Stability and Sensitivity Analysis of Stochastic Programs with Second Order Dominance Constraints

Yongchao Liu[1]
School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China
(Y.Liu@soton.ac.uk)

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

June 17, 2010

Abstract. In this paper we present stability and sensitivity analysis of a stochastic optimization problem with stochastic second order dominance constraints. We consider perturbation of the underlying probability measure in the space of regular measures equipped with pseudometric discrepancy distance ([30]). By exploiting a result on error bound in semi-infinite programming due to Gugat [13], we show under the Slater constraint qualification that the optimal value function is Lipschitz continuous and the optimal solution set mapping is upper semicontinuous with respect to the perturbation of the probability measure. In particular, we consider the case when the probability measure is approximated by empirical probability measure and show the exponential rate of convergence of optimal solution obtained from solving the approximation problem. The analysis is extended to the stationary points when the objective function is nonconvex.

Key words. Second order dominance, stochastic semi-infinite programming, error bound, Lipschitz stability, sample average approximation

1 Introduction

Stochastic dominance is a fundamental concept in decision theory and economics [22]. For two random variables $\xi_1(\omega)$ and $\xi_2(\omega)$ with finite expected values, $\xi_1(\omega)$ is said to dominate $\xi_2(\omega)$ in the second order, denoted by $\xi_1(\omega) \succeq_2 \xi_2(\omega)$, if

$$\int_{-\infty}^{\eta} \mathrm{Prob}\{\xi_1(\omega) \le \eta\} d\eta \le \int_{-\infty}^{\eta} \mathrm{Prob}\{\xi_2(\omega) \le \eta\} d\eta, \quad \forall \eta \in \mathbb{R}. \tag{1}$$

By changing the order of integration in (1), the second order dominance can be mathematically reformulated as:

$$\mathbb{E}_P[(t - \xi_1(\omega))_+] \le \mathbb{E}_P[(t - \xi_2(\omega))_+], \qquad \forall t \in \mathbb{R}. \tag{2}$$

Here and later on $(\cdot)_+$ denotes the plus function, that is, $(x)_+ = \max\{x, 0\}$.

In this paper, we consider the following stochastic programs with stochastic second order

[1]The work of this author is carried out while he is visiting the second author in the School of Mathematics, University of Southampton sponsored by China Scholarship Council.

dominance (SSD) constraints:

$$\begin{array}{ll}\min\limits_{x} & \mathbb{E}_P[f(x,\xi(\omega))] \\ \text{s.t} & G(x,\xi(\omega)) \succeq_2 Y(\xi(\omega)), \\ & x \in X,\end{array} \tag{3}$$

where X is a nonempty compact subset of $\mathbb{R}^n$, $\xi : \Omega \to \Xi$ is a vector of random variables defined on probability $(\Omega, \mathcal{F}, P)$ with support set $\Xi \subset \mathbb{R}^m$, $f, G : \mathbb{R}^n \times \Xi \to \mathbb{R}$ are Lipschitz continuous functions and for every $\xi \in \Xi$, $G(\cdot, \xi) : \mathbb{R}^n \to \mathbb{R}$ is concave; $Y(\xi(\omega))$ is a random variable, and $\mathbb{E}_P[\cdot]$ denotes the expected value with respect to the probability (P) distribution of ξ. For the simplicity of discussion, we make a blanket assumption that f and G are P-integrable.

Using the equivalent formulation of the second order dominance constraint (2), problem (3) can be written as a stochastic semi-infinite programming (SSIP) problem:

$$\begin{array}{ll}\min\limits_{x} & \mathbb{E}_P[f(x,\xi(\omega))] \\ \text{s.t} & \mathbb{E}_P[(t - G(x,\xi(\omega)))_+] \leq \mathbb{E}_P[(t - Y(\xi(\omega)))_+], \quad \forall t \in \mathbb{R}, \\ & x \in X.\end{array} \tag{4}$$

Stochastic optimization models with SSD constraints were introduced by Dentcheva and Ruszczyński [8,9]. Over the past few years, there has been increasing discussions on the subject ranging from optimization theory, numerical methods and practical applications, see [7–12, 17, 20] and references therein.

It is well-known that the SSIP problem above does not satisfy the well-known Slater's constraint qualification, a condition that a stable numerical method may rely on. Subsequently, a so-called relaxed form of the SSIP is proposed:

$$\begin{array}{ll}\min\limits_{x} & \mathbb{E}_P[f(x,\xi(\omega))] \\ \text{s.t.} & \mathbb{E}_P[H(x,t,\xi(\omega))] \leq 0, \quad \forall t \in T, \\ & x \in X,\end{array} \tag{5}$$

where

$$H(x,t,\xi(\omega)) := (t - G(x,\xi(\omega)))_+ - (t - Y(\xi(\omega)))_+$$

and T is a compact subset of $\mathbb{R}$. In the literature [8–10], T is a closed interval or the union of a finite number of closed intervals in $\mathbb{R}$.

Our focus in this paper is on the stability analysis of problem (5). Specifically, we are concerned with the impact of the changes of probability measure P in the problem on optimal values and optimal solutions. The analysis is inspired by a recent work [7] on the stability and sensitivity analysis of optimization problems with first order stochastic dominance constraints and is in line with the traditional stability analysis in the literature of deterministic nonlinear programming and stochastic programming [14, 15, 18, 19, 25–27, 30, 31].

From practical viewpoint, this kind of stability analysis is motivated by the fact that the underlying probability distribution P is often incompletely known in applied models, and the stability behavior of problem (5) when changing P is important, see [25]. A particularly interesting case is when the probability measure is approximated by empirical probability measure. In such a case, expected value of the underlying functions are approximated through sample averaging. As far as we are concerned, the contribution of this work can be summarized as follows.

- We carry out stability analysis for problem (5). Specifically, we consider the case when the underlying probability measure P is approximated by a set of probability measures under pseudometric. By exploiting an error bound in semi-infinite programming due to Gugat [13], we show under the Slater condition that the feasible solution set mapping is Lipschitz continuous and further that the optimal solution set-mapping is upper semicontinuous, and the optimal value function is Lipschitz-like (calm). Moreover, when the objective function satisfies certain growth condition, we show the quantitative upper semi-continuity property of the optimal set-valued mapping. This complements the existing research [7] which focuses on the stability analysis of optimization problems with first order dominance constraints.

- We consider a special case when the probability measure P is approximated by empirical probability measure (which is also known as sample average approximation (SAA)) and present a detailed analysis on the convergence of optimal solution and stationary point obtained from solving the sample average approximate optimization problems as sample size increases. Specifically, we show the exponential rate of convergence of optimal solution and almost sure convergence of stationary point as sample size increases. SAA is a popular method in stochastic programming, but there seems to be few discussions on SAA for stochastic programs with SSD constraints. The only exception is a recent work by Homem-de-Mello [17] which discusses the cutting plane method for sample average approximated optimization problems with SSD constraints.

- Our convergence analysis is carried out through exact penalization of (5) (see (11)). The penalty formulation may provide a potential numerical framework for solving (5).

The rest of the paper are organized as follows. In section 2, we investigate the stability and sensitivity of the set of optimal solutions and optimal value as probability measure changes. In section 3, we consider a special case when the original probability measure is approximated by a sequence of empirical probability measures.

Throughout this paper, we use the following notation. For vectors $a, b \in \mathbb{R}^n$, a^Tb denotes the scalar product, $\|\cdot\|$ denotes the Euclidean norm of a vector, $\|\cdot\|_\infty$ denotes the maximum norm of continuous functions defined over set T. $d(x,\mathcal{D}) := \inf_{x'\in\mathcal{D}} \|x-x'\|$ denotes the distance from a point x to a set $\mathcal{D}$. For two compact sets $\mathcal{C}$ and $\mathcal{D}$,

$$\mathbb{D}(\mathcal{C},\mathcal{D}) := \sup_{x\in\mathcal{C}} d(x,\mathcal{D})$$

denotes the deviation of $\mathcal{C}$ from $\mathcal{D}$ and $\mathbb{H}(\mathcal{C},\mathcal{D}) := \max\left(\mathbb{D}(\mathcal{C},\mathcal{D}), \mathbb{D}(\mathcal{C},\mathcal{D})\right)$ denotes the Hausdorff distance between $\mathcal{C}$ and $\mathcal{D}$. Moreover, $\mathcal{C}+\mathcal{D}$ denotes the Minkowski addition of the two sets, that is, $\{C+D : C\in\mathcal{C}, D\in\mathcal{D}\}$, $\mathcal{B}(x,\gamma)$ denotes the closed ball with center x and radius γ, $\mathcal{B}$ denotes the closed unit ball in the respective space.

2 Sensitivity analysis

Let $\mathscr{P}(\Omega)$ denote the set of all Borel probability measures. For $Q \in \mathscr{P}(\Omega)$, let $\mathbb{E}_Q[\xi] = \int_\Omega \xi(\omega)dQ(\omega)$ denote the expected values of random variable ξ with respect to the distribution of Q. Assuming Q is close to P under some metric to be defined shortly, we investigate in this section the following optimization problem:

$$\begin{array}{ll} \min\limits_{x} & \mathbb{E}_Q[f(x,\xi(\omega))] \\ \text{s.t.} & \mathbb{E}_Q[H(x,t,\xi(\omega))] \le 0, \quad \forall t \in T, \\ & x \in X. \end{array} \tag{6}$$

which is regarded as a perturbation of (5). Specifically, we study the relationship between the perturbed problem (6) and true problem (5) in terms of optimal values and optimal solutions when Q is close to P.

