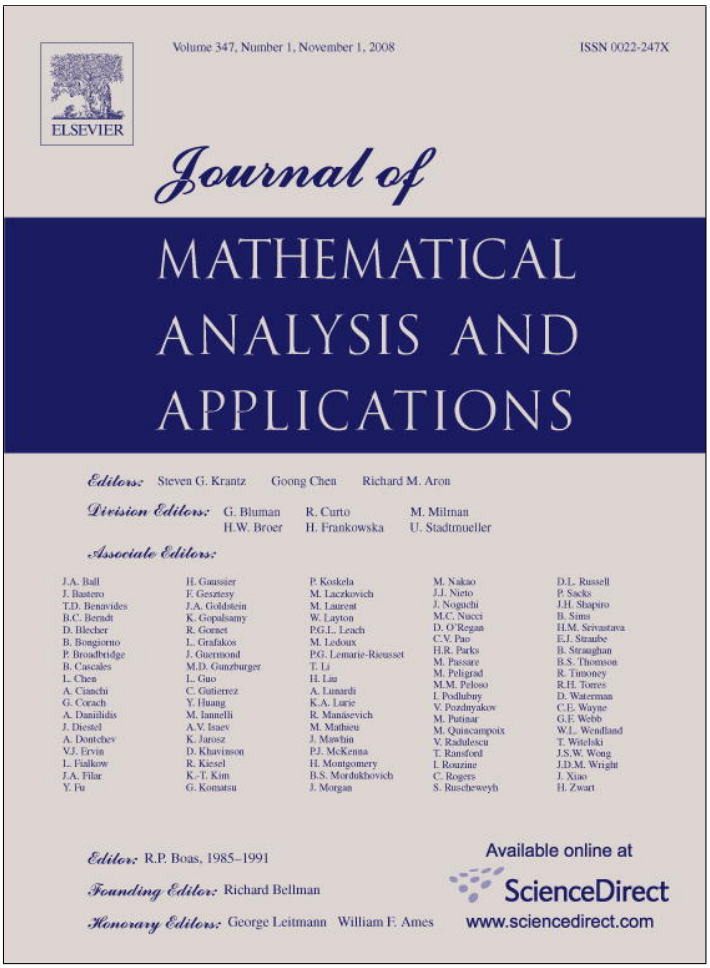

J. Math. Anal. Appl. 347 (2008) 333–343

Contents lists available at ScienceDirect

J. Math. Anal. Appl.

www.elsevier.com/locate/jmaa

Approximating stationary points of stochastic optimization problems in Banach space

Ramamurthy Balaji [a,1], Huifu Xu [b,*]

[a] Department of Mathematics and Statistics, University of Hyderabad, Hyderabad 46, India
[b] School of Mathematics, University of Southampton, Highfield Southampton, UK

ARTICLE INFO

Article history:
Received 15 February 2007
Available online 15 June 2008
Submitted by V. Pozdnyakov

Keywords:
Sample average approximation
Stationary point
Law of large numbers
Exponential convergence
Metric regularity

ABSTRACT

In this paper, we present a uniform strong law of large numbers for random set-valued mappings in separable Banach space and apply it to analyze the sample average approximation of Clarke stationary points of a nonsmooth one stage stochastic minimization problem in separable Banach space. Moreover, under Hausdorff continuity, we show that with probability approaching one exponentially fast with the increase of sample size, the sample average of a convex compact set-valued mapping converges to its expected value uniformly. The result is used to establish exponential convergence of stationary sequence under some metric regularity conditions.

1. Introduction

We consider the following stochastic minimization problem:

$$\min_{x\in\mathcal{X}} \mathbb{E}\big[f\big(x,\xi(\omega)\big)\big], \tag{1}$$

where $f : X \times \mathcal{R}^k \to \mathcal{R}$ is locally Lipschitz continuous, X is a separable Banach space, $\mathcal{X}$ is a subset of X and $\xi : \Omega \to \mathcal{R}^k$ is a random vector defined on a nonatomic probability space $(\Omega, \mathcal{F}, P)$ with support set $\Xi \subset \mathcal{R}^k$. To ease the notation, we will write $\xi(\omega)$ as ξ and this should be distinguished from ξ being a deterministic vector of Ξ in a context.

