Convergence Analysis of Stationary Points in Sample Average Approximation of Stochastic Programs with Second Order Stochastic Dominance Constraints[1]

Dedicated to Professor Jon Borwein on the occasion of his 60th birthday

Hailin Sun[2] and Huifu Xu[3]

May 5, 2012

Abstract

Sample average approximation (SAA) method which is also known under various names such as Monte Carlo method, sample path optimization and stochastic counterpart has recently been applied to solve stochastic programs with second order stochastic dominance (SSD) constraints. In particular, Hu et al [19] presented a detailed convergence analysis of ϵ-optimal values and optimal solutions of sample average approximated stochastic programs with polyhedral SSD constraints. In this paper, we complement the existing research by presenting convergence analysis of stationary points when SAA is applied to a class of stochastic minimization problems with SSD constraints. Specifically, under some moderate conditions we prove that optimal solutions and stationary points obtained from solving sample average approximated problems converge with probability one (w.p.1) to their true counterparts. Moreover, by exploiting some recent results on large deviation of random functions and sensitivity analysis of generalized equations, we derive exponential rate of convergence of stationary points.

Key words. Second order dominance, stationary points, KKT conditions, exponential convergence

AMS subject classification: 90C15, 90C30, 90C31, 90C34

1 Introduction

In this paper, we consider a stochastic optimization problem with second order dominance constraints

$$\begin{array}{ll} \min\limits_{z} & \mathbb{E}[H(z,\xi(\omega))] \\ \text{s.t.} & G(z,\xi(\omega)) \succeq_2 Y(\xi(\omega)), \\ & z \in Z_0, \end{array} \tag{1.1}$$

[1]The work of this author was carried out while he was visiting the second author in the School of Mathematics, University of Southampton sponsored by China Scholarship Council. It was also partly supported by the National Natural Science Foundation of China No. 11071122, 11171159 and the Specialized Research Fund of Doctoral Program of Higher Education of China No. 20103207110002.

[2]Department of Mathematics, Harbin Institute of Technology, Harbin, 150001, China. Email: mathhlsun@gmail.com.

[3]School of Mathematics, University of Southampton, Southampton, SO17 1BJ, UK. School of Maritime and Transportation, Ningbo University, China. Email: h.xu@soton.ac.uk.

where Z_0 is a closed convex subset of $\mathbb{R}^n$, $G, H : \mathbb{R}^{n+q} \to \mathbb{R}$ are continuously differentiable functions, $\xi : \Omega \to \Xi$ is a vector of random variables defined on a nonatomic probability space $(\Omega, \mathcal{F}, P)$ with support set $\Xi \subset \mathbb{R}^q$ and $\mathbb{E}[\cdot]$ denotes the expected value with respect to probability measure P. To ease notation, we will use ξ to denote the random vector $\xi(\omega)$ and a deterministic vector, depending on the context.

The second order dominance constraint is defined in the following sense: for every $\eta \in \mathbb{R}$,

$$\int_{-\infty}^{\eta} P(G;\tau)d\tau \leq \int_{-\infty}^{\eta} P(Y;\tau)d\tau,$$

where $P(G;\tau) = \text{Prob}\{G(z,\xi) \leq \tau\}$ and $P(Y;\tau) = \text{Prob}\{Y \leq \tau\}$. The model has wide economic interpretations, for instance, $H(z,\xi)$ is a cost function, $G(z,\xi)$ is a profit function and $Y(\xi)$ is benchmark profit at a scenario ξ. In particular, we may set $G = -H$ and $Y = G(y,\xi)$ where y is a fixed decision vector. In this paper, our focus is on numerical methods for solving the problem rather than practical application of the model.

It is well-known that the second order dominance constraint in (1.1) is mathematically equivalent to

$$\mathbb{E}[(\eta - G(z,\xi))_+] \leq \mathbb{E}[(\eta - Y(\xi))_+], \ \forall \eta \in \mathbb{R},$$

see [11] and the references therein. Consequently problem (1.1) can be reformulated as:

$$\begin{array}{ll} \min\limits_{z} & \mathbb{E}[H(z,\xi)] \\ \text{s.t.} & \mathbb{E}[(\eta - G(z,\xi))_+] \leq \mathbb{E}[(\eta - Y(\xi))_+], \ \forall \eta \in \mathbb{R}, \\ & z \in Z_0. \end{array} \tag{1.2}$$

From numerical optimization perspective, (1.2) is not helpful in that it does not satisfy a key constraint qualification (Slater's type), a condition relating to numerical stability. Consequently one often considers a relaxed form of the problem:

$$\begin{array}{ll} \min\limits_{z} & \mathbb{E}[H(z,\xi)] \\ \text{s.t.} & \mathbb{E}[p(\eta - G(z,\xi))] \leq \mathbb{E}[p(\eta - Y(\xi))], \ \forall \eta \in [a,b], \\ & z \in Z_0, \end{array} \tag{1.3}$$

where $[a,b]$ is a bounded closed interval in $\mathbb{R}$ and $p(x) := \max(0,x)$. For the simplicity of notation, let

$$F(z,\eta,\xi) := p(\eta - G(z,\xi)) - p(\eta - Y(\xi)) \tag{1.4}$$

and $f(z,\eta) := \mathbb{E}[F(z,\eta,\xi)]$.

Stochastic programs with dominance constraints have been proposed by Dentcheva and Ruszczyński [10, 11] and have found substantial applications in many areas including finance and energy [14, 9]. Over the past decade, there have been extensive discussions on the optimization theory of the mathematical models particularly concerning optimality and duality of both (1.1) and (1.2), see recent work by Dentcheva and Ruszczyński [10, 11, 12] and the references therein.

In this paper, we consider sample average approximation of (1.3) which is fundamentally a stochastic semi-infinite programming problem. The basic idea of SAA can be described as follows. Let $\xi^1, \cdots, \xi^N$ be an independent and identically distributed (i.i.d.) sampling of ξ. We consider the following sample average approximation problem for (1.3):

$$
\begin{aligned}
\min_{z} \quad & h_N(z) := \tfrac{1}{N}\textstyle\sum_{i=1}^N H(z,\xi^i) \\
\text{s.t.} \quad & f_N(z,\eta) := \frac{1}{N}\sum_{i=1}^N \left[(\eta - G(z,\xi^i))_+ - (\eta - Y(\xi^i))_+\right] \le 0,\ \forall \eta \in [a,b], \\
& z \in Z_0.
\end{aligned} \tag{1.5}
$$

We refer to (1.3) as the *true* problem and (1.5) as the *sample average approximation* (SAA) problem. SAA is a very popular method in stochastic optimization and it is known under various names such as sample path optimization (SPO) method [28], stochastic counterpart and more broadly Monte Carlo method, see [30] for a comprehensive review of the subject by Shapiro. The main benefit of SAA is that one does not have to calculate the expected values.

Hu, Homem-de-Mello and Mehrotra [19] seem to be the first to apply SAA to a class of stochastic programs with polyhedral SSD constraints. They presented a detailed convergence analysis of the SAA scheme in terms of ϵ-optimal values and optimal solutions, proposed a cut-generation algorithm for solving the subsequent sample average approximated problem and derived lower and upper bounds for the true optimal values.

In a more recent development, Liu and Xu [23] studied stability of (1.3) where the true probability measure P is approximated by a class of general probability measures including empirical probability measure (which is equivalent to SAA). The stability analysis is carried out for a penalized problem where the second order dominance constraint of (1.3) is moved to the objective through exact penalization (so that the penalized program has deterministic constraint). Convergence analysis of optimal solution and Clarke stationary points of the sample average approximated penalized program has been investigated. While the penalization scheme simplifies the convergence analysis, it also leaves one with an undesired gap: a stationary point of the penalized program is not necessarily a stationary point of the true program unless some strong conditions are satisfied.

In this paper, we focus on convergence analysis of stationary points of the SAA program (1.3) but through a completely different avenue: we carry out the convergence analysis directly through the Karush-Kuhn-Tucker (KKT) conditions of (1.3) and (1.5) without a penalization formulation. While the analysis is technically more challenging, a clear benefit is that the convergence results are stronger and broader in the sense that they cover the convergence of the stationary points of (1.5) regardless of how the SAA problem is solved, e.g., through penalization or other NLP formulation.

The rest of the paper is organized as follows. In section 2, we discuss the first order optimality conditions of the true problem (1.3) and the SAA problem (1.5). In section 3, we investigate almost sure convergence of stationary points of the SAA problem as sample size increases and extend the discussion in section 4 to the exponential rate of convergence.

2 Optimality conditions

In this section, we review/discuss optimality conditions of the relaxed true problem (1.3) and its sample average approximation (1.5). The former has been well established in a number of papers by Dentcheva and Ruszczyński, see for instances [13].

2.1 Notation and preliminaries

Throughout this paper, we use the following notation. $x^T y$ denotes the scalar product of two vectors x and y, $\|\cdot\|$ denotes the Euclidean norm of a vector and a compact set of vectors. We write $d(x, D) := \inf_{x' \in D} \|x - x'\|$ for the distance from point x to set D. For two sets D_1 and D_2, $\mathbb{D}(D_1, D_2) := \sup_{x \in D_1} d(x, D_2)$ represents the deviation from set D_1 to set D_2. Note that $\mathbb{D}(D_1, D_2) = 0$ if and only if $D_1 \subset D_2$. For a real valued function $h(x)$, we use $\nabla h(x)$ to denote the gradient of h at x. If $h(x)$ is vector valued, then the same notation refers to the classical Jacobian of h at x. Finally, for a sequence of sets $\{A_k\}$, we use $\underline{\lim} A_k$ to denote its lower limit and $\overline{\lim}\, A_k$ the upper limit. We refer readers to monograph [1] for the definition of these limits and upper semicontinuity of a set-valued mapping.

Let $v : \mathbb{R}^n \to \mathbb{R}$ be a locally Lipschitz continuous function. Recall that *Clarke generalized derivative* of v at point x in direction d is defined as

$$v^o(x, d) := \limsup_{y \to x, t \downarrow 0} \frac{v(y + td) - v(y)}{t}.$$

v is said to be Clarke *regular* at x if the usual one sided directional derivative, denoted by $v'(x, d)$, exists for every $d \in \mathbb{R}^n$ and $v^o(x, d) = v'(x, d)$. The *Clarke generalized gradient* (also known as Clarke subdifferential) is defined as

$$\partial v(x) := \{\zeta : \zeta^T d \leq v^o(x, d)\},$$

see [8, Chapter 2].

Some basics about measure theory that we need in this paper are in order. Let $\mathscr{C}([a, b])$ denote the space of continuous functions defined on $[a, b]$ with maximum norm. By the Riesz representation theorem, the space dual to $\mathscr{C}([a, b])$, denoted by $\mathscr{C}^*([a, b])$, is the space of regular countably additive measures on $[a, b]$ having finite variation, see [4, Example 2.63], [10] and the references therein. Let $\mathscr{C}_+^*([a, b])$ denote the subset of $\mathscr{C}^*([a, b])$ of positive measures and $\|\mu\|$ the induced norm of map $\int_a^b \cdot \mu(d\eta) : \mathscr{C}([a, b]) \to \mathbb{R}$. Then for $\mu \in \mathscr{C}^*([a, b])$, $\|\mu\| = \int_a^b \mu(d\eta) = \mu([a, b])$, which is the total variation measure of μ on $[a, b]$, see [16, section 3] and [4, Example 2.63].

