to SMPECs.

5.1. Stochastic program with variational constraints. In this section, we apply the results established in the preceding sections to the following stochastic mathematical programs with boxed constrained variational inequality (BVI) constraints:

$$\min_{x \in \mathcal{X}} \quad \mathbb{E}\left[f(x, y(x,\xi), \xi)\right], \tag{31}$$

where $y(x,\xi)$ is a measurable solution to the VI problem

$$F(x,y,\xi)^T(z-y) \geq 0 \quad \forall z \in \Upsilon, \tag{32}$$

where $\mathcal{X}$ is a nonempty compact subset of $\mathbb{R}^m$, $f : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$ is continuously differentiable, $\xi : \Omega \to \Xi \subset \mathbb{R}^k$ is a vector of random variables defined on probability space $(\Omega, \mathcal{F}, P)$ with nonatomic P, $F : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$ is continuously differentiable, $F(x,\cdot,\xi)$ is a P_0-function for every $(x,\xi) \in \mathcal{X} \times \Xi$, $\Upsilon := \{y \in \mathbb{R}^n \mid a \leq y \leq b\}$, $a \in \{\mathbb{R} \cup \{-\infty\}\}^n$, $b \in \{\mathbb{R} \cup \{\infty\}\}^n$, and $a < b$ (componentwise). Here we assume that Ξ is a compact set. Notice that if we set $a = 0$ and $b = \infty$, then problem (31) is reduced to the stochastic mathematical programs with complementarity constraints. For simplicity in analysis, we assume all components in a or b are finite or infinite simultaneously. In other words, we will focus on the following cases: (i) $a \in \mathbb{R}^n$ and $b \in \mathbb{R}^n$; (ii) $a \in \mathbb{R}^n$, $b = \infty$; (iii) $a = -\infty$, $b \in \mathbb{R}^n$; (iv) $a = -\infty$, $b = \infty$.

For every $(x,\xi) \in \mathcal{X} \times \Xi$, the constraint of the second stage problem (32) is actually a parametric BVI problem. Throughout this section, we assume that the BVI has at least one solution for every $(x,\xi) \in \mathcal{X} \times \Xi$.

Let $\Pi_\Upsilon(y)$ be the Euclidean projection of y onto Υ. Then the parametric BVI can be reformulated as a parameterized normal equation

$$H(x,y,\xi) := F(x, \Pi_\Upsilon(y), \xi) + y - \Pi_\Upsilon(y) = 0, \quad (x,\xi) \in \mathcal{X} \times \Xi, \tag{33}$$

in the sense that if $y(x,\xi)$ is a solution of (33), then $\bar{y}(x,\xi) := \Pi_\Upsilon(y(x,\xi))$ is a solution of the BVI problem, and conversely, if $\bar{y}(x,\xi)$ is a solution of the BVI, then $y(x,\xi) := \bar{y}(x,\xi) - F(x, \bar{y}(x,\xi), \xi)$ is a solution of (33). Consequently, we can reformulate (31) as

$$\min_{x \in \mathcal{X}} \quad \mathbb{E}\left[f(x, \Pi_\Upsilon(y(x,\xi)), \xi)\right], \tag{34}$$

where $y(x,\xi)$ is a measurable solution to the following equation:

$$H(x,y,\xi) = 0. \tag{35}$$

Obviously, H defined in (33) is locally Lipschitz continuous and piecewise smooth with respect to x, y, ξ as Υ is a box. The nonsmoothness of Π_Υ may result in the ill-posedness of (33). Note also that since F is a P_0-function, the parametric BVI may have multiple solutions, and consequently (33) may have multiple solutions for every x and ξ. In what follows, we use a smoothed regularization method to deal with (33). First, we will use the Gabriel–Moré smoothing function to smooth Π_Υ [12], and consequently we get a smooth approximation of (33)

$$\hat{R}(x,v,y,\xi) := F(x, p(v,y), \xi) + y - p(v,y) = 0, \quad (x,\xi) \in \mathcal{X} \times \Xi. \tag{36}$$

Here $p : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable, except at the point $(v,y) \in \mathbb{R}^n \times \mathbb{R}^n$, such that $v_i = 0$ for some $i \in \{1, 2, \ldots, n\}$, and for any $(v,y) \in \mathbb{R}^n \times \mathbb{R}^n$, $p(v,y) \in \Upsilon$. We set $v_i = \mu$ and apply the well-known *Tikhonov* regularization to $\hat{R}$

$$R(x,y,\xi,\mu) := \hat{R}(x, \mu e, y, \xi) + \mu y = F(x, p(\mu e, y), \xi) + y - p(\mu e, y) + \mu y, \tag{37}$$

where $e = (1, 1, \ldots, 1) \in \mathbb{R}^n$ and μ is a positive parameter.

In the following analysis, we use the well-known Chen–Harker–Kanzow–Smale (CHKS) smoothing function [7, 16] to smooth the components of $\Pi_\Upsilon(y)$:

$$\phi(\alpha, c, d, \beta) = (c + \sqrt{(c-\beta)^2 + 4\alpha^2})/2 + (d - \sqrt{(d-\beta)^2 + 4\alpha^2})/2,$$

where $(\alpha, \beta) \in \mathbb{R}_{++} \times \mathbb{R}$ and $(c, d) \in \mathbb{R} \times \mathbb{R}$. Then

$$\begin{aligned} p_i(\mu e, y) &:= \phi(\mu, a_i, b_i, y_i) \\ &= (a_i + \sqrt{(a_i - y_i)^2 + 4\mu^2})/2 + (b_i - \sqrt{(b_i - y_i)^2 + 4\mu^2})/2, \quad i = 1, \ldots, n. \end{aligned}$$

For any $\mu \in \mathbb{R}_{++}$ and $(x, \xi) \in \mathcal{X} \times \Xi$, since $p(\mu e, y)$ is continuously differentiable with respect to y, it follows that $R(x, y, \xi, \mu)$ is continuously differentiable with respect to x, y for almost every ξ. In what follows, we will verify that R defined in (37) satisfies Definition 2.1 for regularization functions. We first state the following result.

LEMMA 5.1. *Given $(x, \xi) \in \mathcal{X} \times \Xi$ and $\mu > 0$, let R be defined as in* (37). *Then the Jacobian $\nabla_y R(x, y, \xi, \mu)$ is nonsingular.*

Proof. First, we have

$$\nabla_y R(x, y, \xi, \mu) = \nabla_y F(x, p(\mu e, y), \xi) D(\mu, y) + \mu I + I - D(\mu, y),$$

where $D(\mu, y) = \mathrm{diag}(d_1(\mu, y), \ldots, d_n(\mu, y))$ and $d_i(\mu, y) = \partial p_i(\mu e, y)/\partial y_i \in [0, 1]$ for every $i \in \{1, \ldots, n\}$. Since $F(x, y, \xi)$ is a P_0-function with respect to y and $p(\mu e, y) \in \Upsilon$, then $\nabla_y F(x, p(\mu e, y), \xi)$ is a P_0-matrix. For $u \in \mathbb{R}^n$, let

$$[\nabla_y F(x, p(\mu e, y), \xi) D(\mu, y) + \mu I + I - D(\mu, y)]u = 0.$$

We claim that $D(\mu, y)u = 0$. Assume that $[D(\mu, y)u]_i \neq 0$ for any i; we then have

$$\begin{aligned} [D(\mu, y)u]_i [\nabla_y F(x, p(\mu e, y), \xi) D(\mu, y) u]_i &= -[D(\mu, y)u]_i [(\mu + 1)u - D(\mu, y)u]_i \\ &= -(\mu + 1) d_i(\mu, y) u_i^2 + d_i^2(\mu, y) u_i^2 < 0, \end{aligned}$$

which contradicts the definition of P_0-matrix of $\nabla_y F(x, p(\mu e, y), \xi)$. So, $D(\mu, y)u = 0$. Hence, $(\mu + 1)u = 0$, which derives $u = 0$. Thus, $\nabla_y R(x, y, \xi, \mu)$ is nonsingular. ◻

