2016

Smoothing and SAA method for stochastic programming problems with non-smooth objective and constraints

Gui-Hua Lin
Shanghai University

Mei-Ju Luo
Liaoning University, meijuluolnu@126.com

Jin Zhang
Hong Kong Baptist University, zhangjin@hkbu.edu.hk

Smoothing and SAA Method for Stochastic Programming Problems with Non-smooth Objective and Constraints*

Gui-Hua Lin[†], Mei-Ju Luo[‡] and Jin Zhang[§]

March 2013, Revised December 2014, August 2015 and January 2016

Abstract. We consider a stochastic non-smooth programming problem with equality, inequality and abstract constraints, which is a generalization of the problem studied by Xu and Zhang (*Mathematical Programming, Vol.119, 371–401, 2009*) where only an abstract constraint is considered. We employ a smoothing technique to deal with the non-smoothness and use the sample average approximation techniques to cope with the mathematical expectations. Then, we investigate the convergence properties of the approximation problems. We further apply the approach to solve the stochastic mathematical programs with equilibrium constraints. In addition, we give an illustrative example in economics to show the applicability of proposed approach.

Key Words. Non-smoothness, smoothing, sample average approximation, stochastic mathematical program with equilibrium constraints.

2010 Mathematics Subject Classification. 90C15, 90C30, 90C33, 65K05.

1 Introduction

In the recent work [29], Xu and Zhang consider the following stochastic programming problem:

$$\begin{aligned} \min \quad & \mathbb{E}[f(x, \xi(\omega))] \\ \text{s.t.} \quad & x \in C, \end{aligned} \tag{1.1}$$

*This work was supported in part by the NSFC (No.11431004, No.11501275), the Humanity and Social Science Foundation of Ministry of Education of China (No. 15YJA630034), the Innovation Program of Shanghai Municipal Education Commission (No. 14ZS086), and the Scientific Research Fund of Liaoning Provincial Education Department (No.L2015199). The third author's work was supported in part by the Hongkong Baptist University FRG1/15-16/027 and RC-NACAN-ZHANG J.

†School of Management, Shanghai University, Shanghai, China. E-mail: guihualin@shu.edu.cn.

‡School of Mathematics, Liaoning University, Liaoning; School of Management, Shanghai University, Shanghai, China. E-mail: meijuluolnu@126.com.

§Department of Mathematics, Hong Kong Baptist University, Hong Kong. E-mail: zhangjin@hkbu.edu.hk.

where C is a closed subset of R^n, $\xi : \Omega \to \Xi \subset R^w$ is a random vector defined on the underlying probability space $(\Omega, \mathcal{F}, P)$, $f : R^n \times R^w \to R$ is a random function, and $\mathbb{E}$ denotes the mathematical expectation. Many problems including the stochastic programming problems with recourse and stochastic min-max problems can be covered by problem (1.1); see [3, 10, 22, 28] for instance. In [29], the function $f(x, \cdot)$ is assumed to be locally Lipschitz continuous but not necessarily continuously differentiable with respect to x. Then the authors employ the smoothing techniques introduced in [19] to present a smooth sample average approximation (SAA) method for solving (1.1). They also investigate the limiting behavior of the smoothed SAA problems. Furthermore, the convergence results are applied to conditional value-at-risk problems and inventory control problems in supply chain. See [29] for details.

In this paper, we consider the following problem where the constraint system includes both the functional constraints as well as the abstract constraint:

$$\begin{aligned} \min \quad & \mathbb{E}[f(x, \xi(\omega))] \\ \text{s.t.} \quad & x \in \mathcal{X}, \\ & g(x) \leq 0, \ h(x) = 0, \end{aligned} \tag{1.2}$$

where $\mathcal{X} \subseteq R^n$ is a closed subset, the constraint functions $g : R^n \to R^p$ and $h : R^n \to R^q$ are also locally Lipschitz continuous but not necessarily continuously differentiable everywhere. Throughout the paper, we suppose that $\mathbb{E}[f(x, \xi(\omega))]$ is well-defined for every $x \in \mathcal{X}$ but cannot be calculated in a closed form. In addition, we write $\xi(\omega)$ as ξ for simplicity.

It is obvious that, when the constraint functions g and h vanish, problem (1.2) reduces to (1.1). However, since some smoothing techniques are only applied to the objective function of (1.1). An implicit assumption in [29] is that the constraint set $\mathcal{C}$ does not involve non-smooth constraints. From this point of view, this paper can be regarded as an extension of [29].

One main motivation to consider (1.2) is the well-known mathematical program with equilibrium constraints (MPEC), which is an optimization problem whose constraints include complementarity or variational inequality systems. MPEC plays a very important role in many fields such as engineering design, economic equilibrium, transportation science and game theory. See [5, 13, 16] for more details about the MPEC theory, algorithms, and applications. MPEC is known to be a difficult optimization problem due to the fact that some usual constraint qualifications such as the linear independence constraint qualification (LICQ) and the Mangasarian-Fromovitz constraint qualification (MFCQ) are violated at any feasible point [33, Proposition 1.1]. As a result, the classical Karush-Kuhn-Tucker (KKT) condition is not always a necessary optimality condition for MPEC. There have been proposed several approaches to deal with MPEC and various stationarity concepts such as the strong stationarity, Mordukhovich stationarity, Clarke stationarity (S-/M-/C-stationarity for short) arise; see [24, 30–32] for detailed discussions. One popular approach in the study of MPEC is to make use of the so-called complementarity or merit functions to reformulate MPEC as an optimization problem with non-smooth constraints. Therefore, problem (1.2) may include the stochastic MPEC (SMPEC) as a special case. See Section 5 given below.

Our purpose is trying to design efficient methods for solving (1.2). There are two main d-ifficulties in dealing with problem (1.2): One is the non-smoothness of the functions $\{f, g, h\}$ and the other is the mathematical expectation operator in the objective function. Our s-trategy is similar to [29]: We employ the smoothing techniques given in [19] to deal with the non-smoothness and use the SAA methods to deal with the mathematical expectation. That is, given a smoothing parameter ϵ, we first construct some smoothed functions $\{\hat{f}(\cdot,\cdot,\epsilon), \hat{g}(\cdot,\epsilon), \hat{h}(\cdot,\epsilon)\}$ of $\{f, g, h\}$ to generate the following approximation of (1.2):

$$\begin{aligned} \min \quad & \mathbb{E}[\hat{f}(x,\xi,\epsilon)] \\ \text{s.t.} \quad & x \in \mathcal{X}, \\ & \hat{g}(x,\epsilon) \leq 0, \;\; \hat{h}(x,\epsilon) = 0. \end{aligned} \tag{1.3}$$

Then, we employ a random number generator to get some independent identically distributed (idd) samples $\{\xi^1, ..., \xi^N\}$ and solve the smoothed SAA problem

$$\begin{aligned} \min \quad & \frac{1}{N}\sum_{i=1}^{N} \hat{f}(x,\xi^i,\epsilon) \\ \text{s.t.} \quad & x \in \mathcal{X}, \\ & \hat{g}(x,\epsilon) \leq 0, \;\; \hat{h}(x,\epsilon) = 0. \end{aligned} \tag{1.4}$$

See [15, 25, 26] for more details about SAA methods. With the increase of the sample size N and the decrease of the smoothing parameter ϵ, we may expect to get a satisfactory approximation solution of the original problem (1.2).

Compared with [29], the main difficulty in dealing with (1.2) is of course the additional non-smooth constraints. Therefore, in order to establish convergence theory, we need to find some appropriate constraint qualifications. In particular, to ease our analysis, we regard the smoothing problem (1.3) as a perturbed optimization problem and, for convenience, we sometimes reformulate problems (1.2) and (1.3) as

$$\begin{aligned} \min \quad & \mathbb{E}[f(x,\xi(\omega))] \\ \text{s.t.} \quad & x \in \mathcal{X}, \\ & 0 \in \varphi(x) + \mathfrak{K} \end{aligned} \tag{1.5}$$

and

$$\begin{aligned} \min \quad & \mathbb{E}[\hat{f}(x,\xi,\epsilon)] \\ \text{s.t.} \quad & x \in \mathcal{X}, \\ & 0 \in \hat{\varphi}(x,\epsilon) + \mathfrak{K} \end{aligned} \tag{1.6}$$

respectively, where $\mathfrak{K} := R_+^p \times \{0\}^q$ is a nonempty closed convex set and

$$\varphi(x) := \begin{pmatrix} g(x) \\ h(x) \end{pmatrix}, \quad \hat{\varphi}(x,\epsilon) := \begin{pmatrix} \hat{g}(x,\epsilon) \\ \hat{h}(x,\epsilon) \end{pmatrix}.$$

We will show that the perturbed problem is stably under some regularity conditions. We will also discuss the limiting behavior of both the smoothed problem (1.3) and the smoothed SAA problem (1.4) as the parameters vary. Furthermore, as an extension, we will discuss a class of SMPEC as well. Finally, we show some preliminary numerical results with a stochastic version of Stackelberg-Nash-Cournot game.

2 Preliminaries

In this section, we introduce some notations and definitions that will be used later on.

Given a compact set $\mathcal{M}$ of vectors, we let $\|\mathcal{M}\| := \max_{M\in\mathcal{M}} \|M\|$, where $\|\cdot\|$ denotes the Euclidean norm of a vector. Given two sets $A, B \subseteq R^n$, we denote by

$$dist(x,A) := \inf_{x'\in A} \|x - x'\|$$

and

$$\mathbb{D}(A,B) := \sup_{x\in A} dist(x,B)$$

the distance form a point $x \in R^n$ to A and the deviation from A to B respectively, and by $intA$ and $convA$ the interior and the convex hull of A respectively. In addition, for a linear operator $\mathcal{A}$, we denote by $\mathcal{A}^*$ its conjugate.

Recall that $f(x, \xi)$ is locally Lipschitz continuous in x. Then, for any fixed $\xi \in \Xi$, the Clarke generalized gradient of $f(x, \xi)$ with respect to x at x_0 is defined in [2] as

$$\partial_x f(x_0, \xi) := \left\{ \zeta \in R^n \mid \zeta^T d \leq f^0(x_0, \xi; d) \text{ for all } d \in R^n \right\},$$

where

$$f^0(x_0, \xi; d) := \limsup_{y \to 0,\, \lambda \downarrow 0} \frac{f(x_0 + \lambda d + y, \xi) - f(x_0 + y, \xi)}{\lambda}$$

is the generalized directional derivative of $f(x, \xi)$ at x_0 in the direction d. In finite dimensional spaces, $\partial_x f(x_0, \xi)$ can be obtained by taking the convex hull of the set of limits of $\nabla_x f(x', \xi)$ as $x' \to x_0$. It is known from [2, Propositions 2.1.2 and 2.1.5] that the Clarke generalized gradient $\partial_x f(x, \xi)$ is a convex compact set and it is upper semi-continuous in x. The measurability and integrability of $\partial_x f(x, \xi)$ with respect to the random variable are established in [29].

Lemma 2.1 *Let $f(x, \xi)$ be locally Lipschitz continuous in both x and ξ. Then the Clarke generalized gradient $\partial_x f(x, \cdot)$ is measurable for every x. Furthermore, given a nonempty and compact subset $\mathcal{C}$ of R^n, if there exists a measurable function $\kappa_0(\xi)$ such that $\mathbb{E}[\kappa_0(\xi)] < \infty$ and*

$$\|\partial_x f(x, \xi)\| \leq \kappa_0(\xi) \tag{2.1}$$

for any $x \in \mathcal{C}$ and $\xi \in \Xi$, then $\mathbb{E}[\partial_x f(x, \xi)]$ is well defined over $\mathcal{C}$.

The following definition was first introduced in [19].

Definition 2.1 Let $F : R^n \to R$ be a locally Lipschitz continuous function and $\epsilon \in R$ be a parameter. A function $\hat{F}(x, \epsilon) : R^n \times R \to R$ is called a smoothing of F if it satisfies

(i) for every $x \in R^n$, $\hat{F}(x, 0) = F(x)$;

(ii) for every $x \in R^n$, $\hat{F}$ is locally Lipschitz at $(x, 0)$;

(iii) $\hat{F}$ is continuous differentiable on $R^n \times (R \setminus \{0\})$.

