Jie Zhang; Li-wei Zhang; Yue Wu

A smoothing SAA method for a stochastic mathematical program with complementarity constraints

*Applications of Mathematics*, Vol. 57 (2012), No. 5, 477–502

Persistent URL: http://dml.cz/dmlcz/142912

# A SMOOTHING SAA METHOD FOR A STOCHASTIC MATHEMATICAL PROGRAM WITH COMPLEMENTARITY CONSTRAINTS*

Jie Zhang, Dalian, Li-wei Zhang, Dalian, Yue Wu, Southampton

(Received May 10, 2010)

*Abstract.* A smoothing sample average approximation (SAA) method based on the log-exponential function is proposed for solving a stochastic mathematical program with complementarity constraints (SMPCC) considered by Birbil et al. (*S. I. Birbil, G. Gürkan, O. Listes*: Solving stochastic mathematical programs with complementarity constraints using simulation, Math. Oper. Res. *31* (2006), 739–760). It is demonstrated that, under suitable conditions, the optimal solution of the smoothed SAA problem converges almost surely to that of the true problem as the sample size tends to infinity. Moreover, under a strong second-order sufficient condition for SMPCC, the almost sure convergence of Karash-Kuhn-Tucker points of the smoothed SAA problem is established by Robinson's stability theory. Some preliminary numerical results are reported to show the efficiency of our method.

*Keywords*: smoothing SAA method, log-exponential function, stochastic mathematical program with complementarity constraints, almost sure convergence

*MSC 2010*: 90C15, 90C33

## 1. Introduction

Our concern in this paper is to investigate the almost sure convergence of a numerical method for the following stochastic mathematical program with complementarity constraints (SMPCC):

$$\begin{aligned} \min \quad & \mathbb{E}[f(x,y,\xi(\omega))] \\ \text{s.t.} \quad & \Psi(x,y) \geqslant 0,\ y \geqslant 0, \\ & \Psi(x,y)^{\mathrm{T}} y = 0, \end{aligned} \tag{1.1}$$

* This research was supported by the National Natural Science Foundation of China under project No. 11071029 and the Fundamental Research Funds for the Central Universities.

where $\Psi(x,y) := \mathbb{E}[F(x,y,\xi(\omega))]$, $F\colon \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}^m$ is a random mapping, $\xi\colon \Omega \to \Xi \subset \mathbb{R}^k$ is a random vector defined on a probability space $(\Omega, \mathcal{F}, \mathrm{P})$, $\mathbb{E}$ denotes the mathematical expectation. Throughout the paper, we assume that $\mathbb{E}[f(x,y,\xi(\omega))]$ and $\mathbb{E}[F(x,y,\xi(\omega))]$ are all well defined and finite for any $(x,y) \in \mathbb{R}^n \times \mathbb{R}^m$. To ease the notation, we write $\xi(\omega)$ as $\xi$ and this should be distinguished from $\xi$ being a deterministic vector of $\Xi$ in a context.

Over the past several decades, the mathematical program with complementarity constraints (MPCC) has been intensively studied for its extensive application in engineering, economics, game theory and networks, see [10] and [11] for systematic expositions, examples, and applications. While in practice, there are some important instances where problem data contain some uncertain factors, and consequently various formulations of SMPCC models are proposed to reflect the uncertainties. See [26], [22], [9], [2], [12], and references therein. Among these formulations, Birbil et al. [2] were the first to treat SMPCC (1.1) and presented the so-called sample-path method for solving it.

Evidently, if the integral involved in the mathematical expectation of problem (1.1) exists or is computable, then problem (1.1) is reduced to the usual MPCC problem. However, in many cases, exact evaluation of the expected value in (1.1) is either impossible or prohibitively expensive. Sample average approximation (SAA) method, also known as sample path optimization (SPO) method [18], is suggested by many authors to handle this difficulty, see the recent works [15], [23], [6] and a comprehensive review by Shapiro [24]. The basic idea of SAA is to generate an independent identically distributed (iid) sample $\xi^1, \ldots, \xi^N$ of $\xi$ and then approximate the expected value by sample average. In this context, let $\xi^1, \ldots, \xi^N$ be an iid sample, then the SMPCC (1.1) is approximated by the following SAA problem:

$$\begin{aligned} \min \quad & \hat{f}_N(x,y) \\ \text{s.t.} \quad & 0 \leqslant y \perp \hat{F}_N(x,y) \geqslant 0, \end{aligned} \tag{1.2}$$

where $\hat{f}_N(x,y) := N^{-1} \sum_{i=1}^{N} f(x,y,\xi^i)$ is the sample-average function of $f(x,y,\xi)$ and $\hat{F}_N(x,y) = N^{-1} \sum_{i=1}^{N} F(x,y,\xi^i)$ is the sample-average mapping of $F(x,y,\xi)$. We refer to (1.1) as the true problem and (1.2) as the SAA problem to (1.1).

In this paper, we propose a smoothing SAA method based on the log-exponential function [20] for solving (1.1). That is, utilizing the properties of the log-exponential function, we reformulate the SAA problem (1.2) as a smooth nonlinear programming (NLP) problem by displacing the difficult equilibrium constraints of (1.2) with a smooth function, and then solve the true problem (1.1) by solving a sequence of such smoothed SAA problem.

Recently, the smoothing SAA method has caught great attention among researchers in SMPCC, see e.g., Shapiro and Xu [22], Xu [26], Xu and Meng [27]. However, most available results discuss an application of the SAA method to SMPCC with the assumption that the constraint of (1.1) has a unique solution for every $x$. Our paper, without such assumption, focuses on the sufficient conditions ensuring the almost sure convergence of the proposed smoothing SAA method. Also notice that although the idea of the SAA method is essentially the same as that of the sample-path method, our method differs from the work of Birbil et at. [2] which discusses the almost sure convergence of the optimal solutions of SAA problem (1.2) without referring to a particular smoothing function.

By the notion of epi-convergence in [20], we establish the almost sure convergence of the optimal solution of a smoothed SAA problem as the sample size tends to infinity. As it is more practical to find a stationary point than a global minimizer of the true problem, under the MPCC strong second order sufficient condition (MPCC-SSOSC) in [21], we investigate the almost sure convergence of the stationary points of smoothed SAA problem.

This paper is organized as follows: Section 2 gives preliminaries needed throughout the whole paper. In Section 3, by introducing an iid sample and the log-exponential function, we formulate the SAA problem (1.2) as a smooth NLP problem. In what follows, we discuss the almost sure convergence of optimal solutions and stationary points of the smoothed SAA problem as the sample size tends to infinity in Section 4. In Section 5, we report some preliminary numerical results.

## 2. Preliminaries

Throughout this paper we use the following notation. Let $\|\cdot\|$ denote the Euclidean norm of a vector or the Frobenius norm of a matrix. For an $m \times n$ matrix $A$, $A_{ij}$ denotes the element of the $i$th row and $j$th column of $A$. We use $\mathbb{B}$ to denote the closed unit ball and $\mathbb{B}(x, \delta)$ the closed ball around $x$ of radius $\delta > 0$. For an extended real-valued function $\varphi\colon \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$, $\operatorname{epi}\varphi$, $\nabla\varphi(x)$, and $\nabla^2\varphi(x)$ denote its epigraph i.e. the set $\{(x, \alpha)\colon \varphi(x) \leqslant \alpha\}$, the gradient of $\varphi$ at $x$, and the Hessian matrix of $\varphi$ at $x$, respectively. If a mapping $F\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ is Fréchet-differentiable at $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$, then we use $\mathcal{J}F(x, y)$ and $\mathcal{J}_x F(x, y)$ to denote respectively the Fréchet-derivative of $F$ at $(x, y)$ and the partial Fréchet-derivative of $F$ at $(x, y)$ with respect to $x$. Moreover, if $F$ is twice Fréchet-differentiable at $(x, y)$, we define

$$\mathcal{J}^2 F(x, y) := \mathcal{J}(\mathcal{J}F)(x, y), \quad \mathcal{J}^2_{xx} F(x, y) := \mathcal{J}_x(\mathcal{J}_x F)(x, y).$$

In the following, we introduce some concepts of the convergence of set sequences and mapping sequences from [20] which will be used in the next section. Define

$$\mathcal{N}_\infty := \{N \subset \mathbb{N}\colon\ \mathbb{N} \setminus N \text{ finite}\} \quad \text{and} \quad \mathcal{N}_\infty^\# := \{N \subset \mathbb{N}\colon\ N \text{ infinite}\},$$

where $\mathbb{N}$ denotes the set of all positive integer numbers.

**Definition 2.1.** For sets $C^\nu$ and $C$ in $\mathbb{R}^n$ with $C$ closed, the sequence $\{C^\nu\}_{\nu\in\mathbb{N}}$ is said to *converge to* $C$ (written $C^\nu \to C$) if

$$\limsup_{\nu\to\infty} C^\nu \subset C \subset \liminf_{\nu\to\infty} C^\nu$$

with

$$\limsup_{\nu\to\infty} C^\nu := \big\{x\colon\ \exists\, N \in \mathcal{N}_\infty^\#,\, \exists\, x^\nu \in C^\nu (\nu \in N) \text{ such that } x^\nu \xrightarrow{N} x\big\},$$
$$\liminf_{\nu\to\infty} C^\nu := \big\{x\colon\ \exists\, N \in \mathcal{N}_\infty,\, \exists\, x^\nu \in C^\nu (\nu \in N) \text{ such that } x^\nu \xrightarrow{N} x\big\}.$$

The continuity properties of a set-valued mapping $S$ can be developed by the convergence of sets.

**Definition 2.2.** A set-valued mapping $S\colon\ \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is *continuous at* $\bar{x}$, symbolized by $\lim\limits_{x\to\bar{x}} S(x) = S(\bar{x})$, if

$$\limsup_{x\to\bar{x}} S(x) \subset S(\bar{x}) \subset \liminf_{x\to\bar{x}} S(x).$$

**Definition 2.3.** Consider now a family of functions $f^\nu\colon\ \mathbb{R}^n \to \bar{\mathbb{R}}$, where $\bar{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$. One says that $f^\nu$ *epi-converges to a function* $f\colon\ \mathbb{R}^n \to \bar{\mathbb{R}}$ as $\nu \to \infty$, written

$$f = e - \lim_{\nu\to\infty} f^\nu,$$

if the sequence of sets $\operatorname{epi} f^\nu$ converges to $\operatorname{epi} f$ in $\mathbb{R}^n \times \mathbb{R}$ as $\nu \to \infty$.

The characterization of the epi-convergence can be described by the following result.

**Proposition 2.1** ([20, Proposition 7.2])**.** *Let $\{f^\nu\}$ be any sequence of functions on $\mathbb{R}^n$ and let $x$ be any point of $\mathbb{R}^n$. Then $f^\nu$ epi-converges to $f$ if and only if at each point $x$, the following two conditions both hold:*

(a) $\liminf\limits_{\nu\to\infty} f^\nu(x^\nu) \geqslant f(x)$ *for every sequence* $x^\nu \to x$,
(b) $\limsup\limits_{\nu\to\infty} f^\nu(x^\nu) \leqslant f(x)$ *for some sequence* $x^\nu \to x$.