LEMMA 5.2. *Let $\{x^k\}$, $\{y^k\}$, $\{\xi^k\}$ be sequences in $\mathcal{X}$, $\mathbb{R}^n$, Ξ and let $\{\mu^k\}$ be a sequence in any closed subset of* $(0, \mu_0)$ *with $\{\|y^k\|\} \to \infty$ as $k \to \infty$, where μ_0 is a small positive number. Then $\|R(x^k, y^k, \xi^k, \mu^k)\| \to \infty$ as $k \to \infty$.*

See a detailed proof in the appendix.

By Lemmas 5.1 and 5.2, we are ready to show that function R constructed in (37) is a regularization of H.

PROPOSITION 5.3. *Let μ_0 be a small positive number. Function R defined in* (37) *is a regularization of H as defined in Definition* 2.1. *Moreover, $\tilde{y}$ is continuously differentiable on $\mathcal{X} \times \Xi \times (0, \mu_0)$.*

The proof is long. We move it to the appendix.

Based on the above discussions, we can convert the true problem (31) (or equivalently, (34)) to the following regularized program:

$$\min_{x \in \mathcal{X}} \quad \mathbb{E}\left[f(x, \Pi_\Upsilon(\tilde{y}(x, \xi, \mu)), \xi)\right], \tag{38}$$

where $\tilde{y}(x, \xi, \mu)$ uniquely solves $R(x, y, \xi, \mu) = 0$, $(x, \xi, \mu) \in \mathcal{X} \times \Xi \times (0, \mu_0)$.

In the next subsection, we will investigate a numerical method for solving (38) by using its SAA.

5.2. SAA program. In this subsection, we consider the SAA program of (38)

$$\begin{aligned} \min_{x\in\mathcal{X},\, y^1,\ldots,y^N} \quad & \frac{1}{N}\sum_{i=1}^{N} f(x,\Pi_\Upsilon(y^i),\xi^i) \\ \text{s.t.} \quad & R(x,y^i,\xi^i,\mu)=0, \quad i=1,\ldots,N, \end{aligned} \tag{39}$$

where μ is a small positive number. Analogous to the discussion in section 4, we can write down the GKKT conditions of (39) as follows:

$$\begin{cases} 0 \in \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} \nabla_x f(x,\Pi_\Upsilon(y^i),\xi^i) + \sum_{i=1}^{N}\lambda^i \nabla_x R(x,y^i,\xi^i,\mu) + \mathcal{N}_\mathcal{X}(x), \\ 0 \in \dfrac{1}{N}\begin{pmatrix} \nabla_y f(x,\Pi_\Upsilon(y^1),\xi^1)\partial\Pi_\Upsilon(y^1) \\ \vdots \\ \nabla_y f(x,\Pi_\Upsilon(y^N),\xi^N)\partial\Pi_\Upsilon(y^N) \end{pmatrix} + \begin{pmatrix} \lambda^1\nabla_y R(x,y^1,\xi^1,\mu) \\ \vdots \\ \lambda^N\nabla_y R(x,y^N,\xi^N,\mu) \end{pmatrix}, \\ 0 = R(x,y^i,\xi^i,\mu), \quad i=1,\ldots,N. \end{cases} \tag{40}$$

Following similar arguments as in section 4, we derive

$$\begin{aligned} 0 \in{}& \frac{1}{N}\sum_{i=1}^{N}\left[\nabla_x f(x,\Pi_\Upsilon(\tilde{y}(x,\xi^i,\mu)),\xi^i) + \Lambda^{\text{reg}}(x,\xi^i,\mu)\nabla_x R(x,\tilde{y}(x,\xi^i,\mu),\xi^i,\mu)\right] \\ &+\mathcal{N}_\mathcal{X}(x), \end{aligned} \tag{41}$$

where

$$\Lambda^{\text{reg}}(x,\xi^i,\mu) = -\nabla_y f(x,\Pi_\Upsilon(\tilde{y}(x,\xi^i,\mu)),\xi^i)\partial\Pi_\Upsilon(\tilde{y}(x,\xi^i,\mu))\nabla_y R(x,\tilde{y}(x,\xi^i,\mu),\xi^i,\mu)^{-1}.$$

Note that for any $(x,y,\xi)\in\mathcal{X}\times\mathbb{R}^n\times\Xi$ and $\mu>0$,

$$\nabla_x H(x,y,\xi) = \nabla_x R(x,y,\xi,\mu) = \nabla_x F(x,\Pi_\Upsilon(y),\xi)$$

and

$$\partial\Pi_\Upsilon(y) \subset \{M\in\mathbb{R}^{n\times n} \mid M=\operatorname{diag}(d_1,\ldots,d_n),\ \ d_i\in[0,1]\}.$$

Obviously, $\partial\Pi_\Upsilon(y)$ is bounded for any $y\in\mathbb{R}^n$. Assume that $\lim_{\mu\downarrow 0}\tilde{y}(x,\xi,\mu)$ exists. Let $y(x,\xi)=\lim_{\mu\downarrow 0}\tilde{y}(x,\xi,\mu)$ on $\mathcal{X}\times\Xi$, and for $(x,\xi)\in\mathcal{X}\times\Xi$

$$\begin{aligned} \Lambda(x,\xi) := \operatorname{conv}\{\lambda(x,\xi)\in\mathbb{R}^n \mid 0\in{}& \nabla_y f(x,\Pi_\Upsilon(y(x,\xi)),\xi)\partial\Pi_\Upsilon(y(x,\xi)) \\ &+\lambda(x,\xi)\pi_y\partial H(x,y(x,\xi),\xi)\}. \end{aligned}$$

Following a similar argument as in Theorem 4.1, we derive convergence results for (39) below.

THEOREM 5.4. *Suppose that there exist a function $\kappa_4(\xi)$ and a constant $\mu_0>0$ such that for all $(x,\xi,\mu)\in\mathcal{X}\times\Xi\times(0,\mu_0)$*

$$\begin{aligned} &\max\{\|\nabla_x f(x,\Pi_\Upsilon(\tilde{y}(x,\xi,\mu),\xi))\|, \|\nabla_x F(x,\Pi_\Upsilon(\tilde{y}(x,\xi,\mu)),\xi)\|, \|\Lambda^{\text{reg}}(x,\xi,\mu)\|\} \\ &\quad\le \kappa_4(\xi) \end{aligned} \tag{42}$$

with $\mathbb{E}[\kappa_4(\xi)]<\infty$. Suppose that $\mathbb{E}[\kappa_4(\xi)(1+\kappa_4(\xi))]<\infty$. Then