The above type of smoothing covers a number of interesting elementary smoothing functions in the literature. For instance, for the well-known non-smooth function

$$\phi(x) := \max(0, z),$$

it is shown in [29] that both

$$\phi_1(z,\epsilon) := \begin{cases} z, & z > \epsilon; \\ \frac{1}{4\epsilon}(z^2 + 2z\epsilon + \epsilon^2), & -\epsilon \le z \le \epsilon; \\ 0, & z < -\epsilon \end{cases}$$

introduced by Alexander et al. [1] and

$$\phi_2(z,\epsilon) := \epsilon \ln(1 + e^{z/\epsilon})$$

introduced by Peng [18] satisfy the conditions in Definition 2.1. In particular, they satisfy the gradient consistency at point $(z, 0)$ and hence satisfy the upper semi-continuity property for the partial generalized gradient map.

Note that the Lipschitz continuity in part (ii) of Definition 2.1 implies that the Clarke generalized gradient $\partial_{(x,\epsilon)}\hat{F}(x, 0)$ is well defined and hence we may compare the generalized gradient of the smoothed function at point $(x, 0)$ with that of the original function. Let

$$\pi_x \partial_{(x,\epsilon)}\hat{F}(x,0) := \Big\{\beta \in R^n \;\Big|\; (\beta,\gamma) \in \partial_{(x,\epsilon)}\hat{F}(x,0) \text{ for some } \gamma \in R\Big\}.$$

The gradient consistency defined by

$$\pi_x \partial_{(x,\epsilon)}\hat{F}(x,0) \subset \partial_x F(x),$$

is a key property used later on. There obviously holds

$$\partial_x F(x) = \partial_x \hat{F}(x,0) \subset \pi_x \partial_{(x,\epsilon)}\hat{F}(x,0),$$

which means that, if $\hat{F}$ satisfies the gradient consistency at $(x, 0)$, we may enlarge the partial generalized gradient set $\partial_x \hat{F}(x, 0)$ so as to equal to the projection $\pi_x \partial_{(x,\epsilon)}\hat{F}(x, 0)$. Since the map $(x, \epsilon) \to \pi_x \partial_{(x,\epsilon)}\hat{F}(x, \epsilon)$ is upper semi-continuous at $(x, 0)$ [2, Proposition 2.1.5], we have the upper semi-continuity of $\partial_x \hat{F}(x, \epsilon)$. As a result, if we regard problem (1.3) as a perturbed problem of (1.2), the smoothing approximation happens to be admissible in the sense of [34, Definition 2.3].

Since both problems (1.2) and (1.3) are generally non-convex, we will focus on the limit relationship between stationary points of the problems. We next introduce some stationarity concepts for the problems. To this end, we let

$$\mathcal{N}_{\mathcal{X}}(x) := \{\zeta \in R^n \mid \zeta^T \eta \le 0 \text{ for all } \eta \in \mathcal{T}_{\mathcal{X}}(x)\}$$

denote the normal cone of $\mathcal{X}$ at x and

$$\mathcal{T}_{\mathcal{X}}(x) := \{\zeta \in R^n \mid \zeta^T \theta \le 0 \text{ for all } \theta \in \partial dist(x, \mathcal{X})\}$$

denote the tangent cone of $\mathcal{X}$ at x. It is well known that the above cones are closed and convex. When $\mathcal{X}$ is convex and $x \in \mathcal{X}$, the normal cone reduces to

$$\mathcal{N}_{\mathcal{X}}(x) = \{\zeta \in R^n \mid \zeta^T(x' - x) \leq 0 \quad \text{for all } x' \in \mathcal{X}\}.$$

Definition 2.2 Let $\bar{x}$ be a feasible point of (1.2).

(1) $\bar{x}$ is called a Karush-Kuhn-Tucker (KKT) point of (1.2) if there exist multipliers $\lambda \in R^p$ and $\mu \in R^q$ such that

$$0 \in \partial\, \mathbb{E}[f(\bar{x}, \xi)] + \partial g(\bar{x})^T \lambda + \partial h(\bar{x})^T \mu + \mathcal{N}_{\mathcal{X}}(\bar{x})$$

and

$$\lambda \geq 0, \quad \lambda^T g(\bar{x}) = 0. \tag{2.2}$$

(2) $\bar{x}$ is called a weak Karush-Kuhn-Tucker (weak KKT) point of (1.2) if there exist multipliers $\lambda \in R^p$ and $\mu \in R^q$ such that

$$0 \in \mathbb{E}[\partial_x f(\bar{x}, \xi)] + \partial g(\bar{x})^T \lambda + \partial h(\bar{x})^T \mu + \mathcal{N}_{\mathcal{X}}(\bar{x}) \tag{2.3}$$

and (2.2) hold.

Suppose that the set of stationary points of the original problem is nonempty and there exists a measurable function $\kappa_0(\xi)$ such that $\mathbb{E}[\kappa_0(\xi)] < \infty$ and (2.1) hold. It follows that $\mathbb{E}[\partial_x f(x, \xi)]$ is well defined. Note that $\partial \mathbb{E}[f(x, \xi)] \subset \mathbb{E}[\partial_x f(x, \xi)]$ holds under some mild conditions on $f(x, \xi)$ (see [2, Hypotheses 2.7.1]). In particular, we have from [2, Theorem 2.7.2] that, when f is Clarke regular on $\mathcal{X}$, $\partial\, \mathbb{E}[f(x, \xi)] = \mathbb{E}[\partial_x f(x, \xi)]$ and hence the set of weak KKT points coincides with the set of KKT points.

Definition 2.3 Let $\bar{x}(\epsilon)$ be a feasible point of (1.3).

(1) $\bar{x}(\epsilon)$ is called a KKT point of (1.3) if there exist multipliers $\lambda(\epsilon) \in R^p$ and $\mu(\epsilon) \in R^q$ such that

$$0 \in \nabla_x \mathbb{E}[\hat{f}(\bar{x}(\epsilon), \xi, \epsilon)] + \nabla_x \hat{g}(\bar{x}(\epsilon), \epsilon)^T \lambda(\epsilon) + \nabla_x \hat{h}(\bar{x}(\epsilon), \epsilon)^T \mu(\epsilon) + \mathcal{N}_{\mathcal{X}}(\bar{x}(\epsilon))$$

and

$$\lambda(\epsilon) \geq 0, \quad \lambda(\epsilon)^T \hat{g}(\bar{x}(\epsilon), \epsilon) = 0. \tag{2.4}$$

(2) $\bar{x}(\epsilon)$ is called a weak KKT point of (1.3) if there exist multipliers $\lambda(\epsilon) \in R^p$ and $\mu(\epsilon) \in R^q$ such that

$$0 \in \mathbb{E}[\nabla_x \hat{f}(\bar{x}(\epsilon), \xi, \epsilon)] + \nabla_x \hat{g}(\bar{x}(\epsilon), \epsilon)^T \lambda(\epsilon) + \nabla_x \hat{h}(\bar{x}(\epsilon), \epsilon)^T \mu(\epsilon) + \mathcal{N}_{\mathcal{X}}(\bar{x}(\epsilon)) \tag{2.5}$$

and (2.4) hold.

By [23, Proposition 2, Chapter 2], if $\nabla_x \hat{f}(x, \xi, \epsilon)$ is of integrable boundedness with probability one, then $\mathbb{E}[\nabla_x \hat{f}(x, \xi, \epsilon)] = \nabla_x \mathbb{E}[\hat{f}(x, \xi, \epsilon)]$ and hence the set of weak KKT points coincides with the set of KKT points for problem (1.3). In what follows, we suppose that the integrable boundedness condition holds for $\nabla_x \hat{f}(x, \xi, \epsilon)$.

We now turn our attention to constraint qualifications for problem (1.2). As known to us, the MFCQ plays an important role in nonlinear programming theory. However, since problem (1.2) contains some non-smooth functions and an abstract constraint, it is necessary to describe new Mangasarian-Fromovitz type constraint qualifications for (1.2). Recall that Pappalardo [17] defines a generalized MFCQ (GMFCQ) for the non-smooth case without the abstract constraint.

Definition 2.4 We say that the GMFCQ holds at $\bar{x}$ if for any $\mathcal{A} \in \partial h(\bar{x})$, the rows of $\mathcal{A}$ are linearly independent and there exists a vector $\zeta \in R^n$ such that

(i) $\nu^T \zeta < 0$ for every $\nu \in \partial g_i(\bar{x})$, where $i \in I(\bar{x}) := \{\imath \mid g_\imath(\bar{x}) = 0\}$, or equivalently

$$g_i^0(\bar{x}; \zeta) < 0, \quad \forall i \in I(\bar{x}); \tag{2.6}$$

(ii) $\mathcal{A}\zeta = 0$, for all $\mathcal{A} \in \partial h(\bar{x})$.

The GMFCQ can be extended to the case where the abstract constraint set $\mathcal{X}$ does not vanish.

Definition 2.5 We say that the extended MFCQ (EMFCQ) holds at $\bar{x}$ if, for any $\mathcal{A} \in \partial h(\bar{x})$, the rows of $\mathcal{A}$ are linearly independent and there exists a vector $\zeta \in int\mathcal{T}_{\mathcal{X}}(\bar{x})$ such that

(i) $\nu^T \zeta < 0$ for every $\nu \in \partial g_i(\bar{x})$, where $i \in I(\bar{x})$;

(ii) $\mathcal{A}\zeta = 0$, for all $\mathcal{A} \in \partial h(\bar{x})$.

The following constraint qualification will also be used later.

Definition 2.6 We say that the Jourani constraint qualification (JCQ) holds at $\bar{x}$ if, for any $(\lambda, \mu) \in R^p_+ \times R^q \backslash \{(0,0)\}$ satisfying $\lambda^T g(\bar{x}) = 0$, there holds

$$0 \notin \partial g(\bar{x})^T \lambda + \partial h(\bar{x})^T \mu + \mathcal{N}_{\mathcal{X}}(\bar{x}).$$

It is shown in [9] that the EMFCQ implies the JCQ and, when $\mathcal{X} = R^n$, the JCQ is equivalent to the GMFCQ.

3 Convergence Analysis for Smoothing Approach

In this section, we study the limiting behavior of the smoothed problem (1.3) as the smoothing parameter tends to zero.

3.1 Stability analysis

First of all, note that (1.6) can be regarded as a perturbed problem of (1.5). Our purpose is to study the continuity of the optimal value function $\mathfrak{V} : R \to R$ defined by

$$\mathfrak{V}(\epsilon) := \inf\{\mathbb{E}[\hat{f}(x, \xi, \epsilon)] : x \in \mathfrak{F}(\epsilon)\},$$

where

$$\begin{aligned} \mathfrak{F}(\epsilon) &:= \{x \in \mathcal{X} \mid 0 \in \hat{\varphi}(x, \epsilon) + \mathfrak{K}\} \\ &= \{x \in \mathcal{X} \mid \hat{g}_i(x, \epsilon) \le 0, i = 1, \cdots, p; \hat{h}_j(x, \epsilon) = 0, j = 1, \cdots, q\}. \end{aligned}$$

We denote the corresponding optimal solution set mapping by

$$\mathfrak{M}(\epsilon) := \{x \in \mathfrak{F}(\epsilon) \mid \mathbb{E}[\hat{f}(x, \xi, \epsilon)] = \mathfrak{V}(\epsilon)\}.$$

Since $\mathcal{X}$ is closed, we have from [21, Example 5.8] that $\mathfrak{F}$ is outer semi-continuous. In particular, we note that, by [21, Commentary of Chapter 5], $\mathfrak{F}$ is upper semi-continuous if we consider in restriction to the compact case.