Next, we recall some basic concepts that are often employed in the literature on the mathematical program with complementarity constraints problem.

Let $(\bar{x}, \overline{y})$ be a feasible point of problem (1.1) and for convenience let us define the index sets

$$\begin{aligned} \overline{\alpha} &= \{i \in \{1, 2, \ldots, m\} \colon\ y_i = 0 < \Psi_i(\bar{x}, \overline{y})\}, \\ \overline{\beta} &= \{i \in \{1, 2, \ldots, m\} \colon\ y_i = 0 = \Psi_i(\bar{x}, \overline{y})\}, \\ \overline{\gamma} &= \{i \in \{1, 2, \ldots, m\} \colon\ y_i > 0 = \Psi_i(\bar{x}, \overline{y})\}. \end{aligned} \tag{2.1}$$

The linear independence constraint qualification for SMPCC is as follows.

**Definition 2.4.** Assume $\Psi$ is continuously differentiable at $(\bar{x}, \overline{y})$. We say the MPCC linear independence constraint qualification (MPCC-LICQ) holds at $(\bar{x}, \overline{y})$ if the vectors of the set

$$\left\{ \begin{pmatrix} 0 \\ e_i \end{pmatrix} \colon\ i \in \overline{\alpha} \cup \overline{\beta} \right\} \cup \{\nabla \Psi_i(\bar{x}, \overline{y}) \colon\ i \in \overline{\beta} \cup \overline{\gamma}\}$$

are linearly independent, where $e_i$ denotes the vector with 1 in the $i$th component but 0's everywhere else.

For a deterministic MPCC, it is well known that the usual nonlinear programming constraint qualifications such as the Mangasarian-Fromovitz constraint qualification (MFCQ) do not hold (see [28, Proposition 1.1]). Since there are several different approaches to reformulate MPCC, various stationarity concepts arise (see e.g. [21] and [29]). We use the following two stationarity concepts for SMPCC.

**Definition 2.5.** Assume $(\bar{x}, \overline{y})$ is a feasible point of SMPCC (1.1), $\Psi(\cdot, \cdot)$, $\mathbb{E}[f(\cdot, \cdot, \xi)]$ are continuously differentiable at $(\bar{x}, \overline{y})$. Suppose there exist vectors $\bar{u} \in \mathbb{R}^m$ and $\overline{v} \in \mathbb{R}^m$ such that $(\bar{x}, \overline{y})$ satisfies the condition

$$0 = \nabla \mathbb{E}[f(\bar{x}, \overline{y}, \xi(\omega))] - \sum_{i \in \overline{\alpha} \cup \overline{\beta}} \bar{u}_i \begin{pmatrix} 0 \\ e_i \end{pmatrix} - \sum_{i \in \overline{\beta} \cup \overline{\gamma}} \overline{v}_i \nabla \Psi_i(\bar{x}, \overline{y}).$$

(i) (C-stationary point) We call $(\bar{x}, \overline{y})$ a *Clarke stationary point of* (1.1) if $\bar{u}_i, \overline{v}_i \geqslant 0$, $i \in \beta$.

(ii) (S-stationary point) We call $(\bar{x}, \overline{y})$ a *strongly stationary point of* (1.1) if $\bar{u}_i \geqslant 0$, $\overline{v}_i \geqslant 0$, $i \in \overline{\beta}$.

By [21, Lemma 1], the C-stationary condition in Definition 2.5 is the nonsmooth Karash-Kuhn-Tucker (KKT) condition using the Clarke generalized gradient [3] by reformulating SMPCC as a nonsmooth stochastic nonlinear programming problem

$$\begin{aligned} \min \quad & \mathbb{E}[f(x,y,\xi(\omega))] \\ \text{s.t.} \quad & y_i = 0,\ i \in \overline{\alpha},\ \ \Psi_i(x,y) = 0,\ i \in \overline{\gamma}, \\ & \min\{y_i, \Psi_i(x,y)\} = 0,\ i \in \overline{\beta}. \end{aligned}$$

The S-stationary condition [13] is the KKT condition for the relaxed SMPCC

$$\begin{aligned} \min \quad & \mathbb{E}[f(x,y,\xi(\omega))] \\ \text{s.t.} \quad & y_i = 0,\ i \in \overline{\alpha},\ \ \Psi_i(x,y) = 0,\ i \in \overline{\gamma}, \\ & y_i \geqslant 0,\ \ \Psi_i(x,y) \geqslant 0,\ i \in \overline{\beta}. \end{aligned}$$

The following upper level strict complementarity condition was used in [21] in the context of sensitivity analysis for MPCC.

**Definition 2.6.** We say that the upper level strict complementarity condition (ULSC) holds at $(\bar{x},\overline{y})$ if $\bar{u}_i$ and $\overline{v}_i$, the multipliers corresponding to $\bar{y}_i$ and $\Psi_i(\bar{x},\overline{y})$, respectively, satisfy $\bar{u}_i\overline{v}_i \neq 0$ for all $i \in \overline{\beta}$.

It is well known that a point $(\bar{x},\overline{y})$ satisfies the lower level strict complementarity condition (LLSC) if $\overline{y}_i + \Psi_i(\bar{x},\overline{y}) > 0$ holds for all $i \in \{1,\ldots,m\}$. We can see from an example from [21] that the ULSC condition is considerably weaker than the LLSC condition, and in practice, it may make more sense than the latter.

We use the following second order condition based on the MPCC-Lagrangian:

$$L(x,y,u,v) = \mathbb{E}[f(x,y,\xi(\omega))] - \sum_{i\in\overline{\alpha}\cup\overline{\beta}} u_i y_i - \sum_{i\in\overline{\beta}\cup\overline{\gamma}} v_i \Psi_i(x,y) \tag{2.2}$$

of (1.1).

**Definition 2.7** ([21]). Let $(\bar{x},\overline{y})$ be an S-stationary point of (1.1) and $(\bar{u},\overline{v})$ the corresponding multiplier at $(\bar{x},\overline{y})$. Suppose $\Psi(\cdot,\cdot)$ and $\mathbb{E}[f(\cdot,\cdot,\xi)]$ are twice continuously differentiable at $(\bar{x},\overline{y})$. We say that the MPCC strong second order sufficient condition (MPCC-SSOSC) holds at $(\bar{x},\overline{y})$ if

$$d^{\mathrm{T}}\nabla^2_{(x,y)}L(\bar{x},\overline{y},\bar{u},\overline{v})d > 0$$

for every nonvanishing $d$ with

$$\begin{aligned} & (0,e_i^{\mathrm{T}})d = 0, \quad i \in \overline{\alpha}, \\ & \nabla\Psi_i(\bar{x},\overline{y})^{\mathrm{T}}d = 0, \quad i \in \overline{\gamma}, \\ & \min\{(0,e_i^{\mathrm{T}})d, \nabla\Psi_i(\bar{x},\overline{y})^{\mathrm{T}}d\} = 0, \quad i \in \overline{\beta}. \end{aligned}$$

Assume $(\bar{x}, \overline{y})$ is an S-stationary point of (1.1) and $(\bar{u}, \overline{v})$ is a corresponding multiplier. Then we know from [21, Theorem 7] that if MPCC-SSOSC holds at $(\bar{x}, \overline{y})$, it is a strict local minimizer of the SMPCC (1.1).

## 3. Formulating the smoothing SAA method

The log-exponential function is a smoothing function for max-type functions. Let $w\colon \mathbb{R}^n \to \mathbb{R}$, $w(x) = \max\{w_1(x), w_2(x), \ldots, w_m(x)\}$, where $w_i$, $i = 1, \ldots, m$, are continuously differentiable functions. It is clear that $w(\cdot)$ is continuous in $\mathbb{R}^n$ but not differentiable everywhere. For any $t > 0$, the log-exponential function of $w(x)$, denoted as $w(t, x)\colon \mathbb{R}^{n+1} \to \mathbb{R}$, is defined by

$$w(t, x) := t \ln \sum_{i=1}^{m} \exp(w_i(x)/t), \tag{3.1}$$

which was studied by many authors, see Rockafellar and Wets [19], [20], Li [7], [8]. An interesting feature of $w(t, x)$ (see [20, Example 1.30]) is that

$$0 \leqslant w(t, x) - w(x) \leqslant t \ln m, \tag{3.2}$$

which implies $\lim\limits_{t \downarrow 0} w(t, x) = w(x)$ and the convergence is uniform with respect to $x$. We know from the definition that $w(t, x)$ is a smoothing function with respect to $x$ for $t > 0$ and hence utilizing this property, over the past decade, some authors have used the log-exponential function to propose smoothing methods for generalized linear complementarity problems, nonlinear complementarity problems, and mathematical programs with complementarity constraints, see [14], [16], [30] and references therein.

Notice that $G(x, y) = \min\{\Psi(x, y), y\} = 0$ can be approximated by the equation

$$G_t(x, y) := \begin{pmatrix} g_t(y_1, \Psi_1(x, y)) \\ \vdots \\ g_t(y_m, \Psi_m(x, y)) \end{pmatrix} = 0 \tag{3.3}$$

in the sense that $\lim\limits_{t \downarrow 0} G_t(x, y) = G(x, y)$, where $g_t(a, b) = -t \ln(\exp(-a/t) + \exp(-b/t))$, $t > 0$ for $a, b \in \mathbb{R}$ and $\Psi_i(x, y)$ is the $i$th component of $\Psi(x, y)$. Therefore, it is natural to define $g_0(y_i, \Psi_i(x, y)) = \min\{y_i, \Psi_i(x, y)\}$, $i = 1, \ldots, m$, and $G_0(x, y) = G(x, y)$. By taking independently and identically distributed random samples $\xi^i$, $i = 1, \ldots, N$, and introducing the smoothing function $G_t(\cdot, \cdot)$ (3.3), we obtain the approximation of problem (1.1)

$$\begin{aligned} \min \quad & \hat{f}_N(x, y) \\ \text{s.t.} \quad & \hat{G}_N(x, y) = 0, \end{aligned} \tag{3.4}$$

where

$$\hat{G}_N(x,y) := \begin{pmatrix} g_{t_N}(y_1, \hat{F}_N(x,y)_1) \\ \vdots \\ g_{t_N}(y_m, \hat{F}_N(x,y)_m) \end{pmatrix}$$

with

$$t_N > 0, \quad \hat{f}_N(x,y) = \frac{1}{N}\sum_{i=1}^{N} f(x,y,\xi^i),$$

$$\hat{F}_N(x,y)_j = \frac{1}{N}\sum_{i=1}^{N} F(x,y,\xi^i)_j, \quad j = 1,\ldots,m,$$

where $F(x,y,\xi^i)_j$ is the $j$th component of $F(x,y,\xi^i)$.

## 4. Convergence of the smoothing SAA method

Throughout the paper, we assume the sample $\xi^1,\ldots,\xi^N$ of the random vector $\xi$ is iid and introduce the following assumptions to make (1.1) more clearly defined and to facilitate the analysis.

