Stability analysis of stochastic programs with second order dominance constraints

Yongchao Liu · Huifu Xu

Abstract In this paper we present a stability analysis of a stochastic optimization problem with stochastic second order dominance constraints. We consider a perturbation of the underlying probability measure in the space of regular measures equipped with pseudometric discrepancy distance (Römisch in Stochastic Programming. Elsevier, Amsterdam, pp 483–554, 2003). By exploiting a result on error bounds in semi-infinite programming due to Gugat (Math Program Ser B 88:255–275, 2000), 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 an empirical probability measure and show an exponential rate of convergence of the sequence of optimal solutions obtained from solving the approximation problem. The analysis is extended to the stationary points.

Keywords Second order dominance · Stochastic semi-infinite programming · Error bound · Lipschitz stability · Sample average approximation

Mathematics Subject Classification 90C15 · 90C31

Y. Liu
Department of Mathematics, Dalian Maritime University, Dalian 116023, China
e-mail: googleliu0717@yahoo.com.cn

H. Xu (✉)
School of Mathematics, University of Southampton, Southampton SO17 1BJ, UK
e-mail: h.xu@soton.ac.uk

1 Introduction

Stochastic dominance is a fundamental concept in decision theory and economics [26]. For two random variables $\xi_1(\omega)$ and $\xi_2(\omega)$ defined on the probability space $(\Omega, \mathcal{F}, P)$ 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

$$\int_{-\infty}^{\eta} P\{\omega \in \Omega : \xi_1(\omega) \leq t\}dt \leq \int_{-\infty}^{\eta} P\{\omega \in \Omega : \xi_2(\omega) \leq t\}dt, \quad \forall \eta \in \mathbb{R}. \quad (1)$$

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

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

We use the notation $(\cdot)_+$ to denote the positive part, that is, $(x)_+ := \max\{x, 0\}$.

In this paper, we consider the following stochastic program with stochastic second order dominance (SSD) constraints:

$$\begin{aligned} \min_x \;& \mathbb{E}_P[f(x, \xi(\omega))] \\ \text{s.t.} \;& G(x, \xi(\omega)) \succeq_2 Y(\xi(\omega)), \\ & x \in X, \end{aligned} \quad (3)$$

where X is a nonempty compact subset of $\mathbb{R}^n$, $\xi : \Omega \to \Xi$ is a vector of random variables defined on the probability space $(\Omega, \mathcal{F}, P)$ 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 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:

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

Stochastic optimization models with SSD constraints were introduced by Dentcheva and Ruszczyński [8,9]. Over the past few years, there has been increasing discussion of this subject covering optimization theory, numerical methods and practical applications, see [7–12,17,19,22] and references therein.

It is well-known that the SSIP problem above does not satisfy Slater's constraint qualification, a condition that is often required for a stable numerical method. Subsequently, a relaxed form of SSIP has been proposed:

$$\min_x \ \mathbb{E}_P[f(x, \xi(\omega))]$$
$$\text{s.t.} \ \ \mathbb{E}_P[H(x, t, \xi(\omega))] \le 0, \quad \forall t \in T, \tag{5}$$
$$x \in X,$$

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 changes in the probability measure P on both 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 [14,15,20,21,29,31,32,35,36].

From a practical viewpoint, this kind of stability analysis is motivated by the fact that in applications either the probability distribution of P is not known or a closed form expression for the expected value of the underlying random functions w.r.t. P is difficult to obtain and consequently the probability measure/distribution may have to be approximated. Stability analysis of problem (5) focuses on the impact on the optimal value and optimal solutions of a perturbation of P [29]. A particularly interesting case is when the probability measure is approximated by an empirical probability measure. In this case, the expected value of the underlying functions are approximated through sample averaging. The contribution of this paper can be summarized as follows.

- In Sect. 2, 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, 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 conditions, we show upper semi-continuity of the optimal set-valued mapping. This complements the existing research [7] which focuses on the stability analysis of stochastic optimization problems with first order dominance constraints.
- In Sect. 3, we consider a special case when the probability measure P is approximated by an 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 Hu et al. [19] which discusses the cutting plane method for sample average
approximated optimization problems with SSD constraints.

• Our convergence analysis is carried out through an exact penalization of (5) [see
(11)]. The penalty formulation may provide a potential numerical framework for
solving (5). During the revision of this paper, some progress in this regard has
been made, see [24].

Throughout this paper, we use the following notation. For vectors $a, b \in \mathbb{R}^n$,
$a^T b$ denotes the scalar product, $\| \cdot \|$ denotes the Euclidean norm of a vector, $\| \cdot \|_\infty$
denotes the maximum norm of continuous functions defined over compact 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 $\mathbb{H}(\mathcal{C}, \mathcal{D}) := \max(\mathbb{D}(\mathcal{C}, \mathcal{D}), \mathbb{D}(\mathcal{D}, \mathcal{C}))$ 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, $\{C + D : C \in \mathcal{C}, D \in \mathcal{D}\}$, $B(x, \gamma)$ denotes the
closed ball with center x and radius γ, $\mathcal{B}$ denotes the closed unit ball in the respective
space.