Before proceeding to detailed discussions of optimality conditions, we list two main assumptions which are needed throughout this section.

Assumption 2.1 *There exist a point $z_0 \in Z_0$ and $\eta_0 \in [a, b]$ such that $\mathbb{E}[|H(z_0, \xi)|] < \infty$ and $\mathbb{E}[(\eta_0 - G(z_0, \xi))_+] < \infty$. Moreover, $\mathbb{E}[\|\nabla_z H(z, \xi)\|] < \infty$ and $\mathbb{E}[\|\nabla_z G(z, \xi)\|] < \infty$ for all $z \in Z_0$.*

Assumption 2.2 Problem (1.3) satisfies the *uniform dominance condition*, that is, there exists a point $z_0 \in Z_0$ such that

$$\sup_{\eta \in [a,b]} f(z_0, \eta) < 0.$$

This condition is also known as Slater's constraint qualification and it has been widely used to derive the optimality conditions of the relaxed problem (1.3).

Definition 2.1 Problem (1.3) is said to satisfy the *differential constraint qualification* at a feasible point z if there exists another feasible point z' and a positive number δ such that

$$\sup_{\zeta \in \partial_z f(z,\eta)} \zeta^T (z' - z) \leq -\delta, \ \forall \eta \in [a, b], \tag{2.6}$$

where $\partial_z f(z, \eta)$ denotes Clarke generalized gradient of $f(z, \eta)$ at point z for a given $\eta \in [a, b]$.

The concept of differential constraint qualification was introduced by Dentcheva and Ruszczyński [13]. When

$$\sup_{\zeta \in \partial_z f(z,\eta)} \zeta^T (z' - z) < 0, \ \forall \eta \in [a, b],$$

and f is continuously differentiable, the constraint qualification (2.6) is known as *extended Mangasarian-Fromowitz Constraint Qualification (MFCQ)*, see [4, page 510] in the context of semi-infinite programming.

It is not difficult to show (through Clarke's mean-value theorem [8, Theorem 2.3.7] and upper semicontinuity of the Clarke subdifferential) that the differential constraint qualification implies the Slater constraint qualification when Z_0 is a convex set.

Proposition 2.1 *Under Assumption 2.1*

(i) $\mathbb{E}[H(z, \xi)]$ *and* $f(z, \eta)$ *are well defined for all* $z \in Z_0$ *and* $\eta \in [a, b]$ *and* $f(z, \eta)$ *is locally Lipschitz continuous w.r.t.* z*, globally Lipschitz continuous w.r.t.* η*,*

$$\partial_z f(z, \eta) = -\mathbb{E}\left[\nabla_z G(z, \xi)^T \partial p(\eta - G(z, \xi))\right], \tag{2.7}$$

where

$$\partial p(\eta - G(z, \xi)) = \begin{cases} 1, & \text{if } \eta - G(z, \xi) > 0, \\ [0, 1], & \text{if } \eta - G(z, \xi) = 0, \\ 0, & \text{if } \eta - G(z, \xi) < 0, \end{cases}$$

and the expected value of the Clarke subdifferential of the random function is Aumann's integral [2];

(ii) $f(z, \eta)$ *is Clarke regular w.r.t.* z *for every fixed* $\eta \in [a, b]$*.*

Proof. Part (i). Verification of the well definedness and Lipschitzness is elementary given the fact that $p(\eta - G(z,\xi))$ is a composition of the max function $p(\cdot)$ and $\eta - G(z,\xi)$. In what follows, we show the calculus of the Clarke subdifferentials. Observe that function $p(\cdot)$ is piecewise linear convex in $\mathbb{R}$, by [8, Proposition 2.3.6], it is Clarke regular at any point in $\mathbb{R}$. Through the chain rule ([8, Theorem 2.3.10]), we obtain

$$\partial_z(p(\eta - G(z,\xi))) = \nabla_z G(z,\xi)^T \partial p(\eta - G(z,\xi)).$$

It is easy to verify that the term at the right hand side of the equality above is bounded by $\|\nabla_z G(z,\xi)\|$ which is integrably bounded under Assumption 2.1. By [8, Proposition 2.2], $\mathbb{E}[\nabla_z G(z,\xi)^T \partial p(\eta - G(z,\xi)]$ is well defined. Finally, the exchange of subdifferential operator ∂_z with mathematical expectation operator $\mathbb{E}[\cdot]$ is well-known when the integrand is Clarke regular, see for instance [35, Section 4] and references therein.

Part (ii). The Clarke regularity of $f(z,\eta)$ follows from that of $p(\eta - G(z,\xi))$. The proof is complete. ∎

2.2 True problem

Let $\mu \in \mathscr{C}_+^*([a,b])$ and define the Lagrange function of problem (1.3):

$$\mathscr{L}(x,\mu) := \mathbb{E}[H(z,\xi)] + \int_a^b f(z,\eta)\mu(d\eta).$$

The following optimality condition follows from a discussion at page 499 in [4] by Bonnans and Shapiro and [4, Theorem 5.107], and it is also established by Dentcheva and Ruszczyński, see e.g. [10, Theorem 4.2].

Theorem 2.1 (Optimality condition) *Consider the relaxed problem (1.3). Assume $H(z,\xi)$ is convex and $G(z,\xi)$ is concave w.r.t. z for almost every $\xi \in \Xi$. Let z^* be an optimal solution of the problem. Under Assumptions 2.1-2.2, there exists a measure $\mu^* \in \mathscr{C}_+^*([a,b])$ such that*

$$\begin{cases} z^* \in \arg\min_{z \in Z_0} \mathscr{L}(z,\mu^*), \\ f(z^*,\eta) \le 0, \ \forall \eta \in [a,b], \\ \int_a^b f(z^*,\eta)\mu(d\eta) = 0. \end{cases} \tag{2.8}$$

The set of measures μ^* satisfying (2.8) is nonempty, convex and bounded, and is the same for any optimal solution of the problem.

It is possible to characterize the optimality conditions (2.8) in terms of derivatives of the underlying functions. Indeed, Dentcheva and Ruszczyński derived this kind of first order optimality conditions in both convex and nonconvex case, see [13]. Here we review some of them that are relevant to this paper.

Recall that the Bouligand tangent cone to a set $X \subset \mathbb{R}^n$ at a point $x \in X$ is defined as follows. Let

$$\mathcal{T}_X(x) := \{u \in \mathbb{R}^n : d(x + tu, X) = o(t), t \ge 0\}$$

denote the tangent cone of X at x. The normal cone to X at x, denoted by $\mathcal{N}_X(x)$, is defined as the polar of the tangent cone:

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

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

Theorem 2.2 (First order necessary conditions) *Let Assumption 2.1 hold. Let $z^* \in Z_0$ be a local optimal solution to the true problem (1.3). Assume that the differential constraint qualification is satisfied at z^*. Then the following assertions hold.*

(i) There exists $\mu^* \in \mathscr{C}_+^*([a,b])$ such that

$$\begin{cases} 0 \in \nabla \mathbb{E}[H(z^*, \xi)] + \int_a^b \partial_z f(z^*, \eta)\mu^*(d\eta) + \mathcal{N}_{Z_0}(z^*), \\ f(z^*, \eta) \leq 0, \forall \eta \in [a,b], \\ \int_a^b f(z^*, \eta)\mu^*(d\eta) = 0, \end{cases} \tag{2.9}$$

where

$$\int_a^b \partial_z f(z,\eta)\mu(d\eta) = \left\{ \int_a^b \phi(\eta)\mu(d\eta) : \phi(\eta) \in \partial_z f(z,\eta) \ \text{ and } \ \phi(\eta) \ \text{ is integrable} \right\}.$$

(ii) Assume in addition that: (a) for every $\eta \in [a,b]$, $f(\cdot, \eta)$ is Clarke regular on Z_0, (b) for every fixed $(z,\eta) \in Z_0 \times [a,b]$ and sequence $\{(z_k, \eta_k)\} \to (z, \eta)$,

$$\text{conv } \underline{\lim}_{(z_k,\eta_k)\to(z,\eta)} \partial_z f(z_k, \eta_k) \subset \partial_z f(z, \eta).$$

Then there exist multipliers $\lambda_i^* \geq 0$, $i = 1, \cdots, m$ not all of them being zero, and points $\eta_i \in [a,b]$, $i = 1, \cdots, m$ with $m \leq n+1$ such that

$$0 \in \nabla \mathbb{E}[H(z^*, \xi)] + \sum_{i=1}^m \lambda_i^* \partial_z f(z^*, \eta_i) + \mathcal{N}_{Z_0}(z^*), \eta_i \in \Delta(z^*), \tag{2.10}$$

where $\Delta(z^*) := \{\eta \in [a,b] : f(z^*, \eta) = 0\}$.

Proof. Part (i) is [13, Theorem 5].

Part (ii). Borwein and Zhu [6, Theorem 3.17] derived the first order optimality conditions similar to (2.10) for a general class of nonsmooth semi-infinite programming problems in terms of Fréchet subdifferentials where the underlying functions are lower semicontinuous. It seems that their results imply (2.10) in that Clarke subdifferential of a locally Lipschitz function coincides with the convex hull of the Fréchet subdifferential of the function. Here we include a proof, which is relatively easy to derive based on [26, Theorem 3.2] in that f is locally Lipschitz continuous.

Let

$$\phi(z) := \max\{f(z, \eta), \eta \in [a,b]\}$$

and

$$\hat{\phi}(z) := \max\{\mathbb{E}[H(z, \xi)] - \mathbb{E}[H(z^*, \xi)], \phi(z)\}.$$

Then z^* is a local optimal solution of problem (1.3) if and only if it is a local minimizer of $\hat{\phi}(z)$ and z^* is feasible. Applying the first order optimality condition to $\hat{\phi}(\cdot)$ at point z^*, we have

$$0 \in \partial\hat{\phi}(z^*) + \mathcal{N}_{Z_0}(z^*). \tag{2.11}$$

Under the conditions of part (ii), it follows by virtue of [26, Theorem 3.2],

$$\partial\phi(z^*) = \text{conv}\left\{\bigcup_{\eta\in\Delta(z^*)} \partial_z f(z^*, \eta)\right\}$$

and through [8, Proposition 2.3.12],

$$\begin{aligned}\partial\hat{\phi}(z^*) &\subset \text{cl}\left\{\text{conv}\left\{\nabla\mathbb{E}[H(z^*,\xi)], \partial\phi(z^*)\right\}\right\} \\ &= \text{cl}\left\{\text{conv}\left\{\nabla\mathbb{E}[H(z^*,\xi)], \bigcup_{\eta\in\Delta(z^*)} \partial_z f(z^*, \eta)\right\}\right\},\end{aligned}$$

where "cl" denotes the closure of a set. The closure can be dropped as the convex hull of the subdifferential is closed in finite dimensional space. Consequently, optimality condition (2.11) implies that

$$0 \in \text{conv}\left\{\nabla\mathbb{E}[H(z^*,\xi)], \bigcup_{\eta\in\Delta(z^*)} \partial_z f(z^*, \eta)\right\} + \mathcal{N}_{Z_0}(z^*),$$

that is, there exists

$$\zeta^* \in \text{conv}\left\{\nabla\mathbb{E}[H(z^*,\xi)], \bigcup_{\eta\in\Delta(z^*)} \partial_z f(z^*, \eta)\right\}$$

such that

$$0 \in \zeta^* + \mathcal{N}_{Z_0}(z^*). \tag{2.12}$$

Since $\text{conv}\left\{\nabla\mathbb{E}[H(z^*,\xi)], \bigcup_{\eta\in\Delta(z^*)} \partial_z f(z^*, \eta)\right\}$ is a convex and compact set in $\mathbb{R}^n$, by the Carathéodory's theorem (convex hull), there exists m vectors $\zeta_1^*, \ldots, \zeta_m^*$ $(m \leq n+1)$ with

$$\zeta_i^* \in \left\{\nabla\mathbb{E}[H(z^*,\xi)], \bigcup_{\eta\in\Delta(z^*)} \partial_z f(z^*, \eta)\right\}, \text{ for } i = 1, \cdots, m \tag{2.13}$$

and nonnegative numbers $\lambda_0, \lambda_1, \cdots, \lambda_m$ such that $\sum_{i=0}^m \lambda_i = 1$ and

$$\zeta^* = \lambda_0 \nabla\mathbb{E}[H(z^*,\xi)] + \sum_{i=1}^m \lambda_i \zeta_i^*.$$

