REGULARIZED MATHEMATICAL PROGRAMS WITH STOCHASTIC EQUILIBRIUM CONSTRAINTS: ESTIMATING STRUCTURAL DEMAND MODELS

XIAOJUN CHEN*, HAILIN SUN†, AND ROGER J-B WETS‡

22 July 2013

Abstract. The article considers a particular class of optimization problems involving set-valued stochastic equilibrium constraints. A solution procedure is developed by relying on an approximation scheme for the equilibrium constraints, based on regularization, that replaces them by equilibrium constraints involving only single-valued Lipschitz continuous functions. In addition, sampling has the further effect of replacing the 'simplified' equilibrium constraints by more manageable ones obtained by implicitly discretizing the (given) probability measure so as to render the problem computationally tractable. Convergence is obtained by relying, in particular, on the graphical convergence of the approximated equilibrium constraints. The problem of estimating the characteristics of a demand model, a widely studied problem in micro-economics, serves both as motivation and illustration of the regularization and sampling procedure.

Key words. Stochastic equilibrium, monotone linear complementarity problem, graphical convergence, sample average approximation, regularization.

AMS subject classifications. 90C33, 90C15.

1. Introduction. Solving mathematical optimization involving equilibrium constraints is generally challenging and the design of solutions procedures to deal with such problems when the equilibrium constraints involve set-valued stochastic mappings brings along a new level of difficulty. This article, considers a particular case which enables us to deal with a specific instance, see §4, of the 'inverse' problem in micro-economics: given that prices and the decisions of the agents can be observed, is it possible to infer their utility functions?

Specifically, we consider the following mathematical program with stochastic equilibrium constraints (MPSEC):

$$
\begin{array}{rl}
\min_{x\in X} & \frac{1}{2}\langle x, Hx\rangle + \langle c, x\rangle \\
\text{subject to} & A_t\mathbb{E}[S_t(\boldsymbol{\xi}, x)] \ni b_t, \quad t = 1, \dots, T,
\end{array}
\tag{1.1}
$$

where $c \in \mathbb{R}^\nu$, $A_t \in \mathbb{R}^{m\times n}$, $b_t \in \mathbb{R}^m$, H is a positive semi-definite $\nu \times \nu$-matrix, $X \subseteq \mathbb{R}^\nu$ is a compact set and $\boldsymbol{\xi} : \Omega \to \Xi \subseteq \mathbb{R}^\ell$ is a random vector with realizations as ξ (without boldface) and $(\Xi, \mathcal{F}, P)$ the induced probability space,

$$
S_t(\xi, x) = \operatorname{argmax}_y \{ \langle y, u_t(\xi, x)\rangle \mid \langle e, y\rangle \le 1,\ y \ge 0 \} \subseteq \mathbb{R}^n,
\tag{1.2}
$$

*Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.(maxjchen@polyu.edu.hk). The author's work was supported in part by Hong Kong Research Grant Council grant PolyU5003/11p.

†School of Economics and Management, Nanjing University of Science and Technology, Nanjing, 210049, China. (mathhlsun@gmail.com). This work was supported in part by a Hong Kong Polytechnic University Research grant.

‡Department of Mathematics, University of California, Davis, CA 95616.(rjbwets@ucdavis.edu). This material is based upon work supported in part by the U. S. Army Research Laboratory and the U. S. Army Research Office under grant number W911NF1010246.

$u_t : \Xi \times \mathbb{R}^\nu \to \mathbb{R}^n$ is a given continuous function, and $e = (1, \dots, 1) \in \mathbb{R}^n$, and

$$\mathbb{E}[S_t(\boldsymbol{\xi}, x)] = \{\mathbb{E}[s(\boldsymbol{\xi}, x)] \,|\, s(\boldsymbol{\xi}, x) \in S_t(\boldsymbol{\xi}, x),\ s(\cdot, x)\ P\text{-summable selection of } S_t(\cdot, x)\}$$

is Aumann's expected value [1] with respect to $\boldsymbol{\xi}$.

This problem-type (1.1) is part of an important family of problems in economics which, in particular, includes the pure characteristics demand model which seeks to estimate the parameters in the consumers' utility function [3, 10, 14]. In such a model, the constraint $A_t\mathbb{E}[S_t(\boldsymbol{\xi}, x)] \ni b_t$, with A_t the identity matrix, represents the market share equations, u_t determines the consumer's utility in market 't' and the jth component(s) of the solution(s) in $S_t(\xi, x)$ of (1.2) is the probability that the consumer purchases product j in market t given the observed environment ξ. The linear program (1.2) models the consumer's decision choice, in market t, to acquire the product or products, that yields the highest utility given environment ξ. The variable x consists of two parts $x^1 \in \mathbb{R}^{\nu_1}$ and $x^2 \in \mathbb{R}^{\nu_2}$: x^1 describes the product characteristics or demand shock that is observed by the providers (firms) and consumers, but not explicitly available in the data, and x^2 models the consumer's preferences or taste for the observed product characteristics and its price. The pure characteristics demand model is to estimate x^2 and minimize the demand shock or error x^1. The objective function in this model has $H = \mathrm{diag}(H_1, H_2)$ and $c = 0$, where H_1 is a $\nu_1 \times \nu_1$ positive definite matrix and H_2 is the $\nu_2 \times \nu_2$ zero matrix.

There are quite a number of challenges one has to deal with to solve such a problem. To begin with the solution of (1.2), for any fixed (ξ, x) *is not necessarily unique*, in fact, in general, it's set-valued. Consider a simple example with $u_t(\xi, x) = (\xi_1 + x, \xi_2) \in \mathbb{R}^2$, where $\xi_1 \in \mathbb{R}$ and $\xi_2 > 0$. The solution set has the form

$$S_t(\xi, x) = \begin{cases} (1, 0) & x > \xi_2 - \xi_1, \\ \{(\alpha, 1-\alpha) \,|\, \alpha \in [0, 1]\} & x = \xi_2 - \xi_1, \\ (0, 1) & x < \xi_2 - \xi_1. \end{cases}$$

One cannot find a single-valued function $s(\xi, x) \in S_t(\xi, x)$ which is continuous with respect to x. The use of a sample average approximation (SAA) scheme to approximate the market share equations as proposed in the existing literature becomes intractable. Another major difficulty comes from the fact that all solution sets $S_t(\xi, x)$, $t = 1, \dots, T$ also share the same x-variables.

Market share equations play an important role in economics [3, 10, 14]. The 'inverse' problem, from consumers choices evince their utility functions is a fundamental issue in economics. In the pure characteristics demand model, even when approximating, the market share equations for the unobserved product characteristics, finding best estimates by relying on a nested fixed-point approach has been proposed in the existing econometrics literature but it is known that such an approach is computationally ineffective. Recently, Pang et al. [14] proposed a mathematical programming with linear complementarity constraints (MPLCC) approach for the pure characteristics demand model with a finite number of observations ξ^i, $i = 1, \dots, N$. Their approach provides a promising computational method to estimate the consumer utility under the additional condition that in any market t, the optimal choice of each individual consumer is guaranteed to purchase just one single product in each ξ-environment. They rely extensively on this property of the (basic) optimal solution of (1.2) in their development. This condition and the use of such basic solution with a

finite number of observations ξ^i, $i = 1, \ldots, N$ for (1.1)-(1.2) can be expresses in terms of the following mathematical program with linear equilibrium constraints
(1.3)

$$\begin{array}{ll} \min_{x \in X} & \frac{1}{2}\langle x, Hx\rangle + \langle c, x\rangle \\ \text{subject to} & A_t \frac{1}{N}\sum_{i=1}^N \hat{S}_t(\xi^i, x) \ni b_t, \quad t = 1, \ldots, T, \quad \xi^i \in \Xi,\ i = 1, \ldots, N, \end{array}$$

where

$$\hat{S}_t(\xi, x) = \{\text{argmin}\|s\|_0 \mid s \in S_t(\xi, x)\},$$

here $\|s\|_0$ denotes the number of nonzero entries of s. With the constraints $\langle e, y\rangle \le 1$ and $y \ge 0$, clearly, the linear program (1.2) has always a basic optimal solution $s(\xi, x)$ and it's taken for granted that any basic optimal solution has just a single variable taking on the value 1 while all others are 0 when $\max_{1 \le i \le n} u_i(\xi, x) > 0$. One could refer to a solution of this type, $s(\xi, x) \in \hat{S}_t(\xi, x)$ as a 'sparse solution' of (1.2). However, the use of such 'sparse solutions' raises questions when there is, in fact, a multiplicity of solutions. For example, when $(u_t(\xi, x))_j = \max_{1 \le i \le n}(u_t(\xi, x))_i$, $j = 1, 2, 3$, why would the probability that a consumer purchases one of the three products be 1 and 0 for the two others? Should the choice probability not be 1/3, for example, for each one of the three products? Other question arise about the consistency of the solutions of the MPLCC problem (1.3) to the given problem (1.1) as the sample size N goes to infinity.

Motivated by the MPLCC approach [14] and the preceding questions, we reformulate problem (1.1) as the following mathematical program with stochastic linear complementarity constraints (MPSLCC)

$$\begin{array}{ll} \min_{x \in X} & \frac{1}{2}\langle x, Hx\rangle + \langle c, x\rangle \\ \text{subject to} & A_t \mathbb{E}[S_t(\boldsymbol{\xi}, x)] \ni b_t, \quad t = 1, \ldots, T, \end{array} \tag{1.4}$$

where $S_t(\xi, x)$ consists of all the solutions to

$$\begin{array}{lllll} 0 & \le & y(\xi, x) \perp & -u_t(\xi, x) + \gamma(\xi, x)e & \ge 0 \\ 0 & \le & \gamma(\xi, x) \perp & 1 - \langle e, y(\xi, x)\rangle & \ge 0 \end{array} \tag{1.5}$$

for some $\gamma(\xi, x) \in \mathbb{R}$ or, equivalently, the linear complementarity problem (LCP):

$$0 \ \le \ \begin{pmatrix} y \\ \gamma \end{pmatrix} \ \perp \ M\begin{pmatrix} y \\ \gamma \end{pmatrix} + \begin{pmatrix} -u_t(\xi, x) \\ 1 \end{pmatrix} \ \ge 0 \tag{1.6}$$

with the positive semidefinite matrix

$$M = \begin{pmatrix} 0 & e \\ -e & 0 \end{pmatrix} \in \mathbb{R}^{(n+1)\times(n+1)}.$$

For fixed (t, ξ, x) and with $q_t(\xi, x) = (-u_t(\xi, x), 1) \in \mathbb{R}^{n+1}$, let's denote the complementarity problem (1.6) by $\text{LCP}(q_t(\xi, x), M)$ and by

$$S(q_t, M) = \{s \in \mathbb{R}^n \mid \text{for some } \gamma \ge 0, (s, \gamma) \text{ solves } \text{LCP}(q_t(\xi, x), M)\},$$

i.e., the solution set projected on the s-space[1].

[1]To lighten up the notation, when no confusion is possible, we usually simply write q_t instead of the more precise, but cumbersome, $q_t(\xi, x)$.

It will be shown that, cf. proof of Theorem 2.3, the solution set $S(q_t, M)$ is bounded. With $M^\varepsilon = M + \varepsilon I, \varepsilon > 0$, it also implies that the LCP$(q_t(\xi, x), M^\varepsilon)$ has a *unique* solution, which is then denoted by $z_t^\varepsilon = (s_t^\varepsilon, \gamma_t^\varepsilon)$. It converges to the least norm solution of the LCP$(q_t(\xi, x), M)$ as $\varepsilon \downarrow 0$ [8, Theorem 5.6.2]. Moreover, for any fixed $\varepsilon > 0$, the function $q_t \mapsto z_t^\varepsilon$ is globally Lipschitz continuous (with $q_t = q_t(\xi, x)$) and continuously differentiable at q_t if and only if for no j, $(z_t^\varepsilon)_j = 0 = (Mz_t^\varepsilon + q_t)_j$ [6],[7, Lemma 2.1]; a nondegenerary condition. These engaging properties motivate us to consider a regularized version of MPSLCC: with z_t^ε the unique solution of the regularized LCP$(q_t(\xi, x), M^\varepsilon)$

$$(1.7) \qquad 0 \leq z \perp M^\varepsilon z + q_t(\xi, x) \geq 0, \qquad \text{where} \quad q_t(\xi, x) = \begin{pmatrix} -u_t(\xi, x) \\ 1 \end{pmatrix},$$

the formulation of our problem becomes,

$$(1.8) \qquad \begin{array}{rl} \min_{x \in X} & \frac{1}{2}\langle x, Hx\rangle + \langle c, x\rangle \\ \text{subject to} & \|A_t \mathbb{E}[s_t^\varepsilon(\boldsymbol{\xi}, x)] - b_t\| \leq r(\varepsilon), \qquad t = 1, \dots, T, \end{array}$$

and the SAA-version of the regularized MPSLCC

$$(1.9) \qquad \begin{array}{rl} \min_{x \in X} & \frac{1}{2}\langle x, Hx\rangle + \langle c, x\rangle \\ \text{subject to} & \|A_t \frac{1}{N}\sum_{i=1}^N s_t^\varepsilon(\xi^i, x) - b_t\| \leq \hat{r}(\varepsilon, N), \quad t = 1, \dots, T, \end{array}$$

where $r(\varepsilon) \downarrow 0$ as $\varepsilon \downarrow 0$, $\hat{r}(\varepsilon, N) \to r(\varepsilon)$ as $N \to \infty$ for any fixed $\varepsilon > 0$.