2 Stability 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 value 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:

$$\begin{aligned}
&\min_x \; \mathbb{E}_Q[f(x, \xi(\omega))] \\
&\text{s.t.} \; \mathbb{E}_Q[H(x, t, \xi(\omega))] \leq 0, \quad \forall t \in T, \\
&\quad\quad x \in X,
\end{aligned} \tag{6}$$

which is regarded as a perturbation of (5). Specifically, we study the relationship
between the perturbed problem (6) and the true problem (5) in terms of optimal values
and optimal solutions when Q is close to P. Of course, we restrict our discussion to
the probability measure Q such that the expected values of the underlying functions
in (6) are well defined for all $x \in X$.

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:

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

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

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

This type of distance is considered by Römisch [35, Section 2.2] for the stability analysis of stochastic programming and is called a pseudometric. It is well-known that $\mathscr{D}$ is nonnegative, symmetric and satisfies the triangle inequality, see [35, Section 2.1] and the references therein. Throughout this section, we use the following notation:

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

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

In what follows, we discuss the continuity of the optimal solution set mapping and optimal value function of problem (6). We do so by applying Klatte's [20,21] earlier stability result of parametric nonlinear programming, which was used by Dentcheva et al. [7] for the stability analysis of optimization problems with first order dominance constraints. A key condition in Klatte's stability result is the pseudo-Lipschitz property of the feasible set mapping. Here we derive the property by exploiting an important result on the 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-infinite programming in [13].

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

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

Definition 2.2 Problem (5) is said to satisfy a *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

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

Definition 2.3 Problem (5) is said to satisfy a *Slater condition* if there exist a positive number $\bar{\delta}$ 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. Since X is a compact set, the Slater condition implies the strong Slater condition and then 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 details about the relationship.

Proposition 2.4 *Assume problem* (5) *satisfies the Slater condition. Let $\bar{\delta}$ be given as in Definition 2.3. Then there exists a positive number ϵ ($\epsilon \leq \bar{\delta}/2$) such that for any $Q \in B(P, \epsilon)$*

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

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

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

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

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

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

The proof is complete. □

By Gugat [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 *Assume problem* (5) *satisfies the Slater condition. Then there exist positive numbers ϵ and β such that for any $Q \in B(P, \epsilon)$, the following error bound holds:*

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

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

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

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

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

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

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

Note that this kind of error bound in well known in convex programming, see a pioneering work by Robinson [30]. Here we follow the recent development in [13] as our problem falls into the framework of semi-infinite programming.

Proposition 2.6 *Assume that problem* (5) *satisfies the Slater condition. Then*

(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 $B(P, \epsilon)$, that is,*

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

where β is a defined in Lemma 2.5.

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

Part (ii). Let $t \in T$ be fixed. It follows by virtue of [35, 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 $x^N \to x^*$. By the Fatou's lemma

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

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

Part (iii). Let ϵ be given by Lemma 2.5 and $Q_1, Q_2 \in B(P, \epsilon)$. 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 $x \in \mathcal{F}(Q_1)$

$$d(x, \mathcal{F}(Q_2)) \leq \beta \left\| \left(\mathbb{E}_{Q_2}[H(x, t, \xi)] \right)_+ \right\|_\infty$$

$$= \beta \left(\max_{t \in T} \left(\mathbb{E}_{Q_2}[H(x, t, \xi)] \right)_+ - \max_{t \in T} \left(\mathbb{E}_{Q_1}[H(x, t, \xi)] \right)_+ \right)$$

$$\leq \beta \max_{t \in T} \left((\mathbb{E}_{Q_2}[H(x, t, \xi)])_+ - (\mathbb{E}_{Q_1}[H(x, t, \xi)])_+ \right)$$

$$\leq \beta \max_{t \in T} \left| \mathbb{E}_{Q_2}[H(x, t, \xi)] - \mathbb{E}_{Q_1}[H(x, t, \xi)] \right|$$

$$\leq \beta \max_{(x,t) \in X \times T} \left| \mathbb{E}_{Q_2}[H(x, t, \xi)] - \mathbb{E}_{Q_1}[H(x, t, \xi)] \right|$$

$$\leq \beta \mathscr{D}(Q_1, Q_2),$$

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

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

$$\leq \beta \mathscr{D}(Q_2, Q_1),$$

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

$$\mathbb{H}(\mathcal{F}(Q_1), \mathcal{F}(Q_2)) = \max \{ \mathbb{D}(\mathcal{F}(Q_1), \mathcal{F}(Q_2)), \mathbb{D}(\mathcal{F}(Q_2), \mathcal{F}(Q_1)) \}$$

$$\leq \beta \mathscr{D}(Q_1, Q_2).$$

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* near y if there exists a constant L such that

$$\mathbb{H}(\Gamma(y'), \Gamma(y'')) \leq L \| y' - y'' \|, \quad \forall y', y'' \in y + \delta \mathcal{B}.$$

See [34, page 368].

Proposition 2.6 (iii) says that the feasible set mapping of problem (6) is Lipschitz continuous with respect to probability measure over set $B(P, \epsilon)$. Using this property, we are ready to establish our main stability results.