Inclusion (2.13) implies that there exist at most m points $\eta_1, \ldots, \eta_m \in \Delta(z^*)$ such that

$$\{\zeta_1^*, \ldots, \zeta_m^*\} \subset \left\{\nabla\mathbb{E}[H(z^*,\xi)], \bigcup_{i=1}^m \partial_z f(z^*, \eta_i)\right\}.$$

Based on the discussions above and the first order optimality condition (2.12), we conclude that

$$0 \in \text{conv}\left\{\nabla\mathbb{E}[H(z^*,\xi)], \bigcup_{i=1}^m \partial_z f(z^*, \eta_i)\right\} + \mathcal{N}_{Z_0}(z^*), \tag{2.14}$$

i.e., there exists constants $\lambda_i^* \geq 0$, $i = 0, 1, \cdots, m$ with $\sum_{i=0}^m \lambda_i^* = 1$ such that

$$0 \in \lambda_0^* \nabla \mathbb{E}[H(z^*, \xi)] + \sum_{i=1}^m \lambda_i^* \partial_z f(z^*, \eta_i) + \mathcal{N}_{Z_0}(z^*).$$

Under the differential constraint qualification at z^*, it is easy to verify that $\lambda_0^* \neq 0$. Therefore the inclusion above implies (2.10). ■

Note that if $H(z, \xi)$ is convex and $G(z, \xi)$ is concave w.r.t. z for almost every $\xi \in \Xi$, then problem (1.3) is convex and the differential constraint qualification in Theorem 2.2 can be weakened to Slater constraint qualification (uniform dominance condition). This is because in such a case the first order necessary condition (2.9) follows straightforwardly from (2.8).

We call a tuple (z^*, μ^*) satisfying (2.9) a *KKT pair* of problem (1.3), z^* a *Clarke stationary point* and μ^* the corresponding Lagrange multiplier.

2.3 SAA problem

We now move on to discuss the optimality conditions for the SAA problem (1.5). We need the following technical results.

Proposition 2.2 *Let Assumption 2.1 hold. Then*

(i) w.p.1 $h_N(z)$ and $\frac{1}{N}\sum_{i=1}^N G(z, \xi^i)$ converge respectively to $\mathbb{E}[H(z, \xi)]$ and $\mathbb{E}[G(z, \xi)]$ uniformly over any compact subset of Z_0 as $N \to \infty$;

(ii) w.p.1 $\frac{1}{N}\sum_{i=1}^N p(\eta - G(z, \xi^i))$ and $\frac{1}{N}\sum_{i=1}^N p(\eta - Y(\xi^i))]$ converge respectively to $\mathbb{E}[p(\eta - G(z, \xi))]$ and $\mathbb{E}[p(\eta - Y(\xi))]$ uniformly on $Z_0 \times [a, b]$ and $[a, b]$ as $N \to \infty$;

(iii) if, in addition, Assumption 2.2 holds, then the SAA problem (1.5) satisfies uniform dominance condition w.p.1 for N sufficiently large, that is, there exists a point $z_0 \in Z_0$ such that

$$\sup_{\eta \in [a,b]} f_N(z_0, \eta) < 0$$

w.p.1 for N sufficiently large.

Proof. Part (i) follows straightforwardly from classical uniform law of large numbers under Assumption 2.1, see [29, Lemma A1]. Part (ii) follows from the same argument in that $p(\eta - G(z, \xi))$ is Lipschitz continuous in (η, z) with integrable modulus $\max_{z \in Z_0}(1 + \|\nabla_z G(z, \xi)\|)$ while $p(\eta - Y(\xi))$ is Lipschitz continuous in η with modulus 1. Part (iii) follows from Part (ii) and Assumption 2.2. ■

Let $\mu \in \mathscr{C}_+^*([a, b])$. Define the Lagrange function of sample average approximation problem (1.5):

$$\mathscr{L}_N(z, \mu) := h_N(z) + \int_a^b f_N(z, \eta)\mu(d\eta),$$

where $h_N(z)$ and $f_N(z, \eta)$ are defined as in (1.5).

From Proposition 2.2, we know that for N sufficiently large, the SAA problem (1.5) satisfies the Slater's constraint qualification w.p.1. By invoking [4, Theorem 5.107], we have the following optimality conditions for problem (1.5).

Theorem 2.3 *Consider the sample average approximation problem (1.5). Suppose that $H(z,\xi)$ is convex and $G(z,\xi)$ is concave w.r.t. z for almost every $\xi \in \Xi$, and (1.3) satisfies Assumptions 2.1 and 2.2. If z_N is an optimal solution of the problem, then w.p.1 there exists $\mu_N \in \mathscr{C}_+^*([a,b])$ such that*

$$\begin{cases} z_N \in \arg\min_{z\in Z_0} \mathscr{L}_N(z,\mu_N), \\ f_N(z_N,\eta) \le 0, \forall \eta \in [a,b], \\ \int_a^b f_N(z_N,\eta)\mu_N(d\eta) = 0. \end{cases} \tag{2.15}$$

The set of measures μ_N satisfying (2.8) is nonempty, convex and bounded, and is the same for any optimal solution of the problem.

In what follows, we derive first order optimality condition for the SAA problem (1.5). We need the following technical result.

Recall that for a set D, the support function of D is defined as

$$\sigma(D,u) = \sup_{d\in D} d^T u.$$

Let D_1, D_2 be two convex and compact subsets of $\mathbb{R}^m$. Let $\sigma(D_1,u)$ and $\sigma(D_2,u)$ denote the support functions of D_1 and D_2 respectively. Then

$$\mathbb{D}(D_1,D_2) = \max_{\|u\|\le 1} (\sigma(D_1,u) - \sigma(D_2,u)). \tag{2.16}$$

The above relationship is known as Hörmander's formula, see [7, Theorem II-18].

Lemma 2.1 *Let $v(x,\xi)$ be a continuous function defined on $\mathbb{R}^n \times \mathbb{R}^k$ and $\xi : \Omega \to \mathbb{R}^k$ be a random vector. Let $\xi^1, \cdots, \xi^N$ be an iid sampling of ξ. Suppose that $v(x,\xi)$ is Lipschitz continuous in x for almost every ξ and its Lipschitz modulus is integrably bounded by $\kappa(\xi)$. Then*

$$\lim_{N\to\infty} \sup_{x\in\mathcal{X},\|u\|\le 1} \left[\frac{1}{N}\sum_{i=1}^N v^o(x,\xi^i;u) - \mathbb{E}[v_\epsilon^o(x,\xi;u)]\right] \le 0, \tag{2.17}$$

where $\mathcal{X}$ is a compact set in $\mathbb{R}^n$, u is a fixed nonzero vector of $\mathbb{R}^n$, and $v_\epsilon^o(x,\xi;u)$ is ϵ-Clarke generalized derivative, that is,

$$v_\epsilon^o(x,\xi;u) = \sup_{\zeta\in\partial_x^\epsilon v(x,\xi)} \zeta^T u,$$

where

$$\partial_x^\epsilon v(x,\xi) = \bigcup_{x'\in x+\epsilon\mathcal{B}} \partial_x v(x',\xi)$$

and $\mathcal{B}$ is the unit ball in $\mathbb{R}^n$.

Proof. For fixed positive number ϵ, it follows by [31, Theorem 2],

$$\lim_{N\to\infty}\sup_{x\in\mathcal{X}}\mathbb{D}\left(\frac{1}{N}\sum_{i=1}^{N}\partial_x v(x,\xi^i),\mathbb{E}[\partial_x^\epsilon v(x,\xi)]\right)=0 \tag{2.18}$$

w.p.1. Using Hörmander's formula, we have

$$\begin{aligned}&\mathbb{D}\left(\frac{1}{N}\sum_{i=1}^{N}\partial_x v(x,\xi^i),\mathbb{E}[\partial_x^\epsilon v(x,\xi)]\right)\\ =&\sup_{\|u\|\le 1}\left[\sigma\left(\frac{1}{N}\sum_{i=1}^{N}\partial_x v(x,\xi^i);u\right)-\sigma(\mathbb{E}[\partial_x^\epsilon v(x,\xi)];u)\right].\end{aligned} \tag{2.19}$$

Using the property of the support function (see e.g. [17]) and the definition of the Clarke generalized gradient, we have

$$\sigma\left(\frac{1}{N}\sum_{i=1}^{N}\partial_x v(x,\xi^i);u\right)=\frac{1}{N}\sum_{i=1}^{N}\sigma(\partial_x v(x,\xi^i);u)=\frac{1}{N}\sum_{i=1}^{N}v^o(x,\xi^i;u). \tag{2.20}$$

On the other hand, by [25, Proposition 3.4],

$$\sigma(\mathbb{E}[\partial_x^\epsilon v(x,\xi)];u)=\mathbb{E}[\sigma(\partial_x^\epsilon v(x,\xi);u)]. \tag{2.21}$$

Combining (2.19)–(2.21) , we have

$$\begin{aligned}\sup_{x\in\mathcal{X}}\mathbb{D}\left(\frac{1}{N}\sum_{i=1}^{N}\partial_x v(x,\xi^i),\mathbb{E}[\partial_x^\epsilon v(x,\xi)]\right) &= \sup_{x\in\mathcal{X}}\sup_{\|u\|\le 1}\left[\frac{1}{N}\sum_{i=1}^{N}v^o(x,\xi^i;u)-\mathbb{E}[\sigma(\partial_x^\epsilon v(x,\xi);u)]\right]\\ &= \sup_{x\in\mathcal{X},\|u\|\le 1}\left[\frac{1}{N}\sum_{i=1}^{N}v^o(x,\xi^i;u)-\mathbb{E}[\sigma(\partial_x^\epsilon v(x,\xi);u)]\right],\end{aligned}$$

which immediately yields (2.17) through (2.18). ■

We are now ready to state the first order optimality conditions (Karush-Kuhn-Tucker conditions) of the SAA problem (1.5) in terms of Clarke subdifferentials.