The advantage of working with (1.8) and (1.9) is that one can replace the set-valued mapping by a single valued function. Problem (1.9) is a mathematical program with a convex quadratic objective function and globally Lipschitz continuous inequality constraints. Moreover, we can have a closed form for z_t^ε, see Lemma 2.2. The main contribution of this paper is to propose an efficient approach, via the SAA-version of the regularized MPSLCC, to find a solution of the mathematical program with stochastic equilibrium constraints (1.4) and to show that a sequence of solutions $\{x_N^\varepsilon\}$ of the SAA regularized stochastic MPSLCC (1.9) converges to a solution of the (given) problem (1.4) as $N \to \infty$ and $\varepsilon \downarrow 0$.

In Section 2, we derive various properties of the solution functions s_t^ε and their convergence to the solution set $S_t(\xi, \bar{x})$ as $\varepsilon \downarrow 0$ and $x \to \bar{x}$. In particular, we provide a closed form of the solution functions which is used to prove the graphical convergence of the real-valued function s_t^ε to the set-valued mapping $S_t(\cdot)$. In Section 3, we prove the existence of solutions to the MPSLCC (1.4) and the SAA regularized MPSLCC (1.9). We show that any sequence of solutions of (1.9) has a cluster point as $\varepsilon \downarrow 0$ and $N \to \infty$, and that any such cluster point is a solution of the MPSLCC (1.4) (a.s.). In Section 4, we use the pure characteristics demand model to illustrate the MPSLCC (1.4) and the SAA regularized method.

Throughout the paper, $\|\cdot\|$ stands for the ℓ_2 norm, e denotes the vector whose elements are all 1, $|\mathcal{J}|$ is the cardinality of a set $\mathcal{J}$ and u_+ denotes the vector whose components $(u_+)_i = \max(0, u_i)$, $i = 1, \cdots, n$.

2. Solution function of the regularized LCP. Let's now concern ourselves with the properties of the function s_t^ε generated from the solution of LCP$(q_t(\xi, x), M^\varepsilon)$.

LEMMA 2.1. *Problems (1.1) and (1.4) are equivalent.*

Proof. The LCP (1.5) consists exactly of the KKT-conditions for (1.2) with solution s_t^ε if and only if $z_t^\varepsilon = (s_t^\varepsilon, \gamma_t^\varepsilon)$ is a solution of (1.5) for some $\gamma_t^\varepsilon \geq 0$. $\square$

For simplicity's sake, in the remainder of this section, we concentrate on the 'solution' function $z^\varepsilon = (s^\varepsilon, \gamma^\varepsilon)$ generated by the regularized linear complementarity problem LCP(q_t, M^ε) with $q_t := (-u_t(\xi, x), 1) \in \mathbb{R}^{n+1}$ for *fixed* t, ξ and x; in this section, we *drop* making reference to these quantities to simplify notations and the presentation. Our first aim will be to show that for given u, $s^\varepsilon(q)$ and $z^\varepsilon(q)$ are uniquely determined and even comes with a closed form expression. Note that $u_i = -q_i, i = 1, \cdots, n$ and $q_{n+1} = 1$, and we have

$$\|(-q)_+\|_1 = -\sum_{q_i \leq 0} q_i = \sum_{u_i \geq 0} u_i = \|u_+\|_1.$$

LEMMA 2.2. *Given $\varepsilon > 0$, the function z^ε is uniquely determined by the solution of the regularized LCP(q, M^ε) and is completely described as follows:*
Let $q_{k_1} \leq q_{k_2} \leq \cdots \leq q_{k_n}$, set

$$\alpha_j = -\sum\nolimits_{i=1}^{j} q_{k_i} + (j + \varepsilon^2) q_{k_j} - \varepsilon, \qquad j = 1, \ldots, n$$

$$\text{and } \mathcal{J} = \{j \,|\, \alpha_j \leq 0, j = 1, \ldots, n\}, \quad J = |\mathcal{J}|, \quad \sigma = \sum_{i=1}^{J} q_{k_i}.$$

(a) If $\|(-q)_+\|_1 \geq \varepsilon$, the solution $(s^\varepsilon, \gamma^\varepsilon)$ of the LCP(q, M^ε) has the form

$$\text{(2.1)} \qquad s_{k_j}^\varepsilon = \begin{cases} \frac{\sigma - (J+\varepsilon^2) q_{k_j} + \varepsilon}{J\varepsilon + \varepsilon^3} & \text{if} \quad j \in \mathcal{J}, \\ 0 & \text{if} \quad j \notin \mathcal{J}, \end{cases} \qquad \gamma^\varepsilon = \frac{-\sigma - \varepsilon}{J + \varepsilon^2}.$$

(b) If $\|(-q)_+\|_1 \leq \varepsilon$, the solution takes the form

$$\text{(2.2)} \qquad \text{for } j = 1, \ldots, n, \qquad s_{k_j}^\varepsilon = \begin{cases} -q_{k_j}/\varepsilon & \text{if} \quad q_{k_j} < 0, \\ 0 & \text{if} \quad q_{k_j} \geq 0, \end{cases} \qquad \gamma^\varepsilon = 0.$$

(c) If $\|(-q)_+\|_1 = \varepsilon$, $\sigma = -\|(-q)_+\|_1$ and $\mathcal{J} = \{j \,\big|\, q_{kj} \leq 0,\ j = 1, \ldots, n\}$, that is, formulas (2.1) and (2.2) are consistent.

Moreover, in all cases, $\sum_{j=1}^n s_j^\varepsilon \leq 1 + \varepsilon \|(-q)_+\|_1$.

Proof. Let $z^\varepsilon = (s^\varepsilon, \gamma^\varepsilon)$, i.e., the solution (vector) of LCP(q, M^ε). Without loss of generality, assume $q_1 \leq q_2 \leq \cdots \leq q_n$, i.e., $k_j = j$.

(**a**) $\|(-q)_+\|_1 \geq \varepsilon$: We show, first, that for $j \in \mathcal{J}$: $s_j^\varepsilon \geq 0$ and $(M^\varepsilon z^\varepsilon)_j + q_j = \varepsilon s_j^\varepsilon + \gamma^\varepsilon + q_j = 0$. From $\alpha_j \leq 0$, $j \leq J$, $q_J - q_j \geq 0$, and $\alpha_J \leq 0$, we have

$$(J\varepsilon + \varepsilon^3) s_j^\varepsilon = \sigma - (J + \varepsilon^2) q_j + \varepsilon + \alpha_J - \alpha_J = (J + \varepsilon^2)(q_J - q_j) - \alpha_J \geq 0$$
$$(J + \varepsilon^2)(\varepsilon s_j^\varepsilon + \gamma^\varepsilon + q_j) = \sigma - (J + \varepsilon^2) q_j + \varepsilon - \sigma - \varepsilon + (J + \varepsilon^2) q_j = 0.$$

The next step is to show that $(M^\varepsilon z^\varepsilon)_j + q_j = \gamma^\varepsilon + q_j > 0$ when $j \notin \mathcal{J}$. By the definition of $\mathcal{J}$ and J, one has $\alpha_{J+1} > 0$, $j \geq J + 1$ and $q_j \geq q_{J+1}$. Hence,

$$\text{(2.3)} \qquad \begin{aligned} (J + \varepsilon^2)(\gamma^\varepsilon + q_j) &= -\sigma - \varepsilon + (J + \varepsilon^2) q_j - \alpha_{J+1} + \alpha_{J+1} \\ &= -\sigma - \varepsilon + (J + \varepsilon^2) q_j + \big(\sigma + q_{J+1} - (J + 1 + \varepsilon^2) q_{J+1} + \varepsilon\big) + \alpha_{J+1} \\ &= (J + \varepsilon^2)(q_j - q_{J+1}) + \alpha_{J+1} > 0. \end{aligned}$$

Finally, we show that $\gamma^\varepsilon \geq 0$ and $(M^\varepsilon z^\varepsilon)_{n+1} + 1 = 1 + \varepsilon\gamma^\varepsilon - \sum_{i=1}^n s_i^\varepsilon = 0$. Let $j_0 = \max\{j \,|\, q_j < 0\}$; such an index is guaranteed to exist since $\|(-q)_+\|_1 \geq \varepsilon$. Actually, we are going to establish that $q_j \leq 0$ for all $j \in \mathcal{J}$. Assume for contradiction purposes that $q_j > 0$ for some $j \in \mathcal{J}$. Of course, then $j > j_0$ and

$$\begin{aligned}\alpha_j &= -\sum_{i=1}^{j_0} q_i - \varepsilon + \sum_{i=j_0+1}^{j} (q_j - q_i) + (j_0 + \varepsilon^2)q_j \\ &= \|q_+\|_1 - \varepsilon + \sum_{i=j_0+1}^{j} (q_j - q_i) - (j_0 + \varepsilon^2)q_j \geq 0,\end{aligned}$$

which contradicts the definition of $\mathcal{J}$. Hence, $q_j \leq 0$ for $j \in \mathcal{J}$, which together with $\|(-q)_+\|_1 \geq \varepsilon$ implies $0 < -\sigma \leq \|(-q)_+\|_1$.

Let's now show that $-\sigma \geq \varepsilon$. Note that $q_j \leq 0$, $j = 1, \ldots, J$ and $J \leq j_0$. If $j_0 = J$, then by definition of j_0, one has $-\sigma = \|(-q)_+\|_1 \geq \varepsilon$. If $j_0 > J$, from $q_{J+1} < 0$ and

$$\begin{aligned}\alpha_{J+1} &= -\textstyle\sum_{i=1}^J q_i - \varepsilon - q_{J+1} + q_{J+1} + (J + \varepsilon^2)q_{J+1} \\ &= -\sigma - \varepsilon + (J + \varepsilon^2)q_{J+1} \geq 0,\end{aligned}$$

and, thus, $-\sigma \geq \varepsilon$. Moreover, $\sum_{i=1}^J q_i = \sigma$ yields $\sum_{i=1}^n s_i^\varepsilon = (J - \varepsilon\sigma)/(J + \varepsilon^2)$ and

$$(J\varepsilon + \varepsilon^3)\big(1 + \varepsilon\gamma^\varepsilon - \sum_{i=1}^n s_i^\varepsilon\big) = J\varepsilon + \varepsilon^3 + \varepsilon^2(-\sigma - \varepsilon) + \varepsilon^2\sigma - J\varepsilon = 0.$$

Hence, the solution has the explicit form (2.1).