Over the past few decades, problem (1) has been extensively investigated for the case when X is a finite dimensional space. For details, see [20] and references therein. In this paper we consider the case when X is a separable Banach space. Our focus is on the sample average approximation of the stochastic minimization problem. Sample Average Approximation (SAA) is a popular method in stochastic programming. It is also known as Sample Path (SP) approximation [17]. For a comprehensive review of SAA, see recent work by Shapiro [22]. The basic idea of SAA is to generate an independently identically distributed (iid) sample of random variables and replace the expected value with its sample average. Let $\xi^1, \ldots, \xi^N$ be an iid sample of ξ. The sample average approximation of the stochastic minimization problem (1) is defined as follows:

$$\min_{x\in\mathcal{X}} \hat{f}_N(x) := \frac{1}{N}\sum_{i=1}^{N} f\big(x,\xi^i\big). \tag{2}$$

* Corresponding author.
E-mail address: h.xu@soton.ac.uk (H. Xu).

[1] The work of this author was supported by a United Kingdom Royal Society International Fellowship and it was carried out while he was working with the second author in the School of Mathematics, University of Southampton, Highfield Southampton, UK.

0022-247X/$ – see front matter
doi:10.1016/j.jmaa.2008.06.015

Most convergence analysis of SAA problems in the literature concerns the convergence of optimal solutions and optimal values [22], that is, if we solve an SAA problem and obtain an optimal solution, then we discuss the convergence of the optimal solution sequence as sample size N increases. Our interest here, however, is on the convergence of stationary points, that is, if we obtain only a stationary point of (2) which is not necessarily an optimal solution, then what is an accumulation point of the sequence of the SAA stationary points? The rational behind this is that when $f(x,\xi)$ is nonconvex in x, the sample average function $\hat{f}_N(x)$ is likely to be nonconvex. Consequently we may only obtain a stationary point rather than an optimal solution in solving (2).

Our analysis is carried out in two steps: first we investigate the convergence of Clarke stationary points of the SAA problem under the condition that $f(x,\xi)$ is locally Lipschitz continuous but it is not necessarily continuously differentiable (Section 4); then we discuss the rate of convergence under the condition that $f(x,\xi)$ is continuously differentiable (Section 5).

The main tool to be used in analysis is a uniform Strong Law of Large Numbers (SLLN) for random set-valued mappings in separable Banach space. This is because the stationary points we consider in this paper are characterized by the Clarke generalized gradient which is a set-valued mapping. In a finite dimensional space, Shapiro and Xu [24] obtained a uniform SLLN for random set-valued mappings and applied it to the convergence analysis of sample average approximation of Clarke stationary points. The uniform SLLN is a generalization of a uniform SLLN for scalar valued random functions (see Rubinstein and Shapiro [19]) and an SLLN for compact random sets by Artstein and Vitale [3]. Here we generalize the uniform SLLN of Shapiro and Xu [24] to a separable Banach space (Section 3). We achieve this by using a Strong Law of Large Numbers established by Artstein and Hansen [2] for random sets in Banach space.

Note that when X is a Banach space, the Clarke generalized gradient of a general locally Lipschitz continuous function is not necessarily a compact convex set. Instead, it is only weak* compact [7, Proposition 2.1.2]. Our focus here, however, is on the case when the Clarke generalized gradient of $f(x,\xi)$ is compact set-valued. On one hand, this gets around some difficulties resulting from weak compactness of random sets in convergence analysis; on the other hand, this class of functions covers a number of practically interesting functions such as composition of a smooth vector valued function from a Banach space to a finite dimensional space and a real valued nonsmooth function from the finite dimensional space.

The main contributions of this paper are as follows: we present a uniform SLLN for a random set-valued mapping in separable Banach space which extends the uniform SLLN of Shapiro and Xu [24] in finite dimensional spaces. Moreover, when the set valued mapping is Lipschitz continuous, we show that with probability approaching one exponentially fast with the increase of sample, the sample average of a convex compact set-valued mapping converges to its expected value uniformly. We then apply the results to analyze the sample average approximation of Clarke stationary points of the nonsmooth stochastic minimization problem (1). In particular, when the underlying function is smooth, we obtain an exponential convergence rate under some metric regularity conditions.