Theorem 2.7 *Assume that problem* (5) *satisfies the Slater condition. Assume also that the Lipschitz modulus of $f(x, \xi)$ w.r.t. x is bounded by an integrable function $\kappa(\xi) > 0$. Then*

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

(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[1] estimation:*

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

Proof Under the Slater condition, it follows from Proposition 2.6 that there exists a positive number ϵ such that the feasible set mapping $\mathcal{F}(\cdot)$ is Lipschitz continuous on $B(P, \epsilon)$. The rest follows straightforwardly from [21, Theorem 1] ([29, Theorem 2.3] or [7, Theorem 2.1] in stochastic programming). The proof is complete. $\square$

Theorem 2.7 says that the optimal solution set mapping $S_{opt}(\cdot)$ is nonempty near P and upper semi-continuous at P. In order to quantify the upper semi-continuity of $S_{opt}(\cdot)$, we need some growth condition of the objective function of problem (5) in a neighborhood of $S_{opt}(P)$. Instead of imposing a specific growth condition, here we consider a general growth function

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

of problem (5), and the associated function

$$\widetilde{\Psi}(v) := v + \Psi^{-1}(2v),$$

where s^* denotes the optimal value of problem (5) and $v, v \in \mathbb{R}_+$. This kind of growth function is well known, see for instance [29,34]. The following corollary quantifies the upper semicontinuity of $S_{opt}(\cdot)$ near P.

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

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

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

We omit the proof as it is similar to that of [29, Theorem 2.4]. See also [34, Theorem 7.64] for earlier discussions about functions $\Psi(\cdot)$ and $\widetilde{\Psi}(\cdot)$. Discussions on a particular form of $\widetilde{\Psi}$ can be found in [4,39] when the growth is of second order. We will come back to this in Lemma 3.8.

---

[1] The property is also known as calmness of ϑ at P, see Section F in [34, Chapter 8] for general discussions on calmness.

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 $\xi^1, \ldots, \xi^N$ 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 [16,33,41,43] and the references therein.

For the simplicity of notation, we use $f_N(x)$ and $H_N(x, t)$ to denote $\mathbb{E}_{P_N}[f(x, \xi)]$ and $\mathbb{E}_{P_N}[H(x, t, \xi)]$. Consequently we consider the following approximation of problem (5):

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

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 will be very complicated if it is carried out on (9) directly because the con-

straints of the SAA problem depend on the 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 exact penalization of the stochastic constraints to the objective. In doing so, the feasible set of the penalized problems are deterministic and we only need to analyze the convergence of the the objective functions.

For the simplicity of notation, let

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

It is easy to observe that

$$h(x, t) = (\mathbb{E}_P[H(x, t, \xi)])_+ \text{ and } \theta(x) = \|(\mathbb{E}_P[H(x, t, \xi)])_+\|_\infty.$$

Consider the exact penalization:

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

where $\rho > 0$ is a penalty parameter. This kind of penalization is well documented in the literature, see for instance [28,42]. 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. 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

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

Let $C := \mathbb{E}_P[\kappa(\xi)]$. Under Assumption 3.1, $C < \infty$ and the Lipschitz modulus of $\mathbb{E}_P[f(x, \xi)]$ is bounded by C. Let ρ be a positive constant such that $\rho > \beta C$. Clarke's exact penalty function theorem [5, Proposition 2.4.3] ensures that the two optimal solution sets, S_{opt} and X_{opt}, coincide. 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

$$h_N(x, t) := (H_N(x, t))_+ = \max \{ H_N(x, t), 0 \}$$

and

$$\theta_N(x) := \max_{t \in T} h_N(x, t) = \|(H_N(x, t))_+\|_\infty. \tag{12}$$

Consider the SAA penalty problem

$$\min_x \ \psi_N(x, \rho_N) := f_N(x) + \rho_N \theta_N(x) \tag{13}$$
$$\text{s.t.} \quad x \in X,$$

where $\rho_N > 0$ is a penalty parameter.

Under Assumption 3.1, we know from [37, Section 6, Proposition 7] that the sample average $H_N(x, t)$ converges to $\mathbb{E}_P[H(x, t, \xi)]$ uniformly over compact set $X \times T$ almost surely. Since the true problem (5) satisfies the Slater condition, there exists a sufficiently large N^* (depending on ω) such that for $N \geq N^*$

$$H_N(\bar{x}, t) \leq -\bar{\delta}/2, \quad \forall t \in T,$$

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

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

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

Proposition 3.3 *Assume that the true problem* (5) *satisfies the Slater condition. Then there exist positive numbers ρ^* and N^* (depending on ω) such that for $\rho > \rho^*$ and $N \geq N^*$, the sets of optimal solutions of problems* (9) *and* (13)*, denoted by S_{opt}^N and X_{opt}^N respectively, coincide.*

Proof Following the discussions above, there exist a positive constant β and a sufficiently large positive integer N_1 (depending on ω) 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_{opt}^N and X_{opt}^N, coincide. Moreover, under Assumption 3.1, C_N converges to the Lipschitz modulus of $\mathbb{E}[f(x, \xi)]$ and is bounded by $\mathbb{E}[\kappa(\xi)]$. This implies that there exists a positive integer $N_2 \geq N_1$ such that when $N \geq N_2$, we have $C_N < C + 1$, that is, C_N is bounded almost surely. The conclusion follows by taking $\rho^* = \beta(C + 1)$ 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 uniform convergence of the objective function of problem (13) to the objective function of problem (11). Asymptotic convergence analysis of optimal values and optimal solutions are well known in stochastic

programming. Our analysis differs from those in the literature in that it is carried out through exact penalization.