(i) *for fixed* $\mu > 0$, *w.p.*1 *an accumulation point of the sequence of the generalized stationary points* $\{x_N(\mu)\}$ *of* (39) *satisfies*

$$0 \in \mathbb{E}\left[\hbar(x,\xi,\mu)\right] + \mathcal{N}_{\mathcal{X}}(x),$$

where

$$\hbar(x,\xi,\mu) := \nabla_x f(x, \Pi_\Upsilon(\tilde{y}(x,\xi,\mu)),\xi) + \Lambda^{\mathrm{reg}}(x,\xi,\mu)\nabla_x F(x, \Pi_\Upsilon(\tilde{y}(x,\xi,\mu)),\xi);$$

(ii) *if* $\mu = \mu_N$, *where* $\mu_N \to 0$ *as* $N \to \infty$, *and* $\{x(\mu_N)\}$ *is a sequence of generalized stationary points of* (39), *then w.p.*1 *an accumulation point of* $\{x(\mu_N)\}$ *satisfies*

$$0 \in \mathbb{E}[\mathcal{M}(x,\xi)] + \mathcal{N}_{\mathcal{X}}(x), \tag{43}$$

where

$$\mathcal{M}(x,\xi) := \nabla_x f(x, \Pi_\Upsilon(y(x,\xi)),\xi) + \Lambda(x,\xi)\nabla_x F(x, \Pi_\Upsilon(y(x,\xi)),\xi).$$

Note that the boundedness condition in (42) on $\|\nabla_x f(x, \Pi_\Upsilon(\tilde{y}(x,\xi,\mu)),\xi)\|$ and $\|\nabla_x F(x, \Pi_\Upsilon(\tilde{y}(x,\xi,\mu)),\xi)\|$ is satisfied if f and F are uniformly globally Lipschitz with respect to x. The boundedness condition on $\Lambda^{\mathrm{reg}}(x,\xi,\mu)$ is satisfied if $f(x,y,\xi)$ is uniformly globally Lipschitz with respect to y and $\pi_y \partial H(x, y(x,\xi),\xi)$ is uniformly nonsingular. In particular, if $H(x,y,\xi)$ is regular in the sense of [20] in y at $y(x,\xi)$, then $\pi_y \partial H(x, y(x,\xi),\xi)$ is nonsingular. See [20] for a detailed discussion in this regard.

Example 5.5. Consider the following stochastic mathematical program:

$$\min_{x\in\mathcal{X}} \quad \mathbb{E}\left[f(x, y(x,\xi),\xi)\right]. \tag{44}$$

Here $f : \mathbb{R} \times \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$ is given as

$$f(x,y,\xi) = 2(y_1 - \arctan y_1) + 4y_1^4 y_2/(1+y_1^2)^2 + 1 + x + \xi,$$

and $y(x,\xi)$ is any measurable solution of the following BVI problem:

$$F(x, y(x,\xi),\xi)^T(z - y(x,\xi)) \geq 0 \quad \forall z \in \Upsilon, \tag{45}$$

where $\Upsilon = \mathbb{R}^2_+$, $\mathcal{X} = [0,1]$, ξ can be any random variable that can take values on the interval $\Xi := [-1, -1/4]$, and $F(x,y,\xi) = (0, y_1 + y_2 + x + \xi - 1)^T$. Evidently, F is continuously differentiable and $F(x,\cdot,\xi)$ is a P_0-function for every $(x,\xi) \in \mathcal{X} \times \Xi$ and

$$\nabla_x F(x,y,\xi) = (0,1)^T. \tag{46}$$

Note that f is continuously differentiable on $\mathbb{R} \times \mathbb{R}^2 \times \mathbb{R}$, $f'_x(x,y,\xi) = f'_\xi(x,y,\xi) = 1$, and

$$\nabla_y f(x,y,\xi) = \left(2y_1^2/(1+y_1^2) + 16y_1^3 y_2/(1+y_1^2)^3, \quad 4y_1^4/(1+y_1^2)^2\right). \tag{47}$$

In this example, we have $H(x,y,\xi) = (y_1 - \max\{0, y_1\},\ \max\{0, y_1\} + y_2 + x + \xi - 1)$. It is not hard to obtain the solution set of the VI problem (45) as follows: for $(x,\xi) \in \mathcal{X} \times \Xi$,

$$\mathcal{Y}(x,\xi) := \{y \in \mathbb{R}^2_+ : y_1 \geq 1 - x - \xi,\ y_2 = 0\} \cup \{y \in \mathbb{R}^2_+ : y_1 + y_2 - 1 + x + \xi = 0\}.$$

Obviously, $\mathcal{Y}$ is a set-valued mapping on $\mathcal{X} \times \Xi$.

We now consider the regularization of the VI problem (45), in which given a regularization parameter, we expect to derive a unique solution function on $\mathcal{X} \times \Xi$. Note here that $\Upsilon = \mathbb{R}^2_+$, we have $a_i = 0$, and $b_i = \infty$, $i = 1, 2$. Then we get the ith component of the smoothing function $p_i(\mu e, y) = \phi(\mu, 0, \infty, y_i)$, $i = 1, 2$, $\mu > 0$, where ϕ is the reduced CHKS smoothing NCP function: $\phi(\alpha, 0, \infty, \beta) = (\sqrt{\beta^2 + 4\alpha^2} + \beta)/2, (\alpha, \beta) \in \mathbb{R}_{++} \times \mathbb{R}$. By definition, we have $p_i(\mu e, y) = (\sqrt{y_i^2 + 4\mu^2} + y_i)/2$, $i = 1, 2, \mu > 0$. And

$$R(x, y, \xi, \mu) = \begin{pmatrix} -(\sqrt{y_1^2 + 4\mu^2} + y_1)/2 + (1+\mu) y_1 \\ (\sqrt{y_1^2 + 4\mu^2} + y_1)/2 + (1+\mu) y_2 + x + \xi - 1 \end{pmatrix}.$$

Evidently, R is continuously differentiable on $\mathcal{X} \times \mathbb{R}^2 \times \Xi \times (0, \infty)$.

After some basic manipulations, we derive the unique solution of $R(x, y, \xi, \mu) = 0$ for any $\mu > 0$, $(x, \xi) \in \mathcal{X} \times \Xi$ as follows:

$$\tilde{y}(x, \xi, \mu) = \left(\sqrt{\mu/(\mu+1)}, \quad (1 - x - \xi)/(1+\mu) - \sqrt{\mu/(\mu+1)} \right)^T.$$

Moreover, for any $x \in \mathcal{X}$ and $\mu > 0$, $\|\tilde{y}(x, \xi, \mu)\| \le \kappa_1(\xi)$ with $\mathbb{E}[\kappa_1(\xi)] < \infty$, where $\kappa_1(\xi) = 7 - \xi$, and

$$\|\tilde{y}(x'', \xi, \mu) - \tilde{y}(x', \xi, \mu)\| \le L(\xi) \|x'' - x'\| \qquad \text{for any } x'', x' \in \mathcal{X},$$

where $L(\xi)$ can be taken as any measurable positive function satisfying $1 \le \mathbb{E}[L(\xi)] < \infty$, say, $L(\xi) = 1 - \xi$. Also, $\lim_{\mu \downarrow 0} \tilde{y}(x, \xi, \mu) = y(x, \xi) = (0, 1 - x - \xi) \in \mathcal{Y}(x, \xi)$, $(x, \xi) \in \mathcal{X} \times \Xi$. Thereby, all conditions in Theorem 2.4 are satisfied. Obviously, $y(x, \xi)$ is measurable for every $x \in \mathcal{X}$ and Lipschitz continuous in x.

In addition, by Proposition 5.3, the regularization R defined above satisfies parts (i)–(iv) of Definition 2.1, where μ_0 can be chosen any small positive number.