Let us start by introducing a distance function for the set $\mathscr{P}(\Omega)$, which is appropriate for our problem. Define the set of functions:

$$\mathscr{G} := \{g(\cdot) = H(x,\cdot,t) : x \in X, t \in T\} \cup \{g(\cdot) = f(x,\cdot) : x \in X\}.$$

The distance function for the elements in set $\mathscr{P}(\Omega)$ is defined as:

$$\mathscr{D}(P,Q) := \sup_{g\in\mathscr{G}} \big|\mathbb{E}_P[g] - \mathbb{E}_Q[g]\big|.$$

This type of distance was introduced by Römisch [30, Section 2.2] for the stability analysis of stochastic programming and was called pseudometric. It is well-known that $\mathscr{D}$ is nonnegative, symmetric and satisfies the triangle inequality, see [30, Section 2.1]. Throughout this section, we use the following notation:

$$\begin{array}{rcl} \mathcal{F}(Q) & := & \big\{x \in X : \mathbb{E}_Q[H(x,t,\xi)] \le 0, \forall t \in T\big\}, \\ \vartheta(Q) & := & \inf\big\{\mathbb{E}_Q[f(x,\xi)] : x \in \mathcal{F}(Q)\big\}, \\ S_{opt}(Q) & := & \big\{x \in \mathcal{F}(Q) : \vartheta(Q) = \mathbb{E}_Q[f(x,\xi)]\big\}, \\ \mathscr{P}_{\mathscr{G}}(\Omega) & := & \left\{Q \in \mathscr{P}(\Omega) : -\infty < \inf\limits_{g(\xi)\in\mathscr{G}} \mathbb{E}_Q[g(\xi)] \text{ and } \inf\limits_{g(\xi)\in\mathscr{G}} \mathbb{E}_Q[g(\xi)] < \infty\right\}. \end{array}$$

It is easy to observe that for $P,Q \in \mathscr{P}_{\mathscr{G}}(\Omega)$, $\mathscr{D}(P,Q) < \infty$. Throughout this section, the perturbation probability measure Q in problem (6) is taken from $\mathscr{P}_{\mathscr{G}}(\Omega)$.

In what follows, we discuss the quantitative continuity properties of optimal solution mapping and optimal value function of problem (6). We do so by applying Klatte's earlier result on stability of a parametric nonlinear programming [18, 19], an approach adopted by Dentcheva, Henrion and Ruszczyński for the stability analysis of optimization problems with first order dominance constraints [7]. A key condition in Klatte's stability result is the pseudo-Lipschitz property of the feasible set and we verify it by employing an important result on error bound in semi-infinite programming established by Gugat in [13]. To this end, we need to introduce some definitions and technical results most of which are translated from deterministic semi-definite programming in [13] to our setting.

Definition 2.1 Problem (5) is said to satisfy *weak Slater condition*, if there exist positive numbers α and M such that for any $x \in X$ with $\max\limits_{t\in T}(\mathbb{E}_P[H(x,t,\xi)])_+ \in (0,M)$ there exists a point x^* with $\mathbb{E}_P[H(x^*,t,\xi)] < \max\limits_{t\in T}(\mathbb{E}_P[H(x,t,\xi)])_+$ for all $t \in T$ and

$$\|x - x^*\| \le \alpha \left[\max(\mathbb{E}_P[H(x,t,\xi)])_+ - \max_{t\in T} \mathbb{E}_P[H(x^*,t,\xi)]\right].$$

Definition 2.2 Problem (5) is said to satisfy *strong Slater condition*, if there exists a positive number γ such that for any feasible point x satisfying $\mathbb{E}_P[H(x,t,\xi)] = 0$ for some $t \in T$ there exists a point $z(x)$ with $\mathbb{E}_P[H(z(x),t,\xi)] < 0$ for all $t \in T$ and

$$\|z - z(x)\| \leq \gamma \min_{t\in T} \left(-\mathbb{E}_P[H(z(x),t,\xi)]\right).$$

Definition 2.3 Problem (5) is said to satisfy *Slater condition* if there exist a positive number δ and a point $\bar{x} \in X$ such that

$$\max_{t\in T} \mathbb{E}_P[H(\bar{x},t,\xi)] \leq -\bar{\delta}.$$

Note that the strong Slater condition implies that the weak Slater condition holds for any $M > 0$ and $\alpha = \gamma$, where M is given in Definition 2.1. Moreover, if X is a compact, then the Slater condition implies the strong Slater condition and the positive number γ in Definition 2.2 can be estimated by

$$\gamma =: \sup_{x\in X} \frac{\|x - \bar{x}\|}{\min_{t\in T} -\mathbb{E}_P[H(\bar{x},t,\xi)]}. \tag{7}$$

See [13, Propositions 1 and 2] for more details about the relationships.

Proposition 2.4 *Assume problem (5) satisfies the Slater condition and X is a compact set. Then there exists a positive number ϵ ($\epsilon \geq \bar{\delta}/2$) such that for any $Q \in \mathcal{B}(P,\epsilon)$*

$$\max_{t\in T} \mathbb{E}_Q[H(\bar{x},t,\xi)] \leq -\bar{\delta}/2,$$

where $\bar{x}$ is given as in Definition 2.3, that is, problem (6) satisfies the Slater condition.

Proof. By the definition of pseudometric distance $\mathscr{D}$,

$$\sup_{t\in T} |\mathbb{E}_P[H(x,t,\xi)] - \mathbb{E}_Q[H(x,t,\xi)]| \leq \mathscr{D}(Q,P), \qquad \forall x \in X.$$

Let $Q \in \mathcal{B}(P,\bar{\delta}/2)$. Then

$$\begin{aligned}
\sup_{t\in T} \mathbb{E}_Q[H(\bar{x},t,\xi)] &\leq \sup_{t\in T} \mathbb{E}_P[H(\bar{x},t,\xi)] + \sup_{t\in T} |\mathbb{E}_P[H(\bar{x},t,\xi)] - \mathbb{E}_Q[H(\bar{x},t,\xi)]| \\
&\leq -\bar{\delta} + \bar{\delta}/2 \\
&= -\bar{\delta}/2.
\end{aligned}$$

The proof is complete. ■

By [13, Lemmas 3 and 6] and Proposition 2.4, we can obtain the following uniform error bound, for the feasible set mapping $\mathcal{F}(Q)$ of problem (6).

Lemma 2.5 *Let the conditions of Proposition 2.4 hold. Then there exist positive numbers ϵ and β such that for any $Q \in \mathcal{B}(P,\epsilon)$, the following error bound holds:*

$$d(x,\mathcal{F}(Q)) \leq \beta \big\|(\mathbb{E}_Q[H(x,t,\xi)])_+\big\|_\infty, \ \forall x \in X,$$

where $\mathcal{F}(Q)$ is the feasible set of problem (6).

Proof. Let ϵ be given as in Proposition 2.4. For any fixed $Q \in \mathcal{B}(P,\epsilon)$, we have by [13, Lemmas 3 and 6] that

$$d(x, \mathcal{F}(Q)) \leq \gamma(Q) \big\| (\mathbb{E}_P[H(x,t,\xi)])_+ \big\|_\infty,$$

where

$$\gamma(Q) =: \sup_{x \in X} \frac{\|x - \bar{x}\|}{\min_{t\in T} -\mathbb{E}_Q[H(\bar{x},t,\xi)]},$$

and $\bar{x}$ is given in Definition 2.3. By Proposition 2.4, for $Q \in \mathcal{B}(P,\epsilon)$, $\mathbb{E}_Q[H(\bar{x},t,\xi)] \leq -\bar{\delta}/2$, where $\bar{\delta}$ is given in Definition 2.3. This gives

$$\gamma(Q) \leq \frac{\max_{x',x''\in X} \|x' - x''\|}{\bar{\delta}/2}.$$

The conclusion follows by setting $\beta := \frac{\max_{x',x''\in X}\|x'-x''\|}{\bar{\delta}/2}$ and the boundedness of X. ■

Proposition 2.6 *Assume the conditions of Proposition 2.4. The following assertions hold:*

(i) the solution set $S_{opt}(P)$ is nonempty and compact;

(ii) the graph of the feasible set mapping $\mathcal{F}(\cdot)$ is closed;

(iii) there exists a positive number ϵ such that the feasible set mapping $\mathcal{F}(Q)$ is Lipschitz continuous on $\mathcal{B}(P,\epsilon)$, that is

$$\mathbb{H}\left(\mathcal{F}(Q_1), \mathcal{F}(Q_2)\right) \leq \beta \mathscr{D}(Q_1,Q_2), \qquad \forall Q_1, Q_2 \in \mathcal{B}(P,\epsilon).$$

Proof. Part (i) follows from the Slater condition and compactness of X.

Part (ii). Let $t \in T$ be fixed. It follows by virtue of [30, Propositions 3 and 4] that $\mathbb{E}_Q[H(x,t,\xi)] : X \times (\mathscr{G}, \mathscr{D}) \to \mathbb{R}$ is lower semicontinuous. Let $Q^N \to Q$, $x^N \in \mathcal{F}(Q_N)$ and $x^N \to x^*$. Then

$$\mathbb{E}_Q[H(x^*,t,\xi)] \leq \liminf_{N\to\infty} \mathbb{E}_{Q_N}[H(x^N,t,\xi)] \leq 0, \quad \forall t \in T,$$

which implies that $x^* \in \mathcal{F}(Q)$.