Proposition 3.4 *Let Assumption* 3.1 *hold. Then*

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

(ii) *if* $\rho_N \to \rho$ *as* $N \to \infty$, *then* $\psi_N(x, \rho_N)$ *converges to* $\psi(x, \rho)$ *with probability 1 uniformly over* X.

Proof Part (i). Under Assumption 3.1, $\mathbb{E}_P[H(x, t, \xi)]$ and $H_N(x, t)$ are Lipschitz continuous with respect to (x, t). Since T is a compact set, by [27, 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. [37, Section 6, Proposition 7]. Since $\rho_N \to \rho$, it suffices to show the uniform convergence of $\theta_N(x)$ to $\theta(x)$. By definition,

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

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 0.

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. 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 0 because $\mathbb{E}[e^{-(1+\kappa(\xi)\tau)}] = e^{-\tau} \mathbb{E}[e^{-\kappa(\xi)\tau}] = e^{-\tau} M_\kappa(\tau)$.

Proposition 3.6 *Let Assumptions 3.1, 3.5 hold and $\rho_N \to \rho$. Then $\psi_N(x, \rho_N)$ converges to $\psi(x, \rho)$ with probability approaching 1 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

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

$$= \text{Prob}\left\{\sup_{x \in X} |f_N(x) + \rho_N \theta_N(x) - (\mathbb{E}_P[f(x, \xi)] + \rho \theta(x))| \geq \alpha\right\}$$

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

Under Assumption 3.5, it follows from [41, Theorem 5.1] that the first term on the right hand side 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 [41], we may estimate the sample size. To see this, 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 $x \in X$,

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

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

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

Note that equality in (16) and (17) holds if random variables $f(x, \xi) - \mathbb{E}_P[f(x, \xi)]$ and $H(x, t, \xi) - \mathbb{E}_P[H(x, t, \xi)]$ follow a normal distribution with variance ϱ^2, see a

discussion in [41, 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 [41] that for

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

where $O(1)$ is a generic constant, we have that

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

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

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

we have

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

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

$$\text{Prob} \left\{ \max_{x \in X} |\psi_N(x, \rho_N) - \psi(x, \rho)| \geq \alpha \right\} \leq \text{Prob} \left\{ \sup_{x \in X} |f_N(x) - \mathbb{E}_P[f(x, \xi)]| \geq \alpha/2 \right\}$$

$$+ \text{Prob} \left\{ \sup_{x \in X} |\rho_N \theta_N(x) - \rho\theta(x)| \geq \alpha/2 \right\}. \quad (22)$$

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 $N \geq N_3$

$$
\text{Prob} \left\{ \sup_{x \in X} |\rho_N \theta_N(x) - \rho\theta(x)| \geq \alpha/2 \right\}
$$

$$
\leq \text{Prob} \left\{ \sup_{x \in X} |\theta_N(x) - \theta(x)| \geq \frac{\alpha}{8\rho} \right\}
$$

$$
\leq \text{Prob} \left\{ \max_{(x,t) \in X \times T} |H_N(x, t) - \mathbb{E}_P[H(x, t, \xi)]| \geq \frac{\alpha}{8\rho} \right\}. \tag{23}
$$

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 $N \geq N(\alpha, \beta)$

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

$$
\leq \text{Prob} \left\{ \sup_{x \in X} |f_N(x) - \mathbb{E}_P[f(x, \xi)]| \geq \alpha/2 \right\}
$$

$$
+ \text{Prob} \left\{ \max_{(x,t) \in X \times T} |H_N(x, t) - \mathbb{E}_P[H(x, t, \xi)]| \geq \frac{\alpha}{8\rho} \right\}
$$

$$
\leq \beta_1 + \beta_2
$$

$$
= \beta.
$$

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

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

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

Lemma 3.8 *Let $\phi : \mathbb{R}^m \to \mathbb{R}$ be a continuous function and $X \subseteq \mathbb{R}^m$ be a closed set, let $\varphi : \mathbb{R}^m \to \mathbb{R}$ be a continuous perturbation of ϕ. Consider the following constrained minimization problem*

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

and its perturbation

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

Let X_ϕ^* and X_φ^* denote the set of optimal solutions to (25) and (26) respectively. Then

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

$$\mathbb{D}(X_\varphi^*, X_\phi^*) \leq \epsilon, \tag{27}$$

   when

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

(ii) *if, in addition, there exists a positive constant ς such that*

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

   then

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

Proof The results are a minor extension of [6, Lemma 3.2] which deals with the case when X_ϕ^* is a singleton and are also similar to [34, 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

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

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

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

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, $\mathbb{D}(X_\varphi^*, X_\phi^*) \leq \epsilon$.