**Assumption 1.** The mappings $f(\cdot,\cdot,\xi)$ and $F(\cdot,\cdot,\xi)$ are twice continuously differentiable on $\mathbb{R}^{n+m}$ a.e. $\xi \in \Xi$.

**Assumption 2.** For any $(\bar{x},\bar{y}) \in \mathbb{R}^{n+m}$, there exist a neighborhood $D$ of $(\bar{x},\bar{y})$ and a nonnegative measurable function $g(\xi)$ such that $\mathbb{E}[g(\xi)] < \infty$ and

$$\sup_{(x,y)\in D} \max\{\|\nabla f(x,y,\xi)\|, \|\nabla^2 f(x,y,\xi)\|, \|\mathcal{J}F(x,y,\xi)\|, \|\mathcal{J}^2 F(x,y,\xi)\|\} \leqslant g(\xi)$$

for all $\xi \in \Xi$.

Assumptions 1–2 are popularly used conditions for the analysis of SAA method for stochastic programming. Under these two assumptions, we know from [24, Chapter 7] that $\mathbb{E}[f(x,y,\xi(\omega))]$ and $\mathbb{E}[F(x,y,\xi(\omega))]$ are twice continuously differentiable on $\mathbb{R}^{n+m}$. In particular, $\nabla\mathbb{E}[f(x,y,\xi(\omega))] = \mathbb{E}[\nabla f(x,y,\xi(\omega))]$ and $\mathcal{J}\mathbb{E}[F(x,y,\xi(\omega))] = \mathbb{E}[\mathcal{J}F(x,y,\xi(\omega))]$.

The following lemma results straightforwardly from the Uniform Laws of Large Numbers in [24, Theorem 6.36].

**Lemma 4.1.** *Suppose that Assumptions* 1–2 *are satisfied. Let* $(\bar{x}_N,\bar{y}_N)$ *be a feasible point of* (3.4). *If the sequence* $(\bar{x}_N,\bar{y}_N)$ *converges to a random vector* $(\bar{x},\bar{y})$ *w.p.* 1

*as $N$ tends to infinity, then we obtain*

$$
\begin{aligned}
\hat{F}_N(\bar{x}_N, \overline{y}_N) &\to \Psi(\bar{x}, \overline{y}) \ \textit{w.p.}\, 1, \\
\mathcal{J}\hat{F}_N(\bar{x}_N, \overline{y}_N) &\to \mathbb{E}[\mathcal{J}F(\bar{x}, \overline{y}, \xi)] \ \textit{w.p.}\, 1, \\
\nabla^2 \hat{f}_N(\bar{x}_N, \overline{y}_N) &\to \mathbb{E}[\nabla^2 f(\bar{x}, \overline{y}, \xi)] \ \textit{w.p.}\, 1, \\
\mathcal{J}^2\hat{F}_N(\bar{x}_N, \overline{y}_N) &\to \mathbb{E}[\mathcal{J}^2 F(\bar{x}, \overline{y}, \xi)] \ \textit{w.p.}\, 1.
\end{aligned} \tag{4.1}
$$

### 4.1. Almost sure convergence of optimal solutions

In this subsection, by using the notion of epi-convergence in [20], we establish the almost convergence of optimal solutions of smoothed SAA problem (3.4) to those of SMPCC (1.1) as the sample size tends to infinity.

Let us introduce some notions:

$$
\begin{gathered}
\mathcal{O}_N := \{(u,v) \in \mathbb{R}^m \times \mathbb{R}^m \colon\ g_{t_N}(u_i, v_i) = 0,\ i = 1, \ldots, m\}, \\
\mathcal{O}_0 := \{(u,v) \in \mathbb{R}^m \times \mathbb{R}^m \colon\ \min\{u_i, v_i\} = 0,\ i = 1, \ldots, m\}, \\
\mathbb{Z}_0 := \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m \colon\ \min\{y_i, \Psi_i(x,y)\} = 0,\ i = 1, \ldots, m\}, \\
\mathbb{Z}_N := \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m \colon\ \hat{G}_N(x,y) = 0\}, \\
\overline{f}_N(x,y) := \hat{f}_N(x,y) + \delta_{\mathbb{Z}_N}(x,y), \\
\overline{f}(x,y) := \mathbb{E}[f(x,y,\xi)] + \delta_{\mathbb{Z}_0}(x,y), \\
\kappa_0 := \inf\{\mathbb{E}[f(x,y,\xi)] \colon\ (x,y) \in \mathbb{Z}_0\}, \\
S_0 = \operatorname{argmin}\{\mathbb{E}[f(x,y,\xi)] \colon\ (x,y) \in \mathbb{Z}_0\}, \\
S_N := \operatorname{argmin}\{\hat{f}_N(x,y) \colon\ (x,y) \in \mathbb{Z}_N\}.
\end{gathered}
$$

**Lemma 4.2.** *Let $t_N \searrow 0$ as $N \to \infty$. Then*

$$
\mathcal{O}_0 \subset \liminf_{N \to \infty} \mathcal{O}_N.
$$

P r o o f. Note that by (3.2), for any $(u,v) \in \mathcal{O}_N$, $(u,v)$ satisfies

$$
0 \leqslant \min\{u_i, v_i\} \leqslant t_N \ln 2, \quad i = 1, \ldots, m.
$$

For any positive numbers $\gamma$ and $\varepsilon$, let $\overline{N} \in \mathbb{N}$ be such that $\sqrt{m} t_N \ln 2 < \varepsilon$ for all $N \geqslant \overline{N}$. Then for all $N \geqslant \overline{N}$ we obtain

$$
\mathcal{O}_0 \cap \gamma\mathbb{B} \subset \mathcal{O}_N + \varepsilon\mathbb{B}.
$$

The conclusion follows. $\square$

Now we give a conclusion about the almost sure convergence of the set $\mathbb{Z}_N$ as $N$ tends to infinity in the following proposition.

**Proposition 4.1.** *Suppose Assumptions* 1–2 *hold and* $\mathcal{J}_x\Psi(x,y)$ *is of maximal rank for any* $(x,y)\in\mathbb{R}^{m+n}$. *If* $t_N\searrow 0$ *as* $N$ *tends to infinity, then*

$$\lim_{N\to\infty}\mathbb{Z}_N=\mathbb{Z}_0\quad w.p.\,1.$$

Proof. We first show that $\limsup_{N\to\infty}\mathbb{Z}_N\subset\mathbb{Z}_0$ w.p. 1. It suffices to prove that for a sequence $\{(x_N,y_N)\}$ satisfying $(x_N,y_N)\in\mathbb{Z}_N$ w.p. 1 for each $N$, if $(x_N,y_N)$ converges to $(\bar{x},\overline{y})$ w.p. 1 as $N\to\infty$, then $(\bar{x},\overline{y})\in\mathbb{Z}_0$ w.p. 1. Indeed, by (3.2) we have w.p. 1 that

$$0\leqslant\min\{(y_N)_i,\hat{F}_N(x_N,y_N)_i\}\leqslant t_N\ln 2,\quad i=1,\dots,m,$$

which, by Lemma 4.1, means that $(\bar{x},\overline{y})\in\mathbb{Z}_0$ w.p. 1.

Let $(\bar{x},\overline{y})\in\mathbb{Z}_0$. Next we show that $(\bar{x},\overline{y})\in\liminf_{N\to\infty}\mathbb{Z}_N$ w.p. 1. Let $\bar{u}=\Psi(x,y)$, $\overline{v}=\overline{y}$, then $(\bar{u},\overline{v})\in\mathcal{O}_0$ and we know from Lemma 4.2 that there exists $(u_N,v_N)\in\mathcal{O}_N$ converging to $(\bar{u},\overline{v})$ as $N\to\infty$. Let

$$\Phi(x,y,u,v)=\begin{bmatrix}\Psi(x,y)-u\\ y-v\end{bmatrix},$$

then $\Phi(\bar{x},\overline{y},\bar{u},\overline{v})=0$ and

$$\mathcal{J}_{(x,y)}\Phi(\bar{x},\overline{y},\bar{u},\overline{v})=\begin{bmatrix}\mathcal{J}_x\Psi(\bar{x},\overline{y}) & \mathcal{J}_y\Psi(\bar{x},\overline{y})\\ 0 & I_m\end{bmatrix}$$

is of maximal rank due to the maximal rank of $\mathcal{J}_x\Psi(x,y)$. Therefore, by Clarke's implicit theorem [3], there exist positive numbers $\varepsilon$, $\delta$ and a Lipschitz function $z(\cdot)=(x(\cdot),y(\cdot))\colon\ \mathbb{B}((\bar{u},\overline{v}),\delta)\to\mathbb{B}((\bar{x},\overline{y}),\varepsilon)$ with modular $c>0$ such that $z(\bar{u},\overline{v})=(\bar{x},\overline{y})$ and for any $(u,v)\in\mathbb{B}((\bar{u},\overline{v}),\delta)$,

$$\Phi(z(u,v),u,v)=0. \tag{4.2}$$

Let

$$H_N(x,y)=\begin{bmatrix}\Psi(x,y)-\hat{F}_N(x,y)+u_N\\ v_N\end{bmatrix}$$

and $\delta'=\{\delta,(2c)^{-1}\varepsilon\}$. Then, using the Uniform Laws of Large Numbers in [24, Theorem 6.36], we have w.p. 1 for $N$ large enough

$$\max_{(x,y)\in\mathbb{B}((\bar{x},\overline{y}),\varepsilon)}\|H_N(x,y)-(\bar{u},\overline{v})\|<\delta'$$

and for any $(x,y) \in \mathbb{B}((\bar{x},\overline{y}),\varepsilon)$,

$$\|z(H_N(x,y)) - z(\bar{u},\overline{v})\| \leqslant c\|H_N(x,y) - (\bar{u},\overline{v})\| < \varepsilon/2 \quad \text{w.p. } 1.$$

Define a function

$$\begin{aligned} \varphi\colon\ \mathbb{B}(0,\delta) &\to \mathbb{B}(0,\delta), \\ (x,y) &\mapsto z(H_N(x,y)). \end{aligned}$$

This is almost surely a continuous mapping from the compact convex set $\mathbb{B}(0,\delta)$ to itself. By Brouwer's fixed point theorem, $\varphi$ has a fixed point w.p. 1. Hence, there exists a vector $(x_N,y_N) \in \mathbb{B}((\bar{x},\overline{y}),\varepsilon)$ w.p. 1 such that $(x_N,y_N) = \varphi(x_N,y_N) = z(H_N(x_N,y_N))$ w.p. 1. Therefore, we have from (4.2) that w.p. 1,

$$0 = \Phi(x_N,y_N,H_N(x_N,y_N)) = \begin{bmatrix} \Psi(x_N,y_N) \\ y_N \end{bmatrix} - H_N(x_N,y_N).$$