2. Preliminaries

Let X be a separable Banach space equipped with norm $\|\cdot\|$. Let $\mathcal{C}\subset X$ be a set and $y\in X$ be a point. We denote a pseudo distance from y to $\mathcal{C}$ by

$$d(y,\mathcal{C}) := \inf_{z\in\mathcal{C}} \|y-z\|.$$

For subsets $\mathcal{C}_1$ and $\mathcal{C}_2$ of X, we define

$$\mathbb{D}(\mathcal{C}_1,\mathcal{C}_2) := \sup_{x\in\mathcal{C}_1} d(x,\mathcal{C}_2)$$

which is known as excess of $\mathcal{C}_1$ over $\mathcal{C}_2$ [13] and the *Hausdorff* distance by

$$\mathbb{H}(\mathcal{C}_1,\mathcal{C}_2) := \max\big\{\mathbb{D}(\mathcal{C}_1,\mathcal{C}_2),\mathbb{D}(\mathcal{C}_2,\mathcal{C}_1)\big\}.$$

For subsets $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_3$ in X,

$$\mathbb{D}(\mathcal{C}_1,\mathcal{C}_3)\leqslant \mathbb{D}(\mathcal{C}_1,\mathcal{C}_2)+\mathbb{D}(\mathcal{C}_2,\mathcal{C}_3). \tag{3}$$

Moreover, for subsets $\mathcal{C}$, $\mathcal{D}$, $\mathcal{C}'$, $\mathcal{D}'$ in X,

$$\mathbb{D}(\mathcal{C}+\mathcal{D},\mathcal{C}'+\mathcal{D}')\leqslant \mathbb{D}(\mathcal{C},\mathcal{C}')+\mathbb{D}(\mathcal{D},\mathcal{D}') \tag{4}$$

where $\mathcal{C}+\mathcal{D}$ denote the Minkowski addition, that is,

$$\mathcal{C}+\mathcal{D}=\{c+d\colon\ c\in\mathcal{C},\ \ d\in\mathcal{D}\}.$$

Both (4) and (3) follow easily from

$$\mathbb{D}(\mathcal{C},\mathcal{D}) = \inf_{t\geqslant 0}\{t\colon\ \mathcal{C}\subset\mathcal{D}+t\mathcal{B}\}$$

where $\mathcal{B}$ denotes the unit ball in X. Note both inequalities hold for the Hausdorff distance and we will use them in later discussions. See elementary properties of $\mathbb{D}$ and $\mathbb{H}$ in [6, pp. 38 and 49].

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{A}: \Omega \to 2^X$ be a nonempty random compact set-valued mapping (note that we are not defining through a random vector ξ here). $\mathcal{A}$ is said to be *measurable* if for all closed subset $\mathcal{C} \subset X$, the set $\{\omega \in \Omega:\ \mathcal{A}(\omega) \in \mathcal{C}\} \in \mathcal{F}$. If $\mathcal{A}$ is measurable, then $\mathcal{A}$ is called a *random set*. The *expectation* of a random set $\mathcal{A}$, denoted by $\mathbb{E}[\mathcal{A}]$, is the set

$$\left\{\int_\Omega \zeta\, dP:\ \zeta \text{ is an integrable selection from } \mathcal{A}\right\}.$$

The integral $\int_\Omega \zeta\, dP$ is in the sense of Bochner (see Diestel and Uhl [8]). $\mathbb{E}[\mathcal{A}]$ is known as Aumann's integral [5] of set-valued mapping $\mathcal{A}$.

We call a random set $\mathcal{A}: \Omega \to 2^X$ *integrably bounded* if there exists $\phi: \Omega \to \mathcal{R}_+$ with $\mathbb{E}[\phi(\omega)] < \infty$ such that

$$\mathbb{H}\big(\mathcal{A}(\omega), \{0\}\big) \leqslant \phi(\omega)$$

for P-almost every $\omega \in \Omega$.