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. $\qquad\square$

Remark 3.9 We have a few comments on Lemma 3.8.

(i) Condition (28) is known as a second order growth condition. Using this condition, Shapiro [38] 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 [38, Lemma 4.1] and [39, Proposition 2.1]. Both the second order growth condition and the variational principal have been widely used for stability and asymptotic analysis in stochastic programming, see [4,38,39]. 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 $x_0 \in X$.

(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

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

and obtain from Lemma 3.8 (i) that

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

Theorem 3.10 *Assume that problem* (5) *satisfies the Slater condition. 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) *with probability 1*

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

where X_{opt} *and* X_{opt}^N *denote the sets of optimal solutions of problem* (11) *and* (13) *respectively. Moreover, if Assumption* 3.5 *holds, then 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*

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

for N sufficiently large.

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

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

where ς *is a positive constant, then* $C_1(\alpha) = C(\frac{1}{3}\varsigma\alpha^2)$ *and* $K_1(\alpha) = K(\frac{1}{3}\varsigma\alpha^2)$ *where* $C(\alpha)$ *and* $K(\alpha)$ *are given in Proposition* 3.6.

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

$$\text{Prob}\left\{\mathbb{D}\left(X_{opt}^N, X_{opt}\right) \geq \alpha\right\} \leq \beta,$$

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

Proof The almost sure 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. Next, we show the exponential convergence. By Lemma 3.8, for any $\alpha > 0$, there exists $\varepsilon(\alpha)$ such that if

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

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

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

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

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

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),

$$\text{Prob}\left\{\mathbb{D}\left(X_{opt}^N, X_{opt}\right) \geq \alpha\right\} \leq \text{Prob}\left\{\sqrt{\frac{3}{\varsigma}\sup_{x \in X} |\psi_N(x, \rho_N) - \psi(x, \rho)|} \geq \alpha\right\}$$

$$= \text{Prob}\left\{\sup_{x \in X} |\psi_N(x, \rho_N) - \psi(x, \rho)| \geq \frac{1}{3}\varsigma\alpha^2\right\}.$$

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 penalized sample average approximation

problem (13). This is motivated by our wish to address the case when $f(x, \xi)$ is not convex w.r.t. x. Convergence analysis of SAA stationary sequence has been well documented, see [43] and the references therein. Our analysis here differs from those in the literature in two ways: (a) We analyze the convergence of SAA stationary point to its true counterpart rather than so-called *weak stationary point* of the true problem [43], the analysis is based on an approximation of the Clarke subdifferential of expected value of a random function rather that of the expected value of the Clarke subdifferential of a random function. Note that this kind of subdifferential approximation can be traced back to the earlier work by Birge and Qi [3] and Artstein and Wets [2]. (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]

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

Here and later on "conv" denotes the convex hull of a set. Since $h(\cdot, t)$ is convex for each t, T is a compact set and for every x, $h(x, \cdot)$ is continuous on T, by Levin–Valadier theorem (see [37, Section 2, Theorem 51]),

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

Let

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

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

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

and $\mathcal{N}_X(x) = \emptyset$ if $x \notin X$. A point $x \in X$ is said to be a *stationary point* of the penalized minimization problem (11) if

$$0 \in \partial \mathbb{E}_P[f(x, \xi)] + \rho \partial \theta(x) + \mathcal{N}_X(x).$$

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

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

and

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

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

$$0 \in \partial f_N(x) + \rho_N \partial \theta_N(x) + \mathcal{N}_X(x).$$

Assumption 3.11 $f(x, \xi)$ and $G(x, \xi)$ are locally Lipschitz continuous w.r.t. x and ξ, and their Lipschitz modulus w.r.t. x are bounded by an integrable function $\kappa(\xi)$ for every $x \in \mathbb{R}^n$.

It is easy to observe that Assumption 3.11 is stronger than Assumption 3.1. Over the past few years, there have been extensive discussions (see [43]) on the convergence of SAA stationary points to the so-called weak stationary points of the true problem, which are defined through the expected value of the subdifferential of the underlying functions of the true problem in the first order optimality condition. A stationary point is a weak stationary point but not vice versa. Analysis of convergence of the SAA stationary point to a weak stationary point of the true problem can be proved under Assumption 3.1, but convergence to a stationary point of the true problem requires Assumption 3.11.

Consider a sequence of functions $\{f_N(x)\}$ defined on $\mathbb{R}^n$. Recall that f_N is said to epiconverge to a function f if and only if the epigraph of f_N converges to the epigraph of f. A necessary and sufficient condition for f_N to epiconverge to f is that for every $x \in \mathbb{R}^n$,

$$\begin{cases} \liminf_N f_N(x^N) \geq f(x) & \text{for every sequence } x^N \to x; \\ \limsup_N f_N(x^N) \leq f(x) & \text{for some sequence } x^N \to x. \end{cases}$$

See [34] for details. For set-valued mappings $\Gamma_N, \Gamma : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$, Γ_N is said to converge graphically to Γ if the graph of Γ_N converges to that of Γ.