(**b**) $\|(-q)_+\|_1 \leq \varepsilon$: If $\|(-q)_+\|_1 = 0$, then $q \geq 0$ and (2.2) holds with $z^\varepsilon = 0$. If $\|(-q)_+\|_1 > 0$, then $j_0 \geq 1$. For $j \leq j_0$, $s_j^\varepsilon = -q_j/\varepsilon > 0$, $\gamma^\varepsilon = 0$ and

$$(M^\varepsilon z^\varepsilon)_j + q_j = \varepsilon s_j^\varepsilon + \gamma^\varepsilon + q_j = -q_j + q_j = 0.$$

For $j > j_0$, one has $q_j \geq 0$, $s_j^\varepsilon = 0$, $\gamma^\varepsilon = 0$ and $(M^\varepsilon z^\varepsilon)_j + q_j = \varepsilon s_j^\varepsilon + \gamma^\varepsilon + q_j \geq 0$, and for $j = n+1$, $\gamma^\varepsilon = 0$ and

$$(M^\varepsilon z^\varepsilon)_{n+1} + 1 = 1 + \varepsilon\gamma^\varepsilon - \sum_{i=1}^n s_i^\varepsilon = 1 + \big(-\sum_{i=1}^{j_0} q_i\big)/\varepsilon \geq 0.$$

(**c**) $\|(-q)_+\|_1 = \varepsilon$: For $j > j_0$,

$$\alpha_j = -\sum_{i=1}^{j_0} q_i - \varepsilon + \sum_{i=j_0+1}^{j} q_i + (j + \varepsilon^2)q_j > 0,$$

and for $j \leq j_0$, $\alpha_j = -\sum_{i=1}^j q_i - \varepsilon + (j + \varepsilon^2)q_j \leq 0$. Hence $\sigma = -\|(-q)_+\|_1$ and $\mathcal{J} = \{j \,|\, q_j \leq 0\}$. Moreover, in this case

$$s_j^\varepsilon = \begin{cases}(\sigma - (J + \varepsilon^2)q_j + \varepsilon)/(J\varepsilon + \varepsilon^3) = -q_j/\varepsilon & \text{if} \quad j \in \mathcal{J}, \\ 0 & \text{if} \quad j \notin \mathcal{J},\end{cases} \quad \gamma^\varepsilon = \frac{-\sigma - \varepsilon}{J + \varepsilon^2} = 0,$$

which implies that formulas (2.1) and (2.2) coincide.

Moreover, in case (a),

$$\sum_{i=1}^{n} s_i^{\varepsilon} \leq 1 + (\varepsilon/J)\|(-q)_+\|_1 \leq 1 + \varepsilon\|(-q)_+\|_1$$

and $\sum_{i=1}^{n} s_j^{\varepsilon} \leq 1$ for (b) and (c) which completes the proof. □

The following theorem shows that the unique solution $z^{\varepsilon}(q)$ of the regularized $\mathrm{LCP}(q, M^{\varepsilon})$ is componentwise monotonically convergent to the least norm solution of the $\mathrm{LCP}(q, M)$ with $O(\varepsilon)$.

THEOREM 2.3. *Let $z^{\varepsilon}(q) = (s^{\varepsilon}(q), \gamma^{\varepsilon}(q))$ be the unique solution of the* $\mathrm{LCP}(q, M^{\varepsilon})$ *and $z(q) = (s(q), \gamma(q))$ be the least norm solution of the* $\mathrm{LCP}(q, M)$. *Then for fixed q, we have $\lim_{\varepsilon \downarrow 0} \|z^{\varepsilon}(q) - z(q)\| = 0$. Moreover, there are positive constants $\bar{\varepsilon}, \kappa_1, \kappa_2$, such that for any $\varepsilon \in (0, \bar{\varepsilon})$,*

$$0 \leq s^{\varepsilon}(q) - s(q) \leq \kappa_1 e \quad \text{and} \quad 0 \leq \gamma(q) - \gamma^{\varepsilon}(q) \leq \kappa_2 \varepsilon. \tag{2.4}$$

Proof. From $\langle e, s(q)\rangle \leq 1$ and $s(q) \geq 0$, we know that $s(q)$ is bounded. When $\gamma(q) > 0$ from the complementarity conditions one must have $1 - \langle e, s(q)\rangle = 0$ which implies that there has to be an entry $s_j(q) > 0$ and $\gamma(q) + q_j = 0$. Hence, the solution set $\mathrm{SOL}(q, M)$ is bounded.

By [8, Theorem 3.1.8], we know that $z^{\varepsilon}(q)$ converges to the least norm solution $z(q)$ of $\mathrm{LCP}(q, M)$ as $\varepsilon \downarrow 0$ since the matrix M is positive semi-definite.

If $\|(-q)_+\|_1 = 0$, then $z^{\varepsilon}(q) = z(q) = 0$ for any $\varepsilon > 0$. Hence (2.4) holds for any $\varepsilon > 0$.

When $\|(-q)_+\|_1 > 0$, let

$$\sigma_1 = \min_{1 \leq j \leq n} q_j, \qquad \sigma_2 = \min\{0, \min_{\substack{1 \leq j \leq n \\ q_j \neq \sigma_1}} q_j\} \quad \text{and} \quad \bar{\varepsilon} := \min\{\frac{-\sigma_1 + \sigma_2}{1 - \sigma_2}, 1\}. \tag{2.5}$$

From $\|(-q)_+\|_1 > 0$, there is $\varepsilon_0 > 0$ such that $\|(-q)_+\|_1 \geq -\sigma_1 > \varepsilon_0$. Thus, for any $\varepsilon \in (0, \bar{\varepsilon})$, $\|(-q)_+\|_1 \geq \varepsilon$ and the solution $z^{\varepsilon}(q)$ has the explicit form (2.1) from Lemma 2.2 and $(-\sigma_1 + \sigma_2)/(1 - \sigma_2) \leq \varepsilon_0$. Our next step is to show that $\alpha_j < 0$ if and only if $q_j = \sigma_1$ for $\varepsilon \in (0, \bar{\varepsilon})$ which implies $\mathcal{J} = \{j \,|\, q_j = \sigma_1, \ j = 1, \dots, n\}$ and $\sigma = J\sigma_1$.

Without loss of generality assume that $q_1 \leq q_2 \leq \cdots \leq q_n$. If $q_j = \sigma_1$, then $\alpha_j = \varepsilon^2\sigma_1 - \varepsilon < 0$. Conversely, when $\alpha_j < 0$, from the definition of $\{\alpha_j\}$, one has

$$\alpha_{j+1} - \alpha_j = -q_{j+1} + (j + 1 + \varepsilon^2)q_{j+1} - (j + \varepsilon^2)q_j = (j + \varepsilon^2)(q_{j+1} - q_j) \geq 0. \tag{2.6}$$

Hence, it suffices to show that $\alpha_j \geq 0$ for $q_j \geq \sigma_2$. If $\varepsilon < 1 \leq (-\sigma_1 + \sigma_2)/(1 - \sigma_2)$, then $-\sigma_1 \geq 1 - 2\sigma_2$. For $q_j \geq \sigma_2$, recalling $\sigma_2 \leq 0$,

$$\alpha_j \geq \sum_{i=1}^{j}(\sigma_2 - q_i) + \varepsilon^2\sigma_2 - \varepsilon \geq \sigma_2 - \sigma_1 + \varepsilon^2\sigma_2 - \varepsilon > -\sigma_1 - 1 + 2\sigma_2 \geq 0. \tag{2.7}$$

If $\varepsilon < (-\sigma_1 + \sigma_2)/(1-\sigma_2) \le 1$, then

$$\alpha_j \ge \sum_{i=1}^{j}(\sigma_2 - q_i) + \varepsilon^2\sigma_2 - \varepsilon > \sigma_2 - \sigma_1 + (\frac{\sigma_2 - \sigma_1}{1-\sigma_2})^2\sigma_2 - \frac{\sigma_2 - \sigma_1}{1-\sigma_2}$$

$$= \frac{\sigma_2 - \sigma_1}{1-\sigma_2}(1 - \sigma_2 + \frac{\sigma_2 - \sigma_1}{1-\sigma_2}\sigma_2 - 1) \ge 0. \tag{2.8}$$

Hence, $\alpha_j < 0$ if and only if $q_j = \sigma_1$ for any $\varepsilon \in (0, \bar\varepsilon)$. By Lemma 2.2, for $\varepsilon \in (0, \bar\varepsilon)$, the solution $z^\varepsilon(q)$ of LCP(q, M^ε) has the form

$$s_j^\varepsilon(q) = \begin{cases} (1-\varepsilon\sigma_1)/(J+\varepsilon^2) & \text{if} \quad j \in \mathcal{J}, \\ 0 & \text{if} \quad j \notin \mathcal{J}, \end{cases} \qquad \gamma^\varepsilon(q) = (-J\sigma_1 - \varepsilon)/(J+\varepsilon^2). \tag{2.9}$$

The least norm solution of the LCP(q, M) is the minimizer of the quadratic program

$$\min_{z\ge 0} \tfrac{1}{2}\|z\|^2 \text{ subject to } \sum_{j\in\mathcal{J}} z_j = 1, \;\; z_j = 0,\, j \notin \mathcal{J}, \;\; z_{n+1} = \gamma = -\sigma_1.$$

This least norm solution has the form (cf. the first order optimality conditions):

$$s_j(q) = \begin{cases} J^{-1} & \text{if} \quad j \in \mathcal{J}, \\ 0 & \text{if} \quad j \notin \mathcal{J}, \end{cases} \qquad \gamma(q) = -\sigma_1. \tag{2.10}$$

From (2.9) and (2.10), we easily see that

$$0 \le s_j^\varepsilon(q) - s_j(q) \le (-\sigma_1\varepsilon)/J^2, \quad \text{for } j = 1, \ldots, n,$$

and

$$0 \le \gamma(q) - \gamma^\varepsilon(q) \le (1-\varepsilon\sigma_1)(\varepsilon/J) \le (1-\sigma_1)(\varepsilon/J).$$

Hence (2.4) holds with $\kappa_1 = (-\sigma_1)/J^2$ and $\kappa_2 = (1-\sigma_1)/J$. $\square$

THEOREM 2.4. *For any fixed q, if $\|(-q)_+\|_1 > 0$ then there is $\bar\varepsilon > 0$ such that for any $\varepsilon \in (0, \bar\varepsilon)$, z^ε is differentiable at q. Moreover, if $\min_{1\le i\le n} q_i = q_{i_1}$ is unique then there exists $\hat\varepsilon > 0$ and a neighborhood $\mathcal{N}_q$ of q such that for any $\varepsilon \in (0, \hat\varepsilon)$, $q \mapsto z^\varepsilon(q)$ is linear on $\mathcal{N}_q$. When z^ε is differentiable at q, one has*

$$\nabla z^\varepsilon(q) = -(I - D + DM^\varepsilon)^{-1}D, \tag{2.11}$$

where D is a $n \times n$ diagonal matrix with diagonal entries

$$d_{ii} = \begin{cases} 1 & \text{if } z_i^\varepsilon(q) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. $\|(-q)_+\|_1 > 0$ means there is an $\varepsilon_0 > 0$ such that $\|(-q)_+\|_1 \ge -\sigma_1 > \varepsilon_0$. Consider $\varepsilon \in (0, \bar\varepsilon)$ with $\bar\varepsilon$ as defined by (2.5). From (2.9),

$$z_j^\varepsilon(q) > 0, \quad \text{for} \quad j \in \mathcal{J} \cup \{n+1\}$$

and from (2.3),

$$(M^\varepsilon z^\varepsilon(q))_j + q_j = \gamma^\varepsilon(q) + q_j > 0, \quad \text{for} \quad j \notin \mathcal{J}.$$

Hence the strictly complementarity condition holds at $z^\varepsilon(q)$, that is, there is no j such that $z^\varepsilon_j(q) = (M^\varepsilon z^\varepsilon(q))_j + q_j = 0$. Differentiability of z^ε at q follows from [7, Lemma 2.1].