Theorem 2.4 (First order necessary conditions) *Let $z_N \in Z_0$ be a local optimal solution to the sample average approximation problem (1.5). Let $\hat{Z}$ denote a compact set which contains all cluster points of $\{z_N\}$. Suppose: (a) Assumption 2.1 holds, (b) problem (1.3) satisfies the differential constraint qualification at every point $z \in \hat{Z}$; (c) $\partial_z f(\cdot,\eta)$ is upper semicontinuous uniformly w.r.t. η. Then w.p.1 problem (1.5) satisfies the differential constraint qualification at z_N for N sufficiently large and there exists $\mu_N \in \mathscr{C}_+^*([a,b])$ such that*

$$\left\{\begin{array}{l}0\in\nabla h_N(z_N)+\int_a^b\partial_z f_N(z_N,\eta)\mu_N(d\eta)+\mathcal{N}_{Z_0}(z_N),\\ f_N(z_N,\eta)\le 0,\forall\eta\in[a,b],\\ \int_a^b f_N(z_N,\eta)\mu_N(d\eta)=0.\end{array}\right. \tag{2.22}$$

Proof. Under the differential constraint qualification at z, there exist a constant $\delta > 0$ and a vector $u \neq 0$ (which depends on z) such that

$$\sup_{\zeta\in\partial_z f(z,\eta)}\zeta^T u = f^o(z,\eta;u)\le -\delta.$$

In what follows, we show that

$$\sup_{\eta\in[a,b]}\sup_{\zeta\in\partial_z f_N(z_N,\eta)} \zeta^T u \le -\delta/2 \tag{2.23}$$

w.p.1 for N sufficiently large, i.e., problem (1.5) satisfies the differential constraint qualification. Let $z \in \hat{Z}$,

$$\partial_z^\epsilon f(z,\eta) := \bigcup_{z'\in z+\epsilon\mathcal{B}} \partial_z f(z',\eta)$$

and

$$f_\epsilon^o(z,\eta;u) := \sup_{\zeta\in\partial_z^\epsilon f(z,\eta)} \zeta^T u.$$

The uniform upper semicontinuity of $\partial_z f(\cdot,\eta)$ allows us to find a sufficiently small ϵ (depending on z) such that

$$f_\epsilon^o(z',\eta;u) \le -\frac{3}{4}\delta, \forall \eta\in[a,b]$$

for all $z' \in B(z,\rho)$, where $B(z,\rho)$ denotes a closed neighborhood of z relative to Z_0 and ρ depends on z. Let $\tilde{Z}$ denote the closed ρ neighborhood of $\hat{Z}$ relative to Z_0. Applying Lemma 2.1 to $\tilde{Z}$, we can find an N_0 such that

$$f_N^o(z',\eta;u) - f_\epsilon^o(z',\eta;u) \le \frac{\delta}{2}, \forall z'\in\tilde{Z} \quad \text{w.p.1}$$

for $N \ge N_0$. Using this inequality, we have

$$\sup_{\eta\in[a,b]}\sup_{\zeta\in\partial_z f_N(z_N,\eta)} \zeta^T u = \sup_{\eta\in[a,b]} f_N^o(z_N,\eta;u) \le \sup_{\eta\in[a,b]} f_\epsilon^o(z_N,\eta;u) + \frac{\delta}{2} \le -\frac{\delta}{4}$$

w.p.1 as long as $z_N \in B(z,\rho)$.

The discussions above show that for N sufficiently large, problem (1.5) satisfies the differential constraint qualification. By [13, Theorem 4], we obtain the first order optimality conditions (2.22). ■

Note that in the case when $H(z,\xi)$ is convex and $G(z,\xi)$ is concave w.r.t. z for almost every $\xi\in\Xi$, problems (1.3) and (1.5) become convex. Subsequently the differential constraint qualification can be weakened to Slater constraint qualification. This is because Proposition 2.1 (iii) ensures the SAA problems satisfying Slater constraint qualification w.p.1 for N sufficiently large and the first order necessary condition (2.22) follows straightforwardly from (2.15).

We call a tuple $(z_N, \mu_N(\cdot))$ satisfying (2.22) a *KKT pair* of problem (1.5), z_N a *Clarke stationary point* and $\mu_N(\cdot)$ the corresponding Lagrange multiplier.

We make a blanket assumption that throughout the rest of the paper the conditions of Theorems 2.2 and 2.4 hold.

3 Almost sure convergence

Consider the sample average approximation problem (1.5). Assume that for each given sampling, we solve the problem and obtain a stationary point z_N which satisfies (2.22), we investigate the convergence of z_N as N increases. Note that there is a significant difference between the convergence analysis here and that in [23] in that the latter was based on a penalized formulation of problem (1.3). Generally speaking, there is a gap between stationarity of (1.3) and that of its penalized problem, see a discussion at the end of [23, Theorem 3.13].

Assumption 3.1 *There exists a compact subset* $\mathcal{Z} \times \hat{\mathscr{C}}_+^*([a,b]) \subset Z_0 \times \mathscr{C}_+^*([a,b])$ *and a positive number* N_0 *such that w.p.1 problem (1.5) has a KKT pair* $(z_N, \mu_N(\cdot)) \in Z_0 \times \hat{\mathscr{C}}_+^*([a,b])$ *for* $N \geq N_0$.

Assumption 3.1 is crucial in our main convergence result, Theorem 3.1. Therefore it would be helpful to discuss how strong the assumption is.

First, let us consider the boundedness of z_N. In many practical applications particularly in portfolio optimization, Z_0 is a compact set and hence z_N is bounded. In the case when Z_0 is unbounded, sufficient conditions for the existence and boundedness of z_N may be derived from the property of the underlying functions, e.g., uniform coercivity. We will not discuss this as it is not our main focus here. In the case when $\{z_N\}$ has a bounded subsequence w.p.1, we may carry out our convergence analysis by focusing on the subsequence. We omit this for simplicity of notation and clarity of presentation.

Next, we discuss the boundedness of μ_N. To simplify the discussion, we assume Z_0 is compact. We say problem (1.2) satisfies the *no nonzero abnormal multipliers constraint qualification* (NNAMCQ) at point $z \in Z_0$ if there is no multiplier $\mu \in \mathscr{C}_+^*([a,b])$ such that $\|\mu\| \neq 0$ and

$$\left\{ \begin{array}{ll} 0 \in & \int_a^b \partial_z f(z,\eta)\mu(d\eta) + \mathcal{N}_{Z_0}(z), \\ 0 \geq & f(z,\eta), \ \forall \eta \in [a,b], \\ 0 = & \int_a^b f(z,\eta)\mu(d\eta). \end{array} \right. \tag{3.24}$$

NNAMCQ is well known, see for instance Borwein [5]. In particular, Ye extensively exploited the constraint qualification for studying the first order optimality conditions of mathematical programs with equilibrium constraints (MPEC), see [37].

Here, we claim that the NNAMCQ is implied by the well-known extended MFCQ. To see this, let $z \in Z_0$ be a feasible point such that problem (1.2) satisfies the extended MFCQ at z. Assume for the sake of a contradiction that there exists $\mu \in \mathscr{C}_+^*([a,b])$ such that $\|\mu\| \neq 0$ and (3.24) holds. Then there exists $w(\eta) \in \partial_z f(z,\eta)$ and $u \in \mathcal{N}_{Z_0}(z)$ such that

$$0 = \int_a^b w(\eta)\mu(d\eta) + u.$$

Existence of $w(\eta)$ follows from the definition of Aumann's integral [2]. By the definition of extended MFCQ, there exists $d \in \mathcal{T}_{Z_0}(z)$ such that $d^T w(\eta) < 0$ for all $\eta \in [a,b]$. Since $d^T u \leq 0$, we have from above equation that

$$d^T 0 = \int_a^b d^T w(\eta)\mu(d\eta) + d^T u < 0,$$

a contradiction.

The proposition below says that NNAMCQ guarantees the boundedness of μ_N.

Proposition 3.1 *Consider the SAA problem (1.5). Assume: (a) Z_0 is compact, (b) problem (1.3) satisfies NNAMCQ at any cluster point of $\{z_N\}$ and $f(z,\eta)$ is continuously differentiable w.r.t. z at the cluster points for every $\eta \in [a,b]$. Then the sequence of the Langrange multipliers $\{\mu_N\}$ is bounded w.p.1.*

Proof. Since $\{(z_N,\mu_N)\}$ is a sequence of KKT pairs of problem (1.5), then

$$0 \in \nabla h_N(z_N) + \int_a^b \partial_z f_N(z_N,\eta)\mu_N(d\eta) + \mathcal{N}_{Z_0}(z_N), \tag{3.25}$$

$f_N(z_N,\eta) \le 0$ for all $\eta \in [a,b]$ and $\int_a^b f_N(z_N,\eta)\mu_N(d\eta) = 0$, where $\mu_N \in \mathscr{C}_+^*([a,b])$. Assume for the sake of a contradiction that $\{\mu_N\}$ is unbounded, i.e., $\{\|\mu_N\|\}$ is unbounded. Then $\{\|\mu_N\|\}$ has a subsequence which goes to infinity. Let $\hat{\mu}_N = \mu_N/\|\mu_N\|$. Then $\|\hat{\mu}_N\| = 1$. Since $\hat{\mu}_N$ is a Borel measure defined on compact set $[a,b]$, by Helly-Bray's theorem (see Theorems 9.2.1-9.2.3 and Remark 9.2.1 in [3]), it has a weakly convergent subsequence. Moreover, since Z_0 is a compact set, z_N has a subsequence converging to some point $z^* \in Z_0$. By taking a subsequence if necessary, we may assume for the simplicity of notation that $z_N \to z^*$ and $\|\mu_N\| \to \infty$ and $\hat{\mu}_N$ converges weakly to μ^*. Dividing both sides of (3.25) by $\|\mu_N\|$, we have

$$0 \in \nabla h_N(z_N)/\|\mu_N\| + \int_a^b \partial_z f_N(z_N,\eta)\hat{\mu}_N(d\eta) + \mathcal{N}_{Z_0}(z_N). \tag{3.26}$$

By the definition of Aumann's integral [2], there exists an integral selection $\zeta_N(\eta) \in \partial_z f_N(z_N,\eta)$ such that

$$0 \in \nabla h_N(z_N)/\|\mu_N\| + \int_a^b \zeta_N(\eta)\hat{\mu}_N(d\eta) + \mathcal{N}_{Z_0}(z_N). \tag{3.27}$$

Under Assumption 2.1 (a generic assumption we made at the end of Section 2), it follows from Proposition 2.2 that $\nabla h_N(z_N) \to \nabla \mathbb{E}[H(z^*,\xi)]$ and hence

$$\lim_{N\to\infty} \nabla h_N(z_N)/\|\mu_N\| = 0.$$

Moreover, since Z_0 is convex and closed, the normal cone operator $\mathcal{N}_{Z_0}(\cdot)$ is upper semicontinuous, i.e.,

$$\overline{\lim_{N\to\infty}}\, \mathcal{N}_{Z_0}(z_N) \subset \mathcal{N}_{Z_0}(z^*).$$

In what follows, we show

$$\delta^N := \mathbb{D}\left(\int_a^b \partial_z f_N(z_N,\eta)\tilde{\mu}_N(d\eta), \int_a^b \partial_z f(z^*,\eta)\mu^*(d\eta)\right) \to 0. \tag{3.28}$$

as $N \to \infty$. In doing so, we will arrive at

$$0 \in 0 + \int_a^b \partial_z f(z^*,\eta)\mu^*(d\eta) + \mathcal{N}_{Z_0}(z^*),$$

a desired contradiction to our assumption that the true problem satisfies NNAMCQ at z^*.