Next, we investigate the boundedness condition (42). After some simple calculations, we can see that $\tilde{y}(x, \xi, \mu) \in \mathbb{R}^2_{++}$ for any $(x, \xi, \mu) \in \mathcal{X} \times \Xi \times (0, \hat{\mu})$, where $\hat{\mu} = (\sqrt{5} - 2)/4$. So, $\Pi_\Upsilon(\tilde{y}(x, \xi, \mu)) = \tilde{y}(x, \xi, \mu)$ and

$$\partial \Pi_\Upsilon(\tilde{y}(x, \xi, \mu)) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

On the other hand, we have

$$\nabla_y R(x, y, \xi, \mu)^{-1} = \varpi(y, \mu)^{-1} \begin{pmatrix} 1 + \mu & 0 \\ -\dfrac{1}{2}\left(1 + \dfrac{y_1}{\sqrt{y_1^2 + 4\mu^2}}\right) & -\dfrac{1}{2}\left(1 + \dfrac{y_1}{\sqrt{y_1^2 + 4\mu^2}}\right) + \mu \end{pmatrix},$$

where $\varpi(y, \mu) = (1+\mu)[\frac{1}{2}(1 - y_1/\sqrt{y_1^2 + 4\mu^2}) + \mu]$. Then it follows that

$$\nabla_y R(x, \tilde{y}(x, \xi, \mu), \xi, \mu)^{-1} = \frac{1 + 2\mu}{2\mu(1+\mu)^2} \begin{pmatrix} 1 + \mu & 0 \\ -\dfrac{1+\mu}{1+2\mu} & \dfrac{2\mu^2 - 1}{1 + 2\mu} \end{pmatrix}.$$

In addition, we have

$$\begin{aligned}&\nabla_y f(x, \tilde{y}(x,\xi,\mu), \xi)\\&= \frac{2\mu}{2\mu+1}\left((1+8\sqrt{\mu(\mu+1)})\left(\frac{1-x-\xi-\sqrt{\mu(\mu+1)}}{(2\mu+1)^2}\right), \quad \frac{2\mu}{2\mu+1}\right).\end{aligned}$$

Then, for $\mu \in (0,\hat{\mu})$ and $(x,\xi) \in \mathcal{X} \times \Xi$, it follows that

$$\begin{aligned}\Lambda^{\text{reg}}(x,\xi,\mu) &= -\nabla_y f(x, \Pi_\Upsilon(\tilde{y}(x,\xi,\mu)),\xi)\partial\Pi_\Upsilon(\tilde{y}(x,\xi,\mu))\nabla_y R(x,\tilde{y}(x,\xi,\mu),\xi,\mu)^{-1}\\&= -\left(\frac{8\sqrt{\mu}(1-x-\xi)}{\sqrt{1+\mu}(1+2\mu)^2} + \frac{1-6\mu-4\mu^2}{(1+\mu)(1+2\mu)^2}, \quad \frac{2\mu(2\mu^2-1)}{(1+\mu)^2(1+2\mu)^2}\right).\end{aligned}$$

By applying some basic operations, we have $\|\Lambda^{\text{reg}}(x,\xi,\mu)\| < \varrho(\xi)$ for any $(x,\mu) \in \mathcal{X} \times (0,\mu_0)$, where $\varrho(\xi) := \sqrt{2} + 8\sqrt{2\hat{\mu}}(1-\xi)$. Assume that ξ follows a uniform distribution with parameters -1 and $-1/4$, i.e., $\xi \sim U(-1,-1/4)$. Then $\mathbb{E}[\varrho(\xi)] = \sqrt{2} + 13\sqrt{2\hat{\mu}} < \infty$ and $\mathbb{E}[\varrho(\xi)(1+\varrho(\xi))] < \infty$. Hence, we can choose $\kappa_4(\xi) = \varrho(\xi)$ in Theorem 5.4. This, together with (46) and (47), shows that the boundedness condition (42) holds in this example. ∎

Note that in this example, one may ask why the objective function f is not chosen in a simpler form, say, a linear function of y rather than in a complex form as it stands. The answer is that we use this example not only to illustrate how regularization works for this particular SMPEC problem but also to demonstrate how the boundedness conditions (19) in Lemma 3.10 and (42) of Theorem 5.4 can be satisfied. In this particular example, if f is made linear in y, then we are not able to guarantee the boundedness of the set Λ^{reg}, although this does not mean the method will not work.

6. Preliminary numerical results. We have carried out numerical tests on the regularized SAA scheme for stochastic problems with VI constraints. In this section, we report some preliminary numerical results. Such stochastic problems are artificially made by ourselves, since there are few test problems on SMPECs in the literature. The tests are carried out by implementing mathematical programming codes in MATLAB 6.5 installed in a PC with Windows XP operating system. We use the MATLAB built-in solver *fmincon* for solving the regularized SAA problems.

6.1. Estimating the optimal value of the regularized problem. The following methodology of constructing statistical lower and upper bounds was suggested in [28]. Given $\mu > 0$, let $v(\mu)$ denote the optimal value of the regularized problem (38) and $\tilde{v}_N(\mu)$ the optimal value of (39). It is known [28] that $\mathbb{E}[\tilde{v}_N(\mu)] \le v(\mu)$. To estimate the expected value $\mathbb{E}[\tilde{v}_N(\mu)]$, we generate M independent samples of ξ, $\{\xi_j^1,\dots,\xi_j^N\}$, $j = 1,\dots,M$, each of size N. For each sample j, solve the corresponding SAA problem (39), which can be written as

$$\begin{aligned}&\min_{x\in\mathcal{X}\ y^1,\dots,y^N} \quad \frac{1}{N}\sum_{i=1}^N f(x,\Pi_\Upsilon(y^i),\xi_j^i)\\&\qquad\text{s.t.} \quad R(x,y^i,\xi_j^i,\mu) = 0, \quad i = 1,\dots,N.\end{aligned} \tag{48}$$

Let $\tilde{v}_N^j(\mu)$, $j = 1,\dots,M$, denote the corresponding optimal value of problem (48). Compute

$$L_{N,M}(\mu) := \frac{1}{M}\sum_{j=1}^M \tilde{v}_N^j(\mu),$$

which is an unbiased estimate of $\mathbb{E}[\tilde{v}_N(\mu)]$. Then $L_{N,M}(\mu)$ provides a statistical lower bound for $v(\mu)$. An estimate of variance of the estimator $L_{N,M}(\mu)$ can be computed as

$$s_L^2(M;\mu) := \frac{1}{M(M-1)} \sum_{j=1}^{M} \left(\tilde{v}_N^j(\mu) - L_{N,M}(\mu) \right)^2.$$