Part (iii). Let ϵ be given by Lemma 2.5 and $Q_1, Q_2 \in \mathcal{B}(P,\epsilon)$. Observe that for any $x \in \mathcal{F}(Q_1)$, $(\mathbb{E}_{Q_1}[H(x,t,\xi)])_+ = 0$, for all $t \in T$. By Lemma 2.5, there exists a positive constant β such that for any $x \in \mathcal{F}(Q_1)$

$$\begin{aligned}
d(x, \mathcal{F}(Q_2)) &\leq \beta \left\| (\mathbb{E}_{Q_2}[H(x,t,\xi)])_+ \right\|_\infty \\
&= \beta \left(\max_{t\in T} (\mathbb{E}_{Q_2}[H(x,t,\xi)])_+ - \max_{t\in T} (\mathbb{E}_{Q_1}[H(x,t,\xi)])_+ \right) \\
&\leq \beta \max_{t\in T} \left((\mathbb{E}_{Q_2}[H(x,t,\xi)])_+ - (\mathbb{E}_{Q_1}[H(x,t,\xi)])_+ \right) \\
&\leq \beta \max_{t\in T} \left| \mathbb{E}_{Q_2}[H(x,t,\xi)] - \mathbb{E}_{Q_1}[H(x,t,\xi)] \right| \\
&\leq \beta \max_{(x,t)\in X\times T} \left| \mathbb{E}_{Q_2}[H(x,t,\xi)] - \mathbb{E}_{Q_1}[H(x,t,\xi)] \right| \\
&= \beta \mathscr{D}(Q_1,Q_2),
\end{aligned}$$

which implies $\mathbb{D}(\mathcal{F}(Q_1), \mathcal{F}(Q_2)) \le \beta\mathscr{D}(Q_1, Q_2)$. In the same manner, we can show that for any $x \in \mathcal{F}(Q_2)$,

$$d(x, \mathcal{F}(Q_1)) \le \beta \left(\max_{t\in T} (\mathbb{E}_{Q_1}[H(x,t,\xi)])_+ - \max_{t\in T} (\mathbb{E}_{Q_2}[H(x,t,\xi)])_+ \right) \le \beta\mathscr{D}(Q_2, Q_1),$$

which yields $\mathbb{D}(\mathcal{F}(Q_2), \mathcal{F}(Q_1)) \le \beta\mathscr{D}(Q_1, Q_2)$. Summarizing the discussions above, we have

$$\mathbb{H}\big(\mathcal{F}(Q_1), \mathcal{F}(Q_2)\big) = \max\left\{\mathbb{D}\big(\mathcal{F}(Q_1), \mathcal{F}(Q_2)\big), \mathbb{D}\big(\mathcal{F}(Q_2), \mathcal{F}(Q_1)\big)\right\} \le \beta\mathscr{D}(Q_1, Q_2).$$

The proof is complete. ■

Recall that a set-valued mapping $\Gamma : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ is said to be *upper semi-continuous* at y in the sense of Berge if for any $\epsilon > 0$, there exists a number $\delta > 0$ such that

$$\Gamma(y') \subseteq \Gamma(y) + \epsilon\mathcal{B}, \quad \forall y' \in y + \delta\mathcal{B},$$

where $\mathcal{B}$ denotes the closed unit ball in the respective space. It is said to be *Lipschitz continuous* if there exists a constant L such that

$$\mathbb{H}(\Gamma(y'), \Gamma(y'')) \le L\|y' - y''\|,$$

see [29, page 368]. Proposition 2.6 (iii) shows that the feasible set mapping of problem (6) is Lipschitz continuous with respect to probability measure over $\mathcal{B}(P, \epsilon)$. Using this property, we are ready to establish our main stability results.

Theorem 2.7 *Assume the conditions of Proposition 2.4. Assume also that the Lipschitz modulus of $f(x, \xi)$ w.r.t. x is bounded by an integrable function $\kappa(\xi) > 0$. Then the following stability properties hold true:*

(i) there exists $\epsilon' > 0$ such that the optimal solution set of problem (6), denoted by $S_{opt}(Q)$, is not empty for $\forall Q \in \mathcal{B}(P, \epsilon')$;

(ii) $S_{opt}(\cdot)$ is upper semi-continuous at point P in the sense of Berge;

(iii) there exist positive numbers ϵ^* and L^* such that the optimal value function of problem (6), denoted by $\vartheta(Q)$, is continuous at point P and satisfies the following Lipschitz-like[2] estimation:

$$|\vartheta(Q) - \vartheta(P)| \le L^*\mathscr{D}(Q, P), \qquad \forall Q \in \mathcal{B}(P, \epsilon^*).$$

Proof. Under the Slater condition, it follows from Proposition 2.6 that there exists positive number ϵ such that the feasible set mapping $\mathcal{F}(\cdot)$ is Lipschitz continuous on $\mathcal{B}(P, \epsilon)$. The rest follows straightforwardly from [19, Thoerem 1] ([25, Theorem 2.3] or [7, Thoerem 2.1] in stochastic programming). The proof is complete. ■

Theorem 2.7 asserts that the optimal solution set mapping $S_{opt}(\cdot)$ is nonempty near P and upper semi-continuous at P. In order to quantify this upper semi-continuity property, we need a growth condition on the objective function in a neighborhood of the optimal solution set $S_{opt}(P)$

[2]The property is also known as calmness of ϑ at P, see Section F in Chapter 8 [29] for general discussions on calmness.

to problem (5). Instead of imposing a specific growth condition, here we consider the growth function (see [25, 29]):

$$\Psi(\nu) := \min\{\mathbb{E}_P[f(x,\xi)] - s^* : d(x, S_{opt}(P)) \geq \nu, x \in X\} \tag{8}$$

of problem (5), where s^* denotes the optimal value of problem (5), and the associated function,

$$\widetilde{\Psi}(v) := v + \Psi^{-1}(2v), \quad \kappa \geq 0.$$

We have the following result.

Corollary 2.8 *Let the assumptions of Theorem 2.7 hold. Then there exist positive constants L and ϵ such that*

$$\emptyset \neq S_{opt}(Q) \subseteq S_{opt}(P) + \widetilde{\Psi}\left(L\mathscr{D}(Q,P)\right)\mathcal{B},$$

for any $Q \in \mathcal{B}(P,\epsilon)$, where $\mathcal{B}$ denotes the closed unit ball.

Corollary 2.8 provides a quantitative upper semi-continuity of the set of optimal solutions, see [25, Theorem 2.4] for a detailed proof and [29, Theorem 7.64] for earlier discussions about functions $\Psi(\cdot)$ and $\widetilde{\Psi}(\cdot)$. Discussions on a particular when the growth is of second order can be found in [3, 34].

3 Empirical probability measure

In this section, we consider a special case when the probability measure P is approximated by a sequence of empirical measures P_N defined as

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

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

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

In this case

$$\mathbb{E}_{P_N}[f(x,\xi)] = \frac{1}{N}\sum_{k=1}^{N} f(x,\xi^k)$$

and

$$\mathbb{E}_{P_N}[H(x,t,\xi)] = \frac{1}{N}\sum_{k=1}^{N} H(x,t,\xi^k).$$

It follows from the classical law of large numbers in statistics, $\mathbb{E}_{P_N}[f(x,\xi)]$ and $\mathbb{E}_{P_N}[H(x,t,\xi)]$ converge to $\mathbb{E}_P[f(x,\xi)]$ and $\mathbb{E}_P[H(x,t,\xi)]$ respectively as N increases. This kind of approximation is well-known in stochastic programming under various names such as sample average approximation, Monte Carlo method, sample path optimization, stochastic counterpart etc, see [16, 28, 35, 38] and the references therein.

We use $f_N(x)$ and $H_N(x,t)$ to denote $\mathbb{E}_{P_N}[f(x,\xi)]$ and $\mathbb{E}_{P_N}[H(x,t,\xi)]$ for the simplicity of notation as well as the fact that the latter two quantities have little to do with measures. Consequently we may consider the following approximation of problem (5):

$$\begin{array}{ll} \min\limits_{x} & f_N(x) := \frac{1}{N}\sum_{k=1}^{N} f(x,\xi^k) \\ \text{s.t.} & H_N(x,t) := \frac{1}{N}\sum_{k=1}^{N} H(x,t,\xi^k) \leq 0, \quad \forall t \in T, \\ & x \in X. \end{array} \tag{9}$$

We call (9) the *SAA* problem and (5) the *true* problem.

Assuming that we can obtain an optimal solution, denoted by x^N, by solving the SAA problem, we analyze the convergence of x^N as the sample size increases. The analysis would be very complicated if it was carried out on (9) directly in that the constraints of the SAA problem depend on sampling. To get around the difficulty as well as the infinite number of constraints, we consider a reformulation of both the true and the SAA problem through the exact penalization so that the feasible set of the reformulated problems are deterministic and our analysis focuses on the approximation of the objective functions.

For the simplicity of notation, let

$$h(x,t) := \max\left\{\mathbb{E}_P[H(x,t,\xi)], 0\right\}, \qquad \theta(x) := \max_{t\in T} h(x,t). \tag{10}$$

Consider the exact penalization:

$$\begin{array}{ll} \min\limits_{x} & \psi(x,\rho) := \mathbb{E}_P[f(x,\xi)] + \rho\theta(x) \\ \text{s.t} & x \in X, \end{array} \tag{11}$$

where ρ is a penalty parameter. This kind of penalization is well documented in the literature, see for instance [24, 36]. In what follows, we establish the equivalence between (5) and (11) in the sense of optimal solutions. We do so by exploiting the error bound established in Lemma 2.5 and a well-known result by Clarke [5, Proposition 2.4.3]. We need the following assumptions.

