AN APPROXIMATION SCHEME FOR STOCHASTIC PROGRAMS WITH SECOND ORDER DOMINANCE CONSTRAINTS

YONGCHAO LIU*

Department of Mathematics
Dalian Maritime University, Dalian 116026, China
and
School of Mathematics, University of Southampton
SO17 1BJ, Southampton, UK

HAILIN SUN

School of Economics and Management
Nanjing University of Science and Technology
Nanjing, 210049, China

HUIFU XU

School of Mathematics, University of Southampton
SO17 1BJ, Southampton, UK

(Communicated by Song Wang)

Abstract. It is well known that second order dominance relation between two random variables can be described by a system of stochastic semi-infinite inequalities indexed by $\mathcal{R}$, the set of all real numbers. In this paper, we show the index set can be reduced to the support set of the dominated random variable strengthening a similar result established by Dentcheva and Ruszczyński [9] for discrete random variables. Viewing the semi-infinite constraints as an extreme robust risk measure, we relax it by replacing it with entropic risk measure and regarding the latter as an approximation of the former in an optimization problem with second order dominance constraints. To solve the entropic approximation problem, we apply the well known sample average approximation method to discretize it. Detailed analysis is given to quantify both the entropic approximation and sample average approximation for various statistical quantities including the optimal value, the optimal solutions and the stationary points obtained from solving the sample average approximated problem. The numerical scheme provides an alternative to the mainstream numerical methods for this important class of stochastic programs.

2010 Mathematics Subject Classification. Primary: 90C15, 90C31.

Key words and phrases. Second order dominance, entropic approximation, stochastic semi-infinite programming, sample average approximation.

The work of the first author is supported by NSFC grant 11571056, LNSFC grant 2014025019 and FRFCU grant 3132016220. The work of the second author is supported by NSFC grant 11401308 and JSSFC grant BK20140768.

* Corresponding author: Yongchao Liu.
1. Introduction. Stochastic optimization models with stochastic second order dominance (SSD) constraints introduced by Dentcheva and Ruszczyński [9] provide important optimization models with effective constraints of risk control. Over the past decade, there has been increasing research on the subject ranging from the theory of optimality and duality to numerical methodology and practical applications, see [7, 8, 9, 10, 12, 16, 17, 21, 23] and references therein.

Stochastic dominance is a fundamental concept in decision theory and economics [26]. For random variable \(\xi(\omega)\) with finite first moment, denote \(F_1(\xi(\omega), s) := \text{Prob}\{\xi(\omega) \leq s\}\) and \(F_2(\xi(\omega), t) := \int_{-\infty}^t F_1(\xi(\omega), s) ds\). Then \(\xi_1(\omega)\) is said to do min\(\xi_2(\omega)\) in the second order, denoted by \(\xi_1(\omega) \succeq_2 \xi_2(\omega)\), if

\[
F_2(\xi_1(\omega), t) \leq F_2(\xi_2(\omega), t) \quad \forall t \in \mathcal{R}. 
\]

The objective of this paper is to propose an approximation scheme for the following stochastic program with stochastic second order dominance constraints

\[
\begin{align*}
\min_{x \in X} \quad & E[f(x, \xi(\omega))] \\
\text{s.t.} \quad & G(x, \xi(\omega)) \succeq_2 Y(\xi(\omega)),
\end{align*}
\]

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

By changing the order of integration, (1) may be reformulated as:

\[
E[(t - \xi_1(\omega))_+] \leq E[(t - \xi_2(\omega))_+], \quad \forall t \in \mathcal{R}. 
\]

Here and later on we write \((a)_+\) for \(\max\{a, 0\}\). Then, problem (2) can be expressed as a stochastic semi-infinite programming (SSIP) problem:

\[
\begin{align*}
\min_{x \in X} \quad & E[f(x, \xi)] \\
\text{s.t.} \quad & E[(t - G(x, \xi))_+] \leq E[(t - Y(\xi))_+], \quad \forall t \in \mathcal{R}.
\end{align*}
\]

It is well-known that the above SSIP problem does not satisfy the well-known Slater’s constraint qualification, a condition that a stable numerical method often relies on. Consequently, a so-called relaxed form of the SSIP is proposed [9]:

\[
\begin{align*}
\min_{x \in X} \quad & E[f(x, \xi)] \\
\text{s.t.} \quad & E[H(x, t, \xi)] \leq 0, \quad \forall t \in T,
\end{align*}
\]

where

\[
H(x, t, \xi) := (t - G(x, \xi))_+ - (t - Y(\xi))_+
\]

and \(T\) is a compact subset of \(\mathcal{R}\).

The following question is whether there is a subset \(T\) of \(\mathcal{R}\) such that the semi-infinite constraints of problems (4) and (5) are equivalent, in order words, the index set \(\mathcal{R}\) can be reduced to \(T\) without affecting the semi-infinite constraints. This question has been partly addressed by Dentcheva and Ruszczyński [9] where they show the equivalence with \(T = Y(\xi)\) when \(\Xi\) is a finite set.

Here we first extend their result to the case when \(\xi\) is continuously distributed and then propose an approximation scheme fundamentally based on entropic risk measure to approximate the stochastic semi-infinite constraints. Specifically, by treating \(t\) as a random parameter, we use entropic risk measure of \(E[H(x, t, \xi)]\) to
approximate the essential supremum of $\mathbb{E}[H(x, t, \xi)]$ (w.r.t. $t$) and then discretize the latter (which has two layers of mathematical expectations w.r.t. $\xi$ and $t$) by the well known sample average approximation method. The main advantages of the proposed method are two-fold: (a) instead of discretizing $\xi$ and $t$ separately, we do through a single sampling on $\xi$ and then transfer the discretization to $t$; (b) the resulting optimization problem has a single deterministic constraint. To ensure the approximation schemes are theoretically founded, we carry out detailed convergence analysis for the statistical quantities of the optimal value, the optimal solutions and the stationary points obtained from solving the each of the approximated regimes.

The rest of the paper are structured as follows. Section 2 provides a theoretical background for setting index $T$ as the image of random variable $Y(\xi)$ and give sketches of entropic approximation to the semi-infinite constraints. Section 3 details the stability analysis of entropic approximated problem against variation of the entropic parameter. Section 4 discusses sample average approximation of the entropic approximated program and investigates convergence of the optimal value, the optimal solutions, and the stationary points obtained from the sampled averaged approximated regime as sample size increases. Section 5 reports some numerical test results.

Throughout this paper, we use the following notation. For vectors $a, b \in \mathbb{R}^n$, we denote by $a^T b$ for the scalar product, $\| \cdot \|$ for the Euclidean norm of a vector, and $\| \cdot \|_{\infty}$ for the maximum norm of continuous functions. When $D$ is set of $\mathbb{R}^n$, $\|D\| := \sup_{D \in D} \|D\|$. We write $d(x, D)$ for the distance from a point $x$ to a set $D$ and $B(x, \delta)$ for the closed ball centered at $x$ with radius $\delta$, in particular $B$ represents the closed unit ball in a finite dimensional space. For two compact sets $C$ and $D$, $\mathbb{D}(C, D) := \sup_{x \in C} d(x, D)$ denotes the deviation of $C$ from $D$ and $\mathbb{H}(C, D) := \max \{ \mathbb{D}(C, D), \mathbb{D}(D, C) \}$ the Hausdorff distance between $C$ and $D$, $C + D$ means the Minkowski addition of the two sets, i.e., $\{ C + D : C \in C, D \in D \}$. It is well known that for compact sets $A, B, C, D \subseteq \mathbb{R}^m$,

$$\mathbb{D}(A + C, B + D) \leq \mathbb{D}(A, B) + \mathbb{D}(C, D). \quad (6)$$

For a sequence of subsets $\{C_k\}$ in a metric space, we follow the standard notation \cite{30} by using $\limsup_{k \to \infty} C_k$ and $\liminf_{k \to \infty} C_k$ to denote its outer limit and inner limit, that is,

$$\limsup_{k \to \infty} C_k = \{ x : \liminf_{k \to \infty} d(x, C_k) = 0 \}, \quad \liminf_{k \to \infty} C_k = \{ x : \limsup_{k \to \infty} d(x, C_k) = 0 \}.$$ 

Finally, for a set-valued mapping $A : X \to 2^Y$, $A(\cdot)$ is said to be outer semicontinuous at $\bar{x} \in X$ if $\limsup_{x \to \bar{x}} A(x) \subseteq A(\bar{x})$, see \cite[Chapter 5]{30}.