If there is a unique entry $q_{i_1} = \min_{1\le i\le n} q_i$, then there is a neighborhood $\mathcal{N}_q$ of q such that for any $p \in \mathcal{N}_q$, $\{i\,|p_i = \min_{1\le j\le n} p_j\} = \{i\,|q_i = \min_{1\le j\le n} q_j\}$. Let

$$\hat\sigma_1 = \min_{p\in\mathcal{N}_q}\min_{1\le j\le n} p_j, \qquad \hat\sigma_2 = \min\{0, \min_{p\in\mathcal{N}_q}\min_{\substack{1\le j\le n\\ p_j\ne\sigma}} p_j\} \quad \text{and} \quad \hat\varepsilon = \min\{\frac{-\hat\sigma_1+\hat\sigma_2}{1-\hat\sigma_2}, 1\}.$$

Then for any $\varepsilon \in (0, \hat\varepsilon)$, the strictly complementarity condition holds at $z^\varepsilon(p)$ for any $p \in \mathcal{N}_q$. Using [7, Lemma 2.1] again, we find that z^ε is differentiable at p and the derivative ∇z^ε in (2.11). Hence, z^ε is a linear mapping on $\mathcal{N}_q$. $\square$

Remark. For any fixed $\varepsilon > 0$, M^ε is positive definite. Hence for any q, the LCP(q, M^ε) has a unique solution $z^\varepsilon(q)$ which defines a globally Lipschitz continuous function z^ε on $\mathbb{R}^{n+1}$ [6, 8]. Moreover, by [7, Theorem 2.1], we know that $1/\varepsilon$ is a Lipschitz constant of the solution function z^ε. The solution function $s^\varepsilon(q)$ can be considered as a smoothing function of the indicator function $\mathbb{1}_{(0,\infty)}(u)$ for any $q = (-u, 1)$. To illustrate this, we consider the LCP (1.6) with $n = 1$. Then, the (first) s^ε component of the solution of LCP(q, M^ε) is

$$s^\varepsilon(q) = \begin{cases} (1+\varepsilon u)/(1+\varepsilon^2) & \text{if } u > \varepsilon,\\ u/\varepsilon & \text{if } u \in (0,\, \varepsilon],\\ 0 & \text{if } u \le 0. \end{cases}$$

It is worth noting that for any fixed u

$$\mathbb{1}_{(0,\infty)}(u) = \lim\nolimits_{\varepsilon\downarrow 0} s^\varepsilon(q) = \begin{cases} 1 & \text{if } u > 0,\\ 0 & \text{otherwise.} \end{cases}$$

Moreover, the solution $z^\varepsilon(q)$ of the regularized LCP can be used for the set-valued constraints. In particular, when $n = 1$,

$$\text{(2.12)} \qquad \mathrm{Lim}_{u\to 0,\varepsilon\downarrow 0}\, s^\varepsilon(q) = [0,1] \quad \text{and} \quad \lim\nolimits_{u\downarrow 0,\varepsilon\downarrow 0} s^\varepsilon(q) = \begin{cases} 1 & \text{if } \varepsilon = o(|u|),\\ 0 & \text{if } |u| = o(\varepsilon). \end{cases}$$

Continuous approximation functions have been used to approximate the indicator function in the study of chance constraints [11, 12, 16]

$$\mathrm{Prob}\{c(\boldsymbol{\xi}, x) \le 0\} = \mathbb{E}[\mathbb{1}_{(-\infty,0)} c(\boldsymbol{\xi}, x)] \le \alpha,$$

where $c : \Xi \times \mathbb{R}^\nu \to \mathbb{R}$ and $\alpha \in (0, 1]$. However, these continuous approximation functions cannot easily be implemented to the vector-valued constraints case [10, 14],

$$\mathrm{Prob}\{c_j(\boldsymbol{\xi}, x) = \max_{1\le i\le n} c_i(\boldsymbol{\xi}, x)\} = \mathbb{E}[\mathbb{1}_{\{\max_{1\le i\le n} c_i(\boldsymbol{\xi},x)\}} c_j(\boldsymbol{\xi}, x)] = b_j, \quad j = 1, \dots, n,$$

$c_j : \Xi \times \mathbb{R}^\nu \to \mathbb{R}$ and $b_j \in (0, 1]$, $j = 1, \dots, n$. The LCP approach and its solution $z^\varepsilon(q)$ of the regularized LCP has the ability to deal with vector-valued constraints.

3. Convergence analysis of the SAA regularized problem. In this section, we study the convergence of the SAA regularized method. The objective function of all three problems (1.4), (1.8) and (1.9) is $f(x) = \frac{1}{2}\langle x, Hx\rangle + \langle c, x\rangle$, their feasible sets,

$$\begin{aligned} D &= \{x \in X \mid A_t\mathbb{E}[S_t(\boldsymbol{\xi}, x)] \ni b_t, \quad t = 1, \dots, T\}, \\ D^\varepsilon &= \{x \in X \mid \|A_t\mathbb{E}[s_t^\varepsilon(\boldsymbol{\xi}, x)] - b_t\| \le r(\varepsilon), \quad t = 1, \dots, T\}, \\ D_N^\varepsilon &= \{x \in X \mid \|A_t\frac{1}{N}\sum\nolimits_{i=1}^N s_t^\varepsilon(\xi^i, x) - b_t\| \le \hat{r}(\varepsilon, N), \quad t = 1, \dots, T\}, \end{aligned}$$

and their solution sets,

$$X^* = \mathrm{argmin}_D\, f, \qquad X^\varepsilon = \mathrm{argmin}_{D^\varepsilon}\, f, \qquad X_N^\varepsilon = \mathrm{argmin}_{D_N^\varepsilon}\, f.$$

$\mathbb{B}_\varepsilon^o = \{\, y \mid \|y\| < \varepsilon\}$ will always denote an open ball centered at 0 with radius ε (in $\mathbb{R}^n$ or $\mathbb{R}^\nu$) and $\mathbb{B}_\varepsilon$ the corresponding closed ball.

In Subsection 3.1, we derive the convergence of X^ε to X^* as $\varepsilon \downarrow 0$, in Subsection 3.2 we obtain the convergence of X_N^ε to X^ε for any fixed $\varepsilon > 0$ as $N \to \infty$ and proceed to deduce the convergence of the solutions of the SAA regularized problems by showing the convergence of X_N^ε to X^* as $\varepsilon \downarrow 0$ and $N \to \infty$.

Denote by $d(v, U) = \inf_{u\in U} \|v - u\|$ the distance from v to a set $U \subseteq \mathbb{R}^n$ and for $U, V \subseteq \mathbb{R}^n$, the excess distance of the set U on V and the Pompeiu-Hausdorff distance between U and V by

$$\mathbb{e}(V, U) = \sup\nolimits_{v\in V} d(v, U) \text{ and } \mathbb{h}(U, V) = \max\big(\mathbb{e}(V, U), \mathbb{e}(U, V)\big).$$

3.1. Problems (1.4) and (1.8). Here, we show the convergence of X^ε to X^* as $\varepsilon \downarrow 0$. For simplicity's sake, in this section and next one, we drop the index t and set $A = A_t$, $S = S_t$ and so on. Moreover, we use $z^\varepsilon(q)$ and $z^\varepsilon(\xi, x)$ to denote $z^\varepsilon(q(\xi, x))$ as well as their components s^ε and γ^ε.

Remember that the solution set $\{z^\varepsilon(\xi, x)\} = \mathrm{SOL}(q(\xi, x), M^\varepsilon)$ is a singleton and the solution set $Z^0(\xi, x) = \mathrm{SOL}(q(\xi, x), M)$ is convex and bounded for any (ξ, x). By Theorem 2.3, for every (ξ, x), one has

$$\lim_{\varepsilon\downarrow 0} \|z^\varepsilon(\xi, x) - \bar{z}^0(\xi, x)\| = 0,$$

where $\bar{z}^0(\xi, x)$ is the least-norm solution of the LCP$(q(\xi, x), M)$, implying the pointwise convergence

$$\lim\nolimits_{\varepsilon\downarrow 0} d(z^\varepsilon(\xi, x), Z^0(\xi, x)) = 0. \tag{3.1}$$

However, from our Remark at the end of the previous section, we already know that for some particular choices of $\varepsilon_k \downarrow 0$, $x^k \to x$, $z^{\varepsilon_k}(\xi, x^k)$ may not converge to $\bar{z}^0$, the least-norm solution of LCP$(q(\xi, x), M)$. Our *predominant motivation*, however, is to show that the solutions of the approximating problems converge to the solutions of the given problem and, in the process, establish the convergence of the feasible sets D^ε and solution sets X^ε to D and X^*. To do this, we are naturally led to study of the graphical convergence of the functions z^ε as $\varepsilon \downarrow 0$ rather than their pointwise convergence.

In first part of the arguments that follow, ξ remains fixed and thus it will be convenient to usually ignore the dependence, on ξ, of the functions u, q and the associated solutions functions $z^\varepsilon = (s^\varepsilon, \gamma^\varepsilon)$ and the solution set Z^0, only the dependence on the pair (x, ε) is relevant.

First, we review the definition of *graphical convergence* [15, Definition 5.32] of the function z^ε as $\varepsilon \downarrow 0$. Let $\mathbb{N} = \{1, 2 \dots\}$ be the set of natural numbers, $\mathcal{N}_\infty^\# = \{\text{all subsequences of } \mathbb{N}\}$ and $\mathcal{N}_\infty = \{\text{all indexes} \geq \text{ some } \bar{k}\}$. We use $(x^k, \varepsilon_k) \underset{N}{\to} (x, 0)$ to denote $\varepsilon_k \downarrow 0$ and $x^k \to x$ when $k \in N$.

Definition 3.1. *For the mappings $z^\varepsilon : X \to \mathbb{R}^{n+1}$, the graphical outer limit, denoted by* $\text{g-limsup}_\varepsilon z^\varepsilon : X \rightrightarrows \mathbb{R}^{n+1}$ *is the mapping having as its graph the set* $\text{Limsup}_\varepsilon \operatorname{gph} z^\varepsilon$*:*

$$\text{g-limsup}_\varepsilon z^\varepsilon(x) = \{z \mid \exists N \in \mathcal{N}_\infty^\#,\ (x^k, \varepsilon_k) \underset{N}{\to} (x, 0),\ z^{\varepsilon_k}(x^k) \underset{N}{\to} z\}.$$

The graphical inner limit, denoted by $\text{g-liminf}_\varepsilon z^\varepsilon$ *is the mapping having as its graph the set* $\text{Liminf}_\varepsilon \operatorname{gph} z^\varepsilon$*:*

$$\text{g-liminf}_\varepsilon z^\varepsilon(x) = \{z \mid \exists\, N \in \mathcal{N}_\infty,\ (x^k, \varepsilon_k) \underset{N}{\to} (x, 0),\ z^{\varepsilon_k}(x^k) \underset{N}{\to} z\}.$$

If the outer and inner limits coincide, the graphical limit $\text{g-lim}_\varepsilon z^\varepsilon$ *exists and, thus,* $Z^0 = \text{g-lim}_\varepsilon z^\varepsilon$ *if and only if*

$$\text{g-limsup}_\varepsilon z^\varepsilon \subseteq Z^0 \subseteq \underset{\varepsilon}{\text{g-liminf}}\, z^\varepsilon$$

and one writes $z^\varepsilon \xrightarrow{g} Z^0$; the mappings z^ε are said to converge graphically *to Z^0.*

Theorem 3.2. *If, given ξ, $x \mapsto u(\xi, x)$ is continuous on X then,* $\text{g-limsup}_\varepsilon s^\varepsilon \subseteq S^0$ *as well as* $\text{g-limsup}_\varepsilon z^\varepsilon \subset Z^0$*. If $u(\xi, \cdot)$ is also surjective, then the functions s^ε converge graphically to S^0, that is,*

$$s^\varepsilon \xrightarrow{g} S^0 \text{ as well as } z^\varepsilon \xrightarrow{g} Z^0. \tag{3.2}$$

Proof. Remember, throughout the proof, $\xi \in \Xi$ remains fixed. For a subsequence $N \in \mathcal{N}_\infty^\#$, let $(x^k, \varepsilon_k) \underset{N}{\to} (x, 0)$ with $\varepsilon_k \downarrow 0$, $u^k = u(\xi, x^k)$, $q^k = (-u^k, 1)$, $z^k = z^{\varepsilon_k}(q^k)$ and suppose $z^k \underset{N}{\to} z^0$; for any pair (ξ, x^k), the vector z^k is uniquely defined, cf. Lemma 2.2. Moreover, for any $\varepsilon_k > 0$, $\text{LCP}(q^k, M^{\varepsilon_k})$ has a unique solution $z^k(\xi, x^k) = (s^k, \gamma^{\varepsilon_k})$ that is also a unique solution of the system of (continuous) equations:

$$\text{Min}\ [\, z,\ M^{\varepsilon_k} z + q^k\,] = 0,$$

where "Min" has to be understood componentwise. Hence, with $M^k = M^{\varepsilon_k}$, one has

$$0 = \lim_k \text{Min}\ [z^k,\ M^k z^k + q^k] = \text{Min}\ [\lim_k z^k,\ \lim_k M^k z^k + q^k] = \text{Min}\ [z^0,\ M z^0 + q(\xi, x)]$$

which means $z^0 \in Z^0$ and consequently $\text{g-limsup}_k z^k \subseteq Z^0$ and, in particular, the same applies to the first n-entries of the z^k and z^0, i.e., $s^0 \in S^0 \supseteq \text{g-limsup}_k \{s^k\}$.