Let $v(x,\xi,\mu) = f(x, \Pi_\Upsilon(\tilde{y}(x,\xi,\mu)), \xi)$ and $\tilde{\vartheta}(x,\mu) = \mathbb{E}[v(x,\xi,\mu)]$. Then an upper bound for the optimal value $v(\mu)$ can be obtained by the fact that $\tilde{\vartheta}(\bar{x},\mu) \geq v(\mu)$ for any $\bar{x} \in \mathcal{X}$. Hence, by choosing $\bar{x}$ to be a near-optimal solution, for example, by solving one SAA problem and using an unbiased estimator of $\tilde{\vartheta}(\bar{x},\mu)$, we can obtain an estimate of an upper bound for $v(\mu)$. To do so, generate M' independent batches of samples: $\{\xi_j^1, \ldots, \xi_j^{N'}\}$, $j = 1, \ldots, M'$, each of size N'. For $x \in \mathcal{X}$, let $\tilde{v}_{N'}^j(x,\mu) := \frac{1}{N'} \sum_{i=1}^{N'} v(x, \xi_j^i, \mu)$. Then $\mathbb{E}[\tilde{v}_{N'}^j(x,\mu)] = \tilde{\vartheta}(x,\mu)$. Compute

$$U_{N',M'}(\bar{x};\mu) := \frac{1}{M'} \sum_{j=1}^{M'} \tilde{v}_{N'}^j(\bar{x},\mu),$$

which is an unbiased estimate of $\tilde{\vartheta}(\bar{x},\mu)$. So, $U_{N',M'}(\bar{x};\mu)$ is an estimate of an upper bound on $v(\mu)$. An estimate of variance of the estimator $U_{N',M'}(\bar{x},\mu)$ can be computed as

$$s_U^2(\bar{x}, M';\mu) := \frac{1}{M'(M'-1)} \sum_{j=1}^{M'} \left(\tilde{v}_{N'}^j(\bar{x},\mu) - U_{N',M'}(\bar{x},\mu) \right)^2.$$

Note that in this part, for each $j = 1, \ldots, M'$ and $i = 1, \ldots, N'$, we need to solve the following repeated subproblems:

$$\begin{array}{ll} \min & f(\bar{x}, \Pi_\Upsilon(y), \xi_j^i) \\ \text{s.t.} & R(\bar{x}, y, \xi_j^i, \mu) = 0; \end{array}$$

then the corresponding optimal value is $v(\bar{x}, \xi_j^i, \mu)$. Hence, we can obtain $\tilde{v}_{N'}^j(\bar{x},\mu)$, $U_{N',M'}(\bar{x};\mu)$, and $s_U^2(\bar{x}, M';\mu)$. Note that, in practice, we may choose $\bar{x}$ to be any of the solutions of the M regularized SAA problems (48) by generating independent samples $\{\xi_j^1, \ldots, \xi_j^N\}$, $j = 1, \ldots, M$. In fact, we will use $\bar{x}_N^j$, the *best* optimal solution which estimates the smallest optimal value $v(\mu)$, to compute the upper bound estimates, and the optimality gap.

Using the lower bound estimate and the objective function value estimate of the optimal value, $v(\mu)$, of the first stage regularized problem as discussed above, we compute an estimate of the optimality gap of the solution $\bar{x}$ and the corresponding estimated variance as follows:

$$Gap_{N,M,N',M'}(\bar{x}) := U_{N',M'}(\bar{x};\mu) - L_{N,M}(\mu), \quad S_{\text{Gap}}^2 := s_L^2(M;\mu) + s_U^2(\bar{x}, M';\mu).$$

6.2. Preliminary computational results. In the following test problem, we choose different values for the regularization parameter μ and sample sizes N, M, N', and M'. We report the lower and upper bounds, $L_{N,M}$ and $U_{N',M'}$, of $v(\mu)$, the sample variances, s_L, s_U, and the estimate of the optimality gap, Gap, of the solution candidate $\bar{x}_N^j$, the variance of the gap estimator S_{Gap}.

TABLE 1
Summary of lower and upper bounds on $v(\mu)$, the optimality gap.

μ	N	M	N'	M'	$L_{N,M}$	s_L	$\bar{x}_N^j$	$U_{N',M'}$	s_U	*Gap*	S_{Gap}
10^{-3}	200	10	200	10	.7345	.0118	.4928	.7632	.0138	.0287	.0181
10^{-4}	200	10	200	10	.7657	.0142	.5056	.7719	.0150	.0062	.0207
10^{-5}	200	10	200	10	.7749	.0138	.4948	.7841	.0127	.0092	.0188
10^{-3}	300	10	300	10	.7295	.0104	.4837	.7406	.0096	.0111	.0141
10^{-4}	300	10	300	10	.7506	.0018	.4988	.7574	.0118	.0069	.0167
10^{-5}	300	10	300	10	.7668	.0120	.5071	.7727	.0149	.0059	.0191

Example 6.1. Consider the following problem:

$$\begin{array}{ll} \min & \mathbb{E}[x^2 + y_2(x,\xi)^2] \\ \text{s.t.} & 0 \le x \le 1, \end{array} \tag{49}$$

where $y(x,\xi)$ is a solution of the following complementarity problem, which is a special case of VI problems:

$$0 \le F(x,y,\xi) \perp y \ge 0, \qquad F(x,y,\xi) = (0, y_1 + y_2 + x + \xi - 1)^T,$$

where ξ is a random variable with truncated standard normal distribution on $[-1, 1]$. Using the regularization scheme, we can convert the above problem into the following problem:

$$\begin{array}{ll} \min\limits_{x,y} & \mathbb{E}[x^2 + (\max(0, y_2))^2] \\ \text{s.t.} & R(x,y,\xi,\mu) = 0, \quad 0 \le x \le 1, \end{array}$$

where μ is a small positive parameter tending to 0 and $R(x,y,\xi,\mu)$ is given in Example 5.5. Note that the limit of the corresponding unique solution function $\tilde{y}(x,\xi,\mu)$ of $R(x,y,\xi,\mu) = 0$ equals $y(x,\xi) := (0, 1 - x - \xi)$. After basic operations, we can derive the optimal solution of problem (49) associated with $y(x,\xi)$ as $x^* = 0.5$, and the optimal value is $f^* = 0.77454$ (obtained from Maple). The test results are displayed in Table 1.

The results show that both optimal solutions and values of the regularized SAA problems approximate those of the true problem very well as sample size increases and the regularization parameter is driven to zero. More numerical tests are needed to evaluate the performance of the proposed method, but this is beyond the scope of this paper.

Appendix. *Proof of Lemma* 5.2. We first define an index set $\mathcal{I}_0^\infty := \{i \mid \{y_i^k\}$ is unbounded, $i = 1, \ldots, n\}$. By assumption, $\mathcal{I}_0^\infty$ is nonempty, and for all $i \in \mathcal{I}_0^\infty$, $|y_i^k| \to \infty$ as $k \to \infty$. In the following analysis, we will consider the following cases: (i) $a = -\infty$, $b = \infty$; (ii) $a \in \mathbb{R}^n$, $b = \infty$; (iii) $a = \infty$, $b \in \mathbb{R}^n$; and (iv) $a, b \in \mathbb{R}^n$.