Definition 3.1 Let $\Gamma : R^n \to 2^{R^n}$ be a set valued mapping. Γ is said to be closed at $\bar{\epsilon}$ if the conditions $\epsilon_N \to \bar{\epsilon}$, $x(\epsilon_N) \in \Gamma(\epsilon_N)$ and $x(\epsilon_N) \to \bar{x}$ imply $\bar{x} \in \Gamma(\bar{\epsilon})$. Γ is said to be uniformly compact near $\bar{\epsilon}$ if there is a neighbourhood $N(\bar{\epsilon})$ of $\bar{\epsilon}$ such that the closure of $\bigcup_{\epsilon \in N(\bar{\epsilon})} \Gamma(\epsilon)$ is compact.

The following result is established by Hogan in [8].

Lemma 3.1 *Let $\Gamma : R^n \to 2^{R^n}$ be uniformly compact near $\bar{\epsilon}$. Then Γ is upper semi-continuous at $\bar{\epsilon}$ if and only if Γ is closed.*

Since $\hat{f}(x, \xi, \epsilon)$ is locally Lipschitz continuous with respect to (x, ϵ), it is globally Lipschitz on $\mathcal{C} \times [0, \epsilon_0]$ for any compact subset $\mathcal{C}$ of $\mathcal{X}$ and positive number $\epsilon_0 > 0$. Therefore. there exists $\tilde{\kappa}_{\mathcal{C}}(\xi) > 0$ such that

$$\left\| \hat{f}(x', \xi, \epsilon') - \hat{f}(x'', \xi, \epsilon'') \right\| \leq \tilde{\kappa}_{\mathcal{C}}(\xi) \Big(\|x' - x''\| + | \, \epsilon' - \epsilon'' \, | \Big) \tag{3.1}$$

holds for any $x', x'' \in \mathcal{C}$, $\epsilon', \epsilon'' \in [0, \epsilon_0]$, and almost every $\xi \in \Xi$.

Theorem 3.1 *Consider problem (1.6). Suppose that*

(i) *there exists a compact subset $\mathcal{C} \subset R^n$ such that $\mathfrak{M}(\epsilon) \cap \mathcal{C} \neq \emptyset$ for every ϵ in a neighborhood of 0;*

(ii) $\mathbb{E}[\tilde{\kappa}_{\mathcal{C}}(\xi)] < \infty$*;*

(iii) *there exists $x_0 \in \mathfrak{V}(0) \cap \mathcal{C}$ such that the map $(x, \epsilon) \to \partial_x \hat{\varphi}(x, \epsilon)$ is upper semi-continuous at $(x_0, 0)$ and*

$$0 \in int\{\varphi(x_0) + \mathcal{A}(\mathcal{T}_{\mathcal{X}}(x_0)) + \mathfrak{K}\}, \qquad \forall \, \mathcal{A} \in \partial\varphi(x_0). \tag{3.2}$$

Then $\mathfrak{V}(\epsilon)$ is continuous at $\epsilon = 0$.

Proof. Taking an intersection $\mathcal{C} \cap \mathcal{X}$ if necessary, we assume for simplicity that the compact set $\mathcal{C}$ locates inside $\mathcal{X}$. Since assumption (i) holds, in order to obtain the continuity of $\mathfrak{V}$ at 0, it is sufficient to consider problem (1.6) in restriction to $\mathcal{C}$ instead of $\mathcal{X}$.

We first show that $\mathfrak{V}$ is lower semi-continuous at $\epsilon = 0$. Suppose to the contrary that there exist $\varepsilon_0 > 0$ and a sequence $\epsilon_k \to 0$ such that

$$\mathfrak{V}(\epsilon_k) \leq \mathfrak{V}(0) - \varepsilon_0, \quad \forall k. \tag{3.3}$$

By condition (i), we can choose

$$x_k \in \mathfrak{M}(\epsilon_k) \cap \mathcal{C}$$

for each k large sufficiently. Since $\mathcal{C}$ is compact, it follows from Lemma 3.1 and the discussion above it that the map $\mathfrak{F}(\cdot)$ is closed. We may suppose $\{x_k\}$ converging to a point $\bar{x} \in \mathfrak{F}(0) \cap \mathcal{C}$. It follows from (3.3) that

$$\mathbb{E}[\hat{f}(x_k, \xi, \epsilon_k)] = \mathfrak{V}(\epsilon_k) \leq \mathfrak{V}(0) - \varepsilon_0.$$

Letting $k \to \infty$, we have from (3.1) and condition (ii) that

$$\mathbb{E}[\hat{f}(\bar{x}, \xi, 0)] \leq \mathfrak{V}(0) - \varepsilon_0,$$

which is a contradiction.

We next show that $\mathfrak{V}$ is upper semi-continuous at 0. By the locally Lipschitz continuity of $\hat{f}$ and assumption (ii), for arbitrary $\varepsilon > 0$, there exist neighborhoods $\mathcal{N}_\varepsilon(x_0)$ of x_0 and $\mathcal{N}_\varepsilon(0)$ of 0 such that

$$\Big|\mathbb{E}[\hat{f}(x, \xi, \epsilon)] - \mathbb{E}[\hat{f}(x_0, \xi, 0)]\Big| \leq \mathbb{E}[\tilde{\kappa}_{\mathcal{C}}(\xi)]\Big(\|x - x_0\| + |\ \epsilon - 0\ |\ \Big) \leq \varepsilon \tag{3.4}$$

holds for every $(x, \epsilon) \in \mathcal{N}_\varepsilon(x_0) \times \mathcal{N}_\varepsilon(0)$. By condition (iii), in a similar way to Theorem 3.1 of [34], we can find a neighborhood $\mathcal{N}(0)$ of 0 such that $\mathcal{N}(0) \subset \mathcal{N}_\varepsilon(0)$ and, for every $\epsilon \in \mathcal{N}(0)$, there exists a vector $x(\epsilon) \in \mathfrak{F}(\epsilon) \cap \mathcal{N}_\varepsilon(x_0)$. It follows from (3.4) that

$$\mathfrak{V}(\epsilon) \leq \mathbb{E}[\hat{f}(x(\epsilon), \xi, \epsilon)] \leq \mathbb{E}[\hat{f}(x_0, \xi, 0)] + \varepsilon, \qquad \forall\, \epsilon \in \mathcal{N}(0).$$

In consequence, we have

$$\limsup_{\epsilon \to 0} \mathfrak{V}(\epsilon) \leq \mathbb{E}[\hat{f}(x_0, \xi, 0)] + \varepsilon = \mathfrak{V}(0) + \varepsilon.$$

This completes the proof. ■

We further have the following result from Theorem 3.1 immediately.

Corollary 3.1 *Let x^* be an optimal solution of (1.5). Suppose that*

(i) *there exists a compact subset $\mathcal{C} \subset R^n$ such that $\mathfrak{M}(\epsilon_N) \cap \mathcal{C} \neq \emptyset$ for every N large enough;*

(ii) $\mathbb{E}[\tilde{\kappa}_{\mathcal{C}}(\xi)] < \infty$*;*

(iii) $x^* \in \mathcal{C}$, $\hat{\varphi}(x, \epsilon)$ *satisfies the gradient consistency at* $(x^*, 0)$*, and*

$$0 \in int\{\varphi(x^*) + \mathcal{A}(\mathcal{T}_{\mathcal{X}}(x^*)) + \mathfrak{K}\}, \qquad \forall\, \mathcal{A} \in \partial\varphi(x^*). \tag{3.5}$$

Then, with the increase of the sample size N, $x(\epsilon_N) \in \mathfrak{M}(\epsilon_N) \cap \mathcal{C}$ yields a 'good' approximate optimal value $\mathfrak{V}(\epsilon_N)$ of the true problem (1.5).

As mentioned in Section 2, the gradient consistency assumption is used to ensure the upper semi-continuity of $\partial_x \hat{\varphi}$ so as to apply Theorem 3.1 of [34].

The regularity assumption (3.5), which was first proposed in [34], can be regarded as an extension of the Robinson constraint qualification into non-convex and non-smooth circumstances. Yen [34] also proved that, in the case that we are considering, it happens to be equivalent to the JCQ, which is introduced in Section 2 and will be used for convergence analysis of stationary points in the forthcoming parts. Moreover, [34, Lemma 2.1] restates condition (3.5) in the following dual equivalent form: There exists a constant $\varsigma > 0$ such that

$$\|\mathcal{A}^* v^* + u^*\| \geq \varsigma, \tag{3.6}$$

holds for any $\mathcal{A} \in \partial\varphi(x^*)$, $v^* \in B \cap [(\varphi(x^*) + \mathfrak{K})^-]$, and $u^* \in \mathcal{N}_{\mathcal{X}}(x^*)$, where B denotes the unit ball in R^n. (3.6) is more suitable for some proofs such as in Theorem 3.1 of [34], which plays an extraordinary role in our stability analysis above. As a matter of fact, it is easy to see the implication that the JCQ holds by (3.6).

Assumption (i) in Theorem 3.1 or Corollary 3.1 implies a property called tameness by Rockafellar in [20]. In particular, [34, Theorem 4.3] provides a sufficient condition of Assumption (i) as follows.

Proposition 3.1 *Let x^* be an optimal solution of (1.5). Suppose that*

(i) *there exists a neighbourhood $N(x^*) \subset \mathfrak{F}(0)$ of x^* such that $\mathbb{E}[\tilde{\kappa}_{N(x^*)}(\xi)] < \infty$;*

(ii) *$\hat{\varphi}(x, \epsilon)$ satisfies the gradient consistency at $(x^*, 0)$ and condition (3.5) holds.*

If

$$\liminf_{\|x\| \to \infty,\, \epsilon \to 0} \mathbb{E}[\hat{f}(x, \xi, \epsilon)] > \mathbb{E}[f(x^*, \xi)],$$

then there exists a compact subset $\mathcal{C} \subset R^n$ such that $\mathfrak{M}(\epsilon_N) \cap \mathcal{C} \neq \emptyset$ for every N large enough.

On the other hand, it is worth noting that, although the optimal value could be approached from Corollary 3.1, $x(\epsilon_N)$ may not be regarded as a 'good' approximate solution because of its potential infeasibility.

3.2 Convergence of weak KKT points

Denote by S and $S(\epsilon)$ the sets of weak KKT stationary points of problems (1.2) and (1.3) respectively.

Lemma 3.2 *Suppose that there exists an integrable function $\kappa(\xi)$ such that the Lipschitz module of $\hat{f}(x, \xi, \epsilon)$ with respect to x is bounded by $\kappa(\xi)$. Suppose that* $\lim_{\epsilon \to 0} x(\epsilon) \to x$ *and $\hat{f}$ satisfies the gradient consistency at $(x, \xi, 0)$, that is,*

$$\pi_x \partial_{(x,\epsilon)} \hat{f}(x, \xi, 0) \subset \partial_x f(x, \xi),$$

for almost every ξ. Then we have

$$\limsup_{\epsilon \to 0} \Big\{ \mathbb{E}[\nabla_x \hat{f}(x(\epsilon), \xi, \epsilon)] \Big\} = \mathbb{E}\Big[\limsup_{\epsilon \to 0} \Big\{ \nabla_x \hat{f}(x(\epsilon), \xi, \epsilon) \Big\} \Big] \subset \mathbb{E}[\partial_x f(x, \xi)].$$

See the proof of Theorem 3.1 in [29] for a proof of the lemma.

Theorem 3.2 *Let the assumptions of Lemma 3.2 hold. Suppose that $S(\epsilon)$ is nonempty. Let $\bar{x}(\epsilon) \in S(\epsilon)$ and $\bar{x}$ be a limit point of $\bar{x}(\epsilon)$. If the JCQ holds at $\bar{x}$ and both $\hat{g}$ and $\hat{h}$ satisfy the gradient consistency at $(\bar{x}, 0)$, then $\bar{x} \in S$.*

Proof. First of all, it follows from the closeness of $\mathcal{X}$ and the definition of the smooth approximation that $\bar{x}$ is feasible to problem (1.2). Since the Lipschitz modulus of $\hat{f}(x, \xi, \epsilon)$ with respect to x is bounded by the integrable function $\kappa(\xi)$, $S(\epsilon)$ coincides with the set of KKT points of the smoothed problem (1.3). Without loss of generality, we assume that $\lim_{\epsilon \to 0} \bar{x}(\epsilon) = \bar{x}$. Let $(\lambda(\epsilon), \mu(\epsilon)) \in R^{p+q}$ be the corresponding multiplier vectors in (2.5).