We now concern ourselves with the second assertion of the lemma. For $\tilde{x} \in X$, let $z(\tilde{x}) = \big(s(\tilde{x}), \gamma(\tilde{x})\big) \in Z^0(\xi, \tilde{x})$. We need to show that $z(\tilde{x}) \in \text{g-liminf}_\varepsilon \{z^\varepsilon\}$. The surjective property of $u(\xi, \cdot) : X \to \mathbb{R}^n$ implies that for any $q^k = (-u^k, 1)$, there is $x^k \in X$ such that $q^k = \big(-u(\xi, x^k)\big, 1)$. Let $\tilde{q} = (-u(\xi, x), 1)$, $\tilde{z} = z(\tilde{q})$ and show that

$$\exists\, N \in \mathcal{N}_\infty,\ (q^k, \varepsilon_k) \underset{N}{\to} (\tilde{q}, 0),\ \ z^{\varepsilon_k}(q^k) = z^k \underset{N}{\to} \tilde{z} = z(\tilde{q}). \tag{3.3}$$

Let $\eta_0 = \max_{1 \leq i \leq n}(u_i(\xi, x) = -\tilde{q}_i)$. To prove (3.3), we examine all three cases: $\eta_0 > 0$, $\eta_0 = 0$ and $\eta_0 < 0$.

Case 1. $\eta_0 > 0$. Without loss of generality, assume

$$\tilde{q}_1 = \cdots = \tilde{q}_J < \tilde{q}_i,\ i = J+1, \ldots, n, \quad \text{and} \quad s_1(\tilde{q}) \geq \ldots \geq s_J(\tilde{q}) \tag{3.4}$$

which implies

$$\sum_{i=1}^{J} s_i(\tilde{q}) = 1, \ \ s_i(\tilde{q}) \geq 0,\ i = 1, \ldots, J, \quad s_i(\tilde{q}) = 0, \quad i = J+1, \ldots, n,$$

and $\gamma(\tilde{q}) = \tilde{q}_1$. Choose a sequence $\varepsilon_k \downarrow 0$. Then, for some $\tilde{k}$,

$$\forall\, k \geq \tilde{k}, \quad J(\tilde{q}_{J+1} - \tilde{q}_1) + \varepsilon_k^2 \tilde{q}_{J+1} - \varepsilon_k > 0 \text{ and } -\tilde{q}_1 > \varepsilon_k, \tag{3.5}$$

which implies $\eta_0 > \varepsilon_k$. Let

$$\begin{aligned} q_i^k &= \tilde{q}_i - \lambda_i \varepsilon_k, \text{ with } \lambda_i = \big(J s_i(\tilde{q}) - 1\big)/J, \quad i = 1, \ldots, J, \\ q_i^k &= \tilde{q}_i, i = J+1, \ldots, n. \end{aligned}$$

From (3.4) and $q_i^k = \tilde{q}_i - \lambda_i \varepsilon_k$, one obtains

$$q_1^k \leq \ldots \leq q_J^k \leq \tilde{q}_1 + \varepsilon_k J^{-1} \leq -\eta_0 + \varepsilon_k < 0.$$

Note that, since $\sum_{i=1}^{J} \lambda_i = 0$,

$$\|(-q^k)_+\|_1 \geq \sum_{i=1}^{J} -q_i^k = -J\tilde{q}_1 + \varepsilon \sum_{i=1}^{J} \lambda_i = -J\tilde{q}_1 \geq \eta_0 > \varepsilon_k.$$

Now, apply Lemma 2.2(a) to obtain the solution $z^{\varepsilon_k}(q^k)$ for $k \geq \tilde{k}$.

Since $\tilde{q}_1 = \cdots = \tilde{q}_J = -\eta_0$, from $\varepsilon_k < \eta_0$, $\sum_{i=1}^{J} \lambda_i = 0$ and $-J^{-1} \leq \lambda_J \leq 0$,

$$\alpha_J^k = \varepsilon_k^2(\tilde{q}_1 - \lambda_J \varepsilon_k) - J\lambda_J \varepsilon_k - \varepsilon_k \leq 0.$$

Moreover, from (3.5), we obtain

$$\alpha_{J+1}^k = J(\tilde{q}_{J+1} - \tilde{q}_1) + \varepsilon_k^2 q_{J+1} - \varepsilon_k > 0.$$

Using $\alpha_1^k \leq \ldots \leq \alpha_n^k$ for $k \geq \tilde{k}$, yields

$$\sigma^k = \sum_{i=1}^{J} q_i^k = J\tilde{q}_1 - \sum_{i=1}^{J} \lambda_i \varepsilon_k = J\tilde{q}_1, \quad k \geq \tilde{k},$$

and for $k \geq \tilde{k}$,

$$s_i^{\varepsilon_k}(q^k) = \frac{J\tilde{q}_1 - (J + \varepsilon_k^2)(\tilde{q}_1 - \lambda_i \varepsilon_k) + \varepsilon_k}{J\varepsilon_k + \varepsilon_k^3} = \frac{J\lambda_i + \varepsilon_k(\tilde{q}_1 - \lambda_i \varepsilon_k) + 1}{J + \varepsilon_k^2},\ i = 1, \ldots, J$$

and

$$s_i^{\varepsilon_k}(q^k) = 0, \quad i = J+1, \ldots, n.$$

When $k \to \infty$, $s_i^{\varepsilon_k}(q^k) \to \lambda_i + J^{-1} = s_i(\tilde{q})$, for $i = 1, \ldots, J$ $s_i^{\varepsilon_k}(q^k) \to 0$, for $i = J+1, \ldots, n$, and $\gamma^{\varepsilon_k}(q^k) = (\sigma^k - \varepsilon_k)/(J + \varepsilon_k^2) \to \tilde{q}_1 = \gamma(\tilde{q})$ i.e., $s^{\varepsilon_k}(q^k) \to s(\tilde{q})$ and $z^{\varepsilon_k}(q^k) \to z(\tilde{q})$.

Case 2. $\eta_0 = 0$. Without loss of generality, assume

$$0 = \tilde{q}_1 = \ldots = \tilde{q}_J < \tilde{q}_i,\ i = J+1, \ldots, n, \quad \text{and} \quad s_1(\tilde{q}) \geq \ldots \geq s_J(\tilde{q}) \tag{3.6}$$

which implies that

$$\sum_{i=1}^{J} s_i(\tilde{q}) \leq 1, s_i(\tilde{q}) \geq 0, \quad i = 1, \ldots, J \quad \text{and} \quad s_i(\tilde{q}) = 0, \quad i = J+1, \ldots, n.$$

Choose $\varepsilon_k \downarrow 0$ and let

$$q_i^k = -s_i(\tilde{q})\varepsilon_k, \quad i = 1, \ldots, J \quad \text{and} \quad q_i^k = \tilde{q}_i, \quad i = J+1, \ldots, n.$$

Since $\sum_{i=1}^{J} s_i(\tilde{q}) \leq 1$, one has $\sum_{i=1}^{J}(-q_i^k) = \varepsilon_k \sum_{i=1}^{J} s_i(\tilde{q}) \leq \varepsilon_k$ and $q_i^k = \tilde{q}_i > 0, i = J+1, \ldots, n$. Now, apply Lemma 2.2(b) to obtain the solution

$$s_i^{\varepsilon_k}(q^k) = (s_i(\tilde{q})\varepsilon_k)/\varepsilon_k, \quad i = 1, \ldots, J \quad \text{and} \quad s_i(\tilde{q}) = 0, \quad i = J+1, \ldots, n.$$

Obviously, when $k \to \infty$, $s_i^{\varepsilon_k}(q^k) \to s_i(\tilde{q})$, $i = 1, \ldots, n$, and $\gamma^{\varepsilon_k} = 0$, entailing $z^{\varepsilon_k}(q^k) \to z(\tilde{q})$.

Case 3. $\eta_0 < 0$. In this case $z(\tilde{q}) = z^{\varepsilon_k}(q^k) = 0$ for $q^k = \tilde{q}$ and $\varepsilon_k > 0$.

Together, cases 1-3 in the second part of the proof, yield $Z^0 \subseteq \text{g} - \text{limsup}_\varepsilon z^\varepsilon$.

Combining the two parts of the proof, yields $z^\varepsilon \xrightarrow{g} Z^0$ and $s^\varepsilon \xrightarrow{g} S^0$. ☐

Note that when the set $\{j \mid u_j(\xi, x) = \max_{1 \leq i \leq n} u_i(\xi, x))\}$ is a singleton, then both sets $\{s^\varepsilon(\xi, x)\}$ and $S^0(\xi, x)$ are singletons. In such a case, $\text{g-lim sup}_\varepsilon\, s^\varepsilon(\xi, x) = S^0(\xi, x)$.

THEOREM 3.3. *Assume $u : \Xi \times X \to \mathbb{R}^n$ is continuous and bounded, then $\mathbb{e}(D^\varepsilon, D) \to 0$ as $\varepsilon \downarrow 0$.*

Proof. Let $z^\varepsilon : \Xi \times X \to \mathbb{R}^{n+1}$ be the single valued function and $Z : \Xi \times X \rightrightarrows \mathbb{R}^{n+1}$ be the set-valued function such that for any (ξ, x), $z^\varepsilon(\xi, x)$ is the unique solution of $\text{LCP}(q(\xi, x), M^\varepsilon)$ and $Z(\xi, x)$ is the solution set of $\text{LCP}(q(\xi, x), M)$. By Theorem 3.2, $\text{g-limsup}_\varepsilon\, z^\varepsilon \subset Z^0$.

Let $\varepsilon_k \downarrow 0$ and $x^k \in D^{\varepsilon_k}$. We establish that any cluster point, say $\bar{x}$, of $\{x^k\}$, is in D. From the boundedness of q and Lemma 2.2, we know that $s^{\varepsilon_k}(\xi, x^k)$ is bounded. Hence,

$$\lim_{\varepsilon_k \downarrow 0} \mathbb{E}[s^{\varepsilon_k}(\boldsymbol{\xi}, x^k)] = \mathbb{E}[\lim_{\varepsilon_k \downarrow 0} s^{\varepsilon_k}(\boldsymbol{\xi}, x^k)] \subseteq \mathbb{E}[S(\boldsymbol{\xi}, \bar{x})], \tag{3.7}$$

where the equality comes from the Dominated Convergence Theorem and the inclusion from Theorem 3.2. The sequence $\{\varepsilon_k\}$ being arbitrary, (3.7) implies

$$\text{g-limsup}_\varepsilon\, \mathbb{E}[s^\varepsilon(\boldsymbol{\xi}, x)] \subseteq \mathbb{E}[S(\boldsymbol{\xi}, x)].$$

From $x^k \in D^{\varepsilon_k}$, $x^k \in X$, $A\mathbb{E}[s^{\varepsilon_k}(\boldsymbol{\xi}, x^k)] + \mathbb{B}^o_{r(\varepsilon_k)} \ni b$, X compact and [15, Theorem 5.37], one has $\bar{x} \in X$ and $A\mathbb{E}[S(\boldsymbol{\xi}, \bar{x})] \ni b$, i.e., $\bar{x} \in D$ and $\mathbb{e}(D^\varepsilon, D) \to 0$ as $\varepsilon \downarrow 0$. ☐

ASSUMPTION 1. *For any $\delta > 0$, there exists an $\bar{\varepsilon} > 0$ such that for any $\varepsilon \in [0, \bar{\varepsilon}]$, $D^\varepsilon \cap (X^* + \mathbb{B}_\delta) \neq \emptyset$.*

THEOREM 3.4. *Suppose Assumption 1 holds and q is bounded. Then we have*

$$\lim_{\varepsilon\downarrow 0}\min_{x\in D^\varepsilon} f(x) = \min_{x\in D} f(x) \tag{3.8}$$

$$\operatorname*{Limsup}_{\varepsilon\downarrow 0} \operatorname*{argmin}_{x\in D^\varepsilon} f(x) \subseteq \operatorname*{argmin}_{x\in D} f(x) \tag{3.9}$$

Proof. The objective function f is a quadratic convex function and independent of ε. We need only consider the limiting behavior of the feasible set D^ε as $\varepsilon \downarrow 0$.

Define a set-valued mapping $\mathcal{D} : [0, \bar{\varepsilon}] \rightrightarrows \mathbb{R}^n$ with $\mathcal{D}(\varepsilon) = D^\varepsilon$ and $\mathcal{D}(0) = D$. Since for every $\varepsilon \in [0, \bar{\varepsilon}]$, D^ε and D are closed, $\mathcal{D}$ is a closed-valued mapping. Moreover, by Theorem 3.3, $\mathcal{D}$ is outer semicontinuous or, equivalently [15, Theorem 5.7], gph$\mathcal{D}$ is closed.