Assumption 3.1 *$f(x,\xi)$ and $G(x,\xi)$ are locally Lipschitz continuous w.r.t. x and their Lipschitz modulus are bounded by an integrable function $\kappa(\xi) > 0$.*

Theorem 3.2 *Assume that the true problem (5) satisfies the Slater condition and X is a compact set. Under Assumption 3.1, there exists a positive number $\bar{\rho}$ such that for any $\rho > \bar{\rho}$, the sets of optimal solutions of problems (5) and (11), denoted by S_{opt} and X_{opt} respectively, coincide.*

Proof. Under the Slater condition, it follows by Lemma 2.5 that there exists a constant $\beta > 0$ such that

$$d(x,\mathcal{F}) \leq \beta\left\|(\mathbb{E}_P[H(x,t,\xi)])_+\right\|_\infty, \quad \forall x \in X.$$

Let C denote the Lipschitz modulus of function $\mathbb{E}_P[f(x,\xi)]$. By [5, Proposition 2.4.3], for any $\rho > \beta C$, the two optimal solution sets, S_{opt} and X_{opt}, coincide. Note that under Assumption 3.1, we can set an $C = \mathbb{E}_P[\kappa(\xi)]$. This shows the existence of a positive constant $\bar{\rho} := \beta C$. The proof is complete. ■

We now move on to discuss the exact penalization of the SAA problem (9). Let

$$h_N(x,t) := \max\big\{H_N(x,t), 0\big\}, \qquad\qquad \theta_N(x) := \max_{t\in T} h_N(x,t). \tag{12}$$

Consider the SAA penalty problem

$$\begin{array}{ll} \min\limits_{x} & \psi_N(x,\rho_N) := f_N(x) + \rho_N\theta_N(x) \\ \text{s.t} & x \in X, \end{array} \tag{13}$$

where ρ_N is the penalty parameter.

Under Assumption 3.1, we have by [32, Section 6, Proposition 7] that $H_N(x,t)$ converges to $\mathbb{E}_P[H(x,t,\xi)]$ uniformly over compact set $X\times T$ with probability one (w.p.1). Since true problem (5) satisfies Slater condition, there exists a sufficiently large N^* such that for any $N \geq N^*$

$$H_N(\bar{x},t) \leq -\bar{\delta}/2, \;\; \forall t \in T, \quad \text{w.p.1},$$

where $\bar{x}$ and $\bar{\delta}$ are given in Definition 2.3. Subsequently, by Lemma 2.5 that, for any $N \geq N^*$,

$$d(x,\mathcal{F}_N) \leq \beta\big\|(H_N(x,t))_+\big\|_\infty, \quad \forall x \in X, \tag{14}$$

w.p.1, where $\mathcal{F}_N$ denotes the feasible set of problem (9).

Proposition 3.3 *Let the assumptions in Theorem 3.2 hold. Then there exist positive numbers ρ^* and N^* such that for any $\rho > \rho^*$ and $N \geq N^*$, the sets of optimal solutions of problems (9) and (13), denoted by S^N_{opt} and X^N_{opt} respectively, coincide w.p.1.*

Proof. Note that there exist a positive constant β and a sufficiently large positive integer N_1 such that for any $N \geq N_1$, (14) holds. Let C_N denote the Lipschitz modulus of function $f_N(x)$. By [5, Proposition 2.4.3], for any $\rho > \beta C_N$, the two optimal solution sets, S^N_{opt} and X^N_{opt}, coincide. Moreover, under Assumption 3.1, C_N converges to the Lipschitz modulus of $\mathbb{E}[f(x,\xi)]$, denoted by C. This implies that there exists a positive integer $N_2 \geq N_1$ such that when $N \geq N_2$, we have $C_N < 2C$. The conclusion follows by taking $\rho^* = 2\beta C$ and $N^* = \max\{N_1, N_2\}$. ■

3.1 Optimal solution

Assuming for every fixed sampling, we can obtain an optimal solution, denoted by x^N, from solving the SAA problem (9), we analyze the convergence of x^N as the sample size N increases. We do so by establishing the uniform convergence (both almost sure and exponential) of the objective function of SAA penalty problem (13) to its true counterpart (11). Over the past few decades, there have been a lot of discussions this kind of analysis in stochastic programming. However, as far as we are concerned, our analysis seems to be the first for the SAA applied to (5) and it is carried out through exact penalization.

Proposition 3.4 *Let X be a compact set and ρ_N tend to ρ as $N \to \infty$. Let Assumption 3.1 hold. Then*

(i) $\psi(x,\rho)$ and $\psi_N(x,\rho_N)$, $N = 1, 2, \cdots$, are Lipschitz continuous;

(ii) $\psi_N(x,\rho_N)$ converges to $\psi(x,\rho)$ uniformly over X.

Proof. Part (i). Note that $\mathbb{E}_P[H(x,t,\xi)]$ and $H_N(x,t)$ are Lipschitz continuous with respect to (x,t) and T is a compact set. By [23, Theorem 3.1], $\theta(x)$ and $\theta_N(x)$ are Lipschitz continuous. Together with the Lipschitz continuity of $\mathbb{E}_P[f(x,\xi)]$ and $f_N(x)$, we conclude that $\psi(x,\rho)$ and $\psi_N(x,\rho_N)$ are Lipschitz continuous.

Part (ii). By Assumption 3.1 and the compactness of X, it is not difficult to show that $f(x,\xi)$ and $H(x,t,\xi)$ are dominated by an integrable function. The uniform convergence of $f_N(x)$ to $\mathbb{E}_P[f(x,\xi)]$ and $H_N(x,t)$ to $\mathbb{E}_P[H(x,t,\xi)]$ follows from classical uniform law of large numbers for random functions, see e.g. [32, Section 6, Proposition 7]. Since $\rho_N \to \rho$, it suffices to show the uniformly convergence of $\theta_N(x)$ to $\theta(x)$. By definition,

$$\begin{aligned}
\max_{x\in X}|\theta_N(x)-\theta(x)| &= \max_{x\in X}\left|\max_{t\in T}(\max\{H_N(x,t),0\}) - \max_{t\in T}(\max\{\mathbb{E}_P[H(x,t,\xi)],0\})\right| \\
&\le \max_{(x,t)\in X\times T}\big|\max\{H_N(x,t),0\} - \max\{\mathbb{E}_P[H(x,t,\xi)],0\}\big| \\
&\le \max_{(x,t)\in X\times T}\big|H_N(x,t) - \mathbb{E}_P[H(x,t,\xi)]\big|. \qquad (15)
\end{aligned}$$

This along with the uniform convergence of $H_N(x,t)$ to $\mathbb{E}_P[H(x,t,\xi)]$ over $X\times T$ gives rise to the assertion. The proof is complete. ■

Assumption 3.5 *Let $f(x,\xi)$ and $H(x,t,\xi)$ be defined as in (5). The following hold.*

(a) for every $x\in X$, the moment generating function

$$M_x(\tau) := \mathbb{E}_P\left[e^{\tau(f(x,\xi)-\mathbb{E}_P[f(x,\xi)])}\right]$$

of random variable $f(x,\xi)-\mathbb{E}_P[f(x,\xi)]$ is finite valued for all τ in a neighborhood of zero;

(b) for every $(x,t)\in X\times T$, the moment generating function

$$M_{(x,t)}(\tau) := \mathbb{E}_P\left[e^{\tau(H(x,t,\xi)-\mathbb{E}_P[H(x,t,\xi)])}\right]$$

of random variable $H(x,t,\xi)-\mathbb{E}_P[H(x,t,\xi)]$ is finite valued for all τ in a neighborhood of zero;

(c) let $\kappa(\xi)$ be given as in Assumption 3.1. The moment generating function $M_\kappa(\tau)$ of $\kappa(\xi)$ is finite valued for all τ in a neighborhood of zero.

Assumption 3.5 (a) means that the random variables $f(x,\xi)-\mathbb{E}_P[f(x,\xi)]$ and $H(x,t,\xi)-\mathbb{E}_P[H(x,t,\xi)]$ do not have a heavy tail distribution. In particular, it holds if the random variable ξ has a bounded support set. Assumption 3.5 (c) is satisfied if $\mathbb{E}_P[\kappa(\xi)]$ is finite. Note that under Assumption 3.1, the Lipschitz modulus of $H(x,t,\xi)$ is bounded by $1+\kappa(\xi)$. Assumption 3.5 (c) implies that the moment generating function of $1+\kappa(\xi)$ is finite valued for τ close to zero.

Proposition 3.6 *Let Assumptions 3.1 and 3.5 hold. Assume that X is a compact set and $\rho_N \to \rho$. Then $\psi_N(x, \rho_N)$ converges to $\psi(x, \rho)$ with probability one at an exponential rate, that is, for any $\alpha > 0$, there exist positive constants $C(\alpha)$, $K(\alpha)$ and independent of N, such that*

$$\text{Prob}\left\{\sup_{x\in X} |\psi_N(x, \rho_N) - \psi(x, \rho)| \geq \alpha\right\} \leq C(\alpha)e^{-NK(\alpha)}$$

for N sufficiently large.

Proof. By definition

$$\begin{aligned}
&\text{Prob}\left\{\sup_{x\in X} |\psi_N(x, \rho_N) - \psi(x, \rho)| \geq \alpha\right\} \\
&= \text{Prob}\left\{\sup_{x\in X} |f_N(x) + \rho_N\theta_N(x) - (\mathbb{E}_P[f(x,\xi)] + \rho\theta(x))| \geq \alpha\right\} \\
&\leq \text{Prob}\left\{\sup_{x\in X} |f_N(x) - \mathbb{E}_P[f(x,\xi)]| \geq \alpha/2\right\} + \text{Prob}\left\{\sup_{x\in X} |\rho_N\theta_N(x) - \rho\theta(x)| \geq \alpha/2\right\}.
\end{aligned}$$

Under Assumption 3.5, it follows from [35, Theorem 5.1] that the first term at the right hand of the inequality above converges to zero at an exponential rate. In the same manner, we can obtain uniform exponential convergence of $H_N(x, t)$ to $\mathbb{E}_P[H(x, t, \xi)]$ and hence $\theta_N(x)$ to $\theta(x)$ taking into account that $\rho_N \to \rho$. The proof is complete. ■

Remark 3.7 Similar to the discussions in [35], we may estimate the sample size. To this end, let us strengthen the conditions in Assumption 3.5 (a) and (b) to the following:

- There exists a constant $\varrho > 0$ such that for every $x \in X$,

$$\mathbb{E}_P\left[e^{\tau(f(x,\xi)-\mathbb{E}_P[f(x,\xi)])}\right] \leq e^{\varrho^2\tau^2/2}, \forall \tau \in \mathbb{R} \tag{16}$$

  and for every $(x, t) \in X \times T$,

$$\mathbb{E}_P\left[e^{\tau(H(x,t,\xi)-\mathbb{E}_P[H(x,t,\xi)])}\right] \leq e^{\varrho^2\tau^2/2}, \forall \tau \in \mathbb{R}. \tag{17}$$

Note that equality in (16) and (17) holds if random variables $f(x, \xi) - \mathbb{E}_P[f(x, \xi)]$ and $H(x, t, \xi) - \mathbb{E}_P[H(x, t, \xi)]$ satisfy normal distribution with variance ϱ^2, see a discussion in [35, page 410]. Let α_1 be a small positive number and $\beta_1 \in (0, 1)$. It follows from (5.14) and (5.15) in [35] that for

$$N \geq N_1(\alpha_1, \beta_1) := \frac{O(1)\varrho^2}{\alpha_1^2}\left[n \log\left(\frac{O(1)D_1\mathbb{E}_P[\kappa_1(\xi)]}{\alpha_1}\right) + \log\left(\frac{1}{\beta_1}\right)\right], \tag{18}$$

we have that

$$\text{Prob}\left\{\sup_{x\in X} |f_N(x) - \mathbb{E}_P[f(x,\xi)]| \geq \alpha_1\right\} \leq \beta_1, \tag{19}$$

where $\kappa_1(\xi)$ is the global Lipschitz modulus of $f(\cdot, \xi)$ over X, $D_1 := \sup_{x',x''\in X} \|x' - x''\|$. Likewise, for given positive numbers α_2 and $\beta_2 \in (0, 1)$, when

$$N \geq N_2(\alpha_2, \beta_2) := \frac{O(1)\varrho^2}{\alpha_2^2}\left[n \log\left(\frac{O(1)D_2\mathbb{E}_P[\kappa_2(\xi)]}{\alpha_2}\right) + \log\left(\frac{1}{\beta_2}\right)\right], \tag{20}$$

we have

$$\text{Prob}\left\{\max_{(x,t)\in X\times T}\left|H_N(x,t)-\mathbb{E}_P[H(x,t,\xi)]\right|\geq\alpha_2\right\}\leq\beta_2. \tag{21}$$

where $\kappa_2(\xi)$ is the global Lipschitz modulus of $H(\cdot,\cdot,\xi)$ over $X\times T$,

$$D_2:=\sup_{w',w\in X\times T}\|w'-w\|\leq D_1+\sup_{t',t''\in T}\|t'-t''\|.$$

Let $\alpha>0$ be a positive number and $\beta\in(0,1)$. Observe that

$$\begin{aligned}\text{Prob}\left\{\max_{x\in X}\left|\psi_N(x,\rho_N)-\psi(x,\rho)\right|\geq\alpha\right\} &\leq \text{Prob}\left\{\sup_{x\in X}|f_N(x)-\mathbb{E}_P[f(x,\xi)]|\geq\alpha/2\right\}\\ &\quad+\text{Prob}\left\{\sup_{x\in X}|\rho_N\theta_N(x)-\rho\theta(x)|\geq\alpha/2\right\}. \end{aligned}\tag{22}$$

Let N_3 be sufficiently large such that $\rho_N\leq 2\rho$ and

$$(\rho_N-\rho)\sup_{x\in X}|\theta(x)|\leq\frac{\alpha}{4}.$$

Then it is easy to verify that for $N\geq N_3$

$$\begin{aligned}\text{Prob}\left\{\sup_{x\in X}|\rho_N\theta_N(x)-\rho\theta(x)|\geq\alpha/2\right\} &\leq \text{Prob}\left\{\sup_{x\in X}|\theta_N(x)-\theta(x)|\geq\frac{\alpha}{8\rho}\right\}\\ &\leq \text{Prob}\left\{\max_{(x,t)\in X\times T}\left|H_N(x,t)-\mathbb{E}_P[H(x,t,\xi)]\right|\geq\frac{\alpha}{8\rho}\right\}. \end{aligned}\tag{23}$$

The last inequality is due to (15). Let

$$N(\alpha,\beta):=\max\left\{N_1\left(\frac{\alpha}{2},\beta_1\right),N_2\left(\frac{\alpha}{8\rho},\beta_2\right),N_3\right\}, \tag{24}$$

where $\beta_1,\beta_2\in(0,1)$ and $\beta_1+\beta_2=\beta$. Combining (19), (21), (22) and (23), we have for $N\geq N(\alpha,\beta)$

$$\begin{aligned}\text{Prob}\left\{\max_{x\in X}\left|\psi_N(x,\rho_N)-\psi(x,\rho)\right|\geq\alpha\right\} &\leq \text{Prob}\left\{\sup_{x\in X}|f_N(x)-\mathbb{E}_P[f(x,\xi)]|\geq\alpha/2\right\}\\ &\quad+\text{Prob}\left\{\max_{(x,t)\in X\times T}\left|H_N(x,t)-\mathbb{E}_P[H(x,t,\xi)]\right|\geq\frac{\alpha}{8\rho}\right\}\\ &\leq \beta_1+\beta_2\\ &= \beta.\end{aligned}$$

The discussion above shows that for given α and β, we can obtain sample size $N(\alpha,\beta)$ such that when $N\geq N(\alpha,\beta)$

$$\text{Prob}\left\{\max_{x\in X}\left|\psi_N(x,\rho_N)-\psi(x,\rho)\right|\geq\alpha\right\}\leq\beta.$$

In what follows, we translate the uniform exponential convergence established in Proposition 3.6 into that of optimal solutions. We need the following intermediate sensitivity result.

Lemma 3.8 *Consider a general constrained minimization problem*

$$\begin{array}{ll} \min & \phi(x) \\ \text{s.t.} & x \in X, \end{array} \tag{25}$$

where $\phi : \mathbb{R}^m \to \mathbb{R}$ *is continuous and* $X \subseteq \mathbb{R}^m$ *is closed, and a perturbed program*

$$\begin{array}{ll} \min & \varphi(x) \\ \text{s.t.} & x \in X, \end{array} \tag{26}$$

where $\varphi : \mathbb{R}^m \to \mathbb{R}$ *is continuous. Let* X^*_ϕ *denote the set of optimal solutions to (25) and* X^*_φ *the set of optimal solutions to (26). Then*

(i) for any $\epsilon > 0$*, there exists a* $\delta > 0$ *(depending on* ϵ*) such that*

$$\mathbb{D}(X^*_\varphi, X^*_\phi) \le \epsilon, \tag{27}$$

when

$$\sup_{x\in X} |\varphi(x) - \phi(x)| \le \delta;$$

(ii) if, in addition,

$$\phi(x) \ge \min_{x\in X} \phi(x) + \varsigma d(x, X^*_\phi)^2, \ \forall x \in X, \tag{28}$$

then

$$\mathbb{D}(X^*_\psi, X^*_\phi) \le \sqrt{\frac{3}{\varsigma} \sup_{x\in X} |\varphi(x) - \phi(x)|}. \tag{29}$$

Proof. The results are minor extension of [6, Lemma 3.2] which deals with the case when X^*_ϕ is a singleton and are also similar to [29, Theorem 7.64]. Here we provide a proof for completeness.

Part (i). Let ϵ be a fixed small positive number and ϕ^* the optimal value of (25). Define

$$R(\epsilon) := \inf_{\{x\in X, d(x,X^*_\phi)\ge\epsilon\}} \phi(x) - \phi^*. \tag{30}$$

Then $R(\epsilon) > 0$. Let $\delta := R(\epsilon)/3$ and φ be such that $\sup_{x\in X} |\varphi(x) - \phi(x)| \le \delta$. Then for any $x \in X$ with $d(x, X^*_\phi) \ge \epsilon$ and any fixed $x^* \in X^*_\phi$,

$$\varphi(x) - \varphi(x^*) \ge \phi(x) - \phi(x^*) - 2\delta \ge R(\epsilon)/3 > 0,$$

which implies that x is not an optimal solution to (26). This is equivalent to $d(x, X^*_\phi) < \epsilon$ for all $x \in X^*_\varphi$, that is, $\mathbb{D}(X^*_\varphi, X^*_\phi) \le \epsilon$.

Part (ii). Under condition (28), it is easy to derive that $R(\epsilon) = \varsigma\epsilon^2$. Let

$$\epsilon := \sqrt{\frac{3}{\varsigma} \sup_{x\in X} |\varphi(x) - \phi(x)|}.$$

From Part (i), we immediately arrive at (29). The proof is complete. ■

Remark 3.9 We have a few comments on Lemma 3.8.

(i) Condition (28) is known as second order growth condition. Using this condition, Shapiro [33] developed a variational principal which gives a bound for $d(x, X_\phi^*)$ in terms of the maximum Lipschitz constant of $\varphi - \phi$ over X, see [33, Lemma 4.1] and [34, Proposition 2.1]. Both the second order growth condition and the variational principal have been widely used for the stability and asymptotic analysis in stochastic programming, see [3,33,34]. Our claim in Lemma 3.8 (ii) strengthens the variational principal in that our bound for $d(x, X_\phi^*)$ is $\sqrt{\frac{3}{\varsigma}\sup_{x\in X}|\varphi(x)-\phi(x)|}$ which tends to zero when the maximum Lipschitz constant of $\varphi(x) - \phi(x)$ over X goes to zero and $\varphi(x_0) - \phi(x_0) = 0$ at some point $x_0 \in X$ but conversely this is not necessarily true.