2. Reformulation. In this section, we first address the question as to whether we can reduce the index set $\mathcal{R}$ to a compact set $T$ in the SSIP without affecting the semi-infinite constraints. The next proposition gives an affirmative answer and says that $T$ can be indeed chosen as the image of $Y(\xi)$ over support set $\Xi$. It is an extension from an earlier result by Dentcheva and Ruszczyński \cite{9} where $\xi$ is a discretely distributed random variable.

**Proposition 1.** Let $G$ and $Y$ be random variables with finite expected values and $Y(\Xi) := \{ Y(\xi) : \xi \in \Xi \}$. Then the semi-infinite system

$$\mathbb{E}[(t - G)_+] \leq \mathbb{E}[(t - Y)_+], \quad t \in \mathcal{R} \quad (7)$$

...
is equivalent to the semi-infinite system
\[ \mathbb{E}[(t - G)_+] \leq \mathbb{E}[(t - Y)_+], \quad t \in Y(\Xi), \] (8)
that is, the index set in system (7) can be effectively reduced to \( Y(\Xi) \) without affecting its solutions.

**Proof.** Given the reformulation (3) of second order dominance, it suffices to show that
\[ F_2(G, t) \leq F_2(Y, t), \quad \forall t \in \mathcal{R} \setminus Y(\Xi). \]

We do so by going through three cases:

Case 1. There exists \( a := \inf Y(\Xi) > -\infty \). We first show that \( F_2(G, a) \leq F_2(Y, a) \). If \( a \notin Y(\Xi) \), then we may find a sequence \( \{a_k\} \subseteq Y(\Xi) \) such that \( a_k \downarrow a \).

Since \( F_2(Y, a_k) \geq F_2(G, a_k) \geq F_2(G, a) \)
and \( F_2(Y, \cdot) \) is continuous, by taking a limit on both sides of the inequality above, we arrive at
\[ F_2(Y, a) \geq F_2(G, a). \]

For \( t < a \),
\[ 0 \leq F_2(G, t) \leq F_2(G, a) \leq F_2(Y, a) = 0 = F_2(Y, t). \]

Case 2. \( b = \sup Y(\Xi) < \infty \). We first show \( F_2(G, b) \leq F_2(Y, b) \). If \( b \notin Y(\Xi) \), then there exists a sequence \( \{b_k\} \subseteq Y(\Xi) \) such that \( b_k \uparrow b \). Since \( F_2(Y, b_k) \geq F_2(Y, b) \geq F_2(G, b_k) \),
and \( F(G, \cdot) \) is continuous, then
\[ F_2(Y, b) \geq F_2(G, b). \]

For \( t > b \),
\[ F_2(Y, t) = F_2(Y, b) + t - b \geq F_2(G, b) + \int_b^t F_1(G, t) dt = F_2(G, t). \]

Case 3. \( t \notin Y(\Xi) \) and \( \inf Y(\Xi) < t < \sup Y(\Xi) \). In that case, there exist two adjacent convex subsets of \( Y(\Xi) \), denoted by \( T_1 \) and \( T_2 \), such that \( T_1 \cap T_2 = \emptyset \) and \( t_1 := \sup T_1 \leq t \leq t_2 := \inf T_2 \).

We first show
\[ F_2(G, t_1) \leq F_2(Y, t_1) \quad \text{and} \quad F_2(G, t_2) \leq F_2(Y, t_2). \]

If \( t_1, t_2 \notin Y(\Xi) \), we may verify the formulae above by similar analysis of case 1 for \( t_2 \) and case 2 for \( t_1 \). For \( t \in (t_1, t_2) \), we have by the convexity of \( F_2(G, \cdot) \) that, for any \( 0 \leq \lambda \leq 1 \),
\[ F_2(G, t) \leq \lambda F_2(G, t_1) + (1 - \lambda)F_2(G, t_2) \leq \lambda F_2(Y, t_1) + (1 - \lambda)F_2(Y, t_2) = F_2(Y, t), \]
the equality holds by taking \( \lambda = (t_2 - t)/(t_2 - t_1) \). Summarizing the analysis above, we conclude that
\[ F_2(Y, t) \geq F_2(G, t), \quad \forall t \in \mathcal{R} \setminus Y(\Xi) \]
holds. The proof is complete. \( \square \)

In what follows, we focus on the case where \( T = Y(\Xi) \) is a compact set. Let \( h(x, t) := \mathbb{E}[H(x, t, \xi)] \), we may rewrite problem (5) as
\[
\begin{align*}
\min_{x \in X} & \quad \mathbb{E}[f(x, \xi)] \\
\text{s.t.} & \quad h(x, t) \leq 0, \quad \forall t \in T,
\end{align*}
\] (9)
and further as

\[
\min_{x \in X} \mathbb{E}[f(x, \xi)] \quad \text{s.t.} \quad \sup_{t \in T} h(x, t) \leq 0. \tag{10}
\]

Of course, writing the semi-infinite constraints as above does not bring us any numerical convenience. Our plan here is to consider an approximation of \( \sup_{t \in T} h(x, t) \) through entropic risk measure so that the constraint can be handled numerically more conveniently. This kind of approximation scheme was recently considered by Anderson et al. [1] using CVaR and Liu and Xu with entropic approximation [25]. The rationale behind the scheme is that if we regard \( \sup_{t \in T} h(x, t) \) as an extremely robust risk measure, then the CVaR of \( h(x, t) \) would be a relaxation depending on the confidence level to be set. This kind of relaxation might not significantly affect the optimal decision through stability analysis but does allow one to take a few more samples near the extremum to calculate the CVaR (as opposed to a single sample at the extremely robust constraint) and in that way “smooth up” or “stabilize” the numerical calculation. Here we take a similar initiative but adopt entropic risk measure approximation in that the latter is a smooth function and fits to a broader class of random functions.

For a random variable \( Z \in L^{\infty} \), the entropic risk measure is defined as

\[
e_{\gamma}(Z) := \rho(Z) = \frac{1}{\gamma} \ln \mathbb{E}[e^{-\gamma Z}].
\]

It is well known (see e.g. [13, formulation (4)]) that \( e_{\gamma}(Z) \) is monotonically increasing in \( \gamma \) and

\[
\lim_{\gamma \to \infty} e_{\gamma}(Z) = \text{ess sup}(-Z)
\]

in the case when \( \text{ess inf} Z > -\infty \), where \( \text{ess sup} \) (inf) denotes the essential supremum (infimum). The following lemma states how entropic risk measure approximation works for a general random function.