For a set-valued mapping $\Gamma: X \to 2^X$, we say Γ is *upper semicontinuous* at $x \in X$ if for every open set $\mathcal{C}$ containing $\Gamma(x)$, there exists an open set $\mathcal{D}$ containing x such that

$$x' \in \mathcal{D} \quad \Rightarrow \quad \Gamma(x') \subset \mathcal{C}.$$

It is easy to verify that if Γ is upper semicontinuous at x and the value of Γ is compact, then

$$\lim_{y \to x} \mathbb{D}\big(\Gamma(y), \Gamma(x)\big) = 0.$$

A few words about notation. We use $B_r(x)$ to denote a closed ball with center x and radius r, that is, $B_r(x) := \{x \in X:\ \|x - x'\| \leqslant r\}$. For a set-valued mapping $\Gamma: X \to 2^X$, we use $\Gamma^r(x)$ to denote the collection of $\Gamma(x')$ for $x' \in B_r(x)$, that is,

$$\Gamma^r(x) := \bigcup_{x' \in B_r(x)} \Gamma(x').$$

For a compact set $\mathcal{C} \in X$, we let

$$\|\mathcal{C}\| = \sup_{c \in \mathcal{C}} \|c\|.$$

3. A uniform SLLN for a random set-valued mapping

In this section, we establish a uniform SLLN for a random compact set-valued mapping in separable Banach space. For this purpose, we need the following result due to Artstein and Hansen [2] and Hess [13, Theorem 5.4].

Lemma 1. *Let $(\Omega, \mathcal{F}, P)$ be a probability space which is nonatomic and $\xi: \Omega \to \mathcal{R}^k$ be a random vector. Let X be a separable Banach space and $\mathcal{A}(\xi(\cdot)): \Omega \to 2^X$ be an integrably bounded random set. Assume that for all $\omega \in \Omega$, $\mathcal{A}(\xi(\omega))$ is nonempty, compact. If ξ^i, $i = 1, \ldots, N$, is an iid sample of ξ, then*

$$\mathcal{S}_N := \frac{1}{N} \sum_{i=1}^N \mathcal{A}\big(\xi^i\big) \to \mathbb{E}\big[\mathcal{A}(\xi)\big]$$

almost surely as N tends to infinity. Here the convergence is in the sense of Hausdorff metric in space 2^X.

In what follows we use Lemma 1 to establish a uniform SLLN for a random set-valued mapping. The result below is essentially a generalization of a uniform SLLN of Shapiro and Xu [24] to a Banach space.

Theorem 1. *Let X be a separable Banach space and $\mathcal{X} \subset X$ be a compact subset of X. Let $(\Omega, \mathcal{F}, P)$ be a probability space which is nonatomic and $\xi: \Omega \to \mathcal{R}^k$ be a random vector with support set Ξ. Let $\mathcal{A}(\cdot, \xi(\cdot)): X \times \Omega \to 2^X$ be a random nonempty compact set-valued mapping. Let $\xi^1, \ldots, \xi^N$ be an iid sample of ξ and*

$$\mathcal{S}_N(x) := \frac{1}{N} \sum_{i=1}^N \mathcal{A}\big(x, \xi^i\big).$$

Suppose that $\mathcal{A}(x, \xi)$ is integrably bounded by $\phi(\xi) > 0$, that is,

$$\big\|\mathcal{A}\big(x, \xi(\omega)\big)\big\| \leqslant \phi\big(\xi(\omega)\big),$$

for all $x \in \mathcal{X}$ and P-almost every $\omega \in \Omega$, and $\mathbb{E}[\phi(\xi)] < \infty$. Suppose also that $\mathcal{A}(\cdot, \xi(\omega))$ is upper semicontinuous on $\mathcal{X}$ for P-almost every $\omega \in \Omega$. Then

$$\sup_{x\in\mathcal{X}} \mathbb{D}\big(\mathcal{S}_N(x), \mathbb{E}\big[\mathcal{A}^r(x,\xi)\big]\big) \to 0 \tag{5}$$

almost surely as $N \to \infty$, where

$$\mathcal{A}^r(x,\xi) = \bigcup_{x'\in B_r(x)\cap\mathcal{X}} \mathcal{A}(x',\xi).$$