Observe first that under condition (b), $\partial_z f(z^*,\eta)$ is a singleton and coincides with $\{\nabla_z f(z^*,\eta)\}$. The latter is continuous w.r.t. η. By [8, Proposition 2.1.2] and the definition of Aumann's integral [2], $\int_a^b \partial_z f_N(z_N,\eta)\hat{\mu}_N(d\eta)$ is convex and compact set-valued. By Hörmander's formula, we can reformulate δ^N as:

$$\delta^N = \max_{\|u\|\le 1}\left[\sigma\left(\int_a^b \partial_z f_N(z_N,\eta)\hat{\mu}_N(d\eta), u\right) - \sigma\left(\int_a^b \nabla_z f(z^*,\eta)\mu^*(d\eta), u\right)\right]$$

where

$$\sigma\left(\int_a^b \nabla_z f(z^*,\eta)\mu^*(d\eta), u\right) = \int_a^b u^T \nabla_z f(z^*,\eta)\mu^*(d\eta).$$

Moreover, it follows by [25, Proposition 3.4] that operation $\sigma(\cdot, u)$ and integration are exchangeable which means

$$\delta^N = \max_{\|u\|\le 1}\left[\int_a^b \sigma\left(\partial_z f_N(z_N,\eta), u\right)\hat{\mu}_N(d\eta) - \int_a^b \sigma\left(\nabla_z f(z^*,\eta), u\right)\mu^*(d\eta)\right].$$

Through a simple rearrangement, we have

$$\begin{aligned}\delta^N \quad &\le \quad \int_a^b \max_{\|u\|\le 1}[(\sigma\left(\partial_z f_N(z_N,\eta), u\right) - \sigma\left(\nabla_z f(z^*,\eta), u\right)]\tilde{\mu}_N(d\eta) \\ &\quad + \left|\max_{\|u\|\le 1}\int_a^b \sigma\left(\nabla_z f(z^*,\eta), u\right)(\hat{\mu}_N - \mu^*)(d\eta)\right|. \end{aligned} \tag{3.29}$$

Observe that $\sigma\left(\nabla_z f(z^*,\eta), u\right) = \nabla_z f(z^*,\eta)^T u$ is bounded and continuous in η uniformly w.r.t. u for $\|u\| \le 1$. Therefore the weak convergence of $\hat{\mu}_N(\cdot)$ to $\mu^*(\cdot)$ implies

$$\lim_{N\to\infty}\max_{\|u\|\le 1}\int_a^b \sigma\left(\nabla_z f(z^*,\eta), u\right)(\hat{\mu}_N - \mu^*)(d\eta) = 0. \tag{3.30}$$

In what follows, we show that

$$\lim_{N\to\infty}\int_a^b \max_{\|u\|\le 1}\left[\sigma\left(\partial_z f_N(z_N,\eta), u\right) - \sigma\left(\partial_z f(z^*,\eta), u\right)\right]\hat{\mu}_N(d\eta) \le 0, \tag{3.31}$$

which is adequate to complete our proof in that (3.28) follows from a combination of (3.29)-(3.31). From (2.7), we obtain through [25, Proposition 3.4]

$$\sigma(\nabla_z f(z^*,\eta), u) \quad = \quad \mathbb{E}\left[\sigma(-\nabla_z G(z^*,\xi)^T \partial p(\eta - G(z^*,\xi)), u)\right].$$

On the other hand, since $p(\eta - G(z,\xi)) - p(\eta - Y(\xi))$ is Clarke regular in z for any ξ,

$$\partial_z f_N(z_N,\eta) = \frac{1}{N}\sum_{i=1}^N -\nabla_z G(z_N,\xi^i)^T \partial p(\eta - G(z_N,\xi^i)).$$

Using the property of the support function (see e.g. [17]), we have

$$\sigma\left(\partial_z f_N(z_N,\eta), u\right) = \frac{1}{N}\sum_{i=1}^N \sigma(-\nabla_z G(z_N,\xi^i)^T \partial p(\eta - G(z_N,\xi^i)), u).$$

By virtue of [34, Theorem 4],

$$\lim_{N\to\infty}\sup_{(z,\eta)\in\mathcal{Z}\times[a,b]} \mathbb{D}(\partial_z f_N(z,\eta), \mathbb{E}[-\nabla_z G(z,\xi)^T \partial p(\eta - G(z,\xi))]) = 0,$$

which implies

$$\lim_{N\to\infty}\sigma\left(\partial_z f_N(z_N,\eta), u\right) - \mathbb{E}\left[\sigma(-\nabla_z G(z^*,\xi)^T \partial p(\eta - G(z^*,\xi)), u)\right] = 0$$

and hence (3.31). The proof is complete. ■

Proposition 3.1 gives a sufficient condition for Assumption 3.1. Xu and Zhang considered a similar conditions in [36, Propossion 2.2] in the case when the number of constraints is finite.

We are now ready to state the main result in this section.

Theorem 3.1 *Let $\{(z_N, \mu_N(\cdot))\}$ be a sequence of KKT pairs of problem (1.5) and $(z^*, \mu^*(\cdot))$ be a cluster point w.p.1. Suppose: (a) Assumptions 2.1 and 3.1 hold, (b) $\nabla^2 H(z,\xi)$ is integrably bounded, (c) the support set Ξ of random vector ξ is bounded, (d) $f(z,\eta)$ is continuously differentiable w.r.t. z for every $\eta \in [a,b]$. Then w.p.1 $(z^*, \mu^*(\cdot))$ is a KKT pair of the true problem (1.3), which satisfies the KKT system (2.9).*

Proof. Assumption 3.1 guarantees the boundedness of sequence $\{(z_N, \mu_N(\cdot))\}$. Assume without loss of generality that $(z_N, \mu_N(\cdot))$ converges to $(z^*, \mu^*(\cdot))$ as $N \to \infty$. In view of the KKT conditions (2.22) and (2.9), it suffices to show that w.p.1

$$\begin{aligned}\overline{\lim_{N\to\infty}} \big[\nabla h_N(z_N) \quad &+ \quad \int_a^b \partial_z f_N(z_N,\eta)\mu_N(d\eta) + \mathcal{N}_{Z_0}(z_N)\big] \\ &\subset \quad \mathbb{E}[\nabla H(z,\xi)] + \int_a^b \partial_z f(z^*,\eta)\mu^*(d\eta) + \mathcal{N}_{Z_0}(z^*),\end{aligned} \tag{3.32}$$

$$\lim_{N\to\infty} f_N(z_N,\eta) = f(z^*,\eta) \tag{3.33}$$

and

$$\lim_{N\to\infty} \int_a^b f_N(z_N,\eta)\mu_N(d\eta) = \int_a^b f(z^*,\eta)\mu^*(d\eta). \tag{3.34}$$

Under Assumption 2.1, it follows from Proposition 2.2 that $f_N(z,\eta)$ converges uniformly to $f(z,\eta)$ over $Z_0 \times [a,b]$, which implies (3.33). Moreover, since $\mu_N(\cdot)$ is bounded and $f(z^*,\eta)$ is continuous, it is easy to observe that (3.33) implies (3.34).

Let us look at (3.32). Under condition (a), the classical uniform law of large numbers ensures that $\nabla h_N(z)$ converges uniformly to $\mathbb{E}[\nabla_z H(z,\xi)]$ over any compact subset of Z_0, which implies

$$\lim_{N\to\infty} \nabla h_N(z_N) = \nabla \mathbb{E}[H(z^*,\xi)], \text{ w.p.1.}$$

On the other hand, upper semicontinuity of the normal cone implies

$$\mathcal{N}_{Z_0}(z_N) \subset \mathcal{N}_{Z_0}(z^*).$$

Therefore we are left to show

$$\overline{\lim_{N\to\infty}} \int_a^b \partial_z f_N(z_N,\eta)\mu_N(d\eta) \subset \int_a^b \partial_z f(z^*,\eta)\mu^*(d\eta) \tag{3.35}$$

w.p.1. This is similar to the proof of (3.28) in the proof of Proposition 3.1. We omit the details. ■

Note that if Assumption 3.1 in Theorem 3.1 is dropped, then sequence of the KKT pair of the SAA problem may not have a cluster pair. However, if it does have a cluster pair, then we can show that it is a KKT pair of the true problem w.p.1. This can be achieved by taking a subsequence in the proof. We leave these to interested readers.

4 Exponential rate of convergence

In this section, we take one step further from the analysis in preceding section to look into the exponential rate of convergence of KKT pairs defined by (2.22). To this end, we investigate uniform exponential rate of convergence of the set-valued mapping in the sample average approximated KKT system (2.22). One of the main technical difficulties that we need to tackle with is the exponential rate of convergence of $\partial_z f_N(z,\eta)$. We need some intermediate concepts and results.

Definition 4.1 (Almost H-calmness) Let $\phi : \mathbb{R}^n \times \Xi \to \mathbb{R}$ be a real valued function and $\xi : \Omega \to \Xi \subset \mathbb{R}^k$ be a random vector defined on probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{X} \subset \mathbb{R}^n$ be a closed subset of $\mathbb{R}^n$ and $x \in \mathcal{X}$ be a fixed point. ϕ is said to be *almost H-calm at* x with modulus $\kappa_x(\xi)$ and order γ_x if for any $\epsilon > 0$, there exist a (measurable) function $\kappa_x : \Xi \to \mathbb{R}_+$, positive numbers γ_x, $\delta_x(\epsilon)$, C and an open set $\Delta_x(\epsilon) \subset \Xi$ such that

$$\text{Prob}(\xi \in \Delta_x(\epsilon)) \leq C\epsilon \tag{4.36}$$

and

$$|\phi(x', \xi) - \phi(x, \xi)| \leq \kappa_x(\xi)||x' - x||^{\gamma_x}$$

for all $\xi \in \Xi \backslash \Delta_x(\epsilon)$ and all $x' \in B(x, \delta_x(\epsilon)) \bigcap \mathcal{X}$. Here and later on, $B(x, \delta)$ denotes the δ-neighborhood of x.

Let $z \in \mathbb{R}^n$ and G be defined as in (1.1). Define set

$$\Xi(z, \eta) := \{\xi : G(z, \xi) = \eta, \xi \in \Xi\}.$$

Let F be defined by (1.4). Obviously, $\Xi(z, \eta)$ consists of the set of ξ such that $F(\cdot, \eta, \xi)$ is not differentiable at z.

Lemma 4.1 *Let $\mathcal{Z} \subset \mathbb{R}^n$ be a compact subset of $\mathbb{R}^n$, ξ be a continuous random variable and ν be the Lebesgue measure relative to Ξ. Assume that $\Xi(z, \eta)$ is compact and*

$$\nabla_\xi G(z, \xi) \neq 0 \tag{4.37}$$

for $\xi \in \Xi(z, \eta)$. Then $\nu(\Xi(z, \eta)) = 0$. Moreover, for any $\epsilon > 0$ and any fixed $(z, \eta) \in \mathcal{Z} \times [a, b]$, there exists an open set $\Xi^\epsilon(z, \eta)$ (depending on all z, η and ϵ) such that $\Xi(z, \eta) \subset \Xi^\epsilon(z, \eta)$ and $\nu(\Xi^\epsilon(z, \eta) \cap \Xi) \leq \epsilon$.