(ii) Lemma 3.8 (ii) may be extended to a general case when $R(\epsilon)$ is monotonically increasing on $\mathbb{R}_+$. In such a case, we may set

$$\epsilon := R^{-1}(3\sup_{x\in X}|\varphi(x)-\phi(x)|)$$

and obtain from Lemma 3.8 (i) that

$$\mathbb{D}(X_\varphi^*, X_\phi^*) \le R^{-1}(3\sup_{x\in X}|\varphi(x)-\phi(x)|).$$

Theorem 3.10 *Assume that problem (5) satisfies the Slater condition and X is a compact set. Let $\{\rho_N\}$ be a sequence of positive numbers such that $\rho_N \to \rho$, where ρ is given in Theorem 3.2. Then*

(i) w.p.1

$$\lim_{N\to\infty} \mathbb{D}\left(X_{opt}^N, X_{opt}\right) = 0, \tag{31}$$

where X_{opt} and X_{opt}^N are the sets of optimal solutions of problem (11) and (13) respectively. Moreover, if Assumption 3.5 holds, the convergence rate is exponential, that is, for any $\alpha > 0$, there exist positive constants $C_1(\alpha)$, $K_1(\alpha)$ and independent of N, such that

$$\mathrm{Prob}\left\{\mathbb{D}\left(X_{opt}^N, X_{opt}\right) \ge \alpha\right\} \le C_1(\alpha)e^{-NK_1(\alpha)}$$

for N sufficiently large.

(ii) If the objective function of the true penalty problem (11) satisfies the second order growth condition:

$$\psi(x,\rho) \ge \min_{x\in X}\psi(x,\rho) + \varsigma d(x, X_{opt})^2,\ \forall x \in X, \tag{32}$$

then the $C_1(\alpha) = C\left(\frac{1}{3}\varsigma\alpha^2\right)$ and $K_1(\alpha) = K\left(\frac{1}{3}\varsigma\alpha^2\right)$ where $C(\alpha)$ and $K(\alpha)$ are given in Proposition 3.6.

(iii) Let $N(\alpha,\beta)$ be defined as in (24). For $N \ge N\left(\frac{1}{3}\varsigma\alpha^2, \beta\right)$, we have

$$\mathrm{Prob}\left\{\mathbb{D}\left(X_{opt}^N, X_{opt}\right) \ge \alpha\right\} \le \beta,$$

where $\beta \in (0,1)$.

Proof. The almost assure convergence follows straightforwardly from Proposition 3.4 that $\psi_N(x,\rho_N)$ converges to $\psi(x,\rho)$ uniformly over X and Lemma 3.8. In the next, we show the exponential convergence. By Lemma 3.8, for any $\alpha > 0$, there exists $\varepsilon(\alpha)$ such that if

$$\sup_{x\in X} |\psi_N(x,\rho_N) - \psi(x,\rho)| \le \varepsilon(\alpha),$$

then $\mathbb{D}\left(X_{opt}^N, X_{opt}\right) \le \alpha$. Subsequently,

$$\text{Prob}\left\{\mathbb{D}\left(X_{opt}^N, X_{opt}\right) \ge \alpha\right\} \le \text{Prob}\left\{\sup_{x\in X} |\psi_N(x,\rho_N) - \psi(x,\rho)| \ge \varepsilon(\alpha)\right\}.$$

By Proposition 3.6 and the formula above there exist positive constants $C_1(\alpha)$ and $K_1(\alpha)$, independent of N such that

$$\text{Prob}\left\{\mathbb{D}\left(X_{opt}^N, X_{opt}\right) \ge \alpha\right\} \le C_1(\alpha)e^{-NK_1(\alpha)},$$

for N sufficiently large.

Part (ii). Under the second growth condition, it is easy to derive that $R(\epsilon) = \varsigma\epsilon^2$, where $R(\epsilon)$ is given in Lemma 3.8. By (29) in Lemma 3.8 (ii),

$$\begin{aligned}\text{Prob}\left\{\mathbb{D}\left(X_{opt}^N, X_{opt}\right) \ge \alpha\right\} &\le \left\{\sqrt{\frac{3}{\varsigma}\sup_{x\in X} |\psi_N(x,\rho_N) - \psi(x,\rho)|} \ge \alpha\right\} \\ &= \left\{\sup_{x\in X} |\psi_N(x,\rho_N) - \psi(x,\rho)| \ge \frac{1}{3}\varsigma\alpha^2\right\}.\end{aligned}$$

The rest follows from Part (i).

Part (iii) follows from (24) and Part (ii). The proof is complete. ■

Let us make some comments on the second order growth condition (32). Since $G(\cdot,\xi)$ is assumed to be concave, it is easy to verify that $\theta(x)$ is a convex function. If $f(\cdot,\xi)$ is convex for almost every ξ, then $\psi(\cdot,\rho)$ is convex. The second order growth condition is fulfilled if the latter happens to be strongly convex.

3.2 Stationary point

We now move on to investigate the case when we only obtain a stationary point rather than an optimal solution from solving the sample average approximate penalty problem (13). This is motivated to address the case when $f(x,\xi)$ is not convex in x. Convergence analysis of SAA stationary sequence has been well documented, see [38] and the references therein. Our analysis here differs from those in the literature on twofold: (a) We analyze the convergence of SAA stationary point to its true counterpart rather than so-called weak stationary point of the true problem [38], the analysis is based on the uniform convergence of the subdifferential of the sample average random functions, Lemma 4.1, which is established recently in [21] rather than sample average of the subdifferential of random functions as opposed to the weak case, and also the convergence result is stronger. Note that this kind of subdifferential approximation can be traced back to the earlier work by Birge and Qi [2] and Artstein and Wets [1]. (b) We provide an effective approach to tackle the specific challenges and complications arising from the second order dominance constraints.

We start by defining the stationary points of (11) and (13). Let $h(x,t) = (\mathbb{E}_P[H(x,t,\xi)])_+$ be defined as in (10). For any fixed $x \in X$, let $T^*(x)$ denote the set of $\bar{t} \in T$ such that $h(x,\bar{t}) = \max_{t\in T} h(x,t)$. Since $G(\cdot,\xi)$ is concave, then $\mathbb{E}_P[H(x,t,\xi)]$ is convex in x and hence it is Clarke regular (see [5, Proposition 2.3.6]). By [5, Proposition 2.3.12]

$$\partial_x h(x,t) = \begin{cases} 0, & \mathbb{E}_P[H(x,t,\xi)] < 0, \\ \mathrm{conv}\{0, \partial_x \mathbb{E}_P[H(x,t,\xi)]\}, & \mathbb{E}_P[H(x,t,\xi)] = 0, \\ \partial_x \mathbb{E}_P[H(x,t,\xi)], & \mathbb{E}_P[H(x,t,\xi)] > 0. \end{cases} \tag{33}$$

Here and later on "conv" denotes the convex hull of a set. By Levin-Valadier Theorem (see [32, Section 2, Theorem 51]),

$$\partial\theta(x) = \mathrm{conv}\left\{ \bigcup_{t\in T^*(x)} \partial_x h(x,t) \right\}. \tag{34}$$

A point $x \in X$ is said to be a *stationary point* of the penalty problem (11) if

$$0 \in \partial_x \psi(x,\rho) + \mathcal{N}_X(x),$$

where

$$\partial_x \psi(x,\rho) := \partial \mathbb{E}_P[f(x,\xi)] + \rho \partial\theta(x)$$

and $\mathcal{N}_X(x)$ denotes the Clarke normal cone to X at x, that is, for $x \in X$,

$$\mathcal{N}_X(x) = \left\{ \zeta \in \mathbb{R}^n \, : \, \zeta^T \eta \le 0, \forall \eta \in \mathcal{T}_X(x) \right\},$$

where

$$\mathcal{T}_X(x) = \liminf_{t\to 0,\, X \ni x' \to x} \frac{1}{t}(X - x')$$

and $\mathcal{N}_X(x) = \emptyset$ when $x \notin X$.

Likewise, for any fixed $x \in X$, let $T^N(x)$ denote the set of $\bar{t} \in T$ such that $h_N(x,\bar{t}) = \max_{t\in T} h_N(x,t)$. Then

$$\partial_x h_N(x,t) = \begin{cases} 0, & H_N(x,t) < 0, \\ \mathrm{conv}\{0, \partial_x H_N(x,t)\}, & H_N(x,t) = 0, \\ \partial_x H_N(x,t), & H_N(x,t) > 0 \end{cases} \tag{35}$$

and

$$\partial\theta_N(x) = \mathrm{conv}\left\{ \bigcup_{t\in T^N(x)} \partial_x h_N(x,t) \right\}. \tag{36}$$

A point $x \in X$ is said to be a *stationary point* of the penalized SAA problem (13) if

$$0 \in \partial_x \psi_N(x,\rho_N) + \mathcal{N}_X(x),$$

where

$$\partial_x \psi_N(x,\rho_N) := \partial f_N(x) + \rho_N \partial\theta_N(x).$$

In order to analyze the convergence of stationary points, we need the following condition.

Assumption 3.11 *Let $f(x,\xi)$ and $G(x,\xi)$ are locally Lipschitz continuous w.r.t. x, and their Lipschitz modulus are bounded by a positive constant C.*

It is easy to observe that Assumption 3.11 is stronger than Assumption 3.1. Over the past few years, there has been extensive discussions on the convergence of SAA stationary points to the so-called weak stationary points of the true problem which is defined through the expected value of the subdifferential of the underlying functions of the true problem in the first order optimality condition, see [38] for the recent discussion. A stationary point is a weak stationary point but not vice versa. Analysis of the convergence of SAA stationary point to a weak stationary point of the true problem can be achieved under Assumption 3.1 but convergence to a stationary point of the true problem requires Assumption 3.11.

Proposition 3.12 *Let $\theta(x)$ and $\theta_N(x)$ be defined as in (10) and (12) respectively. Let $\{x^N\} \subset X$ be a sequence which converges to x^*. Under Assumption 3.11*

$$\lim_{N\to\infty} \mathbb{D}(\partial\theta_N(x^N), \partial\theta(x^*)) = 0 \tag{37}$$

w.p.1.