Note that $D^\varepsilon \subseteq X$ for all $\varepsilon > 0$ and X is compact. Hence, from Assumption 1, we obtain the assertions (3.8) and (3.9) from [5, Proposition 4.4]. $\square$

Theorem 3.4 means that under Assumption 1, the optimal value function $v_\varepsilon := \min_{x\in D^\varepsilon} f(x)$ is continuous at $\varepsilon = 0$ and the optimal solution set X^ε is outer semicontinuous at $\varepsilon = 0$. Assumption 1 is related to Robinson's constraint qualification and often used in perturbation analysis of optimization problem [5]. In the following, we present a sufficient condition for Assumption 1, and the existence of solutions of the MPSLCC (1.4), the regularized problem (1.8) and its associated SAA problem (1.9).

For a fixed feasible solution $\hat{x}$ of problem (1.4), let's define

$$\sigma_1(\xi) := \min_{1\le j\le n} q_j(\xi, \hat{x}), \qquad \sigma_2(\xi) := \min\{0, \min_{\substack{1\le j\le n \\ q_j(\xi,\hat{x})\neq\sigma_1(\xi)}} q_j(\xi, \hat{x})\},$$

and

$$\Xi_\varepsilon := \{\xi \in \Xi \,|\, \sigma_2(\xi) - \sigma_1(\xi) \ge \varepsilon(1 + \tau_0) \ \text{ or } \ \sigma_1(\xi) \ge 0\},$$

where $\tau_0 := -\min_{\xi\in\Xi}\{\sigma_1(\xi), 0\}$. By the continuity of u, the functions σ_1, σ_2 and $\sigma_2 - \sigma_1$ are continuous on Ξ. Note that the measure $P(\Xi_0) = P(\Xi)$ and $P(\Xi_\varepsilon)$ is continuous at $\varepsilon = 0$ when the density function is continuous or the support of $\boldsymbol{\xi}$ is finite, i.e., $|\Xi|$ is finite. Hence there is a continuous function $\tilde{r}$ on the interval $[0, \bar{\varepsilon}]$ for sufficiently small $\bar{\varepsilon} > 0$ such that

$$P(\Xi_\varepsilon) \ge 1 - \tilde{r}(\varepsilon), \quad \text{with} \quad \lim_{\varepsilon\downarrow 0} \tilde{r}(\varepsilon) = 0. \tag{3.10}$$

THEOREM 3.5. *Assume that there exists a feasible solution $\hat{x}$ of problem (1.4) such that $A\mathbb{E}[s(\boldsymbol{\xi}, \hat{x})] = b$ and $z(\xi, \hat{x}) = \big(s(\xi, \hat{x}), \gamma(\xi, \hat{x})\big)$ is the least norm solution of the LCP$(q(\xi, \hat{x}), M)$. Then problems (1.4) and (1.8) are solvable with $r(\varepsilon) \ge n\|A\|(\tau_0\varepsilon + 2\tilde{r}(\varepsilon))$, where $\tau_0 = -\min_{\xi\in\Xi}\{\sigma_1(\xi), 0\}$. Moreover,*

$$\min_{x\in D} f(x) \le \liminf_{\varepsilon\downarrow 0} \min_{x\in D^\varepsilon} f(x). \tag{3.11}$$

If the feasible solution $\hat{x}$ is an optimal solution, then Assumption 1 holds.

Proof. For any $\xi \in \Xi$, the solution set of the LCP$(q(\xi, \hat{x}), M)$ is bounded. From Theorem 3.1.8 in [8], the solution $z^\varepsilon(\xi, \hat{x})$ of the LCP$(q(\xi, \hat{x}), M^\varepsilon)$ converges to the least norm solution of LCP$(q(\xi, \hat{x}), M)$ as $\varepsilon \downarrow 0$.

To show $\hat{x} \in D^\varepsilon$, we first prove that for any $\varepsilon \in (0, 1)$,

$$0 \leq s^\varepsilon(\xi, \hat{x}) - \bar{s}(\xi, \hat{x}) \leq (\tau_0 \varepsilon) e, \quad \text{for} \quad \forall\, \xi \in \Xi_\varepsilon. \tag{3.12}$$

We prove (3.12) by consider two cases: $\sigma_1(\xi) < 0$ and $\sigma_1(\xi) \geq 0$.

Case 1. $\sigma_1(\xi) < 0$.

Since $\sigma_1(\xi) < 0$, by definition of $\sigma_1(\xi)$ and $\sigma_2(\xi)$ above, we know that

$$\varepsilon \leq \frac{\sigma_2(\xi) - \sigma_1(\xi)}{1 + \tau_0} \leq \frac{\sigma_2(\xi) - \sigma_1(\xi)}{1 - \sigma_2(\xi)}, \qquad \forall \xi \in \Xi_\varepsilon.$$

Let us define

$$\mathcal{J}(\xi) = \{ j \,| q_j(\xi, \hat{x}) = \sigma_1(\xi), j = 1, \ldots, n \} \quad \text{and} \quad J(\xi) = |\mathcal{J}(\xi)|.$$

Following the proof of Theorem 2.3 (See (2.5) and (2.9)), we can show that the solution $z^\varepsilon(\xi, \hat{x})$ of LCP$(q(\xi, \hat{x}), M^\varepsilon)$ has the following form

$$s_j^\varepsilon(\xi, \hat{x}) = \begin{cases} \frac{1 - \varepsilon \sigma_1(\xi)}{J(\xi) + \varepsilon^2} & \text{if} \quad j \in \mathcal{J}(\xi), \\ 0 & \text{if} \quad j \notin \mathcal{J}(\xi), \end{cases} \qquad \gamma^\varepsilon(\xi, \hat{x}) = \frac{-J(\xi)\sigma_1(\xi) - \varepsilon}{J(\xi) + \varepsilon^2}. \tag{3.13}$$

The least norm solution $\bar{s}(\xi, \hat{x}) = \text{argmin}_{s \in S(\xi, \hat{x})} \|y\|_2$ is the first n-components of the least norm solution of LCP$(q(\xi, \hat{x}), M^\varepsilon)$, which has the form

$$\bar{s}_j(\xi, \hat{x}) = \begin{cases} \frac{1}{J(\xi)}, & \text{if} \quad j \in \mathcal{J}(\xi), \\ 0, & \text{if} \quad j \notin \mathcal{J}(\xi), \end{cases} \qquad \bar{\gamma}(\xi, \hat{x}) = -\sigma_1(\xi).$$

Hence, for $\xi \in \Xi_\varepsilon$, we derive

$$0 \leq s_i^\varepsilon(\xi, \hat{x}) - \bar{s}_i(\xi, \hat{x}) \leq \frac{1 - \varepsilon \sigma_1(\xi)}{J(\xi) + \varepsilon^2} - \frac{1}{J(\xi)} \leq \frac{-\varepsilon \sigma_1(\xi)}{J(\xi)} \leq -\varepsilon \sigma_1(\xi) \leq \tau_0 \varepsilon.$$

For case 2, it is easy to show that the solution $z^\varepsilon(\xi, \hat{x})$ of LCP$(q(\xi, \hat{x}), M^\varepsilon)$ has the following form

$$z_j^\varepsilon(\xi, \hat{x}) = 0, \; j = 1, \ldots, n+1,$$

and it is just the least norm solution of LCP$(q(\xi, \hat{x}), M)$, which means $s_i^\varepsilon(\xi, \hat{x}) - \bar{s}_i(\xi, \hat{x}) = 0$.

Combining cases 1 and 2, we have (3.12).

Now, for sufficiently small ε, we consider the expected value

$$\begin{aligned} & |\mathbb{E}[s^\varepsilon(\boldsymbol{\xi}, \hat{x}) - \bar{s}(\boldsymbol{\xi}, \hat{x})]| \\ & = |\mathbb{E}[\mathbf{1}_{\{\boldsymbol{\xi} \in \Xi_\varepsilon\}}(s^\varepsilon(\boldsymbol{\xi}, \hat{x}) - \bar{s}(\boldsymbol{\xi}, \hat{x}))] + \mathbb{E}[\mathbf{1}_{\{\boldsymbol{\xi} \notin \Xi_\varepsilon\}}(s^\varepsilon(\boldsymbol{\xi}, \hat{x}) - \bar{s}(\boldsymbol{\xi}, \hat{x}))]| \\ & \leq \big(\tau_0 \varepsilon + 2\tilde{r}(\varepsilon)\big) e, \end{aligned}$$

where the last inequality uses the explicit form $s^\varepsilon(\xi,\hat{x})$ in Lemma 2.2, and (3.10) with $0 \le s^\varepsilon(\xi,\hat{x}) \le 2e, 0 \le \bar{s}(\xi,\hat{x}) \le e$ and

$$|s^\varepsilon(\xi,\hat{x}) - \bar{s}(\xi,\hat{x})| \le \max\{s^\varepsilon(\xi,\hat{x}), \bar{s}(\xi,\hat{x})\} \le 2e.$$

Hence, we have

$$\|A\mathbb{E}[s^\varepsilon(\boldsymbol{\xi},\hat{x})] - b\| = \|A\mathbb{E}[s^\varepsilon(\boldsymbol{\xi},\hat{x}) - \bar{s}(\boldsymbol{\xi},\hat{x})]\| \le \|A\|\|e\|(\tau_0\varepsilon + 2\tilde{r}(\varepsilon)) \le r(\varepsilon),$$

which implies that $\hat{x} \in D^\varepsilon$.

Since problems (1.4) and (1.8) have the same objective function that is continuous, the feasibility of the two problems implies their solvability.

If $\hat{x} \in X^*$, then Assumption 1 holds from $\hat{x} \in D^\varepsilon$. □

Remark. The set-valued constraints in (1.2) can also be approximated by a sequence of equality constraints via regularized quadratic programs with unique solutions for fixed $\varepsilon > 0$:

$$\max_y \langle y, u\rangle + \varepsilon\langle y, y\rangle \qquad \text{subject to } \langle e, y\rangle \le 1,\ y \ge 0.$$

However, the solution of the KKT-conditions is not unique and the peremptory results can't be derived from the 'regularized' system

$$\text{Min}\ [\, z,\ \hat{M}^\varepsilon z + q(\xi, x)\,] = 0,$$

where $\hat{M}^\varepsilon = \begin{pmatrix} \varepsilon I & e \\ -e & 0 \end{pmatrix}$ is positive semi-definite and I is $n \times n$ identity matrix. The novel idea in Theorem 3.2 is that we use the well-established theory for monotone LCP to derive the required properties of the regularized solution of z^ε and, in particular, its first n-components s^ε.

3.2. Problems (1.4) and (1.9). In this subsection, we consider the convergence of the solution set X_N^ε of (1.9) to the solution set X^* of problem (1.4) as $\varepsilon \downarrow 0$ and $N \to \infty$. First, we consider the convergence of the solution set X_N^ε of problem (1.9) to the solution set X^ε of problem (1.8) as $N \to \infty$ for a fixed $\varepsilon > 0$. Next, we use this convergence result with Theorem 3.4 to derive the convergence of X_N^ε to X^* as $\varepsilon \downarrow 0$ and $N \to \infty$.

Let v_ε and v_N^ε be the optimal values of problems (1.8) and (1.9).

Assumption 2. *There exists a measurable function $C : \Xi \to (0, +\infty)$ such that $\mathbb{E}[C(\boldsymbol{\xi})^2] < \infty$ and*

$$\|u(\xi, x) - u(\xi, \bar{x})\| \le C(\xi)\|x - \bar{x}\|$$

for all $x, \bar{x} \in X$ and P-almost every $\xi \in \Xi$.

Proposition 3.6. *Let $\hat{r}(\varepsilon, N) := r(\varepsilon) + cN^{-\tau}$, where $\tau \in (0, \frac{1}{2})$ and c is a positive constant. Suppose that the samples are iid and Assumption 2 holds. Moreover, there is η such that $\|(u(\xi, x))_+\|_1 \le \eta$ for $x \in X, \xi \in \Xi$. Then there exists an $\bar{\varepsilon} > 0$ such that the following statements hold for any $\varepsilon \in (0, \bar{\varepsilon}]$.*

(i) For N sufficiently large, $D^\varepsilon \subset D_N^\varepsilon$;

(ii) For any $\Delta > 0$ there exists a sufficiently large N_Δ such that $\mathbb{h}(D^\varepsilon, D^\varepsilon_N) \le \Delta$ holds w.p.1 for $N \ge N_\Delta$;

(iii) $v^\varepsilon_N \to v_\varepsilon$ and $\mathbb{e}(X^\varepsilon_N, X^\varepsilon) \to 0$ w.p.1 as $N \to \infty$.