Case (i). Since $a = -\infty$, $b = \infty$, we have $p(\mu e, y) = y$. Then $R(x,y,\xi,\mu) = F(x,y,\xi) + \mu y$. We now construct a bounded sequence $\{w^k\}$ by letting $w_i^k = 0$ if $i \in \mathcal{I}_0^\infty$ and $w_i^k = y_i^k$ otherwise. Since F is a P_0-function in y, hence for any k,

$$\begin{aligned} 0 &\le \max_{1 \le i \le n} (y_i^k - w_i^k)[F_i(x^k, y^k, \xi^k) - F_i(x^k, w^k, \xi^k)] \\ &= \max_{i \in \mathcal{I}_0^\infty} y_i^k [F_i(x^k, y^k, \xi^k) - F_i(x^k, w^k, \xi^k)] \\ &= y_{i_0}^k [F_{i_0}(x^k, y^k, \xi^k) - F_{i_0}(x^k, w^k, \xi^k)]. \end{aligned} \tag{50}$$

Here i_0 denotes an index in $\mathcal{I}_0^\infty$ at which the maximum value is attained. Without loss of generality, we may assume that the above index i_0 is independent of k. Since $\mathcal{X}$ and Ξ are compact, and $\{w^k\}$ is bounded, hence $\{F_{i_0}(x^k, w^k, \xi^k)\}$ is bounded by virtue of the continuity of F_{i_0}. We now consider two cases: $y_{i_0}^k \to \infty$; $y_{i_0}^k \to -\infty$. In the former case, it follows from (50) that $\{F_{i_0}(x^k, y^k, \xi^k)\}$ does not tend to $-\infty$. Since $\{\mu^k\}$ is contained in a closed interval of $(0, \mu_0)$, hence $F_{i_0}(x^k, y^k, \xi^k) + \mu^k y_{i_0}^k \to \infty$, which implies that $\|F(x^k, y^k, \xi^k) + \mu^k y^k\| \to \infty$. Similarly, in the latter case, we have that $\{F_{i_0}(x^k, y^k, \xi^k)\}$ does not tend to ∞ by (50). Thereby, $F_{i_0}(x^k, y^k, \xi^k) + \mu^k y_{i_0}^k \to -\infty$. Thus, in both cases, we have $\|R(x^k, y^k, \xi^k, \mu^k)\| \to \infty$.

Case (ii). Note that in this case

$$p_i(\mu e, y) = (a_i + \sqrt{(a_i - y_i)^2 + 4\mu^2} + y_i)/2, \quad i = 1, \dots, n.$$

Then it is not hard to show that for each i, and any sequences $\{y_i^l\}$ and $\{\mu^l\}$ satisfying $y_i^l \to \infty$ and μ^l being in a closed subset of $(0, \mu_0)$ for all l, we have

$$\lim_{l\to\infty} [y_i^l - p_i(\mu^l e, y^l)] = 0. \tag{51}$$

Let

$$\mathcal{I}_+^\infty := \{i \in \mathcal{I}_0^\infty \mid \{y_i^k\} \to \infty\} \quad \text{and} \quad \mathcal{I}_-^\infty := \{i \in \mathcal{I}_0^\infty \mid \{y_i^k\} \to -\infty\}.$$

We now consider two cases: $\mathcal{I}_+^\infty = \emptyset$; $\mathcal{I}_+^\infty \neq \emptyset$. In Case (i), we have $\{y_i^k\} \to -\infty$ for all $i \in \mathcal{I}_0^\infty$. Then it is easy to show that $\lim_{k\to\infty} p_i(\mu^k e, y^k) = a_i$ for all $i \in \mathcal{I}_0^\infty$. Thus, $\{p(\mu^k e, y^k)\}$ is bounded. Noticing the boundedness of $\{x^k\}$ and $\{\xi^k\}$ and by virtue of the continuity of F_i, $i \in \mathcal{I}_0^\infty$, it follows that

$$\left|\left(R(x^k, y^k, \xi^k, \mu^k)\right)_i\right| = \left|F_i(x^k, p(\mu^k e, y^k), \xi^k) - p_i(\mu^k e, y^k) + y_i^k + \mu^k y_i^k\right| \to \infty.$$

In Case (ii), evidently, $\lim_{k\to\infty} p_i(\mu^k e, y^k) = \infty$ or a_i for $i \in \mathcal{I}_+^\infty$ or $\mathcal{I}_-^\infty$. We now define a sequence $\{v^k\}$ with $v_i^k := 0$ if $i \in \mathcal{I}_+^\infty$; $v_i^k := p_i(\mu^k e, y^k)$ if $i \in \mathcal{I}_-^\infty$; $v_i^k := p_i(\mu^k e, y^k)$ if $i \notin \mathcal{I}_0^\infty$. Based on the above arguments, evidently, $\{v^k\}$ is bounded. By the notion of P_0-function, we have

$$\begin{aligned} 0 &\le \max_{1\le i\le n} (p_i(\mu^k e, y^k) - v_i^k)[F_i(x^k, p(\mu^k e, y^k), \xi^k) - F_i(x^k, v^k, \xi^k)] \\ &= \max_{i\in\mathcal{I}_+^\infty} p_i(\mu^k e, y^k)[F_i(x^k, p(\mu^k e, y^k), \xi^k) - F_i(x^k, v^k, \xi^k)] \\ &= p_j(\mu^k e, y^k)[F_j(x^k, p(\mu^k e, y^k), \xi^k) - F_j(x^k, v^k, \xi^k)], \end{aligned} \tag{52}$$

where $j \in \mathcal{I}_+^\infty$ such that the maximum value is attained at j, without loss of generality, which is assumed to be independent of k. By assumption, F_j is continuous, and note that $\{x^k\}$, $\{\xi^k\}$, $\{v^k\}$ are bounded; hence $\{F_j(x^k, v^k, \xi^k)\}$ is bounded as well. In addition, since $p_j(\mu^k e, y^k) \to \infty$, thus by (52), $\{F_j(x^k, p(\mu^k e, y^k), \xi^k)\}$ does not tend to $-\infty$. Thereby, $F_j(x^k, p(\mu^k e, y^k), \xi^k) + \mu^k y_j^k \to \infty$. On the other hand, note that

$$\begin{aligned} \left(R(x^k, y^k, \xi^k, \mu^k)\right)_j &= F_j(x^k, p(\mu^k e, y^k), \xi^k) + y_j^k - p_j(\mu^k e, y^k) + \mu^k y_j^k \\ &= F_j(x^k, p(\mu^k e, y^k), \xi^k) + \mu^k y_j^k + y_j^k - p_j(\mu^k e, y^k). \end{aligned}$$

By (51), $y_j^k - p_j(\mu^k e, y^k) \to 0$. So, $\left(R(x^k, y^k, \xi^k, \mu^k)\right)_j \to \infty$. Therefore,

$$\|R(x^k, y^k, \xi^k, \mu^k)\| \to \infty.$$

Case (iii). In this case, the arguments are similar to Case (ii). Here we omit them for brevity.

Case (iv). Note that for any $i \in \mathcal{I}_0^\infty$, $\lim_{k\to\infty} p_i(\mu^k e, y^k) = \lim_{k\to\infty} \phi(\mu^k, a_i, b_i, y_i^k)$ equals b_i if $y_i^k \to \infty$ or a_i if $y_i^k \to -\infty$. Then $\{p(\mu^k e, y^k)\}$ is bounded; thereby, $\{F_i(x^k, p(\mu^k e, y^k), \xi^k)\}$ is bounded as well for $i \in \mathcal{I}_0^\infty$. Hence,

$$\left|\left(R(x^k, y^k, \xi^k, \mu^k)\right)_i\right| = \left|F_i(x, p(\mu^k e, y^k), \xi^k) + y_i^k - p_i(\mu^k e, y^k) + \mu^k y_i^k\right| \to \infty.$$

Thereby, $\|R(x^k, y^k, \xi^k, \mu^k)\| \to \infty$. $\square$

Proof of Proposition 5.3. Note that R is continuous on $\mathcal{X}\times\mathbb{R}^n\times\Xi\times[0,\mu_0]$. We now check parts (i)–(iv) in Definition 2.1. Obviously, part (i) holds, since $p(0,y) = \Pi_\Upsilon(y)$ for any $y \in \mathbb{R}^n$. By [20, Theorem 3.1], $p(\mu e, y)$ is continuously differentiable at any $(\mu, y) \in \mathbb{R}_{++} \times \mathbb{R}^n$. Then R is continuously differentiable on $\mathcal{X} \times \mathbb{R}^n \times \Xi \times (0, \mu_0]$. Note also that $H(x, y, \xi)$ is piecewise smooth; hence R is piecewise smooth on $\mathcal{X} \times \mathbb{R}^n \times \Xi \times [0, \mu_0]$. Thereby, part (ii) holds.