Proof. Let $(\hat{z}, \hat{\eta})$ be any fixed point in $\mathcal{Z} \times [a, b]$ and $\hat{\xi} \in \Xi(\hat{z}, \hat{\eta})$. Then $\hat{\eta} - G(\hat{z}, \hat{\xi}) = 0$ and by assumption, $\nabla_\xi G(\hat{z}, \hat{\xi}) \neq 0$. Note that ξ is a q-dimensional vector in $\mathbb{R}^q$. Let us write $\xi = (\xi_1, \cdots, \xi_q)^T$. Assume without loss of generality that $\frac{\partial G(\hat{z}, \hat{\xi})}{\partial \xi_1} \neq 0$. By the classical implicit function theorem, there exists a neighborhood of point $(\hat{z}, \hat{\eta}, \hat{\xi})$, denoted by $B((\hat{z}, \hat{\eta}, \hat{\xi}), \delta_{\hat{z}, \hat{\eta}, \hat{\xi}})$ and a unique implicit function $\xi_1(z, \eta, \xi_{-1})$, where $\xi_{-1} = (\xi_2, \cdots, \xi_q)^T$, such that $\xi_1(\hat{z}, \hat{\eta}, \hat{\xi}_{-1}) = \hat{\xi}_1$ and

$$\eta' - G(z', \xi_1(z', \eta', \xi'_{-1}), \xi'_{-1}) = 0$$

for (z', η', ξ'_{-1}) in the neighborhood of $(\hat{z}, \hat{\eta}, \hat{\xi}'_{-1})$.

We consider the implicit function $r(\xi_{-1}) := \xi_1(\hat{z}, \hat{\eta}, \xi_{-1})$ defined on Ξ. The graph of the function is a $(q-1)$-dimensional manifold on $\mathbb{R}^q$ relative to Ξ. We claim that under condition (4.37), there exists a finite number of such manifolds in Ξ. Assume for a contradiction that there exists an infinite number of such manifolds, denoted by $\{\xi_1^k(\hat{z}, \hat{\eta}, \cdot)\}$. Let $\xi^k := (\xi_1^k(\hat{z}, \hat{\eta}, \xi_{-1}^k), \xi_{-1}^k)$ be a point at the k-th manifold. Since Ξ is compact, by taking a subsequence if necessary, we may assume for simplicity of notation that $\{\xi_1^k(\hat{z}, \hat{\eta}, \xi_{-1}^k)\} \to \xi_1^*(\hat{z}, \hat{\eta}, \xi_{-1}^*)$, and $\xi^* \in \Xi(\hat{z}, \hat{\eta})$, which means that in a neighborhood of $\xi^* = (\xi_1^*, \xi_{-1}^*)$, there exists an infinite number of manifolds. This is impossible under condition (4.37) as by the implicit function theorem there exists only a unique such manifold in the neighborhood.

The discussion above shows that the Lebesgue measure of $\Xi(\hat{z}, \hat{\eta})$ relative to Ξ is zero as the Lebesgue measure of each manifold is 0 relative to Ξ. Moreover, there exists an open set $\Xi^\epsilon(\hat{z}, \hat{\eta})$ such that $\Xi(\hat{z}, \hat{\eta}) \subset \Xi^\epsilon(\hat{z}, \hat{\eta})$ and $\nu(\Xi^\epsilon(\hat{z}, \hat{\eta})) \to 0$ as $\epsilon \downarrow 0$. Since $(\hat{z}, \hat{\eta})$ can be any point in $\mathcal{Z} \times [a, b]$, we complete the proof. ■

Proposition 4.1 *Let $\mathcal{Z} \subset \mathbb{R}^n$ be a compact subset of $\mathbb{R}^n$, ξ be a continuous random variable and F be defined by (1.4). Assume: (a) $G(z, \xi)$ is twice continuously differentiable w.r.t. z for almost every $\xi \in \Xi$; (b) there exists an integrable function $\kappa : \Xi \to \mathbb{R}$ such that $\nabla_z G(\cdot, \xi)$ is locally Lipschitz continuous with modulus $\kappa(\xi)$ for every $\xi \in \Xi$; (c) $\Xi(z, \eta)$ is compact and (4.37) holds for $\xi \in \Xi(z, \eta)$. Then*

(i) $\mathbb{E}[F_z^o(z, \eta, \xi; u)]$ is continuous w.r.t. (z, η, u) and it is uniformly continuous w.r.t. ξ if $G(x, \xi)$ is uniformly continuous w.r.t. ξ and $Y(\xi)$ is continuous w.r.t. ξ;

(ii) if Ξ is compact, $F_z^o(z, \eta, \xi; u)$ is almost H-calm w.r.t. (z, η, u) with modulus $\kappa(\xi)$ and order 1 on $\mathcal{Z}$.

Proof. Part (i). The continuity argument is a well-known result. Indeed, in this case $\mathbb{E}[F(z, \eta, \xi)]$ is continuously differentiable, see for instance [27, Theorem 1]. The uniform continuity w.r.t ξ following from the fact that the max-function $p(\cdot)$ and the expectation $\mathbb{E}[\cdot]$ preserve the uniform continuity.

Part (ii). Let $\epsilon > 0$ and $(\bar{z}, \bar{\eta}) \in \mathcal{Z} \times [a, b]$ be fixed. Under condition (c), it follows by Lemma 4.1 that there exists an open subset $\Xi^\epsilon(\bar{z}, \bar{\eta})$ such that $\Xi(\bar{z}, \bar{\eta}) \subset \Xi^\epsilon(\bar{z}, \bar{\eta})$ and $\nu(\Xi^\epsilon(\bar{z}, \bar{\eta})) \leq \epsilon$. Let $\bar{\xi} \notin \Xi^\epsilon(\bar{z}, \bar{\eta})$. Then two cases may occur: (a) $\bar{\eta} - G(\bar{z}, \bar{\xi}) > 0$, (b) $\bar{\eta} - G(\bar{z}, \bar{\xi}) < 0$. We only consider the case that $\bar{\eta} - G(\bar{z}, \bar{\xi}) > 0$ as the other cases can be dealt with in a similar way.

Under condition (b), we can find a δ-neighborhood of $(\bar{z}, \bar{\eta}, \bar{\xi})$ (depending on $\bar{z}$, $\bar{\eta}$ and $\bar{\xi}$), denoted by $B((\bar{z}, \bar{\eta}, \bar{\xi}), \delta_{\bar{z}, \bar{\eta}, \bar{\xi}})$, such that for all $(z, \eta, \xi) \in B((\bar{z}, \bar{\eta}, \bar{\xi}), \delta_{\bar{z}, \bar{\eta}, \bar{\xi}}) \cap \mathcal{Z} \times [a, b] \times \Xi$, $\eta - G(z, \xi) > 0$ and

$$|F_z^o(z, \eta, \xi; u) - F_z^o(\bar{z}, \bar{\eta}, \xi; u)| = |\nabla_z G(z, \xi)^T u - \nabla_z G(z, \xi)^T u| \leq \kappa(\xi) ||z - \bar{z}||. \tag{4.38}$$

Since $\Xi \backslash \Xi^\epsilon(\bar{z}, \bar{\eta})$ is compact, we claim through the finite covering theorem that there exists a unified $\delta_{\bar{z}, \bar{\eta}} > 0$ such that (4.38) holds for all $(z, \eta) \in B(\bar{z}, \bar{\eta}, \delta_{\bar{x}, \bar{\eta}})$ and all $\xi \in \Xi \backslash \Xi^\epsilon(\bar{z}, \bar{\eta})$. This shows that $F_z^o(z, \eta, \xi; u)$ is almost H-calm with modulus $\kappa(\xi)$ and order 1 over $\mathcal{Z}$. ■

Proposition 4.1 provides sufficient conditions for almost H-calmness. The key condition is (c). Note that condition $\nabla_\xi G(z,\xi) \neq 0$ is easy to verify. The compactness of $\Xi(z,\eta)$ is satisfied in the following two obvious cases: 1. $G(z,\xi)$ is polynomial in ξ; 2. $G(z,\xi)$ is locally monotonic in ξ for each fixed z. A simple example is portfolio optimization problem with second order dominance constraints, where ξ is a continuous random variable, $G(z,\xi) := z^T\xi$ is a profit function, $Y(\xi) := \bar{z}^T\xi$ is benchmark profit function and $H(z,\xi) := -G(z,\xi)$. It is easy to see that all conditions of Proposition 4.1 are satisfied.

Theorem 4.1 (Uniform exponential convergence) *Let $\mathcal{Z}$ be a nonempty compact subset of Z_0, $F(z,\eta,\xi)$ be defined by (1.4) and $f(z,\eta) = \mathbb{E}[F(z,\eta,\xi)]$. Let $f_N(z,\eta)$ be defined as in (1.5). Suppose, in addition to conditions of Proposition 4.1, that: (a) $\nabla_z H(z,\xi)$ and $\nabla_z G(z,\xi)$ are locally Lipschitz continuous for every ξ with modulus $\kappa(\xi)$, where $\mathbb{E}[\kappa(\xi)] < \infty$, and (b) the support set of ξ is bounded. Let*

$$\vartheta^N(z,\mu) := \mathbb{D}\left(\nabla h_N(z) + \int_a^b \partial_z f_N(z,\eta)\mu(d\eta), \nabla\mathbb{E}[H(z,\xi)] + \int_a^b \partial_z f(z,\eta)\mu(d\eta)\right).$$

Then with probability approaching one exponentially fast with the increase of sample size N, $\sup_{(z,\mu)\in\mathcal{Z}\times\hat{\mathscr{C}}_+^*([a,b])} \vartheta^N(z,\mu)$ tends to 0, where $\hat{\mathscr{C}}_+^*([a,b])$ be a compact subset of $\mathscr{C}_+^*([a,b])$ and $\mu \in \hat{\mathscr{C}}_+^*([a,b])$.