Proof. Since $\mathcal{A}(x)$ is integrably bounded, $\mathbb{E}[\mathcal{A}(x)]$ is well defined for all $x \in \mathcal{X}$. Let k be a positive integer and $V_k(x) := \{y \in \mathcal{X}\colon \|y - x\| \leqslant \frac{1}{k}\}$. By the triangle inequality, we have for any $x', x \in \mathcal{X}$,

$$\mathbb{D}\big(\mathcal{S}_N(x'), \mathcal{S}_N(x)\big) \leqslant \frac{1}{N}\sum_{i=1}^{N} \mathbb{D}\big(\mathcal{A}\big(x',\xi^i\big), \mathcal{A}\big(x,\xi^i\big)\big).$$

Consequently

$$\sup_{x'\in V_k(x)} \mathbb{D}\big(\mathcal{S}_N(x'), \mathcal{S}_N(x)\big) \leqslant \frac{1}{N}\sum_{i=1}^{N} \sup_{x'\in V_k(x)} \mathbb{D}\big(\mathcal{A}\big(x',\xi^i\big), \mathcal{A}\big(x,\xi^i\big)\big).$$

Since $\sup_{x'\in V_k(x)} \mathbb{D}\big(\mathcal{A}(x',\xi), \mathcal{A}(x,\xi)\big)$ is bounded by $2\phi(\xi)$ and by [4, Theorem 8.2.11] it is measurable, then by the strong law of large numbers

$$\frac{1}{N}\sum_{i=1}^{N} \sup_{x'\in V_k(x)} \mathbb{D}\big(\mathcal{A}\big(x',\xi^i\big), \mathcal{A}\big(x,\xi^i\big)\big) \to \mathbb{E}\Big[\sup_{x'\in V_k(x)} \mathbb{D}\big(\mathcal{A}(x',\xi), \mathcal{A}(x,\xi)\big)\Big]$$

almost surely as $N \to \infty$. By the Lebesgue dominated convergence theorem for Bochner integrals and the upper semicontinuity of $\mathcal{A}(\cdot,\xi)$, we conclude that for all $\epsilon > 0$, the right-hand side of the above equation is less than ϵ for all k sufficiently large. Thus there exists a positive number $\delta > 0$ such that for almost every $\omega \in \Omega$, there exists $\bar{N}(\omega)$ such that

$$\sup_{x'\in B_\delta(x)} \mathbb{D}\big(\mathcal{S}_N(x'), \mathcal{S}_N(x)\big) \leqslant \epsilon, \tag{6}$$

for all $N \geqslant \bar{N}(\omega)$. Let $r > 0$ be a positive constant and δ be such that $\delta \leqslant r/2$. By the compactness of $\mathcal{X}$, there exists a finite set of points $x_j \in \mathcal{X}$, $j = 1, \ldots, l$, with respect to neighborhoods $W_j := B_\delta(x_j)$, such that $\mathcal{X} \subset \bigcup_{i=1}^{l} W_i$. Using Lemma 1, we have that by setting $\bar{N}(\omega)$ larger if necessary,

$$\mathbb{D}\big(\mathcal{S}_N(x_j), \mathbb{E}\big[A(x_j,\xi)\big]\big) \leqslant \epsilon, \quad \text{for } j = 1, \ldots, l, \tag{7}$$

for all $N \geqslant \bar{N}(\omega)$. Let $x \in \mathcal{X}$. Then there exists some j such that $x \in W_j$. Consequently we have

$$\mathbb{D}\big(\mathcal{S}_N(x), \mathbb{E}\big[\mathcal{A}^r(x,\xi)\big]\big) \leqslant \mathbb{D}\big(\mathcal{S}_N(x), \mathcal{S}_N(x_j)\big) + \mathbb{D}\big(\mathcal{S}_N(x_j), \mathbb{E}\big[\mathcal{A}(x_j,\xi)\big]\big) + \mathbb{D}\big(\mathbb{E}\big[\mathcal{A}(x_j,\xi)\big], \mathbb{E}\big[\mathcal{A}^r(x,\xi)\big]\big).$$