**Lemma 2.1.** [25, Proposition 2.1] Let \( h : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R} \) be a continuous function and \( X \) be a subset of \( \mathbb{R}^n \). Let \( \xi : \Omega \to \mathbb{R}^k \) be a random variable defined on the probability space \( (\Omega, \mathcal{F}, P) \) with support set \( \Xi \). Let \( \tau(x) \), \( F(\cdot) \) and \( \Omega_x \) denote respectively the essential supremum, the cumulative distribution function and the support set of \(-h(x, \xi)\). Let \( \text{Diam}(\Omega_x) \) denote the diameter of the support set \( \Omega_x \) which is the distance between \( \tau(x) \) and essential infimum of \(-h(x, \xi)\). Assume: (a) \( X \subset \mathbb{R}^n \) is a compact set, (b) for each fixed \( x \in X \),

\[
\inf_{\xi \in \Xi} h(x, \xi) > -\infty.
\]

Then for each fixed \( x \in X \),

\[
\lim_{\gamma \to \infty} e_{\gamma}(h(x, \xi)) = \tau(x).
\]

Assume in addition that (c)

\[
\inf_{x \in X} \inf_{\xi \in \Xi} h(x, \xi) > -\infty,
\]

and (d) for any fixed small positive number \( \epsilon \), there exists \( \delta(\epsilon) \in (0, 1) \) such that

\[
1 - F(\tau(x) - \epsilon) \geq \delta(\epsilon), \quad \forall x \in X, \epsilon,
\]
where \( X := \{ x \in X : \text{Diam}(\Omega_x) > 2\varepsilon \} \). Then

\[
|e_\gamma(h(x, \xi)) - \tau(x)| < 2\varepsilon + \frac{1}{\gamma} \ln \delta(\varepsilon).
\]

Following Lemma 2.1, we may consider an approximation of problem (10) by

\[
\min_{x \in X} \mathbb{E}[f(x, \xi)] \\
\text{s.t. } e_\gamma(x) \leq 0,
\]

where

\[
e_\gamma(x) := e_\gamma(-h(x, t)) = \frac{1}{\gamma} \log \left( \int_T e^{\gamma h(x, t)} P_0(dt) \right).
\]

Here and later on, \( P_0 \) is the distribution of \( Y(\xi) \) induced by the distribution \( P \) of \( \xi \).

3. Stability Analysis. We begin by looking into the impact of parameter \( \gamma \) on the optimal value and the optimal solutions of problem (11) as \( \gamma \) increases. This may be covered by classical stability analysis of parametric programming. Here we include some detailed analysis in that we may exploit the specific structure of the entropic approximation to establish some useful quantitative analysis results.

3.1. Optimal value and optimal solutions. Let \( \mathcal{F}(\gamma) \) and \( \mathcal{F} \) denote the set of feasible solutions to problems (11) and (9) (or (10)) respectively. Since \( e_\gamma(x) \) is monotonically increasing in \( \gamma \) and \( \lim_{\gamma \to \infty} e_\gamma(x) \to \sup_{t \in T} h(x, t) \), \( \mathcal{F} \subseteq \mathcal{F}(\gamma) \).

The following proposition presents a quantitative relationship between of the two feasible sets under Hausdorff distance.

Lemma 3.1. Let \( \zeta \) be a continuously distributed random variable with support set \( T \). Assume: (a) \( X \) is a compact set, (b) \( \inf_{x \in X} \inf_{t \in T} h(x, t) > -\infty \), and for any fixed small positive number \( \varepsilon \), there exists \( \delta(\varepsilon) \in (0, 1) \) such that

\[
1 - F(\tau(x) - \varepsilon) \geq \delta(\varepsilon), \quad \forall x \in X_S
\]

where \( F(\cdot) \) denotes the cumulative distribution function of \( -h(x, \zeta) \), \( X_S := \{ x \in X, \text{Diam}(\Omega_x) \leq \varepsilon \} \) and \( \Omega_x \) is the support set of random \( h(x, t) \), (c) there exist a positive number \( \delta \) and a point \( \bar{x} \in X \) such that

\[
\max_{t \in T} h(\bar{x}, t) \leq -\delta.
\]

Then there exists a positive number \( \tau^* \) such that

\[
\mathbb{H}(\mathcal{F}, \mathcal{F}(\gamma)) \leq \tau^* \Delta_\gamma,
\]

where

\[
\Delta_\gamma := \sup_{x \in X} \sup_{t \in T} h(x, t) - e_\gamma(x),
\]

and \( \Delta_\gamma \to 0 \) as \( \gamma \to \infty \).

Proof. Observe that \( \mathcal{F} \subseteq \mathcal{F}(\gamma) \). Thus by the definition of \( \mathbb{H} \), we only need to estimate \( D(\mathcal{F}(\gamma), \mathcal{F}) \). Since \( G(\cdot, \xi) \) is concave and \( h(x, t) \) is convex in \( x \), we deduce that condition (13) is equivalent to the well known Slater condition. Then though [14, Lemmas 3 and 6], we have

\[
d(x, \mathcal{F}) \leq \tau^* \| (h(x, t))_+ \|_\infty,
\]
where $\tau^* = \sup_{x \in X} \frac{\|x - \bar{x}\|}{\tau}$. The boundedness of $\tau^*$ is ensured by the compactness of the set $X$. Consequently,

$$d(x, F) \leq \tau^* \| (h(x, t))_+ \|_\infty \leq \tau^*(\sup_{t \in T} h(x, t) - e_\gamma(x)), \quad \forall x \in F(\gamma),$$

which implies

$$\mathbb{E}(F(\gamma), F) \leq \tau^* \Delta_\gamma.$$ 

Moreover, under conditions (a)-(b), it follows by [25, Lemma 2.1] that $\Delta_\gamma \to 0$ as $\gamma$ tends to infinity. The proof is complete. \hfill \Box

Lemma 3.1 says that the feasible set $F(\gamma)$ converges to $F$ in Hausdorff distance as $\gamma$ tends to infinity. The condition (12) is similar to the so-called consistent tail behavior condition for CVaR approximation of the essential supremum of a random function in [1], see [25] for more details where a simple example is presented. Condition (13) is needed to ensure the quantitative estimation (14). Using Lemma 3.1, we can present the stability of the optimal value and the optimal solutions to problem (11).

**Theorem 3.2.** Let $X$ be a compact set, $X(\gamma)$ and $X^*$ denote the sets of optimal solutions of problems (11) and (10) respectively, $v(\gamma)$ and $v^*$ be the corresponding optimal values. Then

1. $\lim_{\gamma \to \infty} v(\gamma) = v^*$ and $\limsup_{\gamma \to \infty} X(\gamma) \subseteq X^*$;
2. if, in addition, problem (9) satisfies condition (13) and $f(\cdot, \xi)$ is Lipschitz continuous with integrably bounded modulus $L(\xi)$, then there exists a positive number $K$ such that

$$|v(\gamma) - v^*| \leq K\Delta_\gamma,$$

where $\Delta_\gamma$ is defined as in Lemma 3.1.

**Proof.** The proof is similar to [25, Theorem 1]. Here we give some sketches for completeness.

Part (i). It is sufficient to show convergence of optimal solutions in that the objective function $\mathbb{E}[f(x, \xi)]$ is continuous and independent of $\gamma$. Assume for a contradiction that there exists $x_k \in X(\gamma_k)$ such that $x_k \to x^*$ as $k \to \infty$ and $x^* \notin X^*$. Then $x^* \in F$ and there exists $\bar{x} \in X^*$ such that

$$\mathbb{E}[f(x^*, \xi)] - \mathbb{E}[f(\bar{x}, \xi)] > 0. \quad (15)$$

Since $\bar{x} \in F \subset F(\gamma)$, $\mathbb{E}[f(x_k, \xi)] - \mathbb{E}[f(\bar{x}, \xi)] \leq 0$. Driving $k$ to infinity, we arrive at $\mathbb{E}[f(x^*, \xi)] - \mathbb{E}[f(\bar{x}, \xi)] \leq 0$, which contradicts with (15).