(i) We first show that the set of sequence $\Big\{ (\lambda(\epsilon), \mu(\epsilon)) \Big\}$ is bounded for all ϵ in a neighborhood of $\epsilon = 0$. To this end, we set

$$\tau(\epsilon) := \| \ (\lambda(\epsilon), \mu(\epsilon)). \ \| \tag{3.7}$$

Suppose by contradiction that $\Big\{ (\lambda(\epsilon), \mu(\epsilon)) \Big\}$ is not bounded, which means that there exists a sequence $\Big\{ (\lambda(\epsilon_k), \mu(\epsilon_k) \Big\}$ such that $\lim_{k \to \infty} \tau(\epsilon_k) = +\infty$. We may further assume that the limits $\bar{\lambda} := \lim_{k \to \infty} \frac{\lambda(\epsilon_k)}{\tau(\epsilon_k)}$ and $\bar{\mu} := \lim_{k \to \infty} \frac{\mu(\epsilon_k)}{\tau(\epsilon_k)}$ exist. Obviously, we have

$$\bar{\lambda} \geq 0, \quad \bar{\lambda}^T g(\bar{x}) = 0, \quad \| \ (\bar{\lambda}, \ \bar{\mu}) \ \| = 1. \tag{3.8}$$

Replacing the smoothing parameter ϵ in (2.5) by ϵ_k and dividing by $\tau(\epsilon_k)$, we have

$$\begin{aligned} 0 \ \in \ & \frac{1}{\tau(\epsilon_k)} \mathbb{E}[\nabla_x \hat{f}(\bar{x}(\epsilon_k), \xi, \epsilon_k)] + \nabla_x \hat{g}(\bar{x}(\epsilon_k), \epsilon_k)^T \frac{\lambda(\epsilon_k)}{\tau(\epsilon_k)} \\ & + \nabla_x \hat{h}(\bar{x}(\epsilon_k), \epsilon_k)^T \frac{\mu(\epsilon_k)}{\tau(\epsilon_k)} \ + \mathcal{N}_{\mathcal{X}}(\bar{x}(\epsilon_k)). \end{aligned} \tag{3.9}$$

Consider the right hand side of the above generalized equation.

- Since $\lim_{k\to\infty} \mathbb{E}[\nabla_x \hat{f}(\bar{x}(\epsilon_k), \xi, \epsilon_k)] \subset \mathbb{E}[\partial_x f(\bar{x}, \xi)]$ by Lemma 3.2 and $\partial_x f(\tilde{x}, \xi)$ is well defined by Lemma 2.1, the first term tends to 0 as $k \to \infty$.

- By the gradient consistency assumption, both $\partial_x \hat{g}(x, \epsilon)$ and $\partial_x \hat{h}(x, \epsilon)$ are upper semi-continuous, from which we have that, when ϵ_k is sufficiently small,

$$\nabla_x \hat{g}(x(\epsilon_k), \epsilon_k) \subset \partial g(\bar{x}) + \epsilon_k B, \quad \nabla_x \hat{h}(x(\epsilon_k), \epsilon_k) \subset \partial h(\bar{x}) + \epsilon_k B.$$

- Noting that the normal cone is upper semi-continuous, we have $\mathcal{N}_{\mathcal{X}}(\bar{x}(\epsilon_k)) \subset \mathcal{N}_{\mathcal{X}}(\bar{x}) + \epsilon_k B$ for every ϵ_k small sufficiently.

Letting $k \to \infty$ in (3.9), we obtain

$$0 \in \partial g(\bar{x})^T \bar{\lambda} + \partial h(\bar{x})^T \bar{\mu} + \mathcal{N}_{\mathcal{X}}(\bar{x}),$$

which together with (3.8) contradicts the JCQ assumption.

(ii) Since $\Big\{(\lambda(\epsilon), \mu(\epsilon))\Big\}$ is bounded, we may assume that $\lambda := \lim_{\epsilon\to 0} \lambda(\epsilon)$ and $\mu := \lim_{\epsilon\to 0} \mu(\epsilon)$ exist. Taking a limit in (2.5) and (2.4), we get (2.3) and (2.2) immediately. As a result, $\bar{x}$ is a weak KKT point of problem (1.2), namely, $\bar{x} \in S$. ∎

It is easy to see that the gradient consistency assumption in Theorem 3.2 is a little weaker than Theorem 3.1 in [29].

4 Convergence Analysis for Smoothing SAA Approach

In this section, we focus on the smoothing SAA approach to solve problem (1.2).

4.1 Convergence of KKT points

We first study the convergence of KKT points of the smoothed SAA problem (1.4). We will adopt the standard definition of stationarity for problem (1.4), that is, a feasible point $\bar{x}_N(\epsilon)$ of problem (1.4) is a KKT point if and only if there exist multipliers $\lambda^N(\epsilon)$ and $\mu^N(\epsilon)$ such that

$$\begin{aligned}
0 \in\ & \frac{1}{N}\sum_{i=1}^{N} \nabla_x \hat{f}(\bar{x}_N(\epsilon), \xi^i, \epsilon) + \nabla_x \hat{g}(\bar{x}_N(\epsilon), \epsilon)^T \lambda^N(\epsilon) \\
& + \nabla_x \hat{h}(\bar{x}_N(\epsilon), \epsilon)^T \mu^N(\epsilon) + \mathcal{N}_{\mathcal{X}}(\bar{x}_N(\epsilon)), \qquad (4.1)\\
\lambda^N&(\epsilon) \geq 0, \quad \lambda^N(\epsilon)^T \hat{g}(\bar{x}_N(\epsilon), \epsilon) = 0.
\end{aligned}$$

We assume that, for almost every $\omega \in \Omega$, there exists $N(\omega) > 0$ such that, for all $N > N(\omega)$, (4.1) has a solution (see [26] for details).

There are two ways to set ϵ in (1.4), which will lead different convergence results: One is to fix ϵ; the other is to let ϵ vary as N increases, that is, let $\epsilon_N \to 0$ as $N \to \infty$. In the following, we discuss the convergence for both cases. In analogy to [29], we make the following assumption.

Assumption 4.1 There exists a small positive constant $\epsilon_0 > 0$ and a measurable function $\kappa(\xi)$ such that

$$\sup_{x \in \mathcal{C},\, \epsilon \in [0, \epsilon_0]} \|\partial_x \hat{f}(x, \xi, \epsilon)\| \leq \kappa(\xi)$$

holds for almost every $\xi \in \Xi$, where $\mathbb{E}[\kappa(\xi)] < \infty$ and $\mathcal{C}$ is a compact subset of $\mathcal{X}$.

The condition in the above assumption is equivalent to

$$\sup_{x \in \mathcal{C},\, \epsilon \in (0, \epsilon_0]} \|\nabla_x \hat{f}(x, \xi, \epsilon)\| \leq \kappa(\xi) \tag{4.2}$$

and

$$\sup_{x \in \mathcal{C}} \|\partial_x f(x, \xi)\| \leq \kappa(\xi). \tag{4.3}$$

(4.2) is not difficult to ensure. Here we make some remarks on (4.3). Recall that $f(x, \xi)$ is a locally Lipschitz continuous function in both x and ξ. Suppose that there exists a measurable function $\kappa_0(\xi)$ such that $\mathbb{E}[\kappa_0(\xi)] < \infty$ and (2.1) holds, which implies the well definedness of $\mathbb{E}[\partial_x f(x, \xi)]$, and $\hat{f}$ satisfies the gradient consistency at $(x, \xi, 0)$ for all $x \in \mathcal{C}$ and $\xi \in \Xi$. Then we have (4.3). In fact, to see this, since the gradient consistency guarantees upper semi-continuity of $\partial_x \hat{f}$, we can choose a finite υ-net, say $\{x_\upsilon^1, ..., x_\upsilon^M\}$, in $\mathcal{C}$ such that, for every $x \in \mathcal{C}$, there exists a point x_υ^i, $i \in \{1, ..., M\}$, such that $\|x - x_\upsilon^i\| \leq \upsilon$, and hence we have $\partial_x f(x, \xi) \subset \partial_x f(x_\upsilon^i, \xi) + B$.

Consider the case where $\epsilon > 0$ is fixed. It follows that all constraints of problem (1.4) are continuously differentiable. Denote by $J := \{1, 2, \cdots, q\}$.

Definition 4.1 We say that the MFCQ holds at $\bar{x}(\epsilon)$ if $\{\nabla_x \hat{h}_j(\bar{x}(\epsilon), \epsilon)\}_{j \in J}$ are linearly independent and there exists $\zeta \in int\mathcal{T}_{\mathcal{X}}(\bar{x}(\epsilon))$ such that

- $\zeta^T \nabla_x \hat{h}_j(\bar{x}(\epsilon), \epsilon) = 0$ for $j \in J$;
- $\zeta^T \nabla_x \hat{g}_i(\bar{x}(\epsilon), \epsilon) < 0$ for $i \in I(\bar{x}(\epsilon))$.

Lemma 4.1 (Proposition 7 of [26]) *Let $\mathcal{C}$ be a nonempty compact subset of R^n. Suppose that*

(i) *the function $\Psi(\cdot,\xi)$ is continuous on $\mathcal{C}$ for almost every $\xi \in \Xi$;*

(ii) *$\Psi(x,\xi)$ is dominated by an integrable function over $\mathcal{C}$;*

(iii) *the sample is iid.*

Then the expected value function $\psi(x) := \mathbb{E}[\Psi(x,\xi)]$ is finite valued and continuous on $\mathcal{C}$, and $\hat{\psi}_N(x) := \frac{1}{N}\sum_{i=1}^N \Psi(x,\xi^i)$ converges to $\psi(x)$ uniformly on $\mathcal{C}$ with probability one (w.p.1).

Theorem 4.1 *Let $\epsilon \neq 0$ be fixed and $\{\bar{x}_N(\epsilon)\}$ be a sequence of KKT points of problems (1.4). Let $\bar{x}(\epsilon)$ be an accumulation point of the sequence as N tends to infinity. Suppose that the MFCQ holds at $\bar{x}(\epsilon)$. If there exists a compact set $\mathcal{C} \subset R^n$ such that it contains a neighborhood of $\bar{x}(\epsilon)$ and Assumption 4.1 holds on $\mathcal{C}$, then $\bar{x}(\epsilon) \in S(\epsilon)$ with probability one.*

Proof. Without lose of generality, we assume that $\bar{x}_N(\epsilon)$ tends to $\bar{x}(\epsilon)$ as $N \to \infty$. It is easy to see that $\bar{x}(\epsilon)$ is feasible to problem (1.3). Moreover, since uniform convergence is equivalent to continuous convergence on a compact set [21, Theorem 5.43], we have

$$\lim_{N\to\infty} \left\| \frac{1}{N}\sum_{i=1}^{N} \nabla_x \hat{f}(\bar{x}_N(\epsilon),\xi^i,\epsilon) - \mathbb{E}\Big[\nabla_x \hat{f}(\bar{x}(\epsilon),\xi,\epsilon)\Big] \right\| = 0, \quad w.p.1. \tag{4.4}$$

(i) We first show that the set of Lagrange multipliers $\Big\{(\lambda^N(\epsilon),\mu^N(\epsilon))\Big\}$ is bounded. To this end, we set

$$\tau^N(\epsilon) := \sum_{i=1}^{p} \lambda_i^N(\epsilon) + \sum_{j=1}^{q} \mid \mu_j^N(\epsilon) \mid . \tag{4.5}$$