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^* almost surely. Under Assumption 3.11*

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

almost surely.

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 over X and so are $h_N(x, t)$, $h(x, t)$, $\theta(x)$ and $\theta_N(x)$. Under Assumption 3.11, it follows by [37, Section 6, Proposition 7] that $H_N(x, t, \xi)$ converges to $\mathbb{E}_P[H(x, t, \xi)]$ uniformly over any compact subset of $\mathbb{R}^n \times \mathbb{R}$ almost surely. Subsequently, it is easy to verify that $\theta_N(x)$ converges to $\theta(x)$ uniformly over any compact subset of $\mathbb{R}^n$, which implies, via [34, Proposition 7.15], θ_N epiconverges to θ almost surely. By Attouch's theorem ([34, Theorem 12.35]), the latter convergence implies $\partial\theta_N$ converges to $\partial\theta$ graphically over X and hence (37). 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: (a) Assumption 3.11 holds; (b) for every $\xi \in \Xi$, $f(\cdot, \xi)$ is Clarke regular on X; (c) the probability space is nonatomic. If $\rho_N \to \rho$, then with probability 1 x^* is a stationary point of the true penalty problem (11).*

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

$$\mathbb{D}(A + C, B + D) \leq \mathbb{D}(A + C, B + C) + \mathbb{D}(B + C, B + D)$$
$$\leq \mathbb{D}(A, B) + \mathbb{D}(C, D), \tag{38}$$

where the first inequality follows from the triangle inequality and the second inequality follows from the definition of $\mathbb{D}$. Using the inequality (38), we have

$$\mathbb{D}\left(\partial f_N(x^N) + \rho_N \partial\theta(x^N), \partial\mathbb{E}_P[f(x^*, \xi)] + \rho_N \partial\theta(x^*)\right)$$
$$\leq \mathbb{D}\left(\partial f_N(x^N), \partial\mathbb{E}_P[f(x^*, \xi)]\right)$$
$$+ \mathbb{D}\left(\rho_N \partial\theta_N(x^N), \rho \partial\theta(x^*)\right).$$

In Proposition 3.12, we have shown that

$$\lim_{N \to \infty} \mathbb{D}\left(\rho_N \partial\theta_N(x^N), \rho \partial\theta(x^*)\right) = 0.$$

In what follows, we show

$$\lim_{N \to \infty} \mathbb{D}\left(\partial f_N(x^N), \partial\mathbb{E}_P[f(x^*, \xi)]\right) = 0. \tag{39}$$

Under the Clarke regularity

$$\partial\mathbb{E}_P[f(x^*, \xi)] = \mathbb{E}_P[\partial_x f(x^*, \xi)],$$

where $\mathbb{E}_P[\partial_x f(x^*, \xi)]$ denotes Aumann's [1] integral of the Clarke subdifferential. Under Assumption 3.11, it is well known that $\mathbb{E}_P[\partial_x f(x^*, \xi)]$ is well defined, see for instance [1] and [44, Proposition 2.2]. Moreover, the Clarke regularity implies

$$\partial f_N(x^N) = \frac{1}{N} \sum_{k=1}^{N} \partial_x f(x^N, \xi^k).$$

Therefore it suffices to show that

$$\lim_{N\to\infty} \mathbb{D}\left(\frac{1}{N} \sum_{k=1}^{N} \partial_x f(x^N, \xi^k), \mathbb{E}_P[\partial_x f(x^*, \xi)] \right) = 0 \qquad (40)$$

almost surely. Under Assumption 3.11, we have by virtue of [40, Theorem 2],

$$\lim_{N\to\infty} \sup_{x\in X} \mathbb{D}\left(\frac{1}{N} \sum_{k=1}^{N} \partial_x f(x, \xi^k), \mathbb{E}_P[\partial_x^\delta f(x, \xi)] \right) = 0$$

almost surely, where δ is a positive number which can be arbitrarily small and

$$\partial_x^\delta f(x, \xi) = \bigcup_{x'\in B(x,\delta)} \partial_x f(x', \xi).$$

This implies

$$\lim_{N\to\infty} \mathbb{D}\left(\frac{1}{N} \sum_{k=1}^{N} \partial_x f(x^N, \xi^k), \mathbb{E}_P[\partial_x^\delta f(x^*, \xi)] \right) = 0$$

almost surely and hence

$$0 \in \mathbb{E}_P[\partial_x^\delta f(x^*, \xi)] + \rho \partial \theta(x^*) + \mathcal{N}_X(x^*).$$

By [18, Theorems 2.5] (or [25, Theorem 1.43 (iii)]),

$$\overline{\lim_{\delta\downarrow 0}} \mathbb{E}_P[\partial_x^\delta f(x^*, \xi)] \subset \mathbb{E}_P\left[\overline{\lim_{\delta\downarrow 0}} \partial_x^\delta f(x^*, \xi) \right] = \mathbb{E}_P[\partial_x f(x^*, \xi)].$$