Part (ii). Let $x_\gamma$ and $x_0$ be the optimal solutions of problems (11) and (10) respectively. By the continuity of the feasible set $F(\gamma)$ under Hausdorff distance as demonstrated in Lemma 3.1, there exists $\bar{x}_\gamma \in F$ such that $\|x_\gamma - \bar{x}_\gamma\| \leq \tau^* \Delta_\gamma$. Moreover

$$v^* \leq \mathbb{E}[f(\bar{x}_\gamma, \xi)] = \mathbb{E}[f(x_\gamma, \xi)] + |\mathbb{E}[f(\bar{x}_\gamma), \xi] - \mathbb{E}[f(x_\gamma, \xi)]| \leq v(\gamma) + K \Delta_\gamma,$$

where $K := \tau^* \mathbb{E}[L(\xi)]$. Swapping the position between $F(\gamma)$ and $F$ under Hausdorff distance, we obtain $v(\gamma) \leq v^* + K \Delta_\gamma$. The conclusion follows. \hfill \Box
3.2. Stationary points. We now move on to investigate the case when we only obtain a stationary point rather than an optimal solution from solving the entropic approximation problem (11). This is motivated to address the case when $f(x, \xi)$ is not convex in $x$. We start by defining the stationary points of problems (10) and (11).

Let $\psi : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz continuous function. Recall that Clarke subdifferential of $\psi$ at $x$, denoted by $\partial \psi(x)$, is defined as follows:

$$\partial \psi(x) := \text{conv} \{ \lim_{x' \to x} \nabla \psi(x') \},$$

where $D$ denotes the set of points near $x$ at which $\psi$ is Fréchet differentiable, $\nabla \psi(x)$ denotes the gradient of $\psi$ at $x$ and 'conv' denotes the convex hull of a set, see [6] for details.

Let

$$\theta(x) := \max_{t \in T} h(x, t) \quad \text{and} \quad T^*(x) = \arg \max_{t \in T} h(x, t). \quad (16)$$

Since $G(\cdot, \xi)$ is concave, $h(x, t)$ is convex in $x$ and hence it is Clarke regular (see [6, Proposition 2.3.6]). By [6, Theorem 2.8.2],

$$\partial \theta(x) = \left\{ \int_T \partial_x h(x, t) \mu(dt) : \mu \in \mathcal{P}[T^*(x)] \right\}, \quad (17)$$

where $\mathcal{P}[S]$ signifies the collection of probability measures supported on $S$ and the integral is in the sense of Aumann [2]. By [31, Chapter 2 Theorem 9],

$$\partial_x h(x, t) = \int_{\Xi} \partial_x H(x, t, \xi) P(d\xi)$$

and

$$\partial_x H(x, t, \xi) = \begin{cases} 0, & G(x, \xi) > t, \\ \text{conv}\{0, -\nabla_x G(x, \xi)\}, & G(x, \xi) = t, \\ -\nabla_x G(x, \xi), & G(x, \xi) < t. \end{cases} \quad (18)$$

Obviously, for any $\xi \in \Xi$, $\partial_x H(x, t, \xi)$ is outer semicontinuous in $(x, t)$ and so does $\partial_t h(x, t)$.

A feasible point $x$ is said to be a stationary point of problem (10) if there exists $\rho$ such that

$$0 \in \partial \mathbb{E}[f(x, \xi)] + \rho \partial \theta(x) + \mathcal{N}_X(x),\quad (19)$$

$$0 \leq \rho \perp -\theta(x) \geq 0,\quad (20)$$

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

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

$$T_X(x) = \lim_{\tau \to 0, X \ni x' \to x} \frac{1}{\tau}(X - x')$$

and $\mathcal{N}_X(x) = \emptyset$ when $x \notin X$ and ”$\perp$” denotes perpendicularity of two vectors.

Likewise, a feasible point $x$ is said to be a stationary point of problem (11) if there exists $\rho_\gamma$ such that

$$0 \in \partial \mathbb{E}[f(x, \xi)] + \rho_\gamma \partial e_\gamma(x) + \mathcal{N}_X(x), \quad (21)$$

$$0 \leq \rho_\gamma \perp -e_\gamma(x) \geq 0. \quad (22)$$
Since logarithmic function \(\log(\cdot)\) is strictly differentiable and \(h(x, t)\) is convex in \(x\), by [6, Theorems 2.3.9 and 2.7.2],

\[
\partial e_\gamma(x) = \int_T \mu(x, t; \gamma) \partial_x h(x, t) P_0(dt)
\]

where

\[
\mu(x, t; \gamma) := \frac{e^{\gamma h(x, t)}}{\int_T e^{\gamma h(x, t)} P_0(dt)}.
\]

In what follows, we show the stationary point satisfying (21)-(22) converges to the stationary point satisfying (19)-(20). To this end, we need the following intermediate result which characterizes the asymptotic convergence of subdifferential \(\partial e_\gamma(x^k)\) as \(k \to \infty\).

**Proposition 2.** Let \(\gamma_k \to \infty\) and \(\{x^k\} \in X\) be a sequence converging to \(\bar{x}\). Suppose that the Lipschitz modulus of \(f(\cdot, \xi)\) and \(G(\cdot, \xi)\) are bounded by a positive constant \(L\). Then

\[
\lim_{k \to \infty} \mathbb{D}(\partial e_{\gamma_k}(x^k), \partial \theta(\bar{x})) \to 0,
\]

if one of the following conditions holds:

(a) \(T^*(\bar{x})\) encompasses finite points, where \(T^*(x)\) is defined as in (16);

(b) \(T^*(\bar{x})\) is not a finite set but the probability \(P_0\{T^*(\bar{x})\}\) is strictly positive and \(\partial_x h(\cdot, t)\) is outer semicontinuous at \(x\) uniformly for \(t \in T^*(\bar{x})\).

**Proof.** By the definition of \(h(x, t)\), it is easy to show that \(h(\cdot, t)\) is Lipschitz continuous over \(X\) with modulus \(L\) under the assumption that the Lipschitz modulus of \(f(\cdot, \xi)\) and \(G(\cdot, \xi)\) is bounded by \(L\). Thus

\[
\|\partial_x h(x, t)\| \leq L.
\]

Condition (a). Let us label the points in \(T^*(\bar{x})\) as \(t_1, \cdots, t_M\). Let \(\epsilon\) be a small positive number. Since \(\partial_x h(x, t)\) is outer semicontinuous in \((x, t)\), there exists a positive number \(\delta\) such that

\[
\mathbb{D}(\partial_x h(x, t), \partial_x h(\bar{x}, t_i)) \leq \epsilon, \quad \forall (x, t) \in B(\bar{x}, \delta) \times T(t_i, \delta),
\]

where \(T(t_i, \delta)\) denote the \(\delta\) neighborhood of \(t_i\). Since there are finite number of points in \(T^*(\bar{x})\), we may set \(\delta\) sufficiently small such that \(T(t_i, \delta) \cap \bar{T}(t_j, \delta) = \emptyset\).