The last term in the above equation is zero as $\|x - x_j\| \leqslant r$. Thus for all $N \geqslant \bar{N}(\omega)$, we have from (6) and (7) that

$$\sup_{x\in\mathcal{X}} \mathbb{D}\big(\mathcal{S}_N(x), \mathbb{E}\big[\mathcal{A}^r(x,\xi)\big]\big) \leqslant 2\epsilon.$$

This completes the proof. □

Note that $r > 0$ in (5). In the case when X is a finite dimensional convex combination space, Terán [25] proved recently that

$$\sup_{x\in\mathcal{X}} \mathbb{H}\big(\mathcal{S}_N(x), \mathbb{E}\big[A(x,\xi)\big]\big) \to 0. \tag{8}$$

It is an open question whether (8) holds when X is an infinite dimensional space. However when $\mathcal{A}(\cdot,\xi)$ is continuous in Hausdorff distance with respect to x, we can strengthen the result of Theorem 1 as follows.

Corollary 1. *Assume the conditions of Theorem* 1. *If, in addition, $\mathcal{A}(\cdot, \xi(\omega))$ is continuous (in the Hausdorff metric) on $\mathcal{X}$ for P-almost every $\omega \in \Omega$, then the following holds*:

$$\sup_{x\in\mathcal{X}} \mathbb{H}\big(\mathcal{S}_N(x), \mathbb{E}\big[\mathcal{A}(x,\xi)\big]\big) \to 0 \tag{9}$$

almost surely as $N \to \infty$.

The result is straightforward from the proof of Theorem 1 by replacing $\mathbb{D}$ with $\mathbb{H}$.

Remark 1. In the case when $\mathcal{A}(x,\xi(\omega))$ is single valued, upper semicontinuity implies continuity. Consequently (9) holds.

In what follows, we investigate the rate of convergence when $\mathcal{A}(\cdot,\xi)$ is Lipschitz continuous.

Definition 1. Let $\mathcal{X}$ be a subset of X and $\xi:\Omega\to\mathcal{R}^k$ be a random vector with support set Ξ. A random set-valued mapping $\mathcal{A}(\cdot,\xi(\cdot)):\mathcal{X}\times\Omega\to 2^X$ is said to be Lipschitz continuous with respect to x, if there exists $\kappa(\xi)>0$ such that

$$\mathbb{H}\big(\mathcal{A}(x',\xi(\omega)),\mathcal{A}(x'',\xi(\omega))\big)\leqslant\kappa\big(\xi(\omega)\big)\|x'-x''\| \tag{10}$$

for all $x',x''\in\mathcal{X}$ and P-almost every $\omega\in\Omega$.

Let $\phi(\xi)$ be defined as in Theorem 1 and $\kappa(\xi)$ be defined by (10), let

$$\tilde{\kappa}(\xi)=\max\big(\phi(\xi)+\mathbb{E}\big[\phi(\xi)\big],\kappa(\xi)\|\mathcal{X}\|,\kappa(\xi)\big). \tag{11}$$

Theorem 2. *Assume conditions of Theorem* 1. *Assume further that*:

(a) $\mathcal{A}(\cdot,\xi(\cdot)):\mathcal{X}\times\Omega\to 2^X$ *is convex set-valued and* $\mathcal{X}$ *is a compact subset of* X;
(b) $\mathcal{A}(x,\xi(\omega))$ *is Lipschitz continuous with respect to* x *for* P*-almost every* $\omega\in\Omega$, *that is,* (10) *holds*;
(c) $\mathbb{E}[\tilde{\kappa}(\xi)]<\infty$ *and the moment generating function* $\mathbb{E}[e^{\tilde{\kappa}(\xi)t}]$ *of* $\tilde{\kappa}(\xi)$ *is finite valued for* t *close to zero, where* $\tilde{\kappa}(\xi)$ *is defined by* (11).