Suppose by contradiction that $\Big\{(\lambda^N(\epsilon),\mu^N(\epsilon))\Big\}$ is unbounded. As a result, there exists a subsequence satisfying $\lim_{k\to\infty} \tau^{N_k}(\epsilon) = +\infty$. We may further assume that the limits $\bar{\lambda}_i(\epsilon) := \lim_{k\to\infty} \frac{\lambda_i^{N_k}(\epsilon)}{\tau_i^{N_k}(\epsilon)}$ $(i = 1,2,\cdots,p)$ and $\bar{\mu}_j(\epsilon) := \lim_{k\to\infty} \frac{\mu_j^{N_k}(\epsilon)}{\tau_j^{N_k}(\epsilon)}$ $(j = 1,2,\cdots,q)$ exist. Obviously, we have

$$\sum_{i=1}^{p} \bar{\lambda}_i(\epsilon) + \sum_{j=1}^{q} \mid \bar{\mu}_j(\epsilon) \mid = 1. \tag{4.6}$$

Dividing (4.1) by $\tau^{N_k}(\epsilon)$, we have

$$\begin{aligned} 0 \quad \in \quad & \frac{1}{N_k\tau^{N_k}(\epsilon)} \sum_{i=1}^{N_k} \nabla_x \hat{f}(\bar{x}_{N_k}(\epsilon),\xi^i,\epsilon) + \sum_{i=1}^{p} \frac{\lambda_i^{N_k}(\epsilon)}{\tau^{N_k}(\epsilon)}\ \nabla_x \hat{g}_i(\bar{x}_{N_k}(\epsilon),\epsilon) \\ & + \sum_{j=1}^{q} \frac{\mu_j^{N_k}(\epsilon)}{\tau^{N_k}(\epsilon)}\ \nabla_x \hat{h}_j(\bar{x}_{N_k}(\epsilon),\epsilon) + \mathcal{N}_{\mathcal{X}}(\bar{x}_{N_k}(\epsilon)). \end{aligned} \tag{4.7}$$

By the MFCQ assumption, $\{\nabla_x \hat{h}_j(\bar{x}(\epsilon),\epsilon)\}_{j\in J}$ are linearly independent. Let $J_0 := \{j \in J \mid \bar{\mu}(\epsilon)_j > 0\}$. Then, by Gordan's Theorem, there exists $\rho_0 \in R^n$ such that

$$\rho_0^T \nabla_x \hat{h}_j(\bar{x}(\epsilon),\epsilon) \begin{cases} < 0, & j \in J_0, \\ > 0, & j \notin J_0. \end{cases} \tag{4.8}$$

Let $\zeta \in int\mathcal{T}_{\mathcal{X}}(\bar{x}(\epsilon))$ be the vector satisfying Assumption 4.1. Choose $\beta_0 > 0$ and $\alpha_0 \in (0,1)$ such that

$$(1-\alpha_0)\zeta^T \nabla_x \hat{g}_i(\bar{x}(\epsilon),\epsilon) + \alpha_0 \rho_0^T \nabla_x \hat{g}_i(\bar{x}(\epsilon),\epsilon) \le -\beta_0 < 0, \quad \forall i \in I(\bar{x}(\epsilon)),$$

and

$$\bar{\rho}_0 := (1-\alpha_0)\zeta + \alpha_0\rho_0 \in \mathcal{T}_{\mathcal{X}}(\bar{x}(\epsilon)).$$

Let

$$\vartheta_0 := \min_{j\in J} \mid \rho_0^T \nabla_x \hat{h}_j(\bar{x}(\epsilon),\epsilon) \mid, \quad \gamma_0 := \min(\beta_0, \alpha_0\vartheta_0) > 0.$$

We have

$$\bar{\rho}_0^T \nabla_x \hat{g}_i(\bar{x}(\epsilon),\epsilon) \le -\beta_0 \le -\gamma_0 < 0, \quad i \in I(\bar{x}(\epsilon))$$

and

$$\bar{\rho}_0^T \nabla_x \hat{h}_j(\bar{x}(\epsilon),\epsilon) = \alpha_0 \rho_0^T \nabla_x \hat{h}_j(\bar{x}(\epsilon),\epsilon) \begin{cases} \le -\alpha_0\vartheta_0 \le -\gamma_0, & j \in J_0, \\ \ge \alpha_0\vartheta_0 \ge \gamma_0, & j \notin J_0. \end{cases}$$

Return now to (4.7). As shown in (4.4), there holds

$$\lim_{k\to\infty} \frac{1}{N_k} \sum_{i=1}^{N_k} \nabla_x \hat{f}(\bar{x}_{N_k}(\epsilon), \xi^i, \epsilon) = \mathbb{E}\Big[\nabla_x \hat{f}(\bar{x}(\epsilon), \xi, \epsilon)\Big] \quad w.p.1.$$

Moreover, we have $\lim_{k\to\infty} \nabla_x \hat{g}_i(\bar{x}_{N_k}(\epsilon),\epsilon) = \nabla_x \hat{g}_i(\bar{x}(\epsilon),\epsilon)$ and $\lim_{k\to\infty} \nabla_x \hat{h}_j(\bar{x}_{N_k}(\epsilon),\epsilon) = \nabla_x h_j(\bar{x}(\epsilon),\epsilon)$. In addition, since the normal cone is upper semi-continuous, we have $\mathcal{N}_{\mathcal{X}}(\bar{x}_{N_k}(\epsilon)) \subset \mathcal{N}_{\mathcal{X}}(\bar{x}(\epsilon)) + \epsilon_{N_k} B$ when k is large enough. Thus, since $\bar{\rho}_0 \in \mathcal{T}_{\mathcal{X}}(\bar{x}(\epsilon))$, by multiplying (4.7) with $\bar{\rho}_0$ and taking a limit, we have

$$0 \le \sum_{i=1}^{p} \bar{\lambda}_i(\epsilon) \bar{\rho}_0^T \nabla_x \hat{g}_i(\bar{x}(\epsilon),\epsilon) + \sum_{j=1}^{q} \bar{\mu}_j(\epsilon) \bar{\rho}_0^T \nabla_x \hat{h}_j(\bar{x}(\epsilon),\epsilon) \le -\gamma_0 \Big(\sum_{i=1}^{p} \bar{\lambda}_i(\epsilon) + \sum_{j=1}^{q} \mid \bar{\mu}_j(\epsilon) \mid \Big).$$

Since $\gamma_0 > 0$, this obviously contradicts (4.6) and hence $\Big\{(\lambda^N(\epsilon), \mu^N(\epsilon))\Big\}$ is bounded.

(ii) Since $\Big\{(\lambda^N(\epsilon), \mu^N(\epsilon))\Big\}$ is bounded, we may assume that $\lambda(\epsilon) := \lim_{N\to\infty} \lambda^N(\epsilon)$ and $\mu(\epsilon) := \lim_{N\to\infty} \mu^N(\epsilon)$ exist. Combined with the assertion above, by taking a limit in (4.1), we obtain (2.5) and (2.4) immediately. As a result, $\bar{x}(\epsilon)$ is a weak KKT point of problem (1.3), i.e., $\bar{x}(\epsilon) \in S(\epsilon)$. ■

Combining Theorem 3.2 and Theorem 4.1, we may expect that, when N is sufficiently large and $\epsilon > 0$ is sufficiently small, $x_N(\epsilon)$ is an acceptable approximation weak KKT point of the original problem (1.2).

We now turn to the case where $\epsilon = \epsilon_N \to 0$ as $N \to \infty$. In this case, a feasible point $\bar{x}(\epsilon_N)$ is a KKT point of problem (1.4) with $\epsilon = \epsilon_N$ if and only if there exist multipliers $\lambda(\epsilon_N)$ and $\mu(\epsilon_N)$ such that

$$\begin{aligned} 0 \in & \frac{1}{N}\sum_{i=1}^{N} \nabla_x \hat{f}(\bar{x}(\epsilon_N), \xi^i, \epsilon_N) + \nabla_x \hat{g}(\bar{x}(\epsilon_N), \epsilon_N)^T \lambda(\epsilon_N) \\ & + \nabla_x \hat{h}(\bar{x}(\epsilon_N), \epsilon_N)^T \mu(\epsilon_N) + \mathcal{N}_{\mathcal{X}}(\bar{x}(\epsilon_N)), \\ \lambda(\epsilon_N) & \geq 0, \quad \lambda(\epsilon_N)^T \hat{g}(\bar{x}(\epsilon_N), \epsilon_N) = 0. \end{aligned} \tag{4.9}$$

The following result is extracted from the proof of Theorem 4.3 in [29].

Lemma 4.2 *Let V be a compact set and $\mathcal{G}(v,\xi) : V \times \Xi \to 2^{\Re^m}$ be a measurable and compact set-valued mapping that is upper semi-continuous with respect to v on V for almost every ξ. Let $\{\xi^1, \cdots, \xi^N\}$ be independently and identically distributed random samples and $\mathcal{G}_N(v) := \frac{1}{N}\sum_{i=1}^{N} \mathcal{G}(v, \xi^i)$. Suppose that $\mathcal{G}(v,\xi)$ is dominated by an integrable function. Let $\{v^N\}$ be an arbitrary sequence in V and $\bar{v}$ be an accumulation point of $\{v^N\}$. Then*

$$\lim_{N\to\infty} \mathbb{D}\Big(\mathcal{G}_N(v^N), \mathbb{E}\Big[\mathrm{conv}\, \mathcal{G}(\bar{v},\xi)\Big]\Big) = 0.$$

Theorem 4.2 *Let $\{\bar{x}(\epsilon_N)\}$ be a sequence of KKT points of problems (1.4) with $\epsilon = \epsilon_N$. Let $\bar{x}$ be an accumulation point of the sequence and the JCQ hold at $\bar{x}$. Suppose that, for almost every ξ, $\hat{f}$, $\hat{g}$ and $\hat{h}$ satisfy the gradient consistency at $(\bar{x}, \xi, 0)$, $(\bar{x}, 0)$, and $(\bar{x}, 0)$ respectively, that is,*

$$\pi_x \partial_{(x,\epsilon)} \hat{f}(\bar{x}, \xi, 0) \subset \partial_x f(\bar{x}, \xi), \quad \pi_x \partial_{(x,\epsilon)} \hat{g}(\bar{x}, 0) \subset \partial_x g(\bar{x}), \quad \pi_x \partial_{(x,\epsilon)} \hat{h}(\bar{x}, 0) \subset \partial_x h(\bar{x}).$$

If there exists a compact set $\mathcal{C} \subset R^n$ such that it contains a neighborhood of $\bar{x}$ and Assumption 4.1 holds, then we have $\bar{x} \in S$ with probability one.