That is, $\hat{F}_N(x_N,y_N) = u_N$ w.p. 1 and $y_N = v_N$ w.p. 1, which means $(x_N,y_N) \in \mathbb{Z}_N$ w.p. 1 due to $(u_N,v_N) \in \mathcal{O}_N$. As a result, $(\bar{x},\overline{y})$ belongs to $\liminf\limits_{N\to\infty} \mathbb{Z}_N$ w.p. 1 follows from the fact that $(x_N,y_N)$ converges to $(\bar{x},\overline{y})$ w.p. 1 as $N \to \infty$. □

**Lemma 4.3.** *If the conditions in Proposition* 4.1 *hold, then we have*

$$e - \lim_{N\to\infty} \overline{f}_N = \overline{f} \quad \textit{w.p. } 1.$$

P r o o f. Noting that

$$\operatorname{epi}[\delta_{\mathbb{Z}_N}(\cdot,\cdot)] = \{(x,y,\alpha)\colon\ (x,y) \in \mathbb{Z}_N,\ \alpha \geqslant 0\} = \mathbb{Z}_N \times \mathbb{R}_+$$

and that by Proposition 4.1,

$$\lim_{N\to\infty} \mathbb{Z}_N \times \mathbb{R}_+ = \mathbb{Z}_0 \times \mathbb{R}_+ \quad \text{w.p. } 1,$$

we obtain from Definition 2.3 that

$$e - \lim_{N\to\infty} \delta_{\mathbb{Z}_N}(\cdot,\cdot) = \delta_{\mathbb{Z}_0}(\cdot,\cdot) \quad \text{w.p. } 1,$$

which, by Proposition 2.1, means that for any sequence $(x_N,y_N) \to (x,y)$ w.p. 1,

$$\liminf_{N\to\infty} \delta_{\mathbb{Z}_N}(x_N,y_N) \geqslant \delta_{\mathbb{Z}_0}(x,y) \quad \text{w.p. } 1$$

and there exists $(\bar{x}_N, \overline{y}_N) \to (x, y)$ w.p. 1 such that

$$\limsup_{N\to\infty} \delta_{\mathbb{Z}_N}(x_N, y_N) \leqslant \delta_{\mathbb{Z}_0}(x, y) \quad \text{w.p. 1.}$$

We know from Lemma 4.1 that

$$\lim_{N\to\infty} \hat{f}_N(x_N, y_N) = \mathbb{E}[f(x, y, \xi)] \quad \text{w.p. 1 as} \quad N \to \infty,$$

which leads to

$$\liminf_{N\to\infty} [\hat{f}_N(x_N, y_N) + \delta_{\mathbb{Z}_N}(x_N, y_N)] \geqslant \mathbb{E}[f(x, y, \xi)] + \delta_{\mathbb{Z}_0}(x, y) \quad \text{w.p. 1}$$

and there exists $(\bar{x}_N, \overline{y}_N) \to (\bar{x}, \overline{y})$ w.p. 1 such that

$$\limsup_{N\to\infty} [\hat{f}_N(x_N, y_N) + \delta_{\mathbb{Z}_N}(x_N, y_N)] \leqslant \mathbb{E}[f(x, y, \xi)] + \delta_{\mathbb{Z}_0}(x, y) \quad \text{w.p. 1.}$$

Then also by Proposition 2.1, we obtain the conclusion. □

**Theorem 4.1.** *Suppose* $(x_N, y_N)$ *solves* (3.4) *for each* $N$ *and a random vector* $(\bar{x}, \overline{y})$ *is almost surely a cluster point of the sequence* $\{(x_N, y_N)\}$. *If the conditions in Proposition* 4.1 *hold and* $\kappa_0$ *is finite, then* $(\bar{x}, \overline{y})$ *is almost surely an optimal solution of the true problem* (1.1).

Proof. Since $\kappa_0$ is finite, we have $-\infty < \inf \overline{f} < \infty$, which, together with $\overline{f}_N$ epi-converging to $\overline{f}$ by [20, Theorem 7.31], means that

$$\limsup_{N\to\infty} \operatorname{argmin} \overline{f}_N \subset \operatorname{argmin} \overline{f} \quad \text{w.p. 1.} \tag{4.3}$$

Furthermore, $-\infty < \kappa_0 < \infty$ implies that $\mathbb{Z}_0 \neq \emptyset$ and there exists $(\tilde{x}, \tilde{y}) \in \mathbb{Z}_0$ such that $\mathbb{E}[f((\tilde{x}, \tilde{y}), \xi)]$ is finite. Then we have

$$\operatorname{argmin} \overline{f} = S_0 \tag{4.4}$$

and by Proposition 4.1, when $N$ is large enough, $\mathbb{Z}_N \neq \emptyset$ w.p. 1 and there exists $(\tilde{x}_N, \tilde{y}_N) \in \mathbb{Z}_N$ w.p. 1 such that $\hat{f}_N(\tilde{x}_N, \tilde{y}_N))$ is finite almost surely, which leads to the equivalence of $\operatorname{argmin} \overline{f}_N$ and $S_N$ w.p. 1. The conclusion follows from (4.3) and (4.4). □

Remark 4.1. From the above proof, we know that if the condition $\kappa_0$ being finite is replaced by $f$ being proper and $-\infty < \inf \overline{f} < \infty$, the conclusion in Theorem 4.1 can also be obtained.

### 4.2. Almost sure convergence of stationary points

In practice, finding a global minimizer might be difficult and in some cases we might just find a stationary point. As a result, we want to know whether or not a cluster point of the sequence of stationary points is almost surely a kind of stationary point of SMPCC (1.1). For this purpose, we need to investigate the almost sure convergence of stationary points of smoothed SAA problem (3.4) with the sample size tends to infinity.

Notice that (3.4) is a standard nonlinear programming with smooth constraints. If $(x_N, y_N)$ is a local optimal solution of the smoothed SAA problem (3.4), then under some constraint qualifications, $(x_N, y_N, \lambda_N)$ is a stationary point of (3.4) almost surely, namely, there exists a Lagrange multiplier $\lambda_N \in \mathbb{R}^m$ such that the vector $(x_N, y_N, \lambda_N)$ satisfies the following Karash-Kuhn-Tucker (KKT) condition for problem (3.4):

$$0 = \nabla \hat{f}_N(x_N, y_N) - \sum_{i=1}^{m} (\lambda_N)_i \nabla_{(x,y)} g_{t_N}((y_N)_i, \hat{F}_N(x_N, y_N)_i) \quad \text{w.p. 1.} \tag{4.5}$$

Moreover, $(x_N, y_N, \lambda_N)$ satisfies the inequality

$$(d^N)^{\mathrm{T}} \nabla^2_{(x,y)} \hat{L}_N(x_N, y_N, \lambda_N) d^N \geqslant 0 \quad \text{for all} \quad d^N \in T_N(x_N, y_N) \quad \text{w.p. 1,} \tag{4.6}$$

where

$$\hat{L}_N(x, y, \lambda) = \hat{f}_N(x, y) - \sum_{i=1}^{m} \lambda_i g_{t_N}(y_i, \hat{F}_N(x, y)_i)$$

and

$$T_N(x, y) = \{d \in \mathbb{R}^{n+m} \colon \nabla_{(x,y)} g_{t_N}(y_i, \hat{F}_N(x, y)_i)^{\mathrm{T}} d = 0, \ i = 1, \ldots, m\}.$$

Inequality (4.6) is called the *second order necessary condition* of problem (3.4).

From [16, Proposition 3.2] and by simple calculation we get the following properties of $g_t(\cdot, \cdot)$.

**Lemma 4.4.** *Let $(x_N, y_N)$ satisfy $\hat{G}_N(x, y) = 0$. Then under Assumption 1, one has that $g_{t_N}(y_i, \hat{F}_N(x, y)_i)$, $i = 1, \ldots, m$, is twice continuously differentiable at $(x_N, y_N)$ almost surely, and for all $i = 1, 2, \ldots, m$,*

$$\nabla_{(x,y)} g_{t_N}((y_N)_i, \hat{F}_N(x_N, y_N)_i) = \eta_{1i}^N \begin{pmatrix} 0 \\ e_i \end{pmatrix} + \eta_{2i}^N \nabla \hat{F}_N(x_N, y_N)_i \quad \text{w.p. 1} \tag{4.7}$$

*and*

$$
\begin{aligned}
(4.8)\quad & \nabla^2_{(x,y)} g_{t_N}((y_N)_i, \hat{F}_N(x_N,y_N)_i) \\
&= \eta^N_{2i} \nabla^2 \hat{F}_N(x_N,y_N)_i \\
&\quad - \frac{1}{t_N}\eta^N_{1i}\eta^N_{2i}\left[\begin{pmatrix}0\\ e_i\end{pmatrix}\begin{pmatrix}0\\ e_i\end{pmatrix}^{\mathrm{T}} + \nabla\hat{F}_N(x_N,y_N)_i \nabla\hat{F}_N(x_N,y_N)_i^{\mathrm{T}}\right] \\
&\quad + \frac{1}{t_N}\eta^N_{1i}\eta^N_{2i}\left[\begin{pmatrix}0\\ e_i\end{pmatrix}\nabla\hat{F}_N(x_N,y_N)_i^{\mathrm{T}} + \nabla\hat{F}_N(x_N,y_N)_i\begin{pmatrix}0\\ e_i\end{pmatrix}^{\mathrm{T}}\right] \quad \textit{w.p.}\,1,
\end{aligned}
$$

*where*

$$
\eta^N_{1i} = \frac{\exp(-(y_N)_i/t_N)}{\exp(-(y_N)_i/t_N) + \exp(-\hat{F}_N(x_N,y_N)_i/t_N)} \in (0,1),
$$

$$
\eta^N_{2i} = \frac{\exp(-\hat{F}_N(x_N,y_N)_i/t_N)}{\exp(-(y_N)_i/t_N) + \exp(-\hat{F}_N(x_N,y_N)_i/t_N)} \in (0,1)
$$

*with* $\eta^N_{1i} + \eta^N_{2i} = 1$.

The following lemma is important for deriving the convergence of our smoothing SAA method for SMPCC.

**Lemma 4.5.** *Suppose Assumptions* 1–2 *hold,* $t_N \searrow 0$ *and the index sets* $\overline{\alpha}$, $\overline{\beta}$, $\overline{\gamma}$ *are defined as in* (2.1). *Let* $(x_N, y_N)$ *be a feasible point of* (3.4) *for each* $N$ *and let the sequence* $\{(x_N,y_N)\}$ *converge to a random vector* $(\bar{x},\overline{y})$ *w.p.* 1 *as* $N$ *tends to infinity. Then*

(i) *For* $i \in \overline{\alpha}$, $\lim\limits_{N\to\infty} \nabla_{(x,y)} g_{t_N}((y_N)_i, \hat{F}_N(x_N,y_N)_i) = \begin{pmatrix}0\\ e_i\end{pmatrix}$ *w.p.* 1 *as* $N \to \infty$.

(ii) *For* $i \in \overline{\gamma}$, $\lim\limits_{N\to\infty} \nabla_{(x,y)} g_{t_N}((y_N)_i, \hat{F}_N(x_N,y_N)_i) = \nabla\Psi_i(\bar{x},\overline{y})$ *w.p.* 1 *as* $N \to \infty$.