Then for every $\epsilon>0$, *there exist positive constants* $C(\epsilon)>0$, $\beta(\epsilon)>0$ *independent of* N *such that*

$$\mathrm{Prob}\Big(\sup_{x\in\mathcal{X}}\mathbb{H}\big(S_N(x),\mathbb{E}\big[\mathcal{A}(x,\xi)\big]\big)\geqslant\epsilon\Big)\leqslant C(\epsilon)e^{-\beta(\epsilon)N}, \tag{12}$$

for N *sufficiently large.*

Proof. Let $\sigma(u,\mathcal{C})$ denote the support function of a set $\mathcal{C}\subset X$, that is, for $u\in X^*$

$$\sigma(u,\mathcal{C}):=\sup_{c\in\mathcal{C}}\langle u,c\rangle$$

where $\langle u,c\rangle$ denotes the duality pairing and X^* for the dual space of X.

Let $\mathcal{B}^*$ denote the unit ball in X^*. For any two compact sets $\mathcal{C}_1,\mathcal{C}_2\subset X$, we have by the Hörmander formula (see for instance [1,6,13]) that

$$\mathbb{H}(\mathcal{C}_1,\mathcal{C}_2)=\sup_{u\in\mathcal{B}^*}\big|\sigma(u,\mathcal{C}_1)-\sigma(u,\mathcal{C}_2)\big|.$$

Using this relationship, it suffices to prove that

$$\mathrm{Prob}\Big(\sup_{u\in\mathcal{B}^*,\,x\in\mathcal{X}}\big|\sigma\big(u,S_N(x)\big)-\sigma\big(u,\mathbb{E}\big[\mathcal{A}(x,\xi)\big]\big)\big|\geqslant\epsilon\Big)\leqslant C(\epsilon)e^{-\beta(\epsilon)N}. \tag{13}$$

Since $\mathcal{A}(x,\xi)$ is convex set-valued for all $\xi\in\Xi$ and $\sigma(p,\mathcal{C})$ is homogeneous and additive in $\mathcal{C}$, then

$$\sigma\big(u,S_N(x)\big)=\frac{1}{N}\sum_{i=1}^{N}\sigma\big(u,\mathcal{A}\big(x,\xi^i\big)\big).$$

Moreover, by [16, Proposition 3.4],

$$\sigma\big(u,\mathbb{E}\big[\mathcal{A}(x,\xi)\big]\big)=\mathbb{E}\big[\sigma\big(u,\mathcal{A}(x,\xi)\big)\big],$$
$$\mathrm{Prob}\left(\sup_{u\in\mathcal{B}^*,\,x\in\mathcal{X}}\left|\frac{1}{N}\sum_{i=1}^{N}\sigma\big(u,\mathcal{A}\big(x,\xi^i\big)\big)-\mathbb{E}\big[\sigma\big(u,\mathcal{A}(x,\xi)\big)\big]\right|\geqslant\epsilon\right)\leqslant C(\epsilon)e^{-\beta(\epsilon)N}. \tag{14}$$

We use [23, Theorem 5.1] to prove (14) and hence (13) (note that the theorem is established in finite dimensional space but it also holds in a separable Banach space). In what follows, we verify the conditions of [23, Theorem 5.1].

First, since $\|\mathcal{A}(x,\xi)\|\leqslant\phi(\xi)$,

$$\big|\sigma\big(u,\mathcal{A}(x,\xi)\big)-\mathbb{E}\big[\sigma\big(u,\mathcal{A}(x,\xi)\big)\big]\big|\leqslant\big\|\mathcal{A}(x,\xi)\big\|+\mathbb{E}\big[\big\|\mathcal{A}(x,\xi)\big\|\big]\leqslant\phi(\xi)+\mathbb{E}\big[\phi(\xi)\big]$$

for all $u\in\mathcal{B}^*$. This shows that the random variable $\sigma(u,\mathcal{A}(x,\xi))-\mathbb{E}[\sigma(u,\mathcal{A}(x,\xi))]$ is bounded by $\tilde{\kappa}(\xi)$. Since by assumption the moment generating function $\mathbb{E}[e^{\tilde{\kappa}(\xi)t}]$ of $\tilde{\kappa}(\xi)$ is finite valued for t close to zero, then the moment generating function $\mathbb{E}[e^{t(\sigma(u,\mathcal{A}(x,\xi))-\mathbb{E}[\sigma(u,\mathcal{A}(x,\xi))])}]$ is finite valued for t close to zero.