Proof. In analogy to [29, Theorem 4.4], we define a set-valued mapping as follows:

$$\mathcal{A}(x,\xi,\epsilon) := \begin{cases} \nabla_x \hat{f}(x,\xi,\epsilon), & \epsilon \neq 0 \\ \mathrm{conv}\Big\{ \limsup\limits_{x' \to x,\, \epsilon' \to 0} \pi_x \partial_{(x,\epsilon)} \hat{f}(x', \xi, \epsilon') \Big\}, & \epsilon = 0. \end{cases}$$

Then (4.9) can be rewritten as

$$\begin{aligned}0 \in\ & \frac{1}{N}\sum_{i=1}^{N}\mathcal{A}(x(\epsilon_N),\xi^i,\epsilon_N) + \partial_x\hat{g}(\bar{x}(\epsilon_N),\epsilon_N)^T\lambda(\epsilon_N) \\ & +\partial_x\hat{h}(\bar{x}(\epsilon_N),\epsilon_N)^T\mu(\epsilon_N) + \mathcal{N}_{\mathcal{X}}(\bar{x}(\epsilon_N)). \end{aligned} \quad (4.10)$$

It follows from the proof of [29, Theorem 4.4], Lemma 4.2, and the gradient consistency of $\hat{f}$ that

$$\limsup_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\mathcal{A}(x(\epsilon_N),\xi^i,\epsilon_N) \subset \mathbb{E}\Big[\pi_x\partial_{(x,\epsilon)}\hat{f}(\bar{x},\xi,0)\Big] \subset \mathbb{E}\Big[\partial_x f(\bar{x},\xi)\Big].$$

Similarly, from the upper semi-continuity derived by the gradient consistency, we have

$$\limsup_{N\to\infty}\partial_x\hat{g}(\bar{x}(\epsilon_N),\epsilon_N) \subset \partial_x\hat{g}(\bar{x},0) = \partial g(\bar{x})$$

and

$$\limsup_{N\to\infty}\partial_x\hat{h}(\bar{x}(\epsilon_N),\epsilon_N) \subset \partial_x\hat{h}(\bar{x},0) = \partial h(\bar{x}).$$

Note that the normal cone is upper semi-continuous. In a similar way to prove Theorem 3.2, we can show that the corresponding multipliers set is bounded and hence $\bar{x}$ is a weak KKT point of problem (1.2). This completes the proof. ■

4.2 Exponential convergence of optimal values

We next discuss the convergence of optimal values of the smoothed SAA problems. Note that problem (1.4) with $\epsilon = \epsilon_N$ can be equivalently rewritten as

$$\begin{aligned} \min \quad & \frac{1}{N}\sum_{i=1}^{N}\hat{f}(x,\xi^i,\epsilon_N) \\ \text{s.t.} \quad & x \in \mathcal{X}, \\ & 0 \in \hat{\varphi}(x,\epsilon_N) + \mathfrak{K}. \end{aligned} \quad (4.11)$$

Define the feasible set mapping $\mathfrak{F}$, the optimal value function $\mathfrak{V}$, and the corresponding optimal solution set mapping $\mathfrak{M}$ as in Section 3.1.

Theorem 4.3 *Let x^* denote an optimal solution of (1.5). Suppose that*

(i) *there exists a compact subset $\mathcal{C} \subset R^n$ such that $\mathfrak{M}(\epsilon_N)\cap\mathcal{C} \neq \emptyset$ for every N large enough w.p.1;*

(ii) $\mathbb{E}[\tilde{\kappa}_{\mathcal{C}}(\xi)] < \infty$, *where* $\tilde{\kappa}_{\mathcal{C}}$ *denotes the control function in (3.1);*

(iii) *the moment generating function* $\mathbb{E}[e^{\tilde{\kappa}_{\mathcal{C}}(\xi)t}]$ *is finite valued for all* t *in a neighborhood of zero;*

(iv) $x^* \in \mathcal{C}$, $\hat{\varphi}(x,\epsilon)$ *satisfies the gradient consistency at* $(x^*, 0)$, *and*

$$0 \in int\{\varphi(x^*) + \mathcal{A}(\mathcal{T}_{\mathcal{X}}(x^*)) + \mathfrak{K}\}, \qquad \forall\, \mathcal{A} \in \partial\varphi(x^*); \tag{4.12}$$

(v) *for any given* $x \in \mathcal{C}$ *and* $\epsilon \in [0, \epsilon_0]$, *the moment generating function* $\mathbb{E}\Big[e^{(\hat{f}(x,\xi,\epsilon) - \mathbb{E}[\hat{f}(x,\xi,\epsilon)])t}\Big]$ *is finite valued for all* t *in a neighborhood of zero.*

Then, with probability approaching one exponentially fast, $x(\epsilon_N) \in \mathfrak{M}(\epsilon_N) \cap \mathcal{C}$ *yields an approximate optimal value* $\mathfrak{V}(\epsilon_N)$ *of problem (1.5) with the increase of the sample size* N.

Proof. As stated in the proof of Theorem 3.1, without loss of generality, we consider problem (1.6) in restriction to $\mathcal{C}$ instead of $\mathcal{X}$.

By the assumptions (i)–(iii) and (v), we have from Theorem 5.1 of [27] that, for any $\delta > 0$, there exist positive constants $C(\delta)$ and $\beta(\delta)$ independent of N such that

$$\mathrm{Prob}\left\{ \sup_{x\in\mathcal{C},\, \epsilon\in[0,\epsilon_0]} \Big| \frac{1}{N}\sum_{i=1}^{N} \hat{f}(x,\xi^i,\epsilon) - \mathbb{E}\Big[\hat{f}(x,\xi,\epsilon)\Big]\Big| > \delta \right\} \le C(\delta)e^{-N\beta(\delta)}. \tag{4.13}$$

Let $\sigma := \mathbb{E}[\tilde{\kappa}(\xi)]$. Since $\epsilon_N \to 0$ as $N \to \infty$, there exists $N(\omega) > 0$ such that $\epsilon_N \le \delta/\sigma$ whenever $N > N(\omega)$. By the assumption (iv), the map $(x,\epsilon) \to \partial_x\hat{\varphi}(x,\epsilon)$ is upper semicontinuous at $(x^*, 0)$. We next show

$$\mathrm{Prob}\left\{ \Big| \frac{1}{N}\sum_{i=1}^{N} \hat{f}(x(\epsilon_N),\xi^i,\epsilon_N) - \mathbb{E}\Big[\hat{f}(x^*,\xi)\Big]\Big| > 3\delta \right\} \le C(\delta)e^{-N\beta(\delta)}. \tag{4.14}$$

(I) We first show

$$\mathrm{Prob}\left\{ \frac{1}{N}\sum_{i=1}^{N} \hat{f}(x(\epsilon_N),\xi^i,\epsilon_N) - \mathbb{E}\Big[f(x^*,\xi)\Big] < -3\delta \right\} \le C(\delta)e^{-N\beta(\delta)}. \tag{4.15}$$

Otherwise, there must exist a subsequence $\{N_k\}$ such that

$$\mathrm{Prob}\left\{ \frac{1}{N_k}\sum_{i=1}^{N_k} \hat{f}(x(\epsilon_{N_k}),\xi^i,\epsilon_{N_k}) - \mathbb{E}\Big[f(x^*,\xi)\Big] < -3\delta \right\} > C(\delta)e^{-N_k\beta(\delta)}, \quad \forall k.$$

Since $\mathcal{C}$ is compact, by Lemma 3.1, the map $\mathfrak{F}$ is closed. Therefore, we may suppose that

$$\lim_{k\to\infty} x(\epsilon_{N_k}) = \bar{x} \in \mathfrak{F}(0) \cap \mathcal{C}.$$

Note that

$$\begin{aligned}\frac{1}{N_k}\sum_{i=1}^{N_k}\hat f(x(\epsilon_{N_k}),\xi^i,\epsilon_{N_k})-\mathbb{E}\Big[f(x^*,\xi)\Big] \quad &= \quad \frac{1}{N_k}\sum_{i=1}^{N_k}\hat f(x(\epsilon_{N_k}),\xi^i,\epsilon_{N_k})-\mathbb{E}\Big[\hat f(x(\epsilon_{N_k}),\xi,\epsilon_{N_k})\Big]\\ &\quad +\mathbb{E}[\hat f(x(\epsilon_{N_k}),\xi,\epsilon_{N_k})]-\mathbb{E}[f(\bar x,\xi)]\\ &\quad +\mathbb{E}[f(\bar x,\xi)]-\mathbb{E}[f(x^*,\xi)].\end{aligned}$$

Since x^* is an optimal solution, there holds $\mathbb{E}[f(\bar x,\xi)]-\mathbb{E}[f(x^*,\xi)]\geq 0$. Choose k sufficiently large so that $\epsilon_{N_k}\leq \delta/\sigma$ and $\|\ \bar x - x^*\ \|\leq \delta/\sigma$. It follows that

$$\Big|\mathbb{E}[\hat f(x(\epsilon_{N_k}),\xi,\epsilon_{N_k})]-\mathbb{E}[f(\bar x,\xi)]\Big|\leq \mathbb{E}[\tilde\kappa_{\mathcal{C}}(\xi)]\Big(\|x(\epsilon_{N_k})-\bar x\|\ +\epsilon_{N_k}\Big)\leq 2\delta.$$

We then have

$$\begin{aligned}&\mathrm{Prob}\Bigg\{\sup_{x\in\mathcal{C},\,\epsilon\in[0,\epsilon_0]}\Big|\frac{1}{N_k}\sum_{i=1}^{N_k}\hat f(x,\xi^i,\epsilon)-\mathbb{E}\Big[\hat f(x,\xi,\epsilon)\Big]\Big|>\delta\Bigg\}\\ &\geq \mathrm{Prob}\Bigg\{\frac{1}{N_k}\sum_{i=1}^{N_k}\hat f(x(\epsilon_{N_k}),\xi^i,\epsilon_{N_k})-\mathbb{E}\Big[\hat f(x(\epsilon_{N_k}),\xi,\epsilon_{N_k})\Big]<-\delta\Bigg\}\\ &\geq \mathrm{Prob}\Bigg\{\frac{1}{N_k}\sum_{i=1}^{N_k}\hat f(x(\epsilon_{N_k}),\xi^i,\epsilon_{N_k})-\mathbb{E}\Big[f(x^*,\xi)\Big]<-3\delta\Bigg\}>C(\delta)e^{-N_k\beta(\delta)}.\end{aligned}$$

This contradicts (4.13) and hence (4.15) is true.

(II) Now we show

$$\mathrm{Prob}\Bigg\{\frac{1}{N}\sum_{i=1}^{N}\hat f(x(\epsilon_N),\xi^i,\epsilon_N)-\mathbb{E}\Big[f(x^*,\xi)\Big]>3\delta\Bigg\}\leq C(\delta)e^{-N\beta(\delta)}. \tag{4.16}$$

Let N be sufficiently large so that $\epsilon_N\leq\delta/\sigma$. By the Lipschitz continuity of $\hat f$, there exists a neighborhood $\mathcal{N}_\delta(x^*)$ of x^* such that

$$\Big|\mathbb{E}[\hat f(x,\xi,\epsilon_N)]-\mathbb{E}[f(x^*,\xi)]\Big|\leq\mathbb{E}[\tilde\kappa(\xi)]\Big(\|x-x^*\|+\mid\epsilon_N-0\mid\Big)\leq 2\delta \tag{4.17}$$

for every $x\in\mathcal{N}_\delta(x^*)$. By the assumption (iv), we can apply [34, Theorem 3.1] to find a neighborhood $\mathcal{N}(0)$ of 0 such that $\mathcal{N}(0)\subset[-\delta/\sigma,\delta/\sigma]$ and there exists a vector $x'(\epsilon_N)\in\mathfrak{F}(\epsilon_N)\cap\mathcal{N}_\delta(x^*)$ for every N. Since $x(\epsilon_N)$ is an optimal solution of (4.11) and

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}\hat f(x(\epsilon_N),\xi^i,\epsilon_N)-\mathbb{E}\Big[f(x^*,\xi)\Big] \quad &= \quad \frac{1}{N}\sum_{i=1}^{N}\hat f(x(\epsilon_N),\xi^i,\epsilon_N)-\frac{1}{N}\sum_{i=1}^{N}\hat f(x'(\epsilon_N),\xi^i,\epsilon_N)\\ &\quad +\frac{1}{N}\sum_{i=1}^{N}\hat f(x'(\epsilon_N),\xi^i,\epsilon_N)-\mathbb{E}[\hat f(x'(\epsilon_N),\xi,\epsilon_N)]\\ &\quad +\mathbb{E}[\hat f(x'(\epsilon_N),\xi,\epsilon_N)]-\mathbb{E}[f(x^*,\xi)],\end{aligned}$$

we have from (4.17) that

$$\begin{aligned}
&\text{Prob}\left\{\frac{1}{N}\sum_{i=1}^{N}\hat{f}(x(\epsilon_N),\xi^i,\epsilon_N)-\mathbb{E}\Big[f(x^*,\xi)\Big]>3\delta\right\}\\
&\le \text{Prob}\left\{\frac{1}{N}\sum_{i=1}^{N}\hat{f}(x'(\epsilon_N),\xi^i,\epsilon_N)-\mathbb{E}[\hat{f}(x'(\epsilon_N),\xi,\epsilon_N)]>\delta\right\}\\
&\le \text{Prob}\left\{\sup_{x\in\mathcal{C},\,\epsilon\in[0,\epsilon_0]}\left|\frac{1}{N}\sum_{i=1}^{N}\hat{f}(x,\xi^i,\epsilon)-\mathbb{E}\Big[\hat{f}(x,\xi,\epsilon)\Big]\right|>\delta\right\}\\
&\le C(\delta)e^{-N\beta(\delta)},
\end{aligned}$$

that is, (4.16) is true.