(iii) *For* $i \in \overline{\beta}$, *if* $\lim\limits_{N\to\infty} \eta^N_{1i} = \overline{\eta}_{1i}$ *and* $\lim\limits_{N\to\infty} \eta^N_{2i} = \overline{\eta}_{2i}$, *we have*

$$
\lim_{N\to\infty} \nabla_{(x,y)} g_{t_N}((y_N)_i, \hat{F}_N(x_N,y_N)_i) = \overline{\eta}_{1i}\begin{pmatrix}0\\ e_i\end{pmatrix} + \overline{\eta}_{2i}\nabla\Psi_i(\bar{x},\overline{y}) \quad \textit{w.p.}\,1.
$$

P r o o f. For $i \in \overline{\alpha}$, we have

$$
\Psi_i(x,y) - y_i > 0. \tag{4.9}
$$

Furthermore, since $(x_N,y_N) \to (\bar{x},\overline{y})$ w.p. 1 as $N \to \infty$, there is a positive number $\varrho$ such that $\{(x_N,y_N)\} \subset \varrho\mathbb{B}$ almost surely. It follows that

$$
\begin{aligned}
&\|\hat{F}_N(x_N,y_N)_i - (y_N)_i - \Psi_i(\bar{x},\overline{y}) + \overline{y}_i\| \\
&\qquad \leqslant \max_{(x,y)\in\varrho\mathbb{B}} \|\hat{F}_N(x,y)_i - \Psi_i(x,y)\| \\
&\qquad\quad + \|\Psi_i(x_N,y_N) - \Psi_i(\bar{x},\overline{y})\| + \|(y_N)_i - \overline{y}_i\| \quad \text{w.p. } 1,
\end{aligned}
$$

which, under Assumptions 1–2, by the Uniform Laws of Large Numbers [24, Theorem 6.36], implies that

$$(\hat{F}_N(x_N, y_N)_i - (y_N)_i) \to (\Psi_i(\bar{x}, \overline{y}) - \overline{y}_i) \quad \text{w.p. 1 as } N \to \infty.$$

This, together with (4.9), leads to

$$\hat{F}_N(x_N, y_N)_i - (y_N)_i > 0, \quad \text{w.p. 1}$$

when $N$ is large enough. Consequently, by the definition of $\eta_{1i}^N$ and $\eta_{2i}^N$ in Lemma 4.4, we obtain

$$\eta_{1i}^N \to 1, \quad \eta_{2i}^N \to 0 \quad \text{w.p. 1 as } N \to \infty,$$

which, by (4.7) in Lemma 4.4, means that (i) holds.

Notice that by Lemma 4.1,

$$\nabla \hat{F}_N(x_N, y_N)_i \to \nabla \Psi_i(\bar{x}, \overline{y}) \quad \text{w.p. 1 as } N \to \infty.$$

Then we can obtain the conclusions (ii) and (iii) in the same way as in the proof of (i). We have completed the proof. □

By Lemma 4.1, we easily get the relationship of MPCC-LICQ between the SAA problem (1.2) and the true problem (1.1) when $N$ is sufficiently large.

**Lemma 4.6.** *Suppose Assumptions* 1–2 *hold. Let* $(x_N, y_N)$ *be a feasible point of* (3.4) *for each* $N$ *and let the sequence* $\{(x_N, y_N)\}$ *converge to a random vector* $(\bar{x}, \overline{y})$ *w.p.* 1 *as* $N$ *tends to infinity. If the MPCC-LICQ (Definition* 2.4*) holds at* $(\bar{x}, \overline{y})$ *w.p.* 1*, then for* $N$ *large enough*

$$\left\{ \begin{pmatrix} 0 \\ e_i \end{pmatrix} : \ i \in \overline{\alpha} \cup \overline{\beta} \right\} \cup \{\nabla \hat{F}_N(x_N, y_N)_i : \ i \in \overline{\beta} \cup \overline{\gamma}\}$$

*are linearly independent almost surely.*

We now prove the almost sure convergence of the smoothing SAA method for SMPCC (1.1).

**Theorem 4.2.** *Suppose Assumptions* 1–2 *hold. Let* $t_N \searrow 0$, $(x_N, y_N, \lambda_N)$ *be a stationary point of problem* (3.4)*, and let the sequence* $\{(x_N, y_N, \lambda_N)\}$ *converge to a random vector* $(\bar{x}, \overline{y}, \bar{\lambda})$ *w.p.* 1 *as* $N \to \infty$*. Then the following statements hold:*

(i) $(\bar{x}, \overline{y})$ *is a C-stationary point of SMPCC* (1.1) *almost surely.*

(ii) *Suppose in addition that* $(x_N, y_N, \lambda_N)$ *satisfies the second-order necessary condition* (4.6) *for each* $N$ *almost surely, and MPCC-LICQ and ULSC (Definition* 2.6*) hold at* $(\bar{x}, \overline{y})$ *almost surely. Then* $(\bar{x}, \overline{y})$ *is an S-stationary point of SMPCC* (1.1) *almost surely.*

P r o o f. Notice that $(x_N, y_N, \lambda_N)$ satisfies

$$0 = \nabla \hat{f}_N(x_N, y_N) - \sum_{i=1}^{m} (\lambda_N)_i \nabla_{(x,y)} g_{t_N}((y_N)_i, \hat{F}_N(x_N, y_N)_i) \quad \text{w.p. 1.} \tag{4.10}$$

For $i \in \overline{\beta}$, due to the boundedness of $\eta_{1i}^N$ and $\eta_{2i}^N$, for simplicity, we assume $\lim\limits_{N\to\infty} \eta_{1i}^N = \overline{\eta}_{1i}$ w.p. 1 and $\lim\limits_{N\to\infty} \eta_{2i}^N = \overline{\eta}_{2i}$ w.p. 1. Then taking the limit in the equation (4.10) and by Lemma 4.1, we have

$$\begin{aligned} 0 = \nabla \mathbb{E}[f(\bar{x}, \overline{y}, \xi(\omega))] - \sum_{i\in\overline{\alpha}} \bar{\lambda}_i \begin{pmatrix} 0 \\ e_i \end{pmatrix} - \sum_{i\in\overline{\gamma}} \bar{\lambda}_i \nabla \Psi_i(\bar{x}, \overline{y}) - \sum_{i\in\overline{\beta}} \bar{u}_i \begin{pmatrix} 0 \\ e_i \end{pmatrix} \\ - \sum_{i\in\overline{\beta}} \overline{v}_i \nabla \Psi_i(\bar{x}, \overline{y}) \quad \text{w.p. 1} \end{aligned}$$

with $\bar{u}_i = \overline{\eta}_{1i}\bar{\lambda}_i$, $\overline{v}_i = \overline{\eta}_{2i}\bar{\lambda}_i$, $i \in \beta$. Since $\bar{u}_i\overline{v}_i \geqslant 0$ for $i \in \beta$, we obtain the conclusion of (i) from Definition 2.5.

Next we prove that under conditions (ii), $\bar{u}_i \geqslant 0$, $\overline{v}_i \geqslant 0$ w.p. 1 for $i \in \beta$. We assume by contradiction that $\bar{u}_j < 0$ w.p. 1 for some $j \in \beta$. Since MPCC-LICQ holds at $(\bar{x}, \overline{y})$ w.p. 1 by Lemma 4.6, we can choose a vector $d^N \in \mathbb{R}^{n+m}$ such that w.p. 1 for $N$ large enough,

$$\begin{gathered} \begin{pmatrix} 0 \\ e_j \end{pmatrix}^{\mathrm{T}} d^N = \eta_{2j}^N, \\ \begin{pmatrix} 0 \\ e_j \end{pmatrix}^{\mathrm{T}} d^N = 0, \quad i \in \overline{\alpha} \cup \overline{\beta} \setminus \{j\}, \\ \nabla \hat{F}_N(x_N, y_N)_j^{\mathrm{T}} d^N = -\eta_{1j}^N, \\ \nabla \hat{F}_N(x_N, y_N)_i^{\mathrm{T}} d^N = 0, \quad i \in \overline{\gamma} \cup \overline{\beta} \setminus \{j\}. \end{gathered} \tag{4.11}$$

Then we obtain w.p. 1,

$$\begin{aligned} &\nabla_{(x,y)} g_{t_N}((y_N)_j, \hat{F}_N(x_N, y_N)_j)^{\mathrm{T}} d^N \\ &\quad = \eta_{1j}^N \begin{pmatrix} 0 \\ e_j \end{pmatrix}^{\mathrm{T}} d^N + \eta_{2j}^N \nabla \hat{F}_N(x_N, y_N)_j^{\mathrm{T}} d^N = 0, \end{aligned}$$

which implies that $d^N \in T_N(x_N, y_N)$ w.p. 1. Notice that w.p. 1,

$$\begin{aligned} &(d^N)^{\mathrm{T}} \nabla_{(x,y)}^2 \hat{L}_N(x_N, y_N, \lambda_N) d^N \\ &\quad = (d^N)^{\mathrm{T}} \Bigg[ \nabla^2 \hat{f}_N(x_N, y_N) \\ &\qquad - \sum_{i=1}^{m} (\lambda_N)_i \nabla_{(x,y)}^2 g_{t_N}((y_N)_i, \hat{F}_N(x_N, y_N)_i) \Bigg] d^N. \end{aligned} \tag{4.12}$$

We know from (4.8) in Lemma 4.4 that w.p. 1,

$$
\begin{aligned}
(4.13)\qquad & (d^N)^{\mathrm{T}}\nabla^2_{(x,y)}g_{t_N}((y_N)_i,\hat{F}_N(x_N,y_N)_i)d^N \\
&= \eta^N_{2i}(d^N)^{\mathrm{T}}\nabla^2\hat{F}_N(x_N,y_N)_i d^N \\
&\quad - \frac{1}{t_N}\eta^N_{1i}\eta^N_{2i}\Bigg[(d^N)^{\mathrm{T}}\begin{pmatrix}0\\ e_i\end{pmatrix}\begin{pmatrix}0\\ e_i\end{pmatrix}^{\mathrm{T}} d^N \\
&\quad + (d^N)^{\mathrm{T}}\nabla\hat{F}_N(x_N,y_N)_i\nabla\hat{F}_N(x_N,y_N)_i^{\mathrm{T}} d^N\Bigg] \\
&\quad + \frac{1}{t_N}\eta^N_{1i}\eta^N_{2i}\Bigg[(d^N)^{\mathrm{T}}\begin{pmatrix}0\\ e_i\end{pmatrix}\nabla\hat{F}_N(x_N,y_N)_i^{\mathrm{T}} d^N \\
&\quad + (d^N)^{\mathrm{T}}\nabla\hat{F}_N(x_N,y_N)_i\begin{pmatrix}0\\ e_i\end{pmatrix}^{\mathrm{T}} d^N\Bigg],
\end{aligned}
$$

which, together with (4.11) and by Lemma 4.4, means that for $i=j$,

$$
\begin{aligned}
(4.14)\qquad & (d^N)^{\mathrm{T}}\nabla^2_{(x,y)}g_{t_N}((y_N)_j,\hat{F}_N(x_N,y_N)_j)d^N \\
&= \eta^N_{2j}(d^N)^{\mathrm{T}}\nabla^2\hat{F}_N(x_N,y_N)_j d^N \\
&\quad - \frac{1}{t_N}\eta^N_{1j}\eta^N_{2j}[(\eta^N_{1j})^2+(\eta^N_{2j})^2] + \frac{1}{t_N}\eta^N_{1j}\eta^N_{2j}[2\eta^N_{1j}(-\eta^N_{2j})] \\
&= \eta^N_{2j}(d^N)^{\mathrm{T}}\nabla^2\hat{F}_N(x_N,y_N)_j d^N - \frac{1}{t_N}\eta^N_{1j}\eta^N_{2j} \quad \text{w.p. 1.}
\end{aligned}
$$