Let \(k_0\) be sufficiently large such that for \(k \geq k_0\), \(x^k \in B(\bar{x}, \delta)\). Let

\[
\alpha_i(x, \gamma, \delta) := \int_{T(t_i, \delta)} \mu(x, t; \gamma) P_0(dt)
\]

and

\[
\alpha_i(\bar{x}) := \lim_{k \to \infty, \delta \to 0} \alpha_i(x^k, \gamma_k, \delta).
\]

By [22, Theorem 5.3], \(\alpha_i(\bar{x}) > 0\) and \(\sum_{i=1}^M \alpha_i(\bar{x}) = 1\). Let \(\mu_i\) be the Dirac probability measure at point \(t_i\) and \(\tilde{\mu} := \sum_{i=1}^M \alpha_i(\bar{x})\mu_i\). Then

\[
\int_T \partial_x h(\bar{x}, t) \tilde{\mu}(dt) \subseteq \partial \theta(\bar{x}).
\]

Let

\[
\Delta(\delta) := \int_{T \setminus \bigcup_{i=1}^M T(t_i, \delta)} \mu(x^k, t; \gamma_k) \partial_x h(x^k, t) P_0(dt).
\]
Then
\[
\mathbb{D}(\partial e_{\gamma_k}(x^k), \partial \theta(\bar{x})) \\
\leq \mathbb{D} \left( \int_T \mu(x^k; t; \gamma_k) \partial_x h(x^k, t) P_0(dt), \int_T \partial_x h(\bar{x}, t) \bar{\mu}(dt) \right) \\
= \mathbb{D} \left( \sum_{i=1}^M \int_{T(t_i, \delta)} \mu(x^k; t; \gamma_k) \partial_x h(x^k, t) P_0(dt) + \Delta(\delta), \int_T \partial_x h(\bar{x}, t) \bar{\mu}(dt) \right) \\
\leq \|\Delta(\delta)\| + \sum_{i=1}^M \mathbb{D} \left( \int_{T(t_i, \delta)} \mu(x^k; t; \gamma_k) \partial_x h(x^k, t) P_0(dt), \alpha_i(\bar{x}) \partial_x h(\bar{x}, t_i) \right).
\]

It is known that
\[
\lim_{k \to \infty} \sup_{t \in T \cup \bigcup_{i=1}^M T(t_i, \delta)} \mu(x^k, t, \gamma_k) = 0,
\]
see for example [22, Theorem 5.3]. Together with (24), we see that the first term \(\|\Delta(\delta)\|\) at the right hand side (rhs for short) of the inequality above goes to zero as \(k \to \infty\). Moreover, by (25)-(26), the second term tends to zero as \(\epsilon \to 0\) (hence \(\delta \to 0\)) and \(k \to \infty\).

Condition (b). Let \(\epsilon\) be a small number. Under the condition there exists a positive number \(\delta\) such that
\[
\mathbb{D}(\partial_x h(x, t), \partial_x h(\bar{x}, t)) \leq \epsilon, \quad \forall (x, t) \in B(\bar{x}, \delta) \times T^* (\bar{x}).
\]
Moreover, \(T^* (\cdot)\) as the set of optimal solution is outer semicontinuous on \(X\), see [30, Theorem 1.17]. Hence
\[
\bigcup_{x \in B(\bar{x}, \delta)} T(x) \subseteq T^* (\bar{x}) + \epsilon B.
\]

Let \(\hat{T} := T \setminus (T^* (\bar{x}) + \epsilon B)\). Then
\[
\mathbb{D}(\partial e_{\gamma_k}(x^k), \partial \theta(\bar{x})) \\
= \mathbb{D} \left( \int_T \mu(x^k; t; \gamma_k) \partial_x h(x^k, t) P_0(dt), \partial \theta(\bar{x}) \right) \\
= \mathbb{D} \left( \int_T \mu(x^k; t; \gamma_k) \partial_x h(x^k, t) P_0(dt) + \int_{T^* (\bar{x}) + \epsilon B \setminus T^* (\bar{x})} \mu(x^k; t; \gamma_k) \partial_x h(x^k, t) P_0(dt) \\
\quad + \int_{T^* (\bar{x})} \mu(x^k; t; \gamma_k) \partial_x h(x^k, t) P_0(dt), \partial \theta(\bar{x}) \right) \\
\leq \left\| \int_T \mu(x^k; t; \gamma_k) \partial_x h(x^k, t) P_0(dt) \right\| + \left\| \int_{(T^* (\bar{x}) + \epsilon B) \setminus T^* (\bar{x})} \mu(x^k; t; \gamma_k) \partial_x h(x^k, t) P_0(dt) \right\| \\
\quad + \mathbb{D} \left( \int_{T^* (\bar{x})} \mu(x, t) \partial_x h(x, t) P_0(dt), \partial \theta(\bar{x}) \right),
\]
where the inequality follows from (6). Let \(R_1, R_2\) and \(R_3\) denote the three items at the rhs of inequality of (28) in sequel. By [22, Theorem 5.3],
\[
\lim_{T \to \infty} \sup_{(x, t) \in B(\bar{x}, \delta) \times \hat{T}} \mu(x, t; \gamma) = 0.
\]
Together with (24), there exists a sufficiently large \(k_1\) such that, when \(k \geq k_1\), \(x^k \in B(\bar{x}, \delta)\) and
\[
R_1 = \left\| \int_T \mu(x^k; t; \gamma_k) \partial_x h(x^k, t) P_0(dt) \right\| \leq \epsilon.
\]
By the continuity of \( \mu(x, t; \gamma) \) and the fact that the probability \( \text{Prob}_{P_0}\{T^*(\bar{x})\} \) is not zero, it follows by [22, Theorem 5.3] that there exist constants \( C > 0 \) and \( k_2 \geq k_1 \) such that, for \( k \geq k_2 \),

\[
\mu(x^k, t; \gamma_k) \leq C, \quad \forall t \in (T^*(\bar{x}) + \epsilon B) \setminus T^*(\bar{x}),
\]

which implies

\[
R_2 = \left\| \int_{(T^*(\bar{x}) + \epsilon B) \setminus T^*(\bar{x})} \mu(x^k, t; \gamma_k) \partial_x h(x^k, t) P_0(dt) \right\| \leq 2CL\epsilon. \tag{32}
\]

Now let us estimate \( R_3 \). Let \( \eta_k := \int_{T^*(\bar{x})} \mu(x^k, t; \gamma_k) P_0(dt) \). By (29) and (31), there exists \( k_3 \geq k_2 \) such that for any \( k \geq k_3 \), \( \eta_k \geq 1 - (C + 1)\epsilon \). Let

\[
\nu^k(t) =: \frac{\mu(x^k, t; \gamma_k)}{\int_{T^*(\bar{x})} \mu(x^k, t; \gamma_k) P_0(dt)}.
\]

Then we may view \( \nu^k(t) \) as a density function of some probability measure over set \( T^*(\bar{x}) \). By the definition of \( \partial \theta(\cdot) \),

\[
\int_{T^*(\bar{x})} \partial_t h(\bar{x}, t) \nu^k(t) P_0(dt) \subseteq \partial \theta(\bar{x})
\]

and hence

\[
R_3 = \mathbb{D} \left( \int_{T^*(\bar{x})} \mu(x^k, t; \gamma_k) \partial_x h(x^k, t) P_0(dt), \partial \theta(\bar{x}) \right) \\
\leq \mathbb{D} \left( \int_{T^*(\bar{x})} \eta_k \partial_x h(x^k, t) \nu^k(t) P_0(dt), \int_{T^*(\bar{x})} \partial_t h(\bar{x}, t) \nu^k(t) P_0(dt) \right) \\
\leq \int_{T^*(\bar{x})} \mathbb{D} \left( \eta_k \partial_x h(x^k, t), \partial_x h(\bar{x}, t) \right) \nu^k(t) P_0(dt). \tag{33}
\]

The last inequality is due to [15, Theorem 3.5] for the exchange of integration and \( \mathbb{D} \). Let us write

\[
\partial h(x^k, t) + \Delta_k = \eta_k \partial_x h(x^k, t).
\]

Then \( \| \Delta_k \| \to 0 \) and

\[
R_3 \leq \int_{T^*(\bar{x})} \mathbb{D} \left( \partial_x h(x^k, t), \partial_x h(\bar{x}, t) \right) \nu^k(t) P_0(dt) + \| \Delta_k \|.
\]

The first term at the rhs of the inequality goes to zero under condition (b).

With the proposition, we are ready to state our main stability result in this section.