This completes the proof of (4.14), which means that, with probability at least $1 - C(\delta)e^{-N\beta(\delta)}$, an optimal solution of (4.11) becomes a (3δ-approximate) optimal solution of problem (1.5). ■

The above theorem gives a result of exponential convergence in probability of the SAA estimators. Actually, by replacing the assumptions with some moderate conditions, we can obtain that the sequence of optimal solutions of (4.11) converges to an optimal solution of (1.5) with probability one. To this end, we make the following assumption:

Assumption 4.2 There exists a constant $\epsilon_0 > 0$ and a measurable function $\bar{\kappa}_{\mathcal{C}}(\xi)$ such that

$$\sup_{x\in\mathcal{C},\,\epsilon\in[0,\epsilon_0]}|\hat{f}(x,\xi,\epsilon)|\le\bar{\kappa}_{\mathcal{C}}(\xi)$$

for all $\xi\in\Xi$, where $\mathbb{E}[\bar{\kappa}_{\mathcal{C}}(\xi)]<\infty$ and $\mathcal{C}$ is a compact subset of $\mathcal{X}$.

Note that, under the above assumption, Lemma 4.1 can be applied to $\frac{1}{N}\sum_{i=1}^{N}\hat{f}(x(\epsilon_N),\xi^i,\epsilon_N)$ and hence we have the following convergence result easily.

Theorem 4.4 *Let x^* denote an optimal solution of (1.5). Suppose that the assumptions (i) and (iv) in Theorem 4.3 and Assumption 4.2 hold. Then, with probability one, $x(\epsilon_N)\in\mathfrak{M}(\epsilon_N)\cap\mathcal{C}$ becomes a 'good' approximate optimal solution of problem (1.5) with the increase of the sample size N.*

Here, we omit its proof because it is similar to Theorem 3.1. We next make some remarks on the assumptions.

In Theorem 4.3, condition (ii) requires the Lipschitz module of the smoothed function to be integrable; conditions (iii) and (v) mean that the probability distribution of the random variables $\tilde{\kappa}_{\mathcal{C}}(\xi)$ and $\hat{f}(x,\xi,\epsilon)$ die exponentially fast in the tails and particularly, they hold if ξ has a distribution supported on a bounded subset of R^k; condition (iv), which generally guarantees the stability of the smooth approximation, has been remarked in Section 3.1. The following proposition gives a sufficient condition for condition (i).

Proposition 4.1 *Let x^* denote an optimal solution of problem (1.5). Suppose that*

(i) *Assumption 4.2 holds in a neighborhood of x^*;*

(ii) *$\hat{\varphi}(x,\epsilon)$ satisfies the gradient consistency at $(x^*,0)$ and condition (3.5) holds.*

If

$$\liminf_{\|x\|\to\infty,\, N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \hat{f}(x,\xi^i,\epsilon_N) > \mathbb{E}[f(x^*,\xi)]$$

holds with probability one, there almost surely exists a compact subset $\mathcal{C} \subset R^n$ such that $\mathfrak{M}(\epsilon_N) \cap \mathcal{C} \neq \emptyset$ for any N large sufficiently.

Since the domination assumption (i) guarantees uniform convergence (and hence continuous convergence), one can prove the above proposition in a similar way to [34, Theorem 4.3].

5 Applications to SMPEC

In this section, we apply the approach discussed in the previous sections to the following SMPEC:

$$\begin{array}{ll} \min & \mathbb{E}[f(x,\xi)] \\ \text{s.t.} & x \in \mathcal{X}, \\ & \mathbb{E}[G(x,\xi)] \geq 0,\ \mathbb{E}[H(x,\xi)] \geq 0,\ \mathbb{E}[G(x,\xi)]^T\mathbb{E}[H(x,\xi)] = 0, \end{array} \tag{5.1}$$

where $\mathcal{X}$ and f are the same as above, $G, H : R^n \to R^m$ are locally Lipschitz and continuously differentiable in the variable x, and $\mathbb{E}[G(x,\xi)]$ and $\mathbb{E}[H(x,\xi)]$ are well defined for every $x \in \mathcal{X}$. By using the Fischer-Burmeister function

$$\varphi(a,b) := a + b - \sqrt{a^2+b^2}$$

given in [4], problem (5.1) can be equivalently written as the nonlinear programming problem

$$\begin{aligned} \min \quad & \mathbb{E}[f(x, \xi(\omega))] \\ \text{s.t.} \quad & x \in \mathcal{X}, \\ & \Phi(x) = 0, \end{aligned} \tag{5.2}$$

where $\Phi : R^n \to R^m$ is defined by

$$\Phi(x) := \Big(\varphi(\mathbb{E}[G_1(x, \xi)], \mathbb{E}[H_1(x, \xi)]), \cdots, \varphi(\mathbb{E}[G_m(x, \xi)], \mathbb{E}[H_m(x, \xi)])\Big)^T.$$

We make extra assumption that $\nabla_x G$ and $\nabla_x H$ are locally Lipschitz continuous with respect to x, which means that $G, H, \nabla_x G$ and $\nabla_x H$ are all globally Lipschitz over any compact set $\mathcal{C}$. That is, there exists $\hat{\kappa}_{\mathcal{C}}(\xi) > 0$ such that

$$\|G(x', \xi) - G(x'', \xi)\| \leq \hat{\kappa}_{\mathcal{C}}(\xi)\|x' - x''\|, \ \|H(x', \xi) - H(x'', \xi)\| \leq \hat{\kappa}_{\mathcal{C}}(\xi)\|x' - x''\|$$

and

$$\|\nabla_x G(x', \xi) - \nabla_x G(x'', \xi)\| \leq \hat{\kappa}_{\mathcal{C}}(\xi)\|x' - x''\|, \ \|\nabla_x H(x', \xi) - \nabla_x H(x'', \xi)\| \leq \hat{\kappa}_{\mathcal{C}}(\xi)\|x' - x''\|$$

for any $x', x'' \in \mathcal{C}$ and almost every $\xi \in \Xi$. In accordance with [23, Proposition 2], these inequalities guarantee the continuous differentiability of $\mathbb{E}[G(x, \xi)]$ and $\mathbb{E}[H(x, \xi)]$. It is not difficult to see that $\varphi(\mathbb{E}[G_l(x, \xi)], \mathbb{E}[H_l(x, \xi)])$ is locally Lipschitz for each $l = 1, 2, \cdots, m$ and so is $\Phi(x)$. Then we may apply the smoothing SAA method discussed in the previous sections to problem (5.2) since, with a little extension, (5.2) is a special case of problem (1.2). The smoothing SAA approximation problem becomes

$$\begin{aligned} \min \quad & \frac{1}{N}\sum_{i=1}^{N} \hat{f}(x, \xi^i, \epsilon_N) \\ \text{s.t.} \quad & x \in \mathcal{X}, \\ & \hat{\Phi}(x, \epsilon_N) = 0, \end{aligned} \tag{5.3}$$

where $\hat{\Phi}$ is defined by

$$\Big(\hat{\varphi}(\frac{1}{N}\sum_{i=1}^{N} G_1(x, \xi^i), \frac{1}{N}\sum_{i=1}^{N} H_1(x, \xi^i), \epsilon_N), \cdots, \hat{\varphi}(\frac{1}{N}\sum_{i=1}^{N} G_m(x, \xi^i), \frac{1}{N}\sum_{i=1}^{N} H_m(x, \xi^i), \epsilon_N)\Big)^T$$

with the smoothing Fischer-Burmeister function [11]

$$\hat{\varphi}(a, b, \epsilon) := a + b - \sqrt{a^2 + b^2 + \epsilon^2}.$$

Let x be feasible to (5.2). Then we have

$$\partial \Phi_i(x) = \begin{cases} \nabla_x \mathbb{E}[G_i(x,\xi)]\Big(1 - \frac{\mathbb{E}[G_i(x,\xi)]}{\sqrt{\mathbb{E}[G_i(x,\xi)]^2 + \mathbb{E}[H_i(x,\xi)]^2}}\Big) + \\ \nabla_x \mathbb{E}[H_i(x,\xi)]\Big(1 - \frac{\mathbb{E}[H_i(x,\xi)]}{\sqrt{\mathbb{E}[G_i(x,\xi)]^2 + \mathbb{E}[H_i(x,\xi)]^2}}\Big), & i \in I_{\mathbb{E}[G]}(x)^c \cup I_{\mathbb{E}[H]}(x)^c; \\ \nabla_x \mathbb{E}[G_i(x,\xi)](1-\alpha) + \nabla_x \mathbb{E}[H_i(x,\xi)](1-\beta), & i \in I_{\mathbb{E}[G]}(x) \cap I_{\mathbb{E}[H]}(x) \end{cases}$$

for each i, where $(\alpha,\beta) \in R^2$ denotes an arbitrary vector with $\|(\alpha,\beta)\| \le 1$. On the other hand, we have

$$\begin{aligned} \pi_x \partial_{(x,\epsilon)} \hat{\Phi}_i(x,0) &= \limsup_{(x',\epsilon)\to(x,0)} \nabla_x \hat{\Phi}_i(x',\epsilon) \\ &= \limsup_{(x',\epsilon)\to(x,0)} \nabla_x \mathbb{E}[G_i(x',\xi)]\Big(1 - \frac{\mathbb{E}[H_i(x',\xi)]}{\sqrt{\mathbb{E}[H_i(x',\xi)]^2 + \mathbb{E}[H_i(x',\xi)]^2 + \epsilon^2}}\Big) \\ &\qquad + \nabla_x \mathbb{E}[G_i(x',\xi)]\Big(1 - \frac{\mathbb{E}[G_i(x',\xi)]}{\sqrt{\mathbb{E}[G_i(x',\xi)]^2 + \mathbb{E}[H_i(x',\xi)]^2 + \epsilon^2}}\Big) \\ &= \partial \Phi_i(x). \end{aligned}$$

From the above discussion, we see that $\hat{\Phi}$ satisfies the gradient consistency at $(x,0)$, which means that $\partial_x \hat{\Phi}(x,\epsilon)$ enjoys the upper semi-continuity. Thus, we can claim safely that the smoothing SAA approach given in the previous sections is admissible.

We now turn to discuss problem (5.2). For coherent adoption, some notation needs to be specialized here. Definition 2.6, i.e., the JCQ can be written as follows.

Definition 5.1 For all $\lambda \in R^m \backslash \{0\}$, there holds

$$0 \notin \partial \Phi(\bar{x})^T \lambda + \mathcal{N}_{\mathcal{X}}(\bar{x}).$$

From the explicit formula of $\partial\Phi(x)$, it is easy to see that the JCQ is equivalent to the MPEC no nonzero abnormal multiplier constraint qualification (MPEC-NNAMCQ) of Clarke (C-) type [7,31]. Sufficient conditions for this condition to hold include the well-known MPEC linear independence constraint qualification (MPEC-LICQ).

Note that $\hat{\Phi}(x,\epsilon)$ satisfies the gradient consistency at $(x,0)$ for all feasible point x. If we apply the smoothing SAA method (5.2), under moderate condition such as the MPEC-LICQ, it is not difficult to obtain some results related to the stability and convergence. We next state a convergence result only and omit its proof here.