Proof. (i) Since X is a compact subset of $\mathbb{R}^\nu$, by the continuity of $u(\xi,\cdot)$, $z^\varepsilon(\xi,\cdot)$ is globally Lipschitz continuous on X for almost every $\xi \in \Xi$. Moreover, by Lemma 2.2, $\|z^\varepsilon(\cdot,\cdot)\| \le 1+\varepsilon\eta$. Then by the classical uniform law of large numbers ([17, Proposition 7, Section 6]), we have $\frac{1}{N}\sum_{i=1}^N s^\varepsilon(\xi^i, x) \to \mathbb{E}[s^\varepsilon(\boldsymbol{\xi}, x)]$ uniformly w.p.1 as $N \to \infty$.

By the Remark following Theorem 2.4 and Assumption 2,

$$\|s^\varepsilon(\xi,x) - s^\varepsilon(\xi,\bar{x})\| \le \frac{1}{\varepsilon}\|u(\xi,x) - u(\xi,\bar{x})\| \le \frac{1}{\varepsilon}C(\xi)\|x - \bar{x}\|$$

and $\mathbb{E}[C(\xi)^2] < \infty$. Moreover, for all $\xi \in \Xi$, $s^\varepsilon(\xi, x)$ is uniformly bounded. Then the mean and variance of random variables $s^\varepsilon(\xi,x)$ are finite for all $x \in X$. By [17, Chapter 6] and the functional central limit theorem [2, Corollary 7.17],

$$\left\| \frac{1}{N}\sum_{i=1}^N s^\varepsilon(\xi^i, x) - \mathbb{E}[s^\varepsilon(\boldsymbol{\xi}, x)] \right\| = O_p(\frac{1}{\sqrt{N}}).$$

By Assumption 1, $D^\varepsilon \ne \emptyset$. Then for all $x \in D^\varepsilon$, there exists sufficiently large N_0, such that, when $N \ge N_0$,

(3.14)
$$\begin{array}{rcl} \left\|\frac{1}{N}A\sum_{i=1}^N s^\varepsilon(\xi^i,x) - b\right\| & \le & \left\|\frac{1}{N}A\sum_{i=1}^N s^\varepsilon(\xi^i,x) - A\mathbb{E}[s^\varepsilon(\boldsymbol{\xi},x)] + A\mathbb{E}[s^\varepsilon(\boldsymbol{\xi},x)] - b\right\| \\ & \le & \left\|\frac{1}{N}A\sum_{i=1}^N s^\varepsilon(\xi^i,x) - A\mathbb{E}[s^\varepsilon(\boldsymbol{\xi},x)]\right\| + r(\varepsilon) \\ & \le & cN^{-\tau} + r(\varepsilon) \\ & = & \hat{r}(\varepsilon, N) \end{array}$$

w.p.1, which implies that $x \in D^\varepsilon_N$ w.p.1.

(ii) Let $\Delta > 0$ and

$$\delta(\Delta) := \inf_{\{x \in X : d(x, D^\varepsilon) \ge \Delta\}} (||A\mathbb{E}[s^\varepsilon(\boldsymbol{\xi},x)] - b|| - r(\varepsilon))_+.$$

By the compactness of X and the continuity of $s^\varepsilon(\xi,\cdot)$, we have $\delta(\Delta) > 0$.

Let N_Δ be sufficiently large such that

$$\sup_{x\in X}\left\|\frac{1}{N}A\sum_{i=1}^N s^\varepsilon(\xi^i,x) - A\mathbb{E}[s^\varepsilon(\boldsymbol{\xi},x)]\right\| \le \frac{\delta(\Delta)}{2}$$

and $cN^{-\tau} < \frac{\delta(\Delta)}{2}$. For any point $x \in X$ with $d(x, D^\varepsilon) \ge \Delta$, we have

$$\|A\mathbb{E}[s^\varepsilon(\boldsymbol{\xi},x)] - b\| \ge r(\varepsilon) + \delta(\Delta),$$

which implies

(3.15)
$$\begin{array}{rl} & ||\frac{1}{N}A\sum_{i=1}^N s^\varepsilon(\xi^i,x) - b|| \\ \ge & ||A\mathbb{E}[s^\varepsilon(\boldsymbol{\xi},x)] - b|| - \left\|\frac{1}{N}A\sum_{i=1}^N s^\varepsilon(\xi^i,x) - A\mathbb{E}[s^\varepsilon(\boldsymbol{\xi},x)]\right\| \\ \ge & r(\varepsilon) + \delta(\Delta) - \frac{\delta(\Delta)}{2} \\ > & r(\varepsilon) + cN^{-\tau} \\ = & \hat{r}(\varepsilon, N). \end{array}$$

This shows that $x \notin D_N^\varepsilon$. Hence for any $x \in D_N^\varepsilon$, $d(x, D^\varepsilon) \le \Delta$, which implies

$$\mathbb{e}(D_N^\varepsilon, D^\varepsilon) \le \Delta.$$

Combining the above result with Part (i), we have $\mathbb{h}(D^\varepsilon, D_N^\varepsilon) \le \Delta$ w.p.1 for $N \ge N_0$.

(iii) We apply [5, Proposition 4.4] to prove this part.

Since problems (1.8) and (1.9) have the same convex quadratic objective function which is independent of ε and ξ, we only need to consider the limit behavior of the feasible set D_N^ε as $N \to \infty$. Let

$$G_\varepsilon(x, N) := \|A \frac{1}{N} \sum_{i=1}^N s^\varepsilon(\xi^i, x) - b\| - r(\varepsilon) - N^{-\tau}.$$

Since $\frac{1}{N}\sum_{i=1}^N s^\varepsilon(\xi^i, x) \to \mathbb{E}[s^\varepsilon(\boldsymbol{\xi}, x)]$ uniformly on X w.p.1 as $N \to \infty$, $G_\varepsilon(x, N) \to \|A\mathbb{E}[s^\varepsilon(\boldsymbol{\xi}, x)] - b\| - r(\varepsilon)$ as $N \to \infty$ uniformly on X w.p.1 and continuous w.r.t. $x \in X$. Hence for any $\bar{x} \in X$,

$$\lim_{N\to\infty, x\to\bar{x}} G_\varepsilon(x, N) = \|A\mathbb{E}[s^\varepsilon(\boldsymbol{\xi}, \bar{x})] - b\| - r(\varepsilon),$$

which implies that the feasible set map D_N^ε with respect to N is closed P-a.s.for sufficiently large N. By part (i) of this proposition, $D^\varepsilon \subseteq X$ and $D_N^\varepsilon \subseteq X$ are nonempty for sufficiently large N. Moreover, from part (ii) of this proposition, we have $\mathbb{h}(D^\varepsilon, D_N^\varepsilon) \to 0$ which implies that for any neighborhood $\mathcal{V}_{X^\varepsilon}$ of X^ε, there exists a sufficiently large N_0 such that for all $N \ge N_0$, $\mathcal{V}_{X^\varepsilon} \cap D_N^\varepsilon \ne \emptyset$. Hence all conditions of [5, Prpposition 4.4] are satisfied, and thus we derive $v_N^\varepsilon \to v_\varepsilon$ and $\mathbb{e}(X_N^\varepsilon, X^\varepsilon) \to 0$. ◻

The proof of part (ii) of Proposition 3.6 is motivated by the proof of [18, Lemma 4.2 (i)].

Now, we are ready to present the convergence of X_N^ε to X^* as $\varepsilon \downarrow 0$ and $N \to \infty$.

Theorem 3.7. *Suppose the conditions of Theorem 3.4 and Proposition 3.6 hold. If the feasible set D is nonempty, then X_N^ε is nonempty and*

$$\lim_{\varepsilon\downarrow 0} \lim_{N\to\infty} \mathbb{e}(X_N^\varepsilon, X^*) = 0 \quad \text{w.p.1.} \tag{3.16}$$

Proof. Since,

$$\begin{aligned} \lim_{\varepsilon\downarrow 0} \lim_{N\to\infty} \mathbb{e}(X_N^\varepsilon, X^*) &\le \lim_{\varepsilon\downarrow 0} \lim_{N\to\infty} \mathbb{e}(X_N^\varepsilon, X^\varepsilon) + \lim_{\varepsilon\downarrow 0} \lim_{N\to\infty} \mathbb{e}(X_\varepsilon, X^*) \\ &= \lim_{\varepsilon\downarrow 0} 0 + \lim_{\varepsilon\downarrow 0} \mathbb{e}(X_\varepsilon, X^*) \quad \text{w.p.1} \\ &= 0 \quad \text{w.p.1}, \end{aligned}$$

the assertion now follows directly from Theorem 3.4 and Proposition 3.6. ◻

4. The pure characteristics demand model. One important application of problem (1.1) is to estimate the parameters of the pure characteristics demand model proposed by Berry and Pakes [3]. Although the model has several advantages in describing markets, it faces serious challenges and difficulties in estimating some key parameters when relying on the generalized method of moments (GMM). Pang, Su and Lee [14] reformulated the GMM estimation problem of the pure characteristics demand model as a computationally tractable quadratic program with linear complementarity constraints; the reformulated GMM estimation problem can be thought as a special case of problem (1.1). To illustrate our SAA regularized approach and the convergence results established in §2 and §3, we consider an example of the pure characteristics demand model:

- T is the number of markets and n the number of products in each market.
- The utility function of product j in market t is:

$$u_t(\xi, x) = c_t\beta(\xi_1, x_1, x_2) - \alpha(\xi_2, x_3)p_t + x_{4t}, \tag{4.1}$$

  where $c_{jt} \in \mathbb{R}^K$, $c_t = (c_{1t}, \ldots, c_{nt}) \in \mathbb{R}^{n\times K}$, $x_{1t} \in \mathbb{R}^n$,
  $x_1 = (x_{11}^T, \ldots, x_{1T}^T) \in \mathbb{R}^{nT}$, $x_2, x_3 \in \mathbb{R}^K$, $x_4 \in \mathbb{R}$, $x = (x_1, \ldots, x_4) \in \mathbb{R}^\nu$.
  Here $\boldsymbol{\xi} = (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) : \Omega \to \Xi \subseteq \mathbb{R}^\ell$ represents a consumer (or, more precisely, a consumer's behavior), which is described as a random vector and $\boldsymbol{\xi}_1$, $\boldsymbol{\xi}_2$ are independent and satisfy standard normal distribution as in [14].

$$\beta(\xi_1, x_2, x_3) = x_2 + x_3\xi_1 \quad \text{and} \quad \alpha(\xi_2, x_4) = \exp(x_4\xi_2).$$

  For product j in market t, we use $c_{jt} \in \mathbb{R}^K$ to denote the K observed product characteristics, $p_{jt} \in \mathbb{R}$ denotes the observed price, and $(x_{1t})_j \in \mathbb{R}$ denotes the demand shock or errors which is not available in the data. We assume that $X := \{x : \underline{x} \le x \le \bar{x}\}$ for given $\underline{x}$, $\bar{x} \in \mathbb{R}^\nu$ and $\nu = 2K + nT + 1$.
- Consumer ξ chooses to purchase product j in market t if and only if

$$u_{jt}(\xi, x) \ge \max_{1\le i\le n}\{u_{it}(\xi, x), 0\}.$$

- $b_t = (b_{jt})_{j=1}^n$ with b_{jt} the observed market share of product j in market t.