**Theorem 3.3.** Let \( X \) be a compact set, \( \gamma_k \to \infty \) and \( (x^k, \rho^k) \) be a KKT pair of problem (11). Let \( (x^*, \rho^*) \) be an accumulation point of sequence \( \{(x^k, \rho_{x_k}^k)\} \). Under the conditions of Proposition 2, \( (x^*, \rho^*) \) is a KKT pair of problem (10).

**Proof.** The conclusion follows from outer semicontinuity of \( \partial \mathbb{D}[f(\cdot, \xi)], \mathcal{N}_X(\cdot) \) and the subdifferential consistency (23).

The theorem assumes convergence of the Lagrange multiplier \( \rho^k \). A sufficient condition for this is condition (13), under which we can easily show that \( \rho^k \) is bounded.
4. Sample average approximation method. In practice, the distribution of $\xi$ is often unknown or it is numerically too expensive to calculate the expected values. Instead it might be possible to obtain a sample of the random vector $\xi$ from past data. This motivates us to find an approximate optimal solution to problem (11) on the basis of the sampling information. This kind of statistical approximation is guaranteed by the classical law of large numbers in statistics.

Let $\xi_1, \cdots, \xi_N$ be an independent and identically distributed sample of $\xi$, let $t^j = Y(\xi^j)$, for $j = 1, \cdots, N$ and $h^N(x, t^i) := \frac{1}{N} \sum_{k=1}^{N} H(x, t^i, \xi^k)$. We consider the following sample average approximation (SAA for short) problem:

$$\min_{x \in X} \ f^N(x) := \frac{1}{N} \sum_{k=1}^{N} f(x, \xi^k)$$

s.t. $\ e^N(x) := \frac{1}{\gamma} \log \left( \frac{1}{N} \sum_{i=1}^{N} e^{\gamma h^N(x, t^i)} \right) \leq 0. \quad (34)$$

We call (34) SAA problem and (11) true problem. Let $x^N$ be an optimal solution of problem (34). We analyze the convergence of $x^N$ as the sample size $N$ increases.

**Theorem 4.1.** Let $X^N(\gamma)$, $X(\gamma)$ and $X^*$ denote the set of optimal solutions of problems (34), (11) and (10) respectively, $v^N(\gamma)$, $v(\gamma)$ and $v^*$ the corresponding optimal values. Assume: (a) $f(x, \xi)$ and $G(x, \xi)$ are uniformly bounded on $(X, \Xi)$; (b) condition (13) holds. We have the following assertions.

(i) For fixed $\gamma$, w.p.1

$$\limsup_{N \to \infty} X^N(\gamma) \subseteq X(\gamma), \quad \lim_{N \to \infty} v^N(\gamma) = v(\gamma);$$

(ii) w.p.1

$$\limsup_{N \to \infty, \gamma \to \infty} X^N(\gamma) \subseteq X^*, \quad \lim_{N \to \infty, \gamma \to \infty} v^N(\gamma) = v^*.$$ 

**Proof.** Under condition (a), we can easily show uniform convergence of $f^N(\cdot)$ to $E[f(\cdot, \xi)]$ and $e^N(\cdot)$ to $e_\gamma(\cdot)$, see for example [11, Corollary 3.7]. The convergence of the optimal value and optimal solutions follows from well known stability result in nonlinear programming (see e.g. [36, Lemma 2.2] and the references therein) and [25, Lemma 3.1] for a similar proof.

We now move on to investigate the case when we obtain a stationary point rather than an optimal solution from solving the sample average approximation problem (34).

A feasible point $x$ is said to be a sample average approximated stationary point (SAA-stationary point) of problem (34) if there exists $\rho^N$ such that

$$0 \in \partial f^N(x) + \rho^N \partial e^N_\gamma(x) + \mathcal{N}_X(x), \quad (35)$$

$$0 \leq \rho^N \perp -e^N_\gamma(x) \geq 0. \quad (36)$$

By [6, Propostion 2.3.3],

$$\partial e^N_\gamma(x) \subseteq \Phi^N(x) := \frac{1}{N} \sum_{i=1}^{N} \phi^N(x, t^i) \partial_x h^N(x, t^i), \quad (37)$$

where

$$\phi^N(x, t) := \frac{e^{\gamma h^N(x, t)}}{N \sum_{i=1}^{N} e^{\gamma h^N(x, t^i)}}, \quad \partial_x h^N(x, t^i) = \frac{1}{N} \sum_{k=1}^{N} \partial_x H(x, t^i, \xi^k)$$

and $\partial_x H(x, t, \xi)$ is defined in (18).
Theorem 4.2. Let \( \{x^N, \rho^N\} \) be a sequence of KKT pairs of problem (34) and \((x^*, \rho^*)\) be an accumulation point. Assume that \( f(x, \xi) \) and \( G(x, \xi) \) are locally Lipschitz continuous w.r.t. \( x \) and \( \xi \), and their Lipschitz modulus are bounded by a positive number \( C \). Then w.p.1 \( x^* \) is an stationary point of problem (11).

**Proof.** Let \( \Phi^N(x) \) be defined as in (37) and \( x^N \in X^N(\gamma) \). Then
\[
0 \in \partial f^N(x^N) + \rho^N \Phi^N(x^N) + \mathcal{N}_X(x^N). \tag{38}
\]
By taking a subsequence if necessary, we assume for the simplicity of notation that \( (x^N, \rho^N) \) converges \((x^*, \rho^*)\). Since \( \limsup_{N \to \infty} \partial f^N(x^N) \subseteq \partial \mathbb{E}[f(x^*, \xi)] \), and \( \mathcal{N}_X(x^N) \subseteq \mathcal{N}_X(x^*) \), if we are able to show that \( \limsup_{N \to \infty} \Phi^N(x^N) \subseteq \partial e_\gamma(x^*) \), then we will arrive at
\[
0 \in \partial \mathbb{E}[f(x^*, \xi)] + \rho^* \partial e_\gamma(x^*) + \mathcal{N}_X(x^*). \tag{39}
\]
To this end, we define
\[
\psi^N(x, t) := \frac{e^{\gamma h(x, \xi)}}{N} \sum_{i=1}^N e^{\gamma h(x_i, \xi)} \quad \text{and} \quad \partial^N_x h(x, t) := \bigcup_{(x', t') \in B_i(x, t, \epsilon)} \partial_x h(x', t'),
\]
where \( \epsilon \) is a positive number. Let
\[
\Psi^N_\epsilon(x) := \frac{1}{N} \sum_{i=1}^N \psi^N(x, t') \partial^N_x h(x, t').
\]
We will demonstrate
\[
\limsup_{N \to \infty} \Phi^N(x^N) \subseteq \lim_{N \to \infty} \Psi^N_\epsilon(x^*) \tag{40}
\]
and then
\[
\lim_{\epsilon \to 0} \limsup_{N \to \infty} \Psi^N_\epsilon(x^*) \subseteq \partial e_\gamma(x^*). \tag{41}
\]
By definition of \( \mathbb{D} \),
\[
\mathbb{D}(\Phi^N(x^N), \Psi^N_\epsilon(x^*))
= \mathbb{D} \left( \frac{1}{N} \sum_{i=1}^N \phi^N(x^N, t^i) \partial_x h^N(x^N, t^i), \frac{1}{N} \sum_{i=1}^N \psi^N(x^*, t^i) \partial^N_x h(x^*, t^i) \right)
\leq \mathbb{D} \left( \frac{1}{N} \sum_{i=1}^N \phi^N(x^N, t^i) \partial_x h^N(x^N, t^i), \frac{1}{N} \sum_{i=1}^N \phi^N(x^N, t^i) \partial^N_x h(x^*, t^i) \right)
+ \mathbb{D} \left( \frac{1}{N} \sum_{i=1}^N \phi^N(x^N, t^i) \partial^N_x h(x^*, t^i), \frac{1}{N} \sum_{i=1}^N \psi^N(x^*, t^i) \partial^N_x h(x^*, t^i) \right).
\]
Let \( R^N_1 \) and \( R^N_2 \) denote the two terms at the rhs of the inequality above in sequel. By the properties of \( \mathbb{D} \),
\[
R^N_1 \leq \frac{1}{N} \sum_{i=1}^N \mathbb{D} \left( \phi^N(x^N, t^i) \partial_x h^N(x^N, t^i), \phi^N(x^N, t^i) \partial^N_x h(x^*, t^i) \right)
\leq \frac{1}{N} \sum_{i=1}^N \| \phi^N(x^N, t^i) \| \mathbb{D} \left( \partial_x h^N(x^N, t^i), \partial^N_x h(x^*, t^i) \right)
\leq \frac{1}{N} \sum_{i=1}^N \| \phi^N(x^N, t^i) \| \mathbb{D} \left( \partial_x h^N(x^* + \delta \mathcal{B}, t^i), \partial^N_x h(x^*, t^i) \right), \tag{42}
\]
where
\[ \partial_x h^N(x^* + \delta B, t^i) := \bigcup_{x \in B(x^*, \delta)} \partial_x h^N(x, t^i). \]