Theorem 5.1 *Suppose that there exist multipliers $\lambda(\epsilon_N) \in R^m$ such that $\bar{x}(\epsilon_N)$ satisfies*

$$0 \in \frac{1}{N}\sum_{i=1}^{N} \partial_x \hat{f}(\bar{x}(\epsilon_N), \xi^i, \epsilon_N) + \partial_x \hat{\Phi}(\bar{x}(\epsilon_N), \epsilon_N)^T \rho(\epsilon_N) + \mathcal{N}_{\mathcal{X}}(\bar{x}(\epsilon_N)),$$

namely, $\bar{x}(\epsilon_N)$ is a stationary point of the smoothing SAA problem (5.1). Let $\epsilon_N \to 0$ as $N \to \infty$ and $\bar{x}$ be an accumulation of the sequence. Suppose that the C-type MPEC-NNAMCQ holds at $\bar{x}$. Suppose also that, for almost every ξ, $\hat{f}$ satisfies the gradient consistency at $(\bar{x}, \xi, 0)$, $(\bar{x}, 0)$ and $(\bar{x}, 0)$ respectively. If there exists a compact set $\mathcal{C} \subset R^n$ containing a neighborhood of $\bar{x}$ and Assumption 4.1 holds, then $\bar{x}$ is a weakly C-stationary point of problem (5.1) with probability one.

In the rest of this section, we use the example given by Lin et al. in [12] to illustrate the smoothing SAA approximation method for solving a stochastic Stackelberg-Nash-Cournot game in the European gas market where a particular gas producer has an opportunity to develop a new and important field. Let x denote the decision variable of the leader, that is, the quantity supplied by the leader to the market, and y_i denote the decision variable of the i-th follower, that is, the quantity supplied by the i-th firm to the market. Then the stochastic Stackelberg-Nash-Cournot equilibrium problem is formulated as the following SMPEC [12]:

$$\begin{aligned} \max \quad & \mathbb{E}\big[xp(x + \mathbf{e}^T y, \xi)\big] - c_0(x) \\ \text{s.t.} \quad & 0 \le x \le L, \\ & 0 \le y \perp \mathbb{E}\big(F(x, y, \xi)\big) \ge 0, \end{aligned} \tag{5.4}$$

where $p(\tau, \xi)$ denotes the inverse demand function with τ to be the total quantity of supply to the market, $L > 0$ is a constant and $c_0(x)$ is the cost for the leader to produce x, $\mathbf{e} := (1, \ldots, 1)^T \in R^m$, $F(x, y, \xi) := -p(x + y^T\mathbf{e}, \xi)\mathbf{e} - p'_\tau(x + y^T\mathbf{e}, \xi)y + \mathbf{c}'(y)$, $\mathbf{c}'(y) := (c'_1(y_1), \ldots, c'_m(y_m))^T$, and $c_i(y_i)$ is the total cost for the i-th firm to produce y_i.

In order to demonstrate the proposed method, we consider three followers and suppose the involved functions to be

$$\begin{aligned} p(\tau, \xi) &:= 20 - (0.002\xi + 0.003)\tau, \\ c_0(x) &:= 9.5x + 60, \\ c_1(y_1) &:= 8.6y_1 + 48, \\ c_2(y_2) &:= 8.9y_2 + 45, \\ c_3(y_3) &:= 9.2y_3 + 75. \end{aligned}$$

Moreover, we suppose that the stochastic variable ξ is uniformly distributed on $[-1, 1]$ and

$L = 1800$. Then the above problem (5.4) becomes

$$\begin{aligned} \max \quad & \mathbb{E}[x(20 - (0.002\xi + 0.003)(x + y_1 + y_2 + y_3) + \xi)] - c_0(x) \\ \text{s.t.} \quad & 0 \leq x \leq 1800, \\ & y \geq 0,\ \mathbb{E}[F(x, y, \xi)] \geq 0, \\ & y^T \mathbb{E}[F(x, y, \xi)] = 0 \end{aligned} \tag{5.5}$$

with $F(x, y, \xi) := N(\xi)x + M(\xi)y + q$ and

$$\begin{aligned} N(\xi) \;&:=\; \begin{pmatrix} 0.002\xi + 0.003 \\ 0.002\xi + 0.003 \\ 0.002\xi + 0.003 \end{pmatrix}, \\ M(\xi) \;&:=\; \begin{pmatrix} 0.004\xi + 0.006, 0.002\xi + 0.003, 0.002\xi + 0.003 \\ 0.002\xi + 0.003, 0.004\xi + 0.006, 0.002\xi + 0.003 \\ 0.002\xi + 0.003, 0.002\xi + 0.003, 0.004\xi + 0.006 \end{pmatrix}, \\ q \;&:=\; -(11.4, 11.1, 10.8)^T. \end{aligned}$$

The solution of this problem is $(x^*, y^*) = (1450, 662.5, 562.5, 462.5)^T$. Since there always holds $\mathbb{E}[F(x^*, y^*, \xi)] = 0$ and the gradients

$$\{\nabla_{(x,y)}\mathbb{E}[F_1(x^*, y^*, \xi)], \nabla_{(x,y)}\mathbb{E}[F_2(x^*, y^*, \xi)], \nabla_{(x,y)}\mathbb{E}[F_3(x^*, y^*, \xi)]\}$$

are linearly independent, the MPEC-LICQ holds at (x^*, y^*).

In our tests, we employed the command "haltonset" in Matlab R2010a to generate random sequences and the solver "fmincon" to solve the smoothing SAA approximation problems. In particular, the optimization algorithm was set to be interior-point algorithm. Moreover, we chose the initial point to be $(20, 20, 20, 20)^T$, set the parameter $\epsilon_N := \frac{1}{N}$, and employed $\hat{\varphi}(a, b, \epsilon)$ as the smoothing function of $\varphi(a, b)$. The computational results are shown in Table 1, in which $\bar{x}(\epsilon_N)$ denotes the solution of the smoothing SAA approximation problems for example (5.5). The results shown in the table reveal that the smoothing SAA method was able to solve this example successfully.

6 Conclusions

We have proposed a smoothing and SAA approximation approach for solving the non-smooth stochastic programming problem (1.2). The approach is similar to the way used in [29]. However, in order to deal with the additional non-smooth constraints, we have presented some appropriate constraint qualifications, which have been used in the convergence analysis.

Table 1: Computational results

N	ϵ_N	$\bar{x}(\epsilon_N)$
10^2	$1/10^2$	$(1546.63, 622.25, 522.37, 422.30)$
10^3	$1/10^3$	$(1600.57, 542.87, 568.24, 432.57)$
10^4	$1/10^4$	$(1685.24, 665.85, 591.38, 458.37)$
10^5	$1/10^5$	$(1584.25, 658.89, 568.54, 468.91)$
10^6	$1/10^6$	$(1436.59, 660.38, 560.85, 459.67)$
10^7	$1/10^7$	$(1450.38, 662.43, 562.68, 462.39)$

We have shown that the perturbed problem (1.3) is stably under some regularity conditions. We have also investigated the limiting behavior of both the smoothed problem (1.3) and the smoothed SAA problem (1.4). Furthermore, we have applied the proposed approach to the SMPEC (5.1).

Acknowledgements. The authors are grateful to two anonymous referees whose helpful suggestions have led to much improvement of the paper.

References

[1] Alexander, S., Coleman, T.F. and Li, Y., *Minimizing CVaR and VaR for a portfolio of derivatives*, Journal of Banking and Finance, 30 (2006), 583–605.

[2] Clarke, F.H., Optimization and Nonsmooth Analysis, Wiley, New York, 1983.

[3] Ermoliev, Y.M., Stochatic Programming Methods, Nauka, Moscow, 1976.

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

[5] Fukushima, M. and Lin, G.H., *Smoothing methods for mathematical programs with equilibrium constraints*, Proceedings of the ICKS'04, IEEE Computer Society, 2004, 206–213.

[6] Gauvin, J., *A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming*, Mathematical Programming, 12 (1977), 136–138.

[7] Guo, L., Lin, G.H., Ye, J.J. and Zhang, J., *Sensitivity analysis of the value function for parametric mathematical programs with equilibrium constraints.* SIAM Journal on Optimization, 24 (2014), 1206–1237.

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

[9] Jourani, A., *Constraint qualifications and Lagrange multipliers in nondifferentiable programming probelms*, Journal of Optimization Theory and Applications, 81 (1994), 533–548.

[10] Kall, P., Stochatic Linear Programming, Springer, Berlin, 1976.

[11] Kanzow, C., *Some noninterior continuation methods for nonlinear complementarity problems*, SIAM Journal on Matrix Analysis and Applications, 17 (1996), 851–868.

[12] Lin, G.H., Xu, H.F. and Fukushima, M., *Monte Carlo and quasi-Monte Carlo sampling methods for a class of stochastic mathematical programs with equilibrium constraints*, Mathematical Methods of Operations Research, 67 (2008), 423–441.

[13] Luo, Z.Q., Pang, J.S. and Ralph, D., Mathematical Programs with Equilibrium Constraints, Cambridge University Press, Cambridge, UK, 1996.

[14] Meng, F.W. and Xu, H.F., *Exponential convergence of sample average approximation methods for a class of stochastic mathematical programs with complementarity constraints*, Journal of Computational Mathematics, 24 (2006), 733–748.

[15] Niederreiter, H., Random Number Generation and Quasi-Monte Carlo Method, SIAM, Philadelphia, 1992.

[16] Outrata, J., Kocvara, M. and Zowe, J., Nonsmooth Approach to Optimization Problems with Equilibrium Constraints: Theory, Applications, and Numerical Results, Kluwer Academic Publishers, Dordrect, Netherlands, 1998.

[17] Pappalardo, M., *Error bounds for generalized Lagrange multipliers in locally Lipschitz programming*, Journal of Optimization Theory and Applications, 73 (1992), 205–210.

[18] Peng, J.M., *A smoothing function and its applications*, Fukushima, M. and Qi, L. eds., Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and smoothing methods, Kluwer, Dordrecht, 1998, 293–316.

[19] Ralph, D. and Xu, H., *Implicit smoothing and its application to optimization with piecewise smooth equality constraints*, Journal of Optimization Theory and Applications, 124 (2005), 673–699.

[20] Rockafellar, R.T., *Lagrange multipliers and subderivatives of optimal value functions in nonlinear programming*, Mathematical Programming Study, 17 (1982), 28–66.

[21] Rochafellar, R.T. and Wets, R.J.-B., Variational Analysis. Springe, Berlin, 1998.

[22] Ruszczyński, A., *A linearization method for nonsmooth stochatic programming problems*, Mathematics of Operations Reseach, 12 (1987), 32–49.

[23] Ruszczyński, A. and Shapiro, A., Optimality and Duality in Stochastic Programming, Ruszczyński, A. and Shapiro, A. eds., Stochastic Programming, Handbook in Operations Research and Management Science, Elsevier, 2003.

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

[25] Shapiro, A., *Monte Carlo sampling approach to stochastic programming*, European Series of Applied and Indusrial Mathematics (ESAIM): Proceedings, 13 (2003), 65–73.

[26] Shapiro, A., Monte Carlo Sampling Methods, Ruszczyński, A. and Shapiro, A. eds., Stochastic Programming, Handbook in Operations Research and Management Science, Elsevier, 2003.

[27] Shapiro, A. and Xu, H.F., *Stochastic mathematics programs with equilibrium constraints, modling and sample average approximation*, Optimization, 57 (2008), 395–418.

[28] Wets, R.J.B., Stochastic Programming, Nemhauser, G.L. et al. eds., Handbooks in Operations Research and Management Science, 573–629, 1989.

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

[30] Ye, J.J., *Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints*, SIAM Journal on Optimization, 10 (2000), 943–962.

[31] Ye, J.J., *Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints*. Journal of Mathematical Analysis and Applications, 307 (2005), 350–369.

[32] Ye, J.J. and Ye, X.Y., *Necessary optimality conditions for optimization problems with variational inequality constraints*, Mathematics of Operations Research, 22 (1997), 977–997.

[33] Ye, J.J., Zhu, D.L. and Zhu, Q.J., *Exact penalization and necessary optimality conditions for generalized bilevel programming problems*, SIAM Journal on Optimization, 7 (1997), 481–507.

[34] Yen, N.D., *Stability of the solution set of perturbed nonsmooth inequality systems and application*, Journal of Optimization Theory and Applications, 93 (1997), 199–225.