We now consider part (iii). Note that

$$\pi_x \partial R(x, y, \xi, \mu) = \partial_x R(x, y, \xi, \mu) = \nabla_x F(x, p(\mu e, y), \xi),$$

$$\pi_x \partial H(x, y, \xi) = \partial_x H(x, y, \xi) = \nabla_x F(x, \Pi_\Upsilon(y), \xi)$$

and $\lim_{\mu\downarrow 0} p(\mu e, y) = \Pi_\Upsilon(y)$. Then we have

$$\lim_{\mu\downarrow 0} \partial_x R(x, y, \xi, \mu) = \lim_{\mu\downarrow 0} \nabla_x F(x, p(\mu e, y), \xi) = \nabla_x F(x, \Pi_\Upsilon(y), \xi) = \partial_x H(x, y, \xi)$$

for any $(x, y, \xi) \in \mathcal{X} \times \mathbb{R}^n \times \Xi$. On the other hand, noticing that $\pi_y \partial H(x, y, \xi) = \partial_y H(x, y, \xi) = (\nabla_y F(x, \Pi_\Upsilon(y), \xi) - I)\partial \Pi_\Upsilon(y) + I$, and by Lemma 5.1, $\pi_y \partial R(x, y, \xi, \mu) = \partial_y R(x, y, \xi, \mu) = \nabla_y R(x, y, \xi, \mu) = (\nabla_y F(x, p(\mu e, y), \xi) - I)D(\mu, y) + \mu I + I$. Hence, to show $\lim_{\mu\downarrow 0} \pi_y \partial R(x, y, \xi, \mu) \subset \pi_y \partial H(x, y, \xi)$, it suffices to prove $\lim_{\mu\downarrow 0} D(\mu, y) \subset \partial \Pi_\Upsilon(y)$. Note that for any $y \in \mathbb{R}^n$,

$$\partial \Pi_\Upsilon(y) = \begin{bmatrix} \partial \Pi_{[a_1,b_1]}(y_1) & \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \cdots & \partial \Pi_{[a_n,b_n](y_n)} \end{bmatrix},$$

where $\partial \Pi_{[a_i,b_i]}(y_i)$ equals 0 if $y_i \in (-\infty, a_i) \cup (b_i, \infty)$; 1 if $y_i \in (a_i, b_i)$; $[0, 1]$ if $y_i = a_i$ or b_i. Then, after some basic manipulations, $\lim_{\mu\downarrow 0} d_i(\mu, y) = \lim_{\mu\downarrow 0} \partial p_i(\mu e, y)/\partial y_i$ equals 0 if $y_i \in (-\infty, a_i) \cup (b_i, \infty)$; 1 if $y_i \in (a_i, b_i)$; 1/2 if $y_i = a_i$ or b_i. Hence, $\lim_{\mu\downarrow 0} d_i(\mu, y) \subset \partial \Pi_{[a_i,b_i]}(y_i)$ for each i; thereby, $\lim_{\mu\downarrow 0} D(\mu, y) \subset \partial \Pi_{[a,b]}(y)(= \partial \Pi_\Upsilon(y))$. Thus, $\lim_{\mu\downarrow 0} \nabla_y R(x, y, \xi, \mu) \subset \partial_y H(x, y, \xi)$ for any $(x, y, \xi) \in \mathcal{X} \times \mathbb{R}^n \times \Xi$. So, part (iii) holds.

Finally, we prove part (iv). Define a mapping $G : \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^k \times \mathbb{R} \to \mathbb{R}^m \times \mathbb{R}^n \times \mathbb{R}^k \times \mathbb{R}$ by $G(x, y, \xi, \mu) := (x, R(x, y, \xi, \mu), \xi, \mu)$. Then G is continuously differentiable on $\mathcal{X}\times\mathbb{R}^n\times\Xi\times(0, \mu_0)$. We first show that G is a diffeomorphism on $\mathcal{X}\times\mathbb{R}^n\times\Xi\times(0, \mu_0)$; that is, G has a differentiable inverse function on $\mathcal{X} \times \mathbb{R}^n \times \Xi \times (0, \mu_0)$. It is well known that a necessary and sufficient condition for function G to be a diffeomorphism is the nonsingularity of the Jacobian ∇G at every point (x, y, ξ, μ) and the closedness of G; that is, the image $G(S)$ of any closed set $S \subset \mathcal{X} \times \mathbb{R}^n \times \Xi \times (0, \mu_0)$ is closed.

We now prove the closedness of G. Let S be a closed subset of $\mathcal{X}\times\mathbb{R}^n\times\Xi\times(0,\mu_0)$. Assume for the sake of a contradiction that $G(S)$ is not closed. Then there exists a convergent sequence $\{w^k\}\subset G(S)$ such that $\lim_{k\to\infty} w^k = w^0$, but $w^0\notin G(S)$. By definition, there exists a sequence $\{z^k\}$ with $z^k=(x^k,y^k,\xi^k,\mu^k)\in S$ such that $w^k = G(z^k)$. We consider two cases: (i) $\{z^k\}$ is bounded; (ii) $\{z^k\}$ is unbounded. In case (i), obviously, there exists a convergent subsequence, $\{z^{k_l}\}$, of $\{z^k\}$ with $\{z^{k_l}\}$ tending to z^0. Hence, $z^0\in S$ given the closeness of S. Thus, $\lim_{l\to\infty} w^{k_l} = \lim_{l\to\infty} G(z^{k_l}) = G(z^0)$. Clearly, $G(z^0)\in G(S)$. Since $w^{k_l}\to w^0$, then $w^0 = G(z^0)\in G(S)$, which leads to a contradiction as desired. In case (ii), without loss of generality, we assume that $\|z^k\|\to\infty$ as $k\to\infty$. With the help of the compactness of $\mathcal{X}$ and Ξ, there exists a subsequence $\{z^{k_l}\}$ of $\{z^k\}$ such that $\{x^{k_l}\}$, $\{\xi^{k_l}\}$, and $\{\mu^{k_l}\}$ are bounded, while $\|y^{k_l}\|\to\infty$ as $l\to\infty$. Then, by Lemma 5.2, $\lim_{l\to\infty}\|R(x^{k_l},y^{k_l},\xi^{k_l},\mu^{k_l})\| = \infty$. Thus,

$$\lim_{l\to\infty}\|w^{k_l}\| = \lim_{l\to\infty}\|G(z^{k_l})\| = \infty$$

by noticing $w^{k_l} = G(z^{k_l}) = \big(x^{k_l}, R(x^{k_l},y^{k_l},\xi,\mu^{k_l}),\xi^{k_l},\mu^{k_l}\big)$. This contradicts the fact that $\lim_{l\to\infty} w^{k_l} = w^0$. Therefore, G is closed on $\mathcal{X}\times\mathbb{R}^n\times\Xi\times(0,\mu_0)$.