Under the uniform Lipschitz continuity of \( G(\cdot, \xi) \) and outer semi-continuity of \( \partial_x H(\cdot, \cdot, \xi) \), it follows by [32, Theorem 2] that
\[ \lim_{N \to \infty} \sup_{(x, t) \in X \times T} \mathbb{D}\left( \partial_x h^N(x, t), \partial_x^* h(x, t) \right) = 0 \] (43)
almost surely (a.s. for short). Thus, for \( N \) sufficiently large,
\[ \partial_x h^N(x^* + \delta B, t^i) \subseteq \partial_x^* h(x^*, t^i), \quad \forall t^i \in T \]
a.s. Combining with (42), we have
\[ R_1^N \leq \frac{1}{N} \sum_{i=1}^N \left\| \phi^N(x^N, t^i) \right\| \mathbb{D}\left( \partial_x^* h(x^*, t^i), \partial_x^* h(x^*, t^i) \right), \]
a.s. for \( N \) sufficiently large. Since \( \left\| \phi^N(x^N, t^i) \right\| \) is bounded, there exists a positive constant \( L_0 \) such that
\[ \lim_{N \to \infty} R_1^N \leq \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^N \left\| \phi^N(x^N, t^i) \right\| \mathbb{D}\left( \partial_x^* h(x^*, t^i), \partial_x^* h(x^*, t^i) \right), \]
\[ \leq L_0 \int_T \Delta_{\epsilon, \delta}(t) P_0(dt), \]
where \( \Delta_{\epsilon, \delta}(t) := \mathbb{D}\left( \partial_x^* h(x^*, t), \partial_x^* h(x^*, t) \right). \) Notice that \( \Delta_{\epsilon, \delta}(t) \) is bounded, by Lebesgue’s Dominated Convergence theorem,
\[ \lim_{N \to \infty} R_1^N \leq \lim_{\delta \to 0} L_0 \int_T \Delta_{\epsilon, \delta}(t) P_0(dt) = L_0 \int_T \lim_{\delta \to 0} \Delta_{\epsilon, \delta}(t) P_0(dt) = 0 \]
a.s.. Next, we estimate \( R_2^N \). By definition
\[ R_2^N \leq \left\| \frac{1}{N} \sum_{i=1}^N \phi^N(x^N, t^i) - \frac{1}{N} \sum_{i=1}^N \psi^N(x^*, t^i) \right\| \cdot \left\| \partial_x h(x^*, t^i) \right\|. \]
Since \( \partial_x^* h(x, t^i) \) is uniformly bounded on \( X \times T \) and by [31, Chapter 6, Proposition 7], \( \frac{1}{N} \sum_{i=1}^N \phi^N(x^N, t^i) \) and \( \frac{1}{N} \sum_{i=1}^N \psi^N(x^*, t^i) \) to the same limit, the rhs of the inequality goes to zero as \( N \to \infty \). By [32, Theorem 1],
\[ \lim_{N \to \infty} \Psi^N(x^N) = \partial_x^* e(x^*). \]

Driving \( \epsilon \) to zero, we obtain (40) as desired. \( \square \)

5. **Numerical results.** In this section, we investigate the numerical performance of the approximation scheme (34). We do so by applying it to an academic example and a practical portfolio selection problem with empirical data. For fixed parameter \( \gamma \) and sample size \( N \), problem (34) is standard nonlinear programming. By exploiting NLP solvers for problem (34), we implement the Algorithm as follows.

**Algorithm 1** (SAA-Entropic method). *Set initial parameter \( \gamma > 0 \), step-size \( \lambda > 1 \) and tolerance \( \epsilon > 0 \).*

**Step 1.** Solve problem (34) and write \( x^N_\gamma \) for the optimal solution.

**Step 2.** If \( \| \max_{t \in T} h^N(x^N_\gamma, t) - e^N_\gamma(x^N_\gamma) \| \leq \epsilon \), stop; otherwise, \( \gamma = \lambda \gamma \), go to Step 1.
Since we use the approximation of the constraints as the termination condition, that is,
\[\max_{t \in T} h_N(x_N^{\gamma}, t) - e_N^{\gamma}(x_N^{\gamma}) \leq \epsilon,\]
then for any given positive number \(\epsilon\), Lemma 2.1 and the proof of Theorem 4.1 ensure that the algorithm 1 should terminate at finite steps. But, the number of iteration will go to infinity as \(\epsilon \to 0\).

We use the random number generator \(\text{rand}\) in Matlab R2012a to generate the samples of \(\xi\) and solver \(\text{fmincon}\) in the Step 1. The tolerance \(\epsilon\), step-size \(\lambda\) and initial value of \(\gamma\) are set \(10^{-5}\), 2 and 100 respectively.

**Example 1.** We consider a variation of [33, Example 5.1]:
\[
\begin{align*}
\text{min} & \quad \mathbb{E}[f(x, \xi)], \\
\text{s.t.} & \quad G(x, \xi) \succeq 2 Y(\xi), \\
& \quad x \in [0, 5],
\end{align*}
\]
where \(f(x, \xi) = -x\xi + \frac{\sqrt{2}}{2}x^2,\) \(G(x, \xi) = -x\xi + \frac{1}{3}x^2,\) \(Y(\xi) = G(1, \xi)\) and \(\xi\) is a random variable uniform distributed over interval \([2, 3]\).

It is rather complicated to work out the feasible set precisely. Here we only need to find out the optimal solution of problem (44). Observe first that \(\mathbb{E}[f(z, \xi)] = \frac{\sqrt{2}}{2}x^2 - \frac{5}{4}x\), the global minimizer of the objective is \(x^* = \frac{5\sqrt{2}}{4}\) if it satisfies the second order dominance constraints. For every \(x \in [1, 3]\), \(G(x, \xi)\) dominates \(G(1, \xi)\) in first order as the cumulative distribution function of the former is strictly smaller than that of the latter. This implies \(G(x, \xi)\) dominates \(G(1, \xi)\) in second order for \(x \in [1, 3]\). Thus \(x^* = \frac{5\sqrt{2}}{4} \approx 1.7678\) is the optimal solution.