Similarly, for $i\neq j$, we have

$$
\begin{aligned}
(4.15)\qquad & (d^N)^{\mathrm{T}}\nabla^2_{(x,y)}g_{t_N}((y_N)_i,\hat{F}_N(x_N,y_N)_i)d^N \\
&= \eta^N_{2i}(d^N)^{\mathrm{T}}\nabla^2\hat{F}_N(x_N,y_N)_i d^N \quad \text{w.p. 1.}
\end{aligned}
$$

As a result, combining (4.12), (4.14), and (4.15), we obtain w.p. 1,

$$
\begin{aligned}
(4.16)\quad & (d^N)^{\mathrm{T}}\nabla^2_{(x,y)}\hat{L}_N(x_N,y_N,\lambda_N)d^N \\
&= (d^N)^{\mathrm{T}}\Bigg[\nabla^2\hat{f}_N(x_N,y_N) - \sum_{i=1}^m(\lambda_N)_i(d^N)^{\mathrm{T}}\eta^N_{2i}\nabla^2\hat{F}_N(x_N,y_N)_i\Bigg]d^N \\
&\quad + \frac{1}{t_N}\,\eta^N_{1j}\eta^N_{2j}(\lambda_N)_j.
\end{aligned}
$$

Since the ULSC condition holds at $(\bar{x},\bar{y})$ w.p. 1, we have $\lim_{N\to\infty}\eta^N_{1j}\eta^N_{2j}>0$ w.p. 1. Moreover, $\bar{u}_j<0$ implies that $\lim_{N\to\infty}(\lambda_N)_j=\bar{\lambda}_j<0$ w.p. 1. Hence, we obtain

$$
(4.17)\qquad \frac{1}{t_N}\eta^N_{1j}\eta^N_{2j}(\lambda_N)_j\to-\infty \quad \text{w.p. 1} \quad \text{as } N\to\infty.
$$

Next we show that $\{d^N\}$ can be chosen bounded w.p. 1 for $d^N$ satisfying equations (4.11) w.p. 1 for each $N$. Let $H(z) = T(z) - p$ and $T(z) = Az - b$, where

$$A := \begin{bmatrix} (0, e_j^{\mathrm{T}}) \\ (0, e_i^{\mathrm{T}})_{\overline{\alpha}\cup\overline{\beta}\setminus\{j\}} \\ \nabla\Psi_j(\bar{x}, \overline{y}) \\ \nabla\Psi_{\overline{\gamma}\cup\overline{\beta}\setminus\{j\}}(\bar{x}, \overline{y}) \end{bmatrix} \quad \text{and} \quad b := \begin{bmatrix} \overline{\eta}_{2j} \\ 0_{\overline{\alpha}\cup\overline{\beta}\setminus\{j\}} \\ -\overline{\eta}_{1j} \\ 0_{\overline{\gamma}\cup\overline{\beta}\setminus\{j\}} \end{bmatrix}.$$

Then we know from Lemma 4.6 that $A$ has full row rank and hence there exists $\bar{d}$ such that $H(\bar{d}, 0) = 0$. By the implicit function theorem, there exist positive numbers $\varepsilon$, $\delta$ and a continuous function $z(\cdot)\colon \varepsilon\mathbb{B} \to \mathbb{B}(\bar{d}, \delta)$ such that $z(0) = \bar{d}$ and $H(z(p), p) = 0$ for $p \in \varepsilon\mathbb{B}$, which means that $T(\cdot)$ is so called subinvertible [5] at $(\bar{d}, 0)$. Moreover, let $Q_N(z) = (A_N - A)z + (b - b_N)$ and $J(Q_N) = \{z\colon\ 0 \in T(z) + Q_N(z)\}$, where

$$A_N := \begin{bmatrix} (0, e_j^{\mathrm{T}}) \\ (0, e_i^{\mathrm{T}})_{\overline{\alpha}\cup\overline{\beta}\setminus\{j\}} \\ \nabla\hat{F}_N(x_N, y_N)_j \\ \nabla\hat{F}_N(x_N, y_N)_{\overline{\gamma}\cup\overline{\beta}\setminus\{j\}} \end{bmatrix} \quad \text{and} \quad b_N := \begin{bmatrix} \eta_{2j}^N \\ 0_{\overline{\alpha}\cup\overline{\beta}\setminus\{j\}} \\ -\eta_{1j}^N \\ 0_{\overline{\gamma}\cup\overline{\beta}\setminus\{j\}} \end{bmatrix}.$$

We have from Lemma 4.1 that $Q_N(z) \to 0$ w.p. 1 as $N \to \infty$, which implies that for a bounded neighborhood $U$ of $\bar{d}$ and any positive number $\sigma$, when $N$ is large enough,

$$\|Q_N\|_U := \sup_{z\in U} \|Q_N(z)\| \leqslant \sigma \quad \text{w.p. 1}, \tag{4.18}$$

which, together with the subinvertibility of $T(\cdot)$, by [5, Proposition 3.1] means that when $N$ is sufficiently large,

$$U \cap J(Q_N) \neq \emptyset \quad \text{w.p. 1}. \tag{4.19}$$

Notice that $z \in J(Q_N)$ w.p. 1 means $A_N z = b_N$ w.p. 1. This, together with (4.19), implies that $\{d^N\}$ can be chosen almost surely bounded. By Lemma 4.1, we obtain

$$\nabla^2 \hat{f}_N(x_N, y_N) \to \mathbb{E}[\nabla^2 f(\bar{x}, \overline{y}, \xi)] \quad \text{w.p. 1} \quad \text{as } N \to \infty$$

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

$$\nabla^2 \hat{F}_N(x_N, y_N)_i \to \mathbb{E}[\nabla^2 F(\bar{x}, \overline{y}, \xi)_i] \quad \text{w.p. 1} \quad \text{as } N \to \infty.$$

This, together with the almost sure boundedness of $\{d^N\}$, $\{\eta_{1i}^N\}$, and $\{\eta_{2i}^N\}$, $i = 1, \ldots, m$, leads to the almost sure boundedness of

$$\{(d^N)^{\mathrm{T}} \nabla^2 \hat{f}_N(x_N, y_N) d^N\}$$

and

$$\{(\lambda_N)_i (d^N)^{\mathrm{T}} \eta_{2i}^N \nabla^2 \hat{F}_N(x_N, y_N)_i d^N\}, \quad i = 1, \ldots, m.$$

As a result, combining (4.16) and (4.17), we can choose a sequence $N_k \subset \mathbb{N}$ such that

$$(d^{N_k})^{\mathrm{T}} \nabla^2_{(x,y)} \hat{L}_{N_k}(x_{N_k}, y_{N_k}, \lambda_{N_k}) d^{N_k} \to -\infty \quad \text{w.p. } 1 \quad \text{as } N \to \infty.$$

This contradicts the condition that $(x_N, y_N, \lambda_N)$ satisfies the second order necessary conditions almost surely. Hence, we have that $\bar{u}_i \geqslant 0$ w.p. 1 holds for all $i \in \overline{\beta}$. Similarly, $\overline{v}_i \geqslant 0$ w.p. 1 for all $i \in \overline{\beta}$. Therefore, we know from Definition 2.5 that $(\bar{x}, \overline{y})$ is w.p. 1 an S-stationary point of SMPCC (1.1). $\square$

Utilizing the MPCC-SSOSC (Definition 2.7), we obtain the following theorem through an application of standard NLP stability theory.

**Theorem 4.3.** *Suppose*

(i) $(\bar{x}, \overline{y})$ *is a C-stationary point of SMPCC* (1.1) *and MPCC-LICQ as well as MPCC-SSOSC hold at* $(\bar{x}, \overline{y})$.

(ii) $t_N \searrow 0$ *and ULSC holds at* $(\bar{x}, \overline{y})$.

(iii) *Assumptions* 1–2 *hold at* $(\bar{x}, \overline{y})$.

*Then there exists* $(x_N, y_N)$ *satisfying the stationary condition* (4.5) *of* (3.4) *w.p.* 1 *for each* $N$ *when* $N$ *is sufficiently large and* $(x_N, y_N) \to (\bar{x}, \overline{y})$ *w.p.* 1 *as* $N \to \infty$.

Proof. Since $(\bar{x}, \overline{y})$ is a C-stationary point of SMPCC, there exist vectors $\bar{u} \in \mathbb{R}^{|\overline{\alpha}|+|\overline{\beta}|}$, $\overline{v} \in \mathbb{R}^{|\overline{\beta}|+|\overline{\gamma}|}$ such that

$$\Phi(\bar{x}, \overline{y}, \bar{u}, \overline{v}) = 0 \tag{4.20}$$

and

$$\bar{u}_i \overline{v}_i \geqslant 0, \quad i \in \overline{\beta}, \tag{4.21}$$

where

$$\Phi(x, y, u, v) = \begin{bmatrix} \nabla_{(x,y)} L(x, y, u, v) \\ \overline{y}_{\overline{\alpha} \cup \overline{\beta}} \\ \Psi_{\overline{\beta} \cup \overline{\gamma}}(x, y) \end{bmatrix}$$

with

$$L(x, y, u, v) = \mathbb{E}[f(x, y, \xi(\omega))] - \sum_{i \in \overline{\alpha} \cup \overline{\beta}} u_i y_i - \sum_{i \in \overline{\beta} \cup \overline{\gamma}} v_i \Psi_i(x, y).$$

Notice that the equation (4.20) can be seen as the KKT condition of the NLP problem

$$\begin{aligned} \min \quad & \mathbb{E}[f(x, y, \xi(\omega))] \\ \text{s.t.} \quad & y_i = 0, \ i \in \overline{\alpha} \cup \overline{\beta}, \\ & \Psi_i(x, y) = 0, \ i \in \overline{\beta} \cup \overline{\gamma}. \end{aligned} \tag{4.22}$$