The GMM estimation problem is aimed at finding optimal parameters x by minimizing the model error, $\|x_1\|_2^2$ subject to the generalized market share equations

$$\mathbb{E}[S_t(\boldsymbol{\xi}, x)] \ni b_t, \quad t = 1, \ldots, T,$$

which can be expressed as a quadratic program with stochastic equilibrium set-valued constraints in the following form

$$\begin{array}{rl} \min_{x\in X} & \frac{1}{2}\langle x_1, x_1\rangle \\ \text{subject to} & \mathbb{E}[S_t(\boldsymbol{\xi}, x)] \ni b_t, \quad t = 1, \ldots, T, \end{array} \tag{4.2}$$

where $S_t(\xi, x)$ consists of all the solutions of the linear program:

$$\max_y \{\, \langle y, u_t(\xi, x)\rangle \,|\, \langle e, y\rangle \le 1,\, y \ge 0 \,\}.$$

Obviously, the GMM estimation problem (4.2) is a special case of problem (1.4). We can apply the SAA regularized method to handle the problem. The convergence results established in §3 are applicable. Specifically, the regularized problem of (4.2) is:

$$\begin{array}{rl} \min_{x\in X} & \frac{1}{2}\langle x_1, x_1\rangle \\ \text{subject to} & \|\mathbb{E}[s_t^\varepsilon(\boldsymbol{\xi}, x)] - b_t\| \le r(\varepsilon), \quad t = 1, \ldots, T. \end{array} \tag{4.3}$$

Let $\{(\xi_1^i, \xi_2^i), i = 1, \ldots, N\}$ be iid observations of $(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$. The SAA regularized problem then reads,

$$
\begin{array}{ll}
\min_{x \in X} & \frac{1}{2}\langle x_1, x_1 \rangle \\
\text{subject to} & \|\frac{1}{N}\sum_{i=1}^N s_t^\varepsilon(\xi^i, x) - b_t\| \leq \hat{r}(\varepsilon, N), \quad t = 1, \ldots, T,
\end{array}
\tag{4.4}
$$

where $\left(s_t^\varepsilon(\xi, x), \gamma_t^\varepsilon(\xi, x)\right)$ is the unique solution of LCP$(q_t(\xi, x), M^\varepsilon)$:

$$
0 \leq \begin{pmatrix} y \\ \gamma \end{pmatrix} \perp M^\varepsilon \begin{pmatrix} y \\ \gamma \end{pmatrix} + \begin{pmatrix} -u_t(\xi, x) \\ 1 \end{pmatrix} \geq 0 \tag{4.5}
$$

for some $\gamma_t^\varepsilon(\xi, x) \in \mathbb{R}$. Let us consider a particular case with $T = 1$, $n = 2$, $K = 1$, $\nu = 5$. For simplicity's sake, we omit t in what follows. Set $b = (\frac{1}{2}, \frac{1}{2})$, $c = (2, 3)$, $p = (1, 2)$, $\tau_0 = 1$, $r(\varepsilon) = 2\tau_0\varepsilon$, $\hat{r}(\varepsilon, N) = r(\varepsilon) + N^{-\frac{2}{5}}$, $\ell = 2$, $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ are independent and satisfy standard normal distribution. We choose $\underline{x}_i = -1, \bar{x}_i = 8, i = 1, \ldots, 5$ and initial point $x_0 = (1, 1, 1, 1, 1)$. It is easy to observe that there is an optimal solution $x^* = (0, 1, 0, 0, 0)$ at which the optimal value is 0, since

$$
\beta(\xi_1, x_2^*, x_3^*) = x_2^* + x_3^*\xi_1 = 1, \quad \alpha(\xi_2, x_4^*) = \exp(x_4^*\xi_2) = 1,
$$

and

$$
u(\xi, x^*) = c\beta(\xi_1, x_2^*, x_3^*) - \alpha(\xi_2, x_4^*)p + x_4 = (1, 1),
$$

which implies the solution set of (1.5) is $S(\xi, x^*) = \{(\lambda, 1 - \lambda) \,\big|\, 0 \leq \lambda \leq 1\}$ for all $\xi \in \Xi$. Especially, the constraint $\mathbb{E}[\bar{s}(\boldsymbol{\xi}, x^*)] = b$ holds with the least norm solution

$$
\bar{z}(\xi, x^*) = \left(\bar{s}(\xi, x^*),\, \gamma(\xi, x^*)\right) = (1/2, 1/2, 1).
$$

Moreover, using this optimal solution x^* as a feasible solution for defining $\tilde{r}(\varepsilon)$ in (3.10), we obtain $\tilde{r}(\varepsilon) = 0$ and $r(\varepsilon) = 2\varepsilon$ in Theorem 3.5. Hence, by Theorem 3.5 and Proposition 3.6, problems (4.3) and (4.4) are solvable and Assumption 1 holds at x^*. The conditions of Theorem 3.7 are satisfied which means our convergence results hold for this problem.

The tests were carried out in MATLAB 8.0 installed on a IBM Notebook PC with Windows 7 operating system, Intel Core i5 processor. We used the Matlab solver "fmincon" to solve problem (4.4) with different values of ε and N, where the closed form of $s^\varepsilon(\xi^i, x)$ derived in Lemma 2.2 has been used in our calculations. We report numerical result for $\varepsilon = 0.2, 0.1, 0.05$ and $N = 500, 800, 1100, 1400$. For each combination of ε and N, 35 independent test cases were carried out, each of which solves the SAA regularized problem and yields an approximating solution x_N^ε. Moreover, we use

$$
\text{error}(x_N^\varepsilon) = \|\frac{1}{N_0}\sum_{i=1}^{N_0} s^*(\xi^i, x_N^\varepsilon) - b\|
$$

to measure the infeasibility of x_N^ε with a large sample size $N_0 = 10000 > N$, where $z^*(\xi, x_N^\varepsilon) = (s^*(\xi, x_N^\varepsilon), \gamma^*(\xi, x_N^\varepsilon))$ is the least norm solution of the LCP (1.5). Table 4.1 presents the means of errors of the approximation solutions and the means of the optimal values of problem (4.4). The table shows the downward trend of the errors when the value of ε gets smaller and the sample size N increases and that the approximation optimal values are almost 0. In Figures 4.1-4.3, we use "boxplot" in Matlab to show the convergence trend of the error when the sample size N increases. Each box in the figures displays the range of errors of the computed solutions generated from 35 independent tests, where the central mark is the median and the edges of the box are the 25th and 75th percentiles.

TABLE 4.1
The means of errors and optimal values with different ε and sample size

N / ε	500		800		1100		1400	
	error	fval	error	fval	error	fval	error	fval
0.2	0.0418	0.0005	0.0327	0.0000	0.0319	0.0000	0.0273	0.0003
0.1	0.0349	0.0004	0.0291	0.0005	0.0278	0.0002	0.0244	0.0010
0.05	0.0342	0.0012	0.0248	0.0008	0.0255	0.0007	0.0234	0.0007

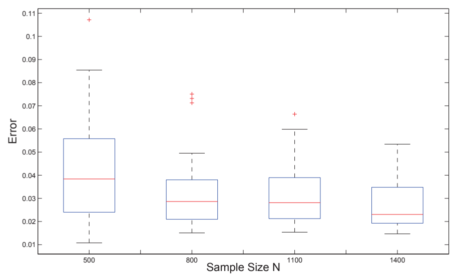

FIG. 4.1. *error(x_N^ε) when $\varepsilon = 0.2$.*

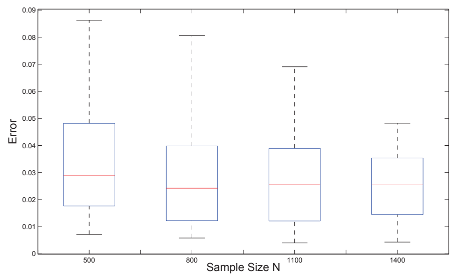

FIG. 4.2. *error(x_N^ε) when $\varepsilon = 0.1$.*

5. Concluding remarks. Mathematical programs with set-valued stochastic equilibrium constraints (1.1) provide a powerful modeling paradigm for many important applications, in particular, in economics. For example, for the estimation of pure characteristics demand models with pricing. However, existing optimization methods with the sample average approximation become intractable for solving such problems. Recently, Pang et al. [14] proposed a mathematical programming with linear complementarity constraints (MPLCC) approach for the pure characteristics demand model with a finite number of observations. Their approach provides a promising computational method to estimate the consumer utility under the following condition:

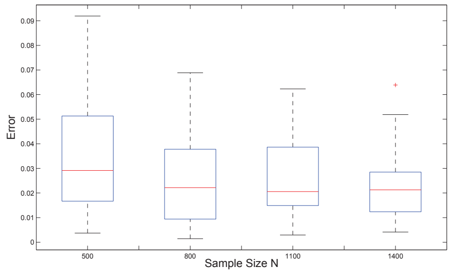

FIG. 4.3. *error*(x_N^ε) *when* $\varepsilon = 0.05$.

Condition 1 In any market t, the optimal choice of each individual consumer is guaranteed to purchase just one single product in each ξ-environment.

Condition 1 and the use of a corresponding basic solution with a finite number of observations can be expresses in terms of mathematical program with linear equilibrium constraints (1.3). This paper is motivated by the MPLCC reformulation proposed by Pang et al [14]. Our main contributions are as follows.

(i) Remove Condition 1. We believe removing Condition 1 is important for real applications in economic. Consider just a simple case with one market and two products. If the value of utility function of the consumer for the two products satisfy $u_1(\xi, x) = u_2(\xi, x) > 0$, the probability that the consumer buy the two products described as a solution $s_1(\xi, x)$ and $s_2(\xi, x)$ of (1.2) satisfy $s_1(\xi, x), s_2(\xi, x) \geq 0$, $s_1(\xi, x) + s_2(\xi, x) = 1$. Under Condition 1, the consumer should buy just one single product. Which solution should the consumer choose with probability one? If we consider $s_1(\xi, x)$ and $s_2(\xi, x)$ as the probability that the consumer buy the products, then the answer is most likely $s_1(\xi, x) = s_2(\xi, x) = \frac{1}{2}$, which is the least norm solution of (1.5). Using graphical convergence for set-valued mappings [15], we can remove Condition 1.

(ii) Develop the SAA regularized method. To handle the set-valued mapping in (1.1), we develop an efficient SAA regularized method using (1.8) and (1.9) which replaces the set-valued mapping by a single valued function. Problem (1.9) is a mathematical program with a convex quadratic objective function and globally Lipschitz continuous inequality constraints. Moreover, we derive a closed form of the solution of the regularized LCP(q, M^ε), which is useful for numerical computation and theoretical analysis. We show that a sequence of solutions $\{x_N^\varepsilon\}$ of the SAA regularized stochastic MPSLCC (1.9) converges to a solution of problem (1.4) as $\varepsilon \downarrow 0$ and $N \to \infty$.

REFERENCES

[1] R.J. AUMANN, Integrals of set-valued functions, J. Math. Anal. Appl., 12(1965), pp. 1-12.
[2] A. ARAUJO AND E. GINÉ, The Central Limit Theorem for Real and Banach Valued Random Variables, Wiley, New York, 1980.
[3] S.T. BERRY AND A. PAKES, The pure characteristics demand model, International Economic Review 48(2007), pp. 1193-1225.
[4] J.R. BIRGE AND F. LOUVEAUX, Introduction to Stochastic Programming, Springer, New York, 1997.
[5] J. F. BONNANS AND A. SHAPIRO, Perturbation Analysis of Optimization Problems, Springer

Series in Operations Research, Springer-Verlag, New York, 2000.
[6] X. Chen and S. Xiang, Perturbation bounds of P-matrix linear complementarity problems, SIAM J. Optim., 19(2007), pp. 1250-1265.
[7] X. Chen and S. Xiang, Newton iterations in implicit time-stepping scheme for differential linear complementarity systems, Math. Program., 138(2013), pp. 579-606.
[8] R.W. Cottle, J.-S. Pang and R. E. Stone, The Linear Complementarity Problem, Academic Press, New York, 1992.
[9] F. Facchinei and J.-S. Pang, Finite-Dimensional Variational Inequalities and Complementarity Problems, Springer-Verlag, New York, 2003.
[10] J.-P. Dubé, J.T. Fox and C.-L. Su, Improving the numerical performance of BLP static and dynamic discrete choice random coeffcients demand estimation, Econometrica, 80(2012), pp. 2231-2267.
[11] L.J. Hong, Y. Yang and L. Zhang, Sequential convex approximations to joint chance constrained programs: A Monte Carlo approach, Oper. Res, 59(2011), pp. 617-630.
[12] A. Nemirovski and A. Shapiro, Convex approximations of chance constrained programs, SIAM J. Optim., 17(2006), pp. 969-996.
[13] Z.Q. Luo, J.-S.Pang and D. Ralph, Mathematical Programs with Equilibrium Constraints. Cambridge University Press, Cambridge (1996).
[14] J.-S. Pang, C.-L. Su and Y.C. Lee, Estimation of pure characteristics demand models with pricing, 2012.
[15] R.T. Rockafellar and R.J-B. Wets, Variational analysis, Springer, Berlin, 1998 (3rd printing 2009).
[16] R.T. Rockafellar and S. Uryasev, Optimization of conditional value-at-risk, J. Risk, 2(2000), pp. 493–517.
[17] A. Ruszczynski and A. Shapiro, Stochastic Programming, Handbooks in Operations Research and Management Science, Elsevier, 2003.
[18] H. Xu, Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming, J. Math. Anal. Appl., 368(2010), pp. 692–710.
[19] H. Xu and D. Zhang, Smooth sample average approximation of stationary points in nonsmooth stochastic optimization and applications, Math. Program., 119(2009), pp. 371–401.