Proof. Since for any $\xi \in \Xi$, $G(\cdot,\xi)$ is concave function, then $H(x,t,\xi)$ is a convex function with respect to (x,t) over $X \times T$ and so are $h_N(x,t)$ and $h(x,t)$. Using the calculus of the subdifferentials of $\theta(x)$ and $\theta_N(x)$ obtained in (34) and (36), it suffices to show that

$$\lim_{N\to\infty} \mathbb{D}\left(\text{conv}\left\{\bigcup_{t\in T^N(x^N)} \partial_x h_N(x^N,t)\right\}, \text{conv}\left\{\bigcup_{t\in T^*(x^*)} \partial_x h(x^*,t)\right\}\right) = 0,$$

which is implied by

$$\lim_{N\to\infty} \mathbb{D}\left(\bigcup_{t\in T^N(x^N)} \partial_x h_N(x^N,t), \bigcup_{t\in T^*(x^*)} \partial_x h(x^*,t)\right) = 0.$$

By the calculus of $\partial h(x)$ and $\partial h_N(x)$ in (33) and (35), this is further implied by the following:

$$\lim_{N\to\infty} \mathbb{D}\left(\bigcup_{t\in T^N(x^N)} \partial_x H_N(x^N,t), \bigcup_{t\in T^*(x^*)} \partial_x \mathbb{E}_P[H(x^*,t,\xi)]\right) = 0, \quad \text{w.p.1.} \tag{38}$$

Since H and H_N are convex function and Lipschitz continuous with respect to (x,t), we have by [5, Propositions 2.3.15 and 2.3.16]

$$\partial_x \mathbb{E}_P[H(x,t,\xi)] = \pi_x \partial \mathbb{E}_P[H(x,t,\xi)], \qquad \partial_x H_N(x,t) = \pi_x \partial H_N(x,t), \tag{39}$$

where π_x denotes the projection of set-valued mapping on x-axis.

Let $\eta^N \in \bigcup_{t\in T^N(x^N)} \partial_x H_N(x^N,t)$. Then there exists $t_N \in T^N(x^N)$ such that $\eta^N \in \partial_x H_N(x^N,t_N)$. By taking a subsequence if necessarily, we assume for the simplicity of notation that

$$\lim_{N\to\infty} \eta^N = \eta^*, \qquad \lim_{N\to\infty} (x^N,t_N) = (x^*,t^*).$$

By Lemma 4.1 in the appendix, $\partial H_N(x,t)$ converges to $\partial \mathbb{E}_P[H(x,t,\xi)]$ uniformly over compact set $X \times T$, which implies

$$\lim_{N\to\infty} \mathbb{D}\left(\partial H_N(x^N,t_N), \partial \mathbb{E}_P[H(x^*,t^*,\xi)]\right) = 0, \quad \text{w.p.1.} \tag{40}$$

On the other hand, by [32, Section 6, Proposition 7] $H_N(x,t,\xi)$ converges to $\mathbb{E}_P[H(x,t,\xi)]$ uniformly over $X \times T$. Then for any fixed $x \in X$, $h_N(x,t)$ converges to $h(x,t)$ uniformly over compact set T. By Lemma 3.8,

$$\lim_{N\to\infty} \mathbb{D}\left(T^N(x^N), T^*(x^*)\right) = 0.$$

This along with (39) and (40) yield $\eta^* \in \left\{\bigcup_{t\in T^*(x^*)} \partial_x \mathbb{E}_P[H(x^*,t,\xi)]\right\}$. Since η^N is taken arbitrarily from $\left\{\bigcup_{t\in T^N(x^N)} \partial_x H_N(x^N,t)\right\}$, the discussion above shows (38) as desired. The proof is complete. ■

Theorem 3.13 *Let $\{x^N\}$ be a sequence of KKT points of problem (13) and x^* be an accumulation point. Suppose Assumption 3.11 holds. If $\rho_N \to \rho$, then w.p.1 x^* is an stationary point of the true penalty problem (11).*

Proof. By taking a subsequence if necessarily we assume for the simplicity of notation that x^N converges to x^*. Observe first that for any compact sets $A, B, C, D \subseteq \mathbb{R}^m$,

$$\mathbb{D}(A+C, B+D) \le \mathbb{D}(A+C, B+C) + \mathbb{D}(B+C, B+D) \le \mathbb{D}(A,B) + \mathbb{D}(C,D), \tag{41}$$

where the first inequality follows from the triangle inequality and the second inequality follows from the definition of $\mathbb{D}$. Using the inequality (41), we have

$$\begin{aligned}\mathbb{D}\left(\partial_x \psi_N(x^N,\rho_N), \partial_x \psi(x^*,\rho)\right) &= \mathbb{D}\left(\partial f_N(x^N) + \rho_N \partial\theta(x^N), \partial \mathbb{E}_P[f(x^*,\xi)] + \rho_N \partial\theta(x^*)\right) \\ &\le \mathbb{D}\left(\partial f_N(x^N), \partial \mathbb{E}_P[f(x^*,\xi)]\right) + \mathbb{D}\left(\rho_N \partial\theta_N(x^N), \rho\partial\theta(x^*)\right).\end{aligned}$$

The first term at the right hand of the inequality of the formula above tends to zero with probability one by the fact that $\partial f_N(x)$ converges to $\partial \mathbb{E}_P[f(x,\xi)]$ uniformly over X, see Lemma 4.1 for the details of the of uniformly convergence; the second term tends to zero with probability one by (37) and $\rho_N \to \rho$. Together with the upper semi-continuity of the Clarke normal cone, we have w.p.1

$$0 \in \partial_x \psi(x^*,\rho) + \mathcal{N}_X(x^*),$$

that is, x^* is a stationary point of the true penalty problem (11). The proof is complete. ■

It might be interesting to ask whether a stationary point of (11) is a stationary point of (5). To answer this question, we need to introduce first order optimality condition for the latter. Let us assume that problem (5) satisfies the Slater condition, X is a compact set and the Lipschitz modulus of $f(x,\xi)$ w.r.t. x is bounded by an integrable function $\kappa(\xi) > 0$. We consider the following first order optimality conditions:

$$\begin{cases} 0 \in \partial \mathbb{E}_P[f(x,\xi)] + \lambda \partial\theta(x), \\ \lambda > 0, \\ \mathbb{E}_P[H(x,t,\xi)] \le 0, \quad \forall t \in T, \\ x \in X. \end{cases} \tag{42}$$

We say a point x^* is a stationary point of (5) if there exists $\lambda^* > 0$ such that (x^*, λ^*) satisfies (42). To justify this definition, we show that every local optimal solution to (5) satisfies (42) (along with some positive number λ). In the case when $\mathbb{E}_P[f(x,\xi)]$ is convex, a point satisfying (42) is a global optimal solution to (5). In what follows, we verify these. Let $\hat{x}$ be a local minimizer of (5). Let

$$\gamma(P) =: \sup_{x\in X} \frac{\|x-\bar{x}\|}{\min_{t\in T} -\mathbb{E}_P[H(\bar{x},t,\xi)]},$$

where $\bar{x}$ is given in Definition 2.3. Then for $\rho > \gamma(P)\mathbb{E}_P[\kappa(\xi)]$, x^* is a local optimal solution of (11). Consequently (x^*, ρ) satisfies (42). Conversely if x^* is a stationary point which means there exists positive number λ^* such that (x^*, λ^*) satisfies (42). If $\mathbb{E}_P[f(x,\xi)]$ is a convex function, then it is easy to see that x^* is a global optimal solution of (11) with $\rho = \lambda^*$. Since x^* is a feasible point of (5), it is not difficult to verify that x^* is a global optimal solution of (5).

Note that Dentcheva and Ruszczyński [12] introduced some first order optimality conditions for a class of semi-infinite programming problems which stem from optimization problems with stochastic second order constraints. Let $\mathscr{M}(T)$ denote the set of regular countably additive measures on T and $\mathscr{M}_+(T)$ its subset of positive measures. Consider the the following Lagrange function of (5):

$$\mathscr{L}(x,\mu) = \mathbb{E}_P[f(x,\xi)] + \int_T \mathbb{E}_P[H(x,t,\xi)]\mu(dt),$$

where $\mu \in \mathscr{M}_+(T)$. Under the so-called differential constraint qualifications, Dentcheva and Ruszczyński showed that if a point x^* is a local optimal solution (5), then there exists $\mu^* \in \mathscr{M}_+(T)$ such that

$$\begin{cases} 0 \in \partial_x \mathscr{L}(x,\mu) = \partial \mathbb{E}_P[f(x^*,\xi)] + \int_T \partial_x \mathbb{E}_P[H(x^*,t,\xi)]\mu^*(dt) + \mathcal{N}_X(x^*), \\ \mathbb{E}_P[H(x^*,t,\xi)] \le 0, \forall t \in T, \\ \int_T \mathbb{E}_P[H(x^*,t,\xi)]\mu^*(dt) = 0, \\ x \in X, \end{cases} \tag{43}$$

see [12, Theorem 4] for details and [12, Definition 2] for the differential constraint qualification. Note that optimality conditions (43) can also be alternatively characterized by some convex functions defined over $\mathbb{R}$. This is achieved by representing the integral with respective to measure μ with some convex functions through Riesz representation theorem, see [8,9] for details. It is an open question as to whether there is some relationship between (42) and (43) or the equivalent conditions of (43) in [8, 9], and this will be the focus of our future work.

References

[1] Z. Artstein and R. J-B. Wets, Approximating the integral of a multifunction, *Multivariate analysis*, Vol. 24, pp. 285–308, 1988.

[2] J. R. Birge and L. Qi, Subdifferential convergence in stochastic programs, *SIAM Journal on Optimization*, Vol. 5, pp. 436-453, 1995.