Second, we show that $\sigma(u, \mathcal{A}(x,\xi))$ is Lipschitz continuous with respect to (u,x) and

$$\left|\sigma\left(u', \mathcal{A}(x',\xi)\right) - \sigma\left(u'', \mathcal{A}(x'',\xi)\right)\right| \leqslant \tilde{\kappa}(\xi)\left(\|u'-u''\|_* + \|x'-x''\|\right), \tag{15}$$

for all $\xi \in \Xi$, $x', x'' \in \mathcal{X}$ and $u', u'' \in \mathcal{B}^*$, where $\|\cdot\|_*$ denotes the norm of X^*. Observe that

$$\begin{aligned}
\sigma\left(u', \mathcal{A}(x',\xi)\right) - \sigma\left(u'', \mathcal{A}(x'',\xi)\right) &\geqslant \sup_{a\in\mathcal{A}(x',\xi)} \langle u', a\rangle - \sup_{a\in\mathcal{A}(x',\xi)+\kappa(\xi)\|x'-x''\|\mathcal{B}} \langle u'', a\rangle \\
&\geqslant \sup_{a\in\mathcal{A}(x',\xi)} \langle u', a\rangle - \sup_{a\in\mathcal{A}(x',\xi)} \langle u'', a\rangle - \kappa(\xi)\|x'-x''\| \\
&\geqslant -\sup_{a\in\mathcal{A}(x',\xi)} \langle u''-u', a\rangle - \kappa(\xi)\|x'-x''\| \\
&\geqslant -\tilde{\kappa}(\xi)\left(\|u''-u'\|_* + \|x'-x''\|\right).
\end{aligned}$$

Swapping position of x', u' with x'', u'', we obtain (15). Since $\tilde{\kappa}(\xi) > 0$ is integrable and by assumption the moment function of $\tilde{\kappa}(\xi)$ is finite valued for all t close to zero, all three conditions in [23, Theorem 5.1] are verified, by the theorem there exist positive constants $C(\epsilon)$ and $\beta(\epsilon)$ such that (14) holds. The conclusion follows. □

4. Convergence of stationary points

In this section, we apply the uniform SLLN, Theorem 1, to analyze the convergence of Clarke stationary points of SAA problem (2). We need to make the following assumptions on f.

Assumption 1. Let $f(x,\xi)$ be defined as in (1). The following hold.

(a) $f(x,\xi)$ is locally Lipschtz in x.
(b) The function $\mathbb{E}[f(x,\xi)]$ is finite valued.
(c) There exists $\kappa(\xi) > 0$ with $\mathbb{E}[\kappa(\xi)] < \infty$ such that

$$\left|f\left(x,\xi(\omega)\right) - f\left(y,\xi(\omega)\right)\right| \leqslant \kappa\left(\xi(\omega)\right)\|x-y\|, \quad \forall x, y \in \mathcal{X} \tag{16}$$

and P-almost every $\omega \in \Omega$.

Obviously Assumption 1(c) implies that $\mathbb{E}[f(x,\xi)]$ is a Lipschitz continuous function. In what follows, we need to consider the Clarke generalized gradient of both $\mathbb{E}[f(x,\xi)]$ and $f(x,\xi)$ with respect to x.

Recall that for a locally Lipschitz function $h(x)$ defined on separable Banach space X, the *Clarke directional derivative* [7] at a point $x \in X$ in the direction d, denoted by $h^\circ_x(x;d)$, is defined as follows:

$$h^\circ_x(x;d) = \limsup_{y\to x,\, t\downarrow 0} \frac{h(y+td)-h(y)}{t},$$

where $y \in X$ and t is a positive scalar. h is said to be *regular* at x, if for all d, the usual one-sided directional derivative $h'(x;d)$ exists and $h^\circ(x;d) = h'(x;d)$. See [7, Definition 2.3.4]