Proof. It is well-known that for sets A, B, $\mathbb{D}(A,B) = \inf_{t>0}\{t : A \subset B + t\mathcal{B}\}$. Using this equivalence definition, one can easily derive that $\mathbb{D}(A,B) \leq \mathbb{D}(A,C) + \mathbb{D}(C,B)$ and $\mathbb{D}(A+C, B+C) \leq \mathbb{D}(A,B)$ for any set C. Consequently, we have

$$\vartheta^N(z,\mu) \leq \mathbb{D}\left(\nabla h_N(z), \nabla\mathbb{E}[H(z,\xi)]\right) + \mathbb{D}\left(\int_a^b \partial_z f_N(z,\eta)\mu(d\eta), \int_a^b \partial_z f(z,\eta)\mu(d\eta)\right).$$

Let

$$\vartheta_f^N(z,\mu) := \mathbb{D}\left(\int_a^b \partial_z f_N(z,\eta)\mu(d\eta), \int_a^b \partial_z f(z,\eta)\mu(d\eta)\right).$$

Since both $\int_a^b \partial_z f_N(z,\eta)\mu(d\eta)$ and $\int_a^b \partial_z f(z,\eta)\mu(d\eta)$ are convex and compact set-valued, we can use the Hörmander's formula (2.16) to reformulate ϑ_f^N as:

$$\vartheta_f^N(z,\mu) = \max_{\|u\|\leq 1}\left[\sigma\left(\int_a^b \partial_z f_N(z,\eta)\mu(d\eta), u\right) - \sigma\left(\int_a^b \partial_z f(z,\eta)\mu(d\eta), u\right)\right].$$

Moreover, it follows by [25, Proposition 3.4], operation $\sigma(\cdot,u)$ and integration are exchangeable which means

$$\begin{aligned}\vartheta_f^N(z,\mu) &= \max_{\|u\|\leq 1}\left[\int_a^b \sigma\left(\partial_z f_N(z,\eta), u\right)\mu(d\eta) - \int_a^b \sigma\left(\partial_z f(z,\eta), u\right)\mu(d\eta)\right] \\ &\leq \int_a^b \sup_{\|u\|\leq 1}\left[\frac{1}{N}\sum_{i=1}^N F_z^o(z,\eta,\xi^i;u) - \mathbb{E}[F_z^o(z,\eta,\xi;u)]\right]\mu(d\eta), \end{aligned} \tag{4.39}$$

where $F_z^o(z,\eta,\xi;u)$ is Clarke generalized directional derivative of $F(z,\eta,\xi)$ at point z in direction u for a given $\eta \in [a,b]$. Let $\Delta_N(z,u,\eta) := \frac{1}{N}\sum_{i=1}^N F_z^o(z,\xi^i,\eta;u) - \mathbb{E}[F_z^o(z,\xi,\eta;u)]$. Since

$\hat{\mathscr{C}}_+^*([a,b])$ is a compact set, there exists a positive number M such that $\sup_{\mu\in\hat{\mathscr{C}}_+^*([a,b])}\int_a^b \mu(d\eta) \leq M$. Consequently

$$\begin{aligned}
&\text{Prob}\left\{\sup_{\mu\in\hat{\mathscr{C}}_+^*([a,b])}\sup_{z\in Z}\int_a^b \sup_{\|u\|\leq 1}\Delta_N(z,u,\eta)\mu(d\eta)\geq\alpha\right\}\\
&\leq \text{Prob}\left\{\sup_{z\in\mathcal{Z},\|u\|\leq 1,\eta\in[a,b]}\Delta_N(z,u,\eta)\sup_{\mu\in\hat{\mathscr{C}}_+^*([a,b])}\int_a^b\mu(d\eta)\geq\alpha\right\}\\
&\leq \text{Prob}\left\{M\sup_{z\in\mathcal{Z},\|u\|\leq 1,\eta\in[a,b]}\Delta_N(z,u,\eta)\geq\alpha\right\}. \qquad (4.40)
\end{aligned}$$

From Proposition 4.1 we know that $F_z^o(\cdot,\eta,\xi;u)$ is almost H-calm with the same modulus $\kappa(\xi)$ as $-\nabla_z G(z,\xi)$ and order 1, and $\mathbb{E}[F_z^o(z,\eta,\xi;u)]$ is a continuous function for every $u\in\mathbb{R}^n$. Under condition (b), the moment generating function

$$M_t(z) := \mathbb{E}\left[e^{(F_z^o(z,\eta,\xi;u)-\mathbb{E}[F_z^o(z,\eta,\xi;u)])t}\right]$$

and

$$M_\kappa(t) := \mathbb{E}\left\{e^{[\kappa(\xi)-\mathbb{E}[\kappa(\xi)]]t}\right\}$$

are finite valued for t close to 0. By [33, Theorem 3.1], for any $\alpha>0$, there exists $c_1(\alpha)>0$ and $\beta_1(\alpha)>0$ (independent of N) such that

$$\text{Prob}\left\{\sup_{z\in\mathcal{Z},\|u\|\leq 1,\eta\in[a,b]}\left[\frac{1}{N}\sum_{i=1}^N F_z^o(z,\eta,\xi^i;u)-\mathbb{E}[F_z^o(z,\eta,\xi;u)]\right]\geq\frac{\alpha}{M}\right\}\leq c_1(\alpha)e^{-N\beta_1(\alpha)}.$$

Combining the inequality above with (4.39) and (4.40), we have

$$\begin{aligned}
&\text{Prob}\left\{\sup_{(z,\mu)\in\mathcal{Z}\times\hat{\mathscr{C}}_+^*([a,b])}\vartheta_f^N(z,\mu)\geq\alpha\right\}\\
&\leq \text{Prob}\left\{\sup_{z\in\mathcal{Z},\|u\|\leq 1,\eta\in[a,b]}\left[\frac{1}{N}\sum_{i=1}^N F_z^o(z,\xi^i,\eta;u)-\mathbb{E}[F_z^o(z,\xi,\eta;u)]\right]\geq\frac{\alpha}{M}\right\}\\
&\leq c_1(\alpha)e^{-\beta_1(\alpha)N}. \qquad (4.41)
\end{aligned}$$

Let

$$\vartheta_H^N(z) := ||\nabla\mathbb{E}[H(z,\xi)]-\nabla h_N(z)||.$$

By [32, Theorem 5.1] and conditions (a) and (b), we have, for any small positive number $\alpha>0$, there exist positive constants $c_2(\alpha)$ and $\beta_2(\alpha)$ (independent of N) such that

$$\begin{aligned}
\text{Prob}\left\{\sup_{z\in\mathcal{Z}}\vartheta_H^N(z)\geq\alpha\right\} &= \text{Prob}\left\{\sup_{z\in\mathcal{Z}}\|\nabla\mathbb{E}[H(z,\xi)]-\nabla h_N(z)\|\geq\alpha\right\}\\
&\leq c_2(\alpha)e^{-\beta_2(\alpha)N}. \qquad (4.42)
\end{aligned}$$

Combining (4.41) with (4.42), we can claim that for any small positive number $\alpha > 0$, there exists $c(\alpha) > 0$ and $\beta(\alpha) > 0$ (independent of N) such that

$$
\begin{aligned}
\text{Prob}\left\{\sup_{(z,\mu)\in\mathcal{Z}\times\hat{\mathscr{C}}_+^*([a,b])} \vartheta^N(z,\mu) \geq 2\alpha\right\} &\leq \text{Prob}\left\{\sup_{(z,\mu)\in\mathcal{Z}\times\hat{\mathscr{C}}_+^*([a,b])} \vartheta_f^N(z,\mu) \geq \alpha\right\} \\
&\quad +\text{Prob}\left\{\sup_{z\in\mathcal{Z}} \vartheta_H^N(z) \geq \alpha\right\} \\
&\leq c(\alpha)e^{-\beta(\alpha)N},
\end{aligned}
$$

where $c(\alpha) := c_1(\alpha) + c_2(\alpha)$ and $\beta(\alpha) := \min(\beta_1(\alpha), \beta_2(\alpha))$. The proof is complete. ■

Similar to the discussions in [32, 35], it is possible to estimate the constants $c(\alpha)$ and $\beta(\alpha)$, and hence a more precise estimate of sample size under some additional conditions on the moment functions. We leave this to interested readers as it involves complex technical details.

In the case when ξ is a discrete random variable, almost H-calmness is equivalent to H-calmness over $\mathcal{X}$; see [35] for the definition of H-calmness. Unfortunately, here $F_z^o(z,\eta,\xi;u)$ is not H-calm over $\mathcal{X}$ and Theorem 4.1 may not be applicable to the discrete case. However the uniform exponential convergence may be established in an entirely different way for a class of random function which is uniformly bounded over a considered compact set. We leave this to interested readers.

In what follows, we translate the uniform exponential convergence established in Theorem 4.1 into exponential convergence of KKT pairs of the SAA problem (1.5). We do so by exploiting a recent result on perturbation analysis of generalized equations [35, Lemma 4.2]. To this end, we reformulate the KKT conditions of both true problem and the SAA problem as a system of generalized equations.

Let

$$
\Gamma(z,\mu) := \begin{pmatrix} \nabla_z \mathbb{E}[H(z,\xi)] + \int_a^b \partial_z f(z,\eta)\mu(d\eta) \\ f(z,\cdot) \\ \int_a^b f(z,\eta)\mu(d\eta) \end{pmatrix} \tag{4.43}
$$

and

$$
\mathcal{G}(z,\mu) := \begin{pmatrix} \mathcal{N}_{Z_0}(z) \\ \mathscr{C}_+^*([a,b]) \\ 0 \end{pmatrix}. \tag{4.44}
$$

Then we can rewrite (2.9) as

$$
0 \in \Gamma(z,\mu) + \mathcal{G}(z,\mu). \tag{4.45}
$$

Likewise, we can rewrite the KKT conditions of the SAA problem (2.22) as follows:

$$
0 \in \hat{\Gamma}^N(z,\mu) + \mathcal{G}(z,\mu), \tag{4.46}
$$

where

$$
\Gamma^N(z,\mu) := \begin{pmatrix} \nabla_z h_N(z) + \int_a^b \partial_z f_N(z,\eta)\mu(d\eta) \\ f_N(z,\cdot) \\ \int_a^b f_N(z,\eta)\mu(d\eta) \end{pmatrix}. \tag{4.47}
$$

Theorem 4.2 *Assume the settings and conditions of Theorem 4.1. Under Assumption 3.1, for any $\alpha > 0$, there exist positive constants $C(\alpha)$ and $\beta(\alpha)$ independent of N such that*

$$\text{Prob}\{d(z_N, Z^*) \geq \alpha\} \leq C(\alpha)e^{-N\beta(\alpha)},$$

where Z^* denotes the set of weak Clarke stationary points characterized by (2.8).

Proof. The thrust of the proof is to apply Theorem 4.1 and [35, Lemma 4.2 (i)] (note that [35, Lemma 4.2 (i)] was presented in finite dimensional space but the conclusion holds in Banach space). To this end, we need to verify the upper semicontinuity of $\Gamma(z,\mu)$. This has been done in [20], here we include proof for completeness.

Observe that $\Gamma(z,\mu)$ consists of three parts: $\nabla_z \mathbb{E}[H(z,\xi)] + \int_a^b \partial_z f(z,\eta)\mu(d\eta)$, $f(z,\cdot)$ and $\int_a^b f(z,\eta)\mu(d\eta)$. It suffices to verify upper semicontinuity of each part.

Let us start with the first part. Since $H(\cdot,\xi)$ is continuously differentiable for every ξ and it is integrably bounded, then $\mathbb{E}[\nabla_z H(z,\xi)]$ is continuous. In what follows, we show upper semicontinuity of $\int_a^b \partial_z f(z,\eta)\mu(d\eta)$.