The last equality is due to the fact that $\overline{\lim}_{\delta\downarrow 0} \partial_x^\delta f(x^*, \xi) = \partial_x f(x^*, \xi)$. Using (38) and the discussions above, we can easily obtain (40) and hence

$$0 \in \mathbb{E}_P[\partial_x f(x^*, \xi)] + \rho \partial \theta(x^*) + \mathcal{N}_X(x^*) = \partial \mathbb{E}_P[f(x^*, \xi)] + \rho \partial \theta(x^*) + \mathcal{N}_X(x^*).$$

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

Note that the Clarke regularity condition used in the theorem may be replaced by other conditions. For instance, if the Lipschitz modulus in Assumption 3.11 is bounded by positive constant and for a small positive constant τ_0, $\frac{1}{\tau}(f(x + \tau u, \xi) - f(x, \xi))$

is uniformly continuous w.r.t. ξ for all $x \in X$, $u \in \mathbb{R}^n$ with $\|u\| \leq 1$ and $\tau \in (0, \tau_0]$, then we may apply [23, Lemma 5.2] to show

$$\lim_{N \to \infty} \sup_{x \in X} \mathbb{D}(\partial f_N(x), \partial \mathbb{E}_P[f(x, \xi)]) \to 0$$

and hence (39). We omit the details.

Note also that it might be interesting to ask whether a stationary point of problem (11) is a stationary point of problem (5). To answer this question, we need to consider the first order optimality conditions for the latter problem. 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 optimality conditions:

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

We say a point x^* is a stationary point of (5) if there exists $\lambda^* > 0$ such that (x^*, λ^*) satisfies (41). To justify this definition, we show that every local optimal solution to problem (5) satisfies (41) (along with some positive number λ). In the case when $\mathbb{E}_P[f(x, \xi)]$ is a convex function, a point satisfying optimality conditions (41) is a global optimal solution to problem (5). In what follows, we verify this. 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). This shows (x^*, ρ) satisfies optimality condition (41). Conversely if x^* is a stationary point, that is, there exists positive number λ^* such that (x^*, λ^*) satisfies optimality conditions (41). 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 problem (5).

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

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

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 of problem (5), then there exists $\mu^* \in \mathscr{M}_+(T)$ such that

$$
\begin{cases}
0 \in \partial_x \mathscr{L}(x, \mu) = \partial \mathbb{E}_P[f(x^*, \xi)] + \int_T \partial_x \mathbb{E}_P[H(x^*, t, \xi)]\mu^*(dt) + \mathcal{N}_X(x^*), \\
\mathbb{E}_P[H(x^*, t, \xi)] \leq 0, \quad \forall t \in T, \\
\int_T \mathbb{E}_P[H(x^*, t, \xi)]\mu^*(dt) = 0, \\
x \in X,
\end{cases}
\tag{42}
$$

see [12, Theorem 4] for details and [12, Definition 2] for the definition of the differential constraint qualification. Note that optimality conditions (42) can also be alternatively characterized by some convex functions defined over $\mathbb{R}$. This can be done by representing the integral w.r.t. measure μ by some convex functions through the Riesz representation theorem, see [8,9] for details. It is an open question as to whether there is some relationship between (41) and (42) or the equivalent conditions of (42) in [8,9], and this will be the focus of our future work.

References

1. Artstein, Z., Vitale, R.A.: A strong law of large numbers for random compact sets. Ann. Probab. **3**, 879–882 (1975)
2. Artstein, Z., Wets, R.J.-B.: Multivar. Anal. Approximating the integral of a multifunction **24**, 285–308 (1988)
3. Birge, J.R., Qi, L.: Subdifferential convergence in stochastic programs. SIAM J. Optim. **5**, 436–453 (1995)
4. Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer, New York (2000)
5. Clarke, F.H.: Optimization and Nonsmooth Analysis. Wiley, New York (1983)
6. Dai, L., Chen, C.H., Birge, J.R.: Convergence properties of two stage stochastic programming. J. Optim. Appl. **106**, 489–509 (2000)
7. Dentcheva, D., Henrion, R., Ruszczynski, A.: Stability and sensitivity of optimization problems with first order stochastic dominance constraints. SIAM J. Optim. **18**, 322–333 (2007)
8. Dentcheva, D., Ruszczyński, A.: Optimization with stochastic dominance constraints. SIAM J. Optim. **14**, 548–566 (2003)
9. Dentcheva, D., Ruszczyński, A.: Semi-infinite probabilistic constraints: optimality and convexification. Optimization **53**, 583–601 (2004)
10. Dentcheva, D., Ruszczyński, A.: Optimality and duality theory for stochastic optimization with nonlinear dominance constraints. Math. Program. **99**, 329–350 (2004)
11. Dentcheva, D., Ruszczyński, A.: Portfolio optimization with stochastic dominance constraints. J. Banking Finance **30**, 433–451 (2006)
12. Dentcheva, D., Ruszczyński, A.: Composite semi-infinite optimization. Control Cybern. **36**, 1–14 (2007)
13. Gugat, M.: Error bounds for infinite systems of convex inequalities without Slater's condition. Math. Program. Ser. B **88**, 255–275 (2000)
14. Henrion, R., Römisch, W.: Metric regularity and quantitative stability in stochastic programs with probabilistic constraints. Math. Program. **84**, 55–88 (1999)
15. Henrion, R., Römisch, W.: Hölder and lipschitz stability of solution sets in programs with probabilistic constraints. Math. Program. **100**, 589–611 (2004)