The MPCC-SSOSC ensures the strong second order sufficient condition for the NLP problem (4.22), which, under MPCC-LICQ, implies the stability of (4.22) in the sense of Robinson [17]. Hence, there exist positive numbers $\varepsilon$, $\delta$, $c$ such that for every $p \in \mathbb{B}(0,\varepsilon)$, the mapping $\Sigma(p) = \{z \in \mathbb{R}^{n+m+|\overline{\alpha}|+2|\overline{\beta}|+|\overline{\gamma}|}\colon\ 0 \in \Phi(z) + p,\ z = (x,y,u,v)\}$ has only one solution $z(p) \in \mathbb{B}(\overline{z},\delta)$, $\overline{z} = (\bar{x},\overline{y},\bar{u},\overline{v}) = z(0)$ and the mapping $z(\cdot)\colon\ \mathbb{B}(0,\varepsilon) \to \mathbb{B}(\overline{z},\delta)$ satisfies

$$\|z(p) - z(p')\| \leqslant c\|p - p'\| \text{ for any } p, p' \in \mathbb{B}(0,\varepsilon). \tag{4.23}$$

Let

$$Q^N(z) = \begin{bmatrix} Q_1^N(z) \\ g_{t_N}(y_i, \hat{F}_N(x,y)_i) - y_i,\ i \in \overline{\alpha} \\ -t_N \ln 2, \quad i \in \overline{\beta} \\ -t_N \ln 2, \quad i \in \overline{\beta} \\ g_{t_N}(y_i, \hat{F}_N(x,y)_i) - \Psi_i(x,y),\ i \in \overline{\gamma} \end{bmatrix}, \tag{4.24}$$

where

$$\begin{aligned} Q_1^N(z) = {} & \nabla \hat{f}_N(x,y) - \nabla \mathbb{E}[f(x,y,\xi(\omega))] \\ & - \sum_{i \in \overline{\alpha}} \frac{u_i}{\eta_{1i}^N(x,y)}\, \eta_{2i}^N(x,y) \nabla \hat{F}_N(x,y)_i \\ & - \sum_{i \in \overline{\gamma}} \frac{v_i}{\eta_{2i}^N(x,y)}\, \eta_{1i}^N(x,y) \begin{pmatrix} 0 \\ e_i \end{pmatrix} \\ & - \sum_{i \in \overline{\beta} \cup \overline{\gamma}} v_i [\nabla \hat{F}_N(x,y)_i - \nabla \Psi_i(x,y)] \end{aligned} \tag{4.25}$$

with

$$\eta_{1i}^N(x,y) = \frac{\exp(-y_i/t_N)}{\exp(-y_i/t_N) + \exp(-\hat{F}_N(x,y)_i/t_N)} \in (0,1),$$
$$\eta_{2i}^N(x,y) = \frac{\exp(-\hat{F}_N(x,y)_i/t_N)}{\exp(-y_i/t_N) + \exp(-\hat{F}_N(x,y)_i/t_N)} \in (0,1).$$

We know from the proof of Lemma 4.5 that when $\delta$ is sufficiently small then for $i \in \overline{\alpha}$,

$$\sup_{z \in \mathbb{B}(\overline{z},\delta)} \frac{\eta_{2i}^N(x,y)}{\eta_{1i}^N(x,y)} \to 0 \quad \text{w.p. } 1 \quad \text{as } N \to \infty \tag{4.26}$$

and for $i \in \overline{\gamma}$,

$$\sup_{z \in \mathbb{B}(\overline{z},\delta)} \frac{\eta_{1i}^N(x,y)}{\eta_{2i}^N(x,y)} \to 0 \quad \text{w.p. } 1 \quad \text{as } N \to \infty. \tag{4.27}$$

By the Uniform Laws of Large Numbers, we have

$$\sup_{z\in\mathbb{B}(\overline{z},\delta)} \|\nabla \hat{f}_N(x,y) - \nabla\mathbb{E}[f(x,y,\xi(\omega))]\| \to 0 \quad \text{w.p. } 1 \quad \text{as } N\to\infty \tag{4.28}$$

and for $i \in m$,

$$\sup_{z\in\mathbb{B}(\overline{z},\delta)} \|\nabla \hat{F}_N(x,y)_i - \nabla\Psi_i(x,y)\| \to 0 \quad \text{w.p. } 1 \quad \text{as } N\to\infty. \tag{4.29}$$

As a result, combining (4.25)–(4.29), we obtain that when $\delta$ is sufficiently small,

$$\|Q_1^N\|_\delta = \sup_{z\in\mathbb{B}(\overline{z},\delta)} \|Q_1^N(z)\| \to 0 \quad \text{w.p. } 1 \quad \text{as } N\to\infty, \tag{4.30}$$

which implies that for $\varepsilon > 0$, when $N$ is sufficiently large, $\|Q_1^N\|_\delta < \varepsilon$ w.p. 1. In addition, we know from the definition and the Uniform Laws of Large Numbers that when $\delta$ is sufficiently small, then

$$\sup_{z\in\mathbb{B}(\overline{z},\delta)} \|g_{t_N}(y_i, \hat{F}_N(x,y)_i) - \min\{y_i, \Psi_i(x,y)\}\| \to 0 \quad \text{w.p. } 1 \quad \text{as } N\to\infty,$$

which implies that for the above $\varepsilon > 0$, when $N$ is sufficiently large,

$$\sup_{z\in\mathbb{B}(\overline{z},\delta)} \|g_{t_N}(y_i, \hat{F}_N(x,y)_i) - y_i\| < \varepsilon \quad \text{w.p. } 1 \quad \text{for } i\in\overline{\alpha} \tag{4.31}$$

and

$$\sup_{z\in\mathbb{B}(\overline{z},\delta)} \|g_{t_N}(y_i, \hat{F}_N(x,y)_i) - \Psi_i(x,y)\| < \varepsilon \quad \text{w.p. } 1 \quad \text{for } i\in\overline{\gamma}. \tag{4.32}$$

Note that $t_N \ln 2 \to 0$ w.p. 1 as $N\to\infty$. Hence, we know from (4.24), (4.30), (4.31), and (4.32) that for the above $\varepsilon > 0$, when $\delta$ is sufficiently small and $N$ sufficiently large, then

$$\|Q^N\|_\delta = \sup_{z\in\mathbb{B}(\overline{z},\delta)} \|Q^N(z)\| < \varepsilon \quad \text{w.p. } 1. \tag{4.33}$$

Applying Brouwer's fixed point theorem to the mapping $z(Q^N(\cdot))\colon\ \mathbb{B}(\overline{z},\delta)\to\mathbb{B}(\overline{z},\delta)$, where $z(\cdot)$ is defined as in (4.23), we conclude that there is at least one fixed point $z_N = (x_N, y_N, u_N, v_N) \in \mathbb{R}^{m+n+|\overline{\alpha}|+2|\overline{\beta}|+|\overline{\gamma}|}$ such that $z_N = z(Q^N(z_N))$ w.p. 1.

Therefore, when $N$ is sufficiently large, there exists $z_N \in \mathbb{B}(\overline{z}, \delta)$ w.p. 1 such that $0 \in \Phi(z_N) + Q^N(z_N)$ w.p. 1, namely,

$$
\begin{aligned}
(4.34) \qquad 0 \in\ & \nabla \hat{f}_N(x_N, y_N) \\
& - \sum_{i \in \overline{\alpha}} \left[ (u_N)_i \begin{pmatrix} 0 \\ e_i \end{pmatrix} + \frac{(u_N)_i}{\eta_{1i}^N(x_N, y_N)} \eta_{2i}^N(x_N, y_N) \nabla \hat{F}_N(x_N, y_N)_i \right] \\
& - \sum_{i \in \overline{\gamma}} \left[ \frac{(v_N)_i}{\eta_{2i}^N(x_N, y_N)} \eta_{1i}^N(x_N, y_N) \begin{pmatrix} 0 \\ e_i \end{pmatrix} + (v_N)_i \nabla \hat{F}_N(x_N, y_N)_i \right] \\
& - \sum_{i \in \overline{\beta}} \left[ (u_N)_i \begin{pmatrix} 0 \\ e_i \end{pmatrix} + (v_N)_i \nabla \hat{F}_N(x_N, y_N)_i \right] \quad \text{w.p. } 1
\end{aligned}
$$

and $g_{t_N}((y_N)_i, \hat{F}_N(x_N, y_N)_i) = 0$. Moreover, combining (4.23) and (4.33), we obtain

$$
(4.35) \qquad z_N \to \overline{z} \quad \text{w.p. } 1 \quad \text{as } N \to \infty.
$$

Under ULSC and condition (4.21), (4.35) leads to $(u_N)_i (v_N)_i > 0$ w.p. 1 for $i \in \overline{\beta}$ when $N$ is sufficiently large. As a result, if

$$
\lambda_N = \left( \frac{(u_N)_i}{\eta_{1i}^N(x_N, y_N)},\ i \in \overline{\alpha};\ \sqrt{\frac{(u_N)_i (v_N)_i}{\eta_{1i}^N(x_N, y_N) \eta_{2i}^N(x_N, y_N)}},\ i \in \overline{\beta};\ \frac{(v_N)_i}{\eta_{2i}^N(x_N, y_N)},\ i \in \overline{\gamma} \right),
$$

then we have from (4.34) that $(x_N, y_N)$ is almost surely a stationary point of (3.4) and $\lambda_N$ is the corresponding multiplier. Furthermore, by (4.35), we have $(x_N, y_N) \to (\bar{x}, \overline{y})$ w.p. 1 as $N \to \infty$. The proof is completed. □

## 5. Numerical results

In this section we present some preliminary numerical results obtained by the smoothing SAA method based on the log-exponential function (3.1). Our numerical experiments have been carried out in Matlab 7.1 running on a PC with Intel Pentium M of 1.60 GHz CPU and our tests are focused on different values of the smoothing parameter $t$ and sample size $N$.

In our experiments, we set the initial values of $N_k$ and $t_k$ as $N_1 = 100$ and $t_1 = 5$, respectively. Then we employed the random number generator `unifrnd` in Matlab 7.1 to generate independently and identically distributed random samples $\{\xi_1, \xi_2, \ldots, \xi_{N_k}\}$. We solved problem (3.4) with $N = N_k$ and $t = t_k$ by the solver `fmincon` in Matlab 7.1 to obtain the approximate optimal solution $(x_{N_k}, y_{N_k})$. The initial point was $(1, \ldots, 1)^{\mathrm{T}}$. In addition, the parameters were updated by

$N_{k+1} := \min\{10N_k, 10^5\}$ and $t_{k+1} := \max\{0.1t_k, 0.005\}$. Throughout the tests, we recorded number of iterations of `fmincon` (`Iter`), the values of the objective function of problem (3.4) at $(x_{N_k}, y_{N_k})$ (`Obj`) and these quantities are displayed in the tables of test results.

The example below is varied from an example in Shapiro and Xu [22].