Let $(z',\mu'), (z,\mu) \in \mathcal{Z} \times \mathscr{C}_+([a,b])$ and (z,μ) be fixed. Then

$$\begin{aligned}\mathbb{D}\left(\int_a^b \partial_z f(z',\eta)\mu'(d\eta), \int_a^b \partial_z f(z,\eta)\mu(d\eta)\right) \;\leq\; & \mathbb{D}\left(\int_a^b \partial_z f(z',\eta)\mu'(d\eta), \int_a^b \partial_z f(z,\eta)\mu'(d\eta)\right) \\ & + \mathbb{D}\left(\int_a^b \partial_z f(z,\eta)\mu'(d\eta), \int_a^b \partial_z f(z,\eta)\mu(d\eta)\right).\end{aligned}$$

By Hörmander's formula (2.16) and [25, Proposition 3.4]

$$\begin{aligned}& \mathbb{D}\left(\int_a^b \partial_z f(z',\eta)\mu'(d\eta), \int_a^b \partial_z f(z,\eta)\mu'(d\eta)\right) \\ = & \sup_{||u||\leq 1} \left(\int_a^b [\sigma(\partial_z f(z',\eta),u) - \sigma(\partial_z f(z,\eta),u)]\mu'(d\eta)\right) \\ \leq & \sup_{||u||\leq 1, \eta\in[a,b]} [\sigma(\partial_z f(z',\eta),u) - \sigma(\partial_z f(z,\eta),u)]\mu'([a,b]) \\ = & \sup_{||u||\leq 1, \eta\in[a,b]} \left(\mathbb{E}[F_z^o(z',\eta,u)] - \mathbb{E}[F_z^o(z,\eta,u)]\right)\mu'([a,b])\end{aligned}$$

By Proposition 4.1 (i), $\mathbb{E}[F_z^o(z,\eta,u)]$ is a continuous function w.r.t. (z,η,u). Therefore

$$\sup_{||u||\leq 1, \eta\in[a,b]} [\sigma(\partial_z f(z',\eta),u) - \sigma(\partial_z f(z,\eta),u)] \to 0$$

as $z' \to z$. Moreover, $\partial_z f(z,\eta)$ is single valued (indeed we could have written it as $\nabla_z f(z,\eta)$and continuous w.r.t. (z,η), therefore $\partial_z f(z,\eta)$ is bounded by a constant over $\mathcal{Z} \times [a,b]$. Further, it is easy to verify that $\partial_z f(z,\eta)$ is uniformly continuous w.r.t. η. By [22, Lemma 5.2],

$$\mathbb{D}\left(\int_a^b \partial_z f(z,\eta)\mu'(d\eta), \int_a^b \partial_z f(z,\eta)\mu(d\eta)\right) \to 0$$

as $\mu' \to \mu$. The discussions above show that $\int_a^b \partial_z f(z,\eta)\mu(d\eta)$ is upper semicontinuous w.r.t. (z,μ). Finally, by the definition of $f(z,\cdot)$, we have

$$\begin{aligned}||f(z,\cdot) - f(z',\cdot)||_\infty \;:=\; & \sup_{\eta\in[a,b]} ||\mathbb{E}[(\eta - G(z,\xi))_+ - (\eta - G(z',\xi))_+]|| \\ \leq\; & \mathbb{E}[|G(z,\xi) - G(z',\xi)|] \\ \leq\; & \mathbb{E}[\kappa(\xi)]||z - z'||\end{aligned}$$

which implies continuity of $f(z, \cdot)$ w.r.t. z. ■

5 Concluding remarks

In this paper, we present a detailed convergence analysis of sample average approximation of stochastic programs with second order dominance constraints. SAA is relevant in a number of cases, e.g., the distribution of random variable is unknown or it is difficult to obtained a closed form of the expected value of random functions in the true problem, or the random variable satisfies a continuous distribution.

Our analysis essentially consists of two parts: almost sure convergence and exponential convergence. Our focus is on stationary points which include optimal solutions. The almost sure convergence results show asymptotic consistency of the statistical estimators of the stationary points. The exponential convergence demonstrates rate of convergence of the sample average approximation. The latter is often used to estimate the sample size in order for the SAA solution to satisfy a specified precision. In other words, the convergence results in this paper give rise to the justification of SAA and demonstrate the efficiency of the approximation method. From numerical perspective, SAA is just a discretization approach. Over the past few years, a number of powerful numerical methods such as cutting plane methods, level function methods have been proposed to solve stochastic programs with second order dominance constraints where the underlying random variable satisfies finite discrete distribution, see for instance [18, 15, 24, 21]. It is unclear whether these numerical methods can be directly applied to the continuous distribution case and this work effectively addresses the gap, that is, these methods can be applied to solve the SAA problem.

Our analysis complements the recent work by Hu, Homem-de-Mello and Mehrotra [19] which showed almost sure convergence of optimal solutions of the SAA problem. In the case when the problem is convex, stationary points coincide with optimal solutions, therefore under the Slater condition, our almost sure convergence result, namely Theorem 3.1, may be recovered by [19, Theorem 3.1]. Note that Hu, Homem-de-Mello and Mehrotra [19] derived exponential convergence of ϵ-feasible set. Our understanding is that under some appropriate conditions (e.g. metric regularity of the constraints), this implies exponential convergence of optimal solution of the SAA problem to the set of ϵ-optimal solutions of the true problem. The exponential convergence result in this paper, namely Theorem 4.1, is derived for stationary points including optimal solutions and we show that the cluster point of the SAA stationary points is *precisely* a stationary point of the true problem, not just a ϵ-stationary point. The set of ϵ-stationary points (optimal solutions) contains the set of stationary points (optimal solutions) but the former might be much larger than the latter in stationary case even when ϵ is small.

Acknowledgements. The authors are grateful to the three anonymous referees for their insightful comments which have significantly helped improve the quality of the paper. They are also thankful to Professor Henry Wolkowicz for organizing an effective review.

References

[1] J.-P. Aubin and H. Frankowska, *Set-Valued Analysis*, Birkhauser, Boston, 1990.

[2] R. J. Aumann, Integrals of set-valued functions, *J. Math. Anal. Appl.*, Vol. 12, pp. 1-12, 1965.

[3] K. B. Athreya and S. N. Lahiri, *Measure Theory and Probability Theory*, Springer texts in statistics, Springer, 2006.

[4] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer series in operations research, Springer-Verlag, 2000.

[5] J. M. Borwein, Stability and regular points of inequality systems, *J. Optim. Theory Appl.*, Vol. 48, pp. 9-52, 1986.

[6] J. M. Borwein and Q. J. Zhu, A survey of subdifferential calculus with applications, *Nonlinear Anal. Ser. A: Theory Methods*, Vol. 38, pp. 687-773, 1999.

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

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

[9] A. Conejo, M. Carrión and J. Morales, *Decision Making Under Uncertainty in Electricity Markets*, Internation Series in Operations Research & Management Science, Springer, New York, 2010.

[10] D. Dentcheva and A. Ruszczyński, Optimization with stochastic dominance constraints, *SIAM J. Optim.*, Vol. 14, pp. 548-566, 2003.

[11] D. Dentcheva and A. Ruszczyński, Optimality and duality theory for stochastic optimization problems with nonlinear dominance constraints, *Math. Program.*, Vol. 99, pp. 329-350, 2004.

[12] D. Dentcheva and A. Ruszczyński, Portfolio optimization with stochastic dominance constraints, *J. Banking Financ*, Vol. 30, pp. 433-451, 2006.

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

[14] D. Dentcheva and A. Ruszczyński, Risk-averse portfolio optimization via stochastic dominance constraints, in C. Lee, A. Lee and J. Lee, editors, *Handbook of Quantitative Finance and Risk Management*, Vol. 1, Springer, NewYork, 2010.

[15] C. Fábián, G. Mitra and D. Roman, Processing second-order stochastic dominance models using cutting-plane representations, *Math. Program.*, Vol. 130, pp. 33-57, 2011.

[16] M. Gugat, A parametric review on the Mangasarian-Fromovitz constraint qualification, *Math. Program.*, Vol. 85, pp. 643-653, 1999.

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

[18] T. Homem-de-Mello and S. Mehrotra, A cutting surface method for uncertain linear programs with polyhedral stochastic dominance constraints, *SIAM J. Optim.*, Vol. 20, pp. 1250-1273, 2009.

[19] J. Hu, T. Homem-de-Mello and S. Mehrotra, Sample average approximation for stochastic dominance constrained programs. To appear in *Math. Program.*, 2010.

[20] Y. Liu, W. Römisch and H. Xu, Quantitative stability analysis of stochastic generalized equations, manuscript, 2011.

[21] G. Rudolf and A. Ruszczyński, Optimization problems with second order stochastic dominance constraints: duality, compact formulations, and cut generation methods, *SIAM J. Optim.*, Vol. 19, pp. 1326-1343, 2008.

[22] Y. Liu, H. Xu and G.H. Lin, Stability analysis of two stage stochastic mathematical programs with complementarity constraints via NLP-Regularization, *SIAM J. Optim.*, Vol. 21, pp. 669-705, 2011.

[23] Y. Liu and H. Xu, Stability and sensitivity analysis of stochastic programs with second order dominance constraints, Stochastic Programming E-pirnt, 2010.

[24] R. Meskarian, H. Xu and J. Fliege, Numerical methods for stochastic programs with second order dominance constraints with applications to portfolio optimization, *European J. Oper. Res.*, Vol. 216, pp. 376-385, 2011.

[25] N.S. Papageorgiou, On the theory of banach space valued multifunctions. I: Integration and conditional expectation, *J. Multivariate Anal.*, Vol. 17, pp. 185-206, 1985.

[26] E. Polak, On the mathematical foundations of nondifferentiable optimization in engineering design, *SIAM Rev.*, Vol. 29, pp. 21-91, 1987.

[27] L. Qi, A. Shapiro and C. Ling, Differentiability and semismoothness properties of integral functions and their applications, *Math. Program.*, Vol. 102, pp. 223-248, 2005.

[28] S. M. Robinson, Analysis of sample-path optimization, *Math. Oper. Res.*, Vol. 21, pp. 513-528, 1996.

[29] A. Rusczyński and A. Shapiro, Stochastic Programming Models, in A. Rusczyński and A.Shapiro, editors, *Stochastic Programming*, Handbooks in OR & MS, Vol. 10, North-Holland Publishing Company, Amsterdam, 2003.

[30] A. Shapiro, Monte Carlo sampling methods, in A. Rusczyński and A.Shapiro, editors, *Stochastic Programming*, Handbooks in OR & MS, Vol. 10, North-Holland Publishing Company, Amsterdam, 2003.

[31] A. Shapiro and H. Xu, Uniform laws of large numbers for set-valued mappings and subdifferentials of random functions, *J. Math. Anal. Appl.*, Vol. 325, pp. 1390-1399, 2007.

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

[33] H. Sun and H. Xu, A note on uniform exponential convergence of sample average approximation of random functions, *J. Math. Anal. Appl.*, Vol. 385, pp. 698-708, 2012.

[34] P. Terán, On a uniform law of large numbers for random sets and subdifferentials of random functions, *Statist. Probab. Lett.*, Vol. 78, pp. 42-49, 2008.

[35] H. Xu, Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming, *J. Math. Anal. Appl.*, Vol. 368, pp. 692-710, 2010.

[36] H. Xu and D. Zhang, Monte Carlo methods for mean-risk optimization and portfolio selection, *Comput. Manag. Sci.*, Vol. 9, pp. 3-29, 2012.

[37] J. J. Ye, Constraint qualifications and necessary optimality conditions for optimization problems with variational inequality constraints, *SIAM J. Optim.*, Vol. 10, pp. 943-962, 2000.