16. Homen-De-Mello, T.: On rates of convergence for stochastic optimization problems under non-independent and identically distributed sampling. SIAM J. Optim. **19**, 524–551 (2008)
17. Homem-de-Mello, T., Mehrotra, S.: A cutting surface method for uncertain linear programs with polyhedral stochastic dominance constraints. SIAM J. Optim. **20**, 1250–1273 (2009)
18. Hiai, F.: Convergence of conditional expectations and strong law of large numbers for multivalued random variables. Trans. Am. Math. Soc. **291**, 613–627 (1985)
19. Hu, J., Homem-de-Mello, T., Mehrotra, S.: Sample average approximation for stochastic dominance constrained programs. Math. Program. **133**, 171–201 (2012)
20. Klatte, D.: On the Stability of Local and Global Optimal Solutions in Parametric Problems of Nonlinear Programming, Part I: Basic Results, Seminarbericht Nr. 75, Sektion Mathematik, pp. 1–21. Humboldt-Universität zu Berlin, Berlin (1985)
21. Klatte, D.: A Note on Quantitative Stability Results in Nonlinear Optimization, Seminarbericht Nr. 90, Sektion Mathematik, pp. 77–86. Humboldt-Universität zu Berlin, Berlin (1987)
22. Levy, H.: Stochastic dominance and expected utility: survey and analysis. Manag. Sci. **38**, 555–593 (1992)
23. Liu, Y., Xu, H., Lin, G.-H.: Stability analysis of two stage stochastic mathematical programs with complementarity constraints via NLP-regularization. SIAM J. Optim. **21**, 609–705 (2011)
24. Meskarian, R., Xu, H., Fliege, J.: Numerical methods for stochastic programs with second order dominance constraints with applications to portfolio optimization. Eur. J. Oper. Res. **216**, 376–385 (2012)
25. Molchanov, I., et al.: Theory of random sets. In: Gani, J. (ed.) Probability and Its Applications. Springer, Berlin (2005)
26. Müller, A., Scarsini, M. (eds.): Stochastic Orders and Decision Under Risk. Institute of Mathematical Statistics, Hayward (1991)
27. Polak, E.: On the mathematical fondations of nondifferentiable optimization in engineering design. SIAM Rev. **29**, 21–89 (1987)
28. Price, C.J., Cooper, I.D.: An exact penalty function algorithm for semi-infinite programmes. BIT Numer. Math. **30**, 723–734 (1990)
29. Rachev, S.T., Römisch, W.: Quantitative stability in stochastic programming: the method of probability metrics. Math. Oper. Res. **27**, 792–818 (2002)
30. Robinson, S.M.: An application of error bounds for convex programming in a linear space. SIAM J. Cont. Optim. **13**, 271–273 (1975)
31. Robinson, S.M.: Stability theory for systems of inequalities. Part II: Differentiable nonlinear systems. SIAM J. Numer. Anal. **13**, 497–513 (1976)
32. Robinson, S.M., Wets, R.J.-B.: Stability in two-stage stochastic programming. SIAM J. Cont. Optim. **25**, 1409–1416 (1987)
33. Robinson, S.M.: Analysis of sample-path optimization. Math. Oper. Res. **21**, 513–528 (1996)
34. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin (1998)
35. Römisch, W.: Stability of stochastic programming problems. In: Ruszczyński, A., Shapiro, A. (eds.) Stochastic Programming, pp. 483–554. Elsevier, Amsterdam (2003)
36. Römisch, W., Schultz, R.: Stability analysis for stochastic programs. Ann. Oper. Res. **30**, 241–266 (1991)
37. Ruszczyński, A., Shapiro, A.: Stochastic Programming, Handbook in Operations Research and Management Science. Elsevier, Amsterdam (2003)
38. Shapiro, A.: Asymptotic properties of statistical estimators in stochastic programming. Ann. Stat. **17**, 841–858 (1989)
39. Shapiro, A.: Quantitative stability in stochastic programming. Math. Program. **67**, 99–108 (1994)
40. Shapiro, A., Xu, H.: Uniform laws of large numbers for set-valued mappings and subdifferentials of random functions. J. Math. Anal. Appl. **325**, 1390–1399 (2007)
41. Shapiro, A., Xu, H.: Stochastic mathematical programs with equilibrium constraints, modeling and sample average approximation. Optimization **57**, 395–418 (2008)
42. Tanaka, Y., Fukushima, M., Ibaraki, T.: A globally convergent SQP method for semi-infinite nonlinear optimization. J. Comput. Appl. Math. **23**, 141–153 (1988)
43. Xu, H.: Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming. J. Math. Anal. Appl. **368**, 692–710 (2010)
44. Xu, H., Zhang, D.: Smooth sample average approximation of stationary points in nonsmooth stochastic optimization and applications. Math. Program. **119**, 371–401 (2009)