[3] J. F. Bonnans and A. Shapiro, *Perturbation analysis of optimization problems*, Springer, New York, 2000.

[4] C. Castaing and M. Valadier, Convex analysis and measurable multifunctions, *Lecture notes in mathematics*, Springer, Berlin, 1977.

[5] F. H. Clarke, *Optimization and nonsmooth analysis*, Wiley, New York, 1983.

[6] L. Dai, C. H. Chen and J. R. Birge, Convergence properties of two stage stochastic programming, *Journal of Optimization and Application*, Vol. 106, pp. 489-509, 2000.

[7] D. Dentcheva, R. Henrion and A. Ruszczynski, Stability and sensitivity of optimization problems with first order stochastic dominance constraints, *SIAM Journal on Optimization*, Vol. 18, pp. 322-333, 2007.

[8] D. Dentcheva and A. Ruszczyński, Optimization with stochastic dominance constraints, *SIAM Journal on Optimization*, Vol. 14, pp. 548-566, 2003.

[9] D. Dentcheva and A. Ruszczyński, Semi-infinite probabilistic constraints: optimality and convexification, *Optimization*, Vol. 53, pp. 583-601, 2004.

[10] D. Dentcheva and A. Ruszczyński, Optimality and duality theory for stochastic optimization with nonlinear dominance constraints, *Mathematical Programming*, Vol. 99, pp. 329-350, 2004.

[11] D. Dentcheva and A. Ruszczyński, Portfolio optimization with stochastic dominance constraints, *Journal of Banking and Finance*, Vol. 30, pp. 433-451, 2006.

[12] D. Dentcheva and A. Ruszczyński, Composite semi-infinite optimization, *Control and Cybernetics*, Vol. 36, pp. 1-14, 2007.

[13] M. Gugat, Error bounds for infinite systems of convex inequalities without Slater's condition, *Mathematical Programming, Ser. B*, Vol. 88, pp. 255-275, 2000.

[14] R. Henrion and W. Römisch, Metric regularity and quantitative stability in stochastic programs with probabilistic constraints, *Mathematical Programming*, Vol. 84, pp. 55-88, 1999.

[15] R. Henrion and W. Römisch, Hölder and lipschitz stability of solution sets in programs with probabilistic constraints, *Mathematical Programming*, Vol. 100, pp. 589-611, 2004.

[16] T. Homen-De-Mello, On rates of convergence for stochastic optimization problems under non-independent and identically distributed sampling, *SIAM Journal on Optimization*, Vol. 19, pp. 524-551, 2008.

[17] T. Homem-de-Mello, A cutting surface method for uncertain linear programs with polyhedral stochastic dominance constraints, *SIAM Journal on Optimization*, 2010.

[18] D. Klatte, On the stability of local and global optimal solutions in parametric problems of nonlinear programming, Part I: Basic results, *Seminarbericht Nr. 75*, Sektion Mathematik, Humboldt-Universität zu Berlin, Berlin, pp. 1-21, 1985.

[19] D. Klatte, A note on quantitative stability results in nonlinear optimization, *Seminarbericht Nr. 90*, Sektion Mathematik, Humboldt-Universität zu Berlin, Berlin, pp. 77-86, 1987.

[20] H. Levy, Stochastic dominance and expected utility: survey and analysis, *Management Science*, Vol. 38, pp. 555-593, 1992.

[21] Y. Liu, H. Xu and G.-H. Lin, Stability analysis of two stage stochastic mathematical programs with complementarity constraints via NLP-regularization, Optimization online, March 2010.

[22] A. Müller and M. Scarsini, Eds., *Stochastic orders and decision under risk, institute of mathematical statistics*, Hayward, CA, 1991.

[23] E. Polak, On the mathematical fondations of nondifferentiable optimization in engineering design, *SIAM Rewiew*, Vol. 29, pp. 21-89, 1987.

[24] C. J. Price and I. D. Cooper, An exact penalty function algorithm for semi-infinite programmes, *BIT Numerical Mathematics*, Vol. 30, pp. 723-734, 1990.

[25] S. T. Rachev and W. Römisch, Quantitative stability in stochastic programming: the method of probability metrics, *Mathematics of Operations Research*, Vol. 27, pp. 792-818, 2002.

[26] S. M. Robinson, Stability theory for systems of inequalities, Part II: Differentiable nonlinear systems, *SIAM Journal on Numerical Analysis*, Vol. 13, pp. 497-513, 1976.

[27] S. M. Robinson and R. J-B. Wets, Stability in two-stage stochastic programming, *SIAM Journal on Control and Optimization*, Vol. 25, pp. 1409-1416, 1987.

[28] S. M. Robinson, Analysis of sample-path optimization, *Mathematics of Operations Research*, Vol. 21, pp. 513–528, 1996.

[29] R. T. Rockafellar and R. J-B. Wets, *Variational analysis*, Springer, Berlin, 1998.

[30] W. Römisch, Stability of stochastic programming problems, in Stochastic Programming, A. Ruszczyński and A. Shapiro, eds., Elsevier, Amsterdam, pp. 483-554, 2003.

[31] W. Römisch and R. Schultz, Stability analysis for stochastic programs, *Annals of Operational Research*, Vol. 30, pp. 241-266, 1991.

[32] A. Ruszczyński and A. Shapiro, *Stochastic programming, handbook in operations research and management science*, Elsevier, 2003.

[33] A. Shapiro, Asymptotic properties of statistical estimators in stochastic programming. *Annals of Statistics*, Vol. 17, pp. 841-858, 1989.

[34] A. Shapiro, Quantitative stability in stochastic programming, *Mathematical Programming*, Vol. 67, pp. 99-108, 1994.

[35] A. Shapiro and H. Xu, Stochastic mathematical programs with equilibrium constraints, modeling and sample average approximation, *Optimization*, Vol. 57, pp. 395-418, 2008.

[36] Y. Tanaka, M. Fukushima and T. Ibaraki, A globally convergent SQP method for semi-infinite nonlinear optimization, *Journal of Computational and Applied Mathematics*, Vol. 23, pp. 141-153, 1988.

[37] P. Terán, On a uniform law of large numbers for random sets and subdifferentials of random functions, *Statistics & Probability letters*, Vol. 78, pp. 42-49, 2008.

[38] H. Xu, Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming, *Journal of Mathematical Analysis and Applications*, Vol. 368, pp. 692-710, 2010.

[39] H. Xu and D. Zhang, Smooth sample average approximation of stationary points in nonsmooth stochastic optimization and applications, *Mathematical Programming*, Vol. 119, pp. 371–401, 2009.

4 Appendix

Lemma 4.1 ([21], Theorem 5.1) *Let $F(x,\xi) : \mathbb{R}^n \times \Xi \to \mathbb{R}^m$ be a continuous function, $\{P_\nu\}$ be a sequence of probability measures and $\mathcal{X}$ be a compact subset. Assume: (a) $F(x,\xi)$ is locally Lipschitz continuous with respect to x for almost every ξ with modulus $L(x,\xi)$ which is bounded by a positive constant C; (b) $\{P_\nu\}$ converges to P in distribution. Then for every fixed x, $\partial\mathbb{E}_{P_\nu}[F(x,\xi)]$ and $\partial\mathbb{E}_P[F(x,\xi)]$ are well-defined and*

$$\lim_{\nu\to\infty} \sup_{x\in\mathcal{X}} \mathbb{H}(\partial\mathbb{E}_{P_\nu}[F(x,\xi)], \partial\mathbb{E}_P[F(x,\xi)]) = 0. \tag{44}$$

Proof. For the simplicity of notation, let $f_{P_\nu}(x) = \mathbb{E}_{P_\nu}[F(x,\xi)]$ and $f_P(x) = \mathbb{E}_P[F(x,\xi)]$. Under condition (a), both $f_{P_\nu}(x)$ and $f_P(x)$ are globally Lipschitz continuous, therefore Clarke's generalized derivatives of $f_{P_\nu}(x)$ and $f_P(x)$, denoted by $f^o_{P_\nu}(x;h)$ and $f^o_P(x;h)$ respectively, are well-defined for any fixed nonzero vector $h \in \mathbb{R}^n$, where

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

and

$$f^o_P(x;h) = \limsup_{x'\to x,\tau\downarrow 0} \frac{1}{\tau}(f_P(x'+\tau h) - f_P(x')).$$

Our idea is to study the Hausdorff distance $\mathbb{H}(\partial f_{P_\nu}(x), \partial f_P(x))$ through certain "distance" of the Clarke generalized derivatives $f^o_{P_\nu}(x;h)$ and $f^o_P(x;h)$. Let D_1, D_2 be two convex and compact subsets of $\mathbb{R}^m$. Let $\sigma(D_1,u)$ and $\sigma(D_2,u)$ denote the support functions of D_1 and D_2 respectively. Then

$$\mathbb{D}(D_1,D_2) = \max_{\|u\|\le 1}(\sigma(D_1,u) - \sigma(D_2,u))$$

and

$$\mathbb{H}(D_1,D_2) = \max_{\|u\|\le 1}|\sigma(D_1,u) - \sigma(D_2,u)|.$$

The above relationships are known as Hömander's formulae, see [4, Theorem II-18]. Applying the second formula to our setting, we have

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

Using the relationship between Clarke's subdifferential and Clarke's generalized derivative, we have that $f^o_{P_\nu}(x;h) = \sigma(\partial f_{P_\nu}(x),h)$ and $f^o_P(x;h) = \sigma(\partial f_P(x),h)$. Consequently,

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

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

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

Using the inequality above, we have

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

Since P_ν converges to P in distribution, and the integrand $\frac{1}{\tau}(F(x'+\tau h, \xi) - F(x', \xi))$ is continuous w.r.t ξ and it is bounded by L, then

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

■