Next, we prove the nonsingularity of ∇G. By Lemma 5.1, we can easily see that $\nabla G(x,y,\mu,\xi)$ is nonsingular at any point $(x,y,\xi,\mu)\in\mathcal{X}\times\mathbb{R}^n\times\Xi\times(0,\mu_0)$. Hence, G is a diffeomorphism on $\mathcal{X}\times\mathbb{R}^n\times\Xi\times(0,\mu_0)$. Let G^{-1} denote its inverse function. For any $(x,y,\xi,\mu)\in\mathcal{X}\times\mathbb{R}^n\times\Xi\times(0,\mu_0)$, we then have $(G^{-1}(x,y,\xi,\mu))_x = x$, $(G^{-1}(x,y,\xi,\mu))_\xi = \xi$, and $(G^{-1}(x,y,\xi,\mu))_\mu = \mu$. Furthermore, for any $(p,t,q)\in\mathcal{X}\times\Xi\times(0,\mu_0)$, equation $G(x,y,\xi,\mu) = (p,0,t,q)$ has a unique solution $(x,y,\xi,\mu) = G^{-1}(p,0,t,q)$. Clearly, $x=p$, $\xi=t$, and $\mu=q$. Let $y=\tilde{y}(p,t,q) := \big(G^{-1}(p\ 0\ t\ q)\big)_y$. By virtue of the arbitrariness of p, t, and q, we obtain the unique solution of $\tilde{y}$ defined on $\mathcal{X}\times\Xi\times(0,\mu_0)$, which satisfies $R(x,\tilde{y}(x,\xi,\mu),\xi,\mu)=0$ for $(x,\xi,\mu)\in\mathcal{X}\times\Xi\times(0,\mu_0)$. Thereby, part (iv) is satisfied. In addition, note that G is continuously differentiable on $\mathcal{X}\times\mathbb{R}^n\times\Xi\times(0,\mu_0)$ by assumption. This leads to the continuous differentiability of $\tilde{y}$ immediately.

In conclusion, based on the above arguments, function R defined in (37) satisfies Definition 2.1. This completes the proof. ∎

Acknowledgments. The authors would like to thank Alexander Shapiro for his constructive comments on an earlier version of this paper. They would also like to thank two anonymous referees for insightful comments which led to a significant improvement of the paper and the associate editor, Andrzej Ruszczyński, for organizing a quick and quality review.

REFERENCES

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

[2] Z. Artstein and S. Hart, *Law of large numbers for random sets and allocation processes*, Math. Oper. Res., 6 (1981), pp. 485–492.

[3] F. Bastin, C. Cirillo, and P. Toint, *Convergence theory for nonconvex stochastic programming with an application to mixed logit*, Math. Program., 108 (2006), pp. 207–234.

[4] S. I. Birbil, G. Gürkan, and O. Listes, *Simulation-Based Solution of Stochastic Mathematical Programs with Complementarity Constraints: Sample-Path Analysis*, working paper, Center for Economic Research, Tilburg University, Tilburg, The Netherlands, 2004.

[5] J. Birge and F. Louveaux, *Introduction to Stochastic Programming*, Springer-Verlag, New York, 1997.

[6] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer-Verlag, New York, 2000.

[7] B. Chen and P. T. Harker, *A non-interior-point continuation method for linear complementarity problems*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 1168–1190.

[8] X. Chen, L. Qi, and D. Sun, *Global and superlinear convergence of the smoothing Newton's method and its application to general box constrained variational inequalities*, Math. Comp., 67 (1998), pp. 519–540.

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

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

[11] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer-Verlag, New York, 2003.

[12] S. A. Gabriel and J. J. Moré, *Smoothing of mixed complementarity problems*, in Complemenarity and Variational Problems: State of the Art, M. C. Ferris and J. S. Pang, eds., SIAM, Philadelphia, 1997, pp. 105–116.

[13] J. L. Higle and S. Sen, *Stochastic Decomposition: A Statistical Method for Large Scale Stochastic Linear Programming*, Kluwer Academic Publishers, Boston, 1996.

[14] J.-B. Hiriart-Urruty, *Refinements of necessary optimality conditions in nondifferentiable programming* I, Appl. Math. Optim., 5 (1979), pp. 63–82.

[15] T. Homem-de-Mello, *Estimation of derivatives of nonsmooth performance measures in regenerative systems*, Math. Oper. Res., 26 (2001), pp. 741–768.

[16] C. Kanzow, *Some noninterior continuation methods for linear complementarity problems*, SIAM J. Matrix Anal. Appl., 17 (1996), pp. 851–868.

[17] G.-H. Lin, X. Chen, and M. Fukushima, *Smoothing Implicit Programming Approaches for Stochastic Mathematical Programs with Linear Complementarity Constraints*, http://www. amp.i.kyoto-u.ac.jp/tecrep/index-e.html (2003).

[18] M. Patriksson and L. Wynter, *Stochastic mathematical programs with equilibrium constraints*, Oper. Res. Lett., 25 (1999), pp. 159–167.

[19] H.-D. Qi, *A regularized smoothing Newton method for box constrained variational inequality problems with P_0-functions*, SIAM J. Optim., 10 (1999), pp. 315–330.

[20] L. Qi, D. Sun, and G. Zhou, *A new look at smoothing Newton methods for nonlinear complementarity problems and box constrained variational inequalities*, Math. Program., 87 (2000), pp. 1–35.

[21] D. Ralph and H. Xu, *Implicit smoothing and its application to optimization with piecewise smooth equality constraints*, J. Optim. Theory Appl., 124 (2005), pp. 673–699.

[22] S. M. Robinson, *Analysis of sample-path optimization*, Math. Oper. Res., 21 (1996), pp. 513–528.

[23] R. T. Rockafellar and R.J.-B. Wets, *Stochastic convex programming: Kuhn-Tucker conditions*, J. Math. Econom., 2 (1975), pp. 349–370.

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

[25] R. Y. Rubinstein and A. Shapiro, *Discrete Events Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Methods*, John Wiley and Sons, New York, 1993.

[26] A. Ruszczyński, *Decomposition methods*, in Stochastic Programming, Handbooks Oper. Res. Management Sci. 10, A. Ruszczyński and A. Shapiro, eds., North–Holland, Amsterdam, 2003, pp. 141–211.

[27] A. Ruszczyński and A. Shapiro, *Stochastic Programming*, Handbooks Oper. Res. Management Sci. 10, North–Holland, Amsterdam, 2003.

[28] T. Santoso, S. Ahmed, M. Goetschalckx, and A. Shapiro, *A stochastic programming approach for supply chain network design under uncertainty*, European J. Oper. Res., 167 (2005), pp. 96–115.

[29] S. Scholtes, *Introduction to Piecewise Smooth Equations*, Habilitation, University of Karlsruhe, Karlsruhe, Germany, 1994.

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

[31] A. Shapiro and T. Homem-de-Mello, *On the rate of convergence of optimal solutions of Monte Carlo approximations of stochastic programs*, SIAM J. Optim., 11 (2000), pp. 70–86.

[32] A. Shapiro and H. Xu, *Stochastic Mathematical Programs with Equilibrium Constraints, Modeling and Sample Average Approximation*, http://www.optimization-online.org/DB_HTML/2005/01/1046.html (2005).

[33] R.J.-B. Wets, *Stochastic programming*, in Optimization, Handbooks Oper. Res. Management Sci. 1, G. L. Nemhauser, A. Rinnooy, and M. Todd, eds., North–Holland, Amsterdam,

1989, pp. 573–629.

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

[35] H. Xu, *An implicit programming approach for a class of stochastic mathematical programs with complementarity constraints*, SIAM J. Optim., 16 (2006), pp. 670–696.

[36] 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., to appear.