We apply Algorithm 1 to solve the problem and report the results in Table 1. Here we use the following notation. \(N\) – the sample size, \(\gamma\) – the parameter in entropy approximation, \(\text{Appr.Sol}\) – the approximate optimal solution and \(\text{Appr.Val}\) – the optimal value, \(t\) – the execution time in seconds.

| \(N\) | \(\gamma\) | \(\text{Appr.Sol}\) | \(\text{Appr.Val}\) | \(t\) |
|-------|---------|----------------|----------------|------|
| 50    | 25600   | 1.7663         | -2.2059        | 0.9804 |
| 100   | 25600   | 1.7664         | -2.2063        | 1.1068 |
| 400   | 25600   | 1.7503         | -2.1663        | 1.2594 |
| 800   | 25600   | 1.7707         | -2.2170        | 1.9361 |

The results show that we can obtain an acceptable approximate optimal solution in a small sample size for this problem because the range of the support set is just 1. Increasing the sample size does not help improve the approximate solution due to computational errors.

**Example 2.** In what follows, we consider a portfolio optimization problem. To simplify the discussions, we ignore the transaction fee, therefore we may denote the total value of portfolio as
\[f(x, r) := r_1x_1 + r_2x_2 + \cdots + r_nx_n,\]
where \(r_i\) is the random return of asset \(i\). Moreover, we set \(G(x, r) = f(x, r)\) and use the equally weighted portfolio as the benchmark strategy \(y(r)\). The corresponding
stochastic second order dominance portfolio optimization problem (SSD) is

\[
\begin{aligned}
\min_{x_1,\ldots,x_n \geq 0} & \quad \mathbb{E}[ -f(x, r)], \\
\text{s.t.} & \quad f(x, r) \preceq 2 \frac{1}{n} \sum_{i=1}^{n} r_i, \\
& \quad \sum_{i=1}^{n} x_i = 1.
\end{aligned}
\] (45)

We carry out numerical tests of problem (45) on a historical data set of 10 assets (Aberdeen Asset Management plc, Admiral Group PLC, AMEC PLC, Anglo American PLC, Antofagasta PLC, AstraZeneca PLC, Aviva PLC, Babcock International Group PLC, BAE Systems PLC and Barclays PL) over a time horizon of 4 years (from 7th Dec 2009 to 18th Nov 2013) with a total of 1001 records on the historical stock returns (these are obtained from http://finance.google.com with adjustment for stock splitting). We have carried out out-of-sample tests with a rolling window of 500 days, that is, we use first 500 data to calculate the optimal portfolio strategy for day 501 and then move on a rolling basis.

Figure 1 depicts the performance of second order dominance model (45). As expected, the cumulative wealth from the optimal strategy performs better than that of the benchmark strategy with an increasing margin from 0 to about 0.15 at the end of the observed time period which is 15% of the total asset in the initial investment.

REFERENCES

[1] E. Anderson, H. Xu and D. Zhang, CVaR approximations for minimax and robust convex optimization, manuscript, 2013.
[2] R. J. Aumann, Integrals of set-valued functions, Journal of Mathematical Analysis and Applications, 12 (1965), 1–12.
[3] J. F. Bonnans and A. Shapiro, Perturbation Analysis of Optimization Problems, Springer Series in Operational Research, Springer, New York, 2000.
[4] G. Calafiore and M. C. Campi, Uncertain convex programs: randomized solutions and confidence levels, Mathematical Programming, 102 (2005), 25–46.
[5] X. Chen, Smoothing methods for nonsmooth, nonconvex minimization, Mathematical Programming, 134 (2012), 71–99.
[6] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Wiley, New York, 1983.

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

[8] D. Dentcheva and W. Römisch, Stability and sensitivity of stochastic dominance constrained optimization models, *SIAM Journal on Optimization*, 23 (2013), 1672–1688.

[9] D. Dentcheva and A. Ruszczyski, Optimization with stochastic dominance constraints, *SIAM Journal on Optimization*, 14 (2003), 548–566.

[10] D. Dentcheva and A. Ruszczyski, Optimality and duality theory for stochastic optimization with nonlinear dominance constraints, *Mathematical Programming*, 99 (2004), 329–350.

[11] Y. Ermoliev and V. Norkin, Sample average approximation method for compound stochastic optimization problems, *SIAM Journal on Optimization*, 23 (2013), 2231–2263.

[12] C. I. Fabián, G. Mitra and D. Roman, Processing second-order stochastic dominance models using cutting-plane representations, *Mathematical Programming*, 130 (2011), 33–57.

[13] H. Föllmer and T. Nijspeil, Entropic risk measures: coherence vs. convexity, model ambiguity, and robust large deviations, *Stochastics and Dynamics*, 11 (2011), 333–351.

[14] M. Gugat, Error bounds for infinite systems of convex inequalities without Slater’s condition, *Mathematical Programming*, 88 (2000), 255–275.

[15] C. Hess, *Set-valued integration and set-valued probability theory: an overview*, Handbook of measure theory, North-Holland, Amsterdam, I-II (2002), 617–673.

[16] T. Homem-de-Mello, A cutting surface method for uncertain linear programs with polyhedral stochastic dominance constraints, *SIAM Journal on Optimization*, 20 (2010), 1250–1273.

[17] J. Hu, T. Homem-De-Mello and S. Mehrotra, Sample average approximation of stochastic dominance constrained programs, *Mathematical Programming*, 133 (2012), 171–201.

[18] P. Kall, *Approximation to Optimization Problems: An Elementary Review*, Mathematics of Operations Research, 11 (1986), 9–18.

[19] 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, (1985) 1–21.

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

[21] H. Levy, Stochastic dominance and expected utility: survey and analysis, *Management Science*, 38 (1992), 555–593.

[22] G. H. Lin, M. W. Xu and J. J. Ye, On solving simple bilevel programs with a nonconvex lower level program, *Mathematical Programming*, 144 (2014), 277–305.

[23] Y. Liu and H. Xu, Stability analysis of stochastic programs with second order dominance constraints, *Mathematical Programming*, 142 (2013), 435–460.

[24] Y. Liu, W. Römisch and H. Xu, Quantitative stability analysis of stochastic generalized equations, *SIAM Journal on Optimization*, 24 (2014), 467–497.

[25] Y. Liu and H. Xu, Entropic approximation for mathematical programs with robust equilibrium constraints, *SIAM Journal on Optimization*, 24Z (2014), 933–958.

[26] A. Müller and M. Scarsini, *Stochastic Orders and Decision Under Risk*, Institute of mathematical statistics, Hayward, CA, 1991.

[27] D. Ralph and H. Xu, Asymptotic analysis of stationary points of sample average two stage stochastic programs: A generalized equation approach, *Mathematics of Operations Research*, 36 (2011), 568–592.

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

[29] S. M. Robinson, Analysis of sample-path optimization, *Mathematics of Operations Research*, 21 (1996), 513–528.

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

[31] A. Ruszczyski and A. Shapiro, *Stochastic Programming*, Handbook in Operations Research and Management Science, Elsevier, 2003.

[32] A. Shapiro and H. Xu, Uniform laws of large numbers for set-valued mappings and subdifferentials of random functions, *Journal of Mathematical Analysis and Applications*, 325 (2007), 1390–1399.

[33] H. Sun, H. Xu, R. Meskarian and Y. Wang, Exact penalization, level function method, and modified cutting-plane method for stochastic programs with second order stochastic dominance constraints, *SIAM Journal on Optimization*, 23 (2013), 602–631.
[34] H. Xu, Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming, *Journal of Mathematical Analysis and Applications*, 368 (2010), 692–710.

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

[36] J. Zhang, H. Xu and L. W. Zhang, Quantitative stability analysis of stochastic quasi-variational inequality problems and applications, Manuscript, 2014.

Received September 2015; 1st revision April 2016; final revision November 2016.

E-mail address: googleliu@dlmu.edu.cn
E-mail address: mathhlsun@gmail.com
E-mail address: H.Xu@soton.ac.uk