Example 1. Consider

$$\begin{aligned} \min \quad & f(x,y) = \mathbb{E}[(x_1-1)^2 + (x_2-1)^2 + 2y_1^2\xi + y_2^2 + 2\xi - 1] \\ \text{s.t.} \quad & 0 \leqslant y \perp \mathbb{E}[F(x,y,\xi)] \geqslant 0, \end{aligned}$$

where $F(x,y,\xi) = (y_1 - x_1 + \xi, y_2 - x_2 + \xi)^{\mathrm{T}}$ and $\xi$ is uniformly distributed on $[0,1]$. The constraint here, which is a complementarity problem, has a unique solution $y = (y_1, y_2)$, where

$$y_i = \begin{cases} x_i - \frac{1}{2}, & x_i \geqslant \frac{1}{2}, \\ 0, & \text{otherwise}, \end{cases}$$

for $i = 1, 2$. Therefore, substituting the above $y_i$ into the objective function, we obtain that $(0.75, 0.75, 0.25, 0.25)$ is the exact optimal solution and $0.25$ is the optimal value. The test results are presented in Tab. 1.

| $N_k$ | $t_{N_k}$ | $(x_{N_k}, y_{N_k})$ | `Obj` | `Iter` |
|---|---|---|---|---|
| $10^2$ | 5 | $(-0.3869\ \ -0.3869\ 3.0425\ 3.0425)$ | 22.3578 | 6 |
| $10^3$ | $5 \times 10^{-1}$ | (0.7231 0.7231 0.4603 0.4603) | 0.6107 | 5 |
| $10^4$ | $5 \times 10^{-2}$ | (0.7495 0.7495 0.2521 0.2521) | 0.2480 | 6 |
| $10^5$ | $5 \times 10^{-3}$ | (0.7497 0.7492 0.2512 0.2507) | 0.2482 | 5 |

Table 1. The computational results for Example 1.

Example 2. Consider

$$\begin{aligned} \min \quad & f(x,y) = \mathbb{E}[2(x_1-2)^2 + 2(x_2-1)^2\xi_2 + y_1\xi_2 + y_2\xi_1] \\ \text{s.t.} \quad & 0 \leqslant y \perp \mathbb{E}[F(x,y,\xi)] \geqslant 0, \end{aligned}$$

where $F(x,y,\xi) = (3x_1 + x_2\xi_1 + 3y_1\xi_1 + 3y_2 - 2\xi_1, -5x_1\xi_1 + x_2\xi_2 + 4y_2\xi_2 - \xi_1)^{\mathrm{T}}$, $\xi = (\xi_1, \xi_2)$ and $\xi_1$, $\xi_2$ are independent random variables having uniform distribution on $[0,1]$. The test results are displayed in Tab. 2.

| $N_k$ | $t_{N_k}$ | $(x_{N_k}, y_{N_k})$ | Obj | Iter |
|---|---|---|---|---|
| $10^2$ | 5 | (1.9038 1.0442 0.1472 3.8332) | 2.1429 | 8 |
| $10^3$ | $5 \times 10^{-1}$ | (1.8471 1.0622 0.0000 2.2959) | 1.1797 | 6 |
| $10^4$ | $5 \times 10^{-2}$ | (1.8439 1.0625 0.0000 2.2892) | 1.1965 | 6 |
| $10^5$ | $5 \times 10^{-3}$ | (1.8434 1.0625 0.0000 2.2886) | 1.1998 | 6 |

Table 2. The computational results for Example 2.

Notice that in Example 2 we can integrate the underlying functions out and obtain a deterministic MPCC problem

$$\begin{aligned} \min \quad & 2(x_1-2)^2 + (x_2-1)^2 + \tfrac{1}{2}y_1 + \tfrac{1}{2}y_2 \\ \text{s.t.} \quad & 0 \leqslant y \perp \overline{F}(x,y) \geqslant 0, \end{aligned}$$

where $\overline{F}(x,y) = (3x_1 + \frac{1}{2}x_2 + \frac{3}{2}y_1 + 3y_2 - 1, -\frac{5}{2}x_1 + \frac{1}{2}x_2 + 2y_2 - \frac{1}{2})^{\mathrm{T}}$. In order to compare the performance of numerical results of the SAA method, we also report the numerical results of the deterministic problem. The test results are stated in Tab. 3.

| $t$ | $(x^*, y^*)$ | Obj | Iter |
|---|---|---|---|
| $5 \times 10^{-2}$ | (1.8437 1.0627 0.0000 2.2881) | 1.1968 | 6 |
| $5 \times 10^{-3}$ | (1.8437 1.0626 0.0000 2.2888) | 1.1971 | 6 |

Table 3. The computational results for deterministic problem of Example 2.

Our preliminary numerical results shown in Tabs. 1 and 2 reveal that our proposed method yields a reasonable solution of the problems considered.

## 6. Conclusion and further remarks

In this paper, we propose a smoothing SAA method for a SMPCC by using the log-exponential function. Utilizing the notion of epi-convergence in variational analysis, we establish the almost sure convergence of optimal solutions generated by the smoothed SAA problem. Moreover, under suitable conditions, we show that any cluster point of the KKT point sequence generated from the smoothed SAA problem is almost surely an S-stationary point of SMPCC as the sample size tends to infinity. The preliminary numerical results indicate that the proposed method is able to solve SMPCC successfully.

Let $\xi^1, \ldots, \xi^N$ be an iid sample of $\xi$. By the flexible structure of the log-exponential function and defining

$$\hat{G}_N(z)_k = -t_N \left( \sum_{i=1}^{l} \exp(-\hat{F}_{ki}^N(z)/t_N) \right), \quad k = 1, \ldots, m,$$

where $\hat{F}_{ki}^N(z) := N^{-1} \sum_{j=1}^{N} F_{ki}(z, \xi^j)$ is the sample average function of $F_{ki}(z, \xi^j)$, the method in our paper can be easily extended to solve the stochastic mathematical program with general vertical complementarity constraints in Birbil et al. [2]:

$$\begin{aligned} \min \quad & \mathbb{E}[f(z, \xi(\omega))] \\ \text{s.t.} \quad & \min\{\mathbb{E}[F_{k1}(z, \xi)], \ldots, \mathbb{E}[F_{kl}(z, \xi)]\} = 0, \quad k = 1, \ldots, m. \end{aligned}$$

However, the smoothing SAA method based on the Fischer-Burmeister function [4] in [22] is difficult to be extended for solving this kind of problems.

Since there is no assumption on measurability of the selections considered, the almost sure convergence appears relatively weak. The uniform almost sure convergence [25] is a more convenient convergence in non-measurable case. Whether the almost sure convergence results in this paper can be extended to uniform almost sure convergence results is one of the important topics in our further study.

*References*


[1] *S. I. Birbil, S.-C. Fang, J. Han*: An entropic regularization approach for mathematical programs with equilibrium constraints. Comput. Oper. Res. *31* (2004), 2249–2262.

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

[3] *F. H. Clarke*: Optimization and Nonsmooth Analysis. John Wiley & Sons, New York, 1983.

[4] *A. Fischer*: A special Newton-type optimization method. Optimization *24* (1992), 269–284.

[5] *A. J. King, R. T. Rockafellar*: Sensitivity analysis for nonsmooth generalized equations. Math. Program. *55* (1992), 193–212.

[6] *J. Linderoth, A. Shapiro, S. Wright*: The empirical behavior of sampling methods for stochastic programming. Ann. Oper. Res. *142* (2006), 215–241.

[7] *X. S. Li*: An aggregate function method for nonlinear programming. Sci. China (Ser. A) *34* (1991), 1467–1473.

[8] *X. S. Li*: An entropy-based aggregate method for minimax optimization. J. Eng. Optim. *18* (1992), 277–185.

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

[10] *Z.-Q. Luo, J.-S. Pang, D. Ralph*: Mathematical Programs with Equilibrium Constraints. Cambridge University Press, Cambridge, 1997.

[11] *J. V. Outrata*: Mathematical programs with equilibrium constraints: Theory and numerical methods. In: Nonsmooth Mechanics of Solids. CISM Courses and Lecture Notes, vol. 485 (J. Haslinger, G. E. Stavroulakis, eds.). Springer, New York, 2006, pp. 221–274.

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

[13] *J. Pang, M. Fukushima*: Complementarity constraint qualifications and simplified $B$-stationary conditions for mathematical programs with equilibrium constraints. Comput. Optim. Appl. *13* (1999), 111–136.

[14] *J. Peng, Z. Lin*: A non-interior continuation method for generalized linear complementarity problems. Math. Program. *86* (1999), 533–563.
[15] *E. L. Plambeck, B.-R. Fu, S. M. Robinson, R. Suri*: Sample-path optimization of convex stochastic performances functions. Math. Program. *75* (1996), 137–176.
[16] *H. Qi, L. Liao*: A smoothing Newton method for extended vertical linear complementarity problems. SIAM J. Matrix Anal. Appl. *21* (1999), 45–66.
[17] *S. M. Robinson*: Strongly regular generalized equations. Math. Oper. Res. *5* (1980), 43–62.
[18] *S. M. Robinson*: Analysis of sample-path optimization. Math. Oper. Res. *21* (1996), 513–528.
[19] *R. T. Rockafellar*: Convex Analysis. Princeton University Press, Princeton, 1970.
[20] *R. T. Rockafellar, R. J. B. Wets*: Variational Analysis. Springer, Berlin, 1998.
[21] *H. Scheel, S. Scholtes*: Mathematical programs with complementarity constraints: stationarity, optimality, and sensitivity. Math. Oper. Res. *25* (2000), 1–22.
[22] *A. Shapiro, H. Xu*: Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation. Optimization *57* (2008), 395–418.
[23] *A. Shapiro, T. Homem-de-Mello*: On the rate of convergence of optimal solutions of Monte Carlo approximations of stochastic programs. SIAM J. Optim. *11* (2000), 70–86.
[24] *A. Shapiro, D. Dentcheva, A. Ruszczyński*: Lectures on Stochastic Programming. Modeling and Theory. SIAM, Philadelphia, 2009.
[25] *A. van der Vaart, J. A. Wellner*: Weak Convergence and Empirical Processes. Springer, New York, 1996.
[26] *H. Xu*: An implicit programming approach for a class of stochastic mathematical programs with equilibrium constraints. SIAM J. Optim. *16* (2006), 670–696.
[27] *H. Xu, F. Meng*: Convergence analysis of sample average approximation methods for a class of stochastic mathematical programs with equality constraints. Math. Oper. Res. *32* (2007), 648–668.
[28] *J. J. Ye, D. L. Zhu, Q. J. Zhu*: Exact penalization and necessary optimality conditions for generalized bilevel programming problems. SIAM J. Optim. *2* (1997), 481–507.
[29] *J. J. Ye*: Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints. J. Math. Anal. Appl. *307* (2005), 350–369.
[30] *H. Yin, J. Zhang*: Global convergence of a smooth approximation method for mathematical programs with complementarity constraints. Math. Methods Oper. Res. *64* (2006), 255–269.

*Authors' addresses*: *J. Zhang*, *Li-wei Zhang*, Institute of ORCT, School of Mathematical Sciences, Dalian University of Technology, Dalian, 116024, P. R. China, e-mail: `zhangjie04212001@163.com`, `lwzhang@dlut.edu.cn`; *Yue Wu*, School of Management, University of Southampton, Highfield Southampton SO17 1BJ, United Kingdom, e-mail: `Y.Wu@soton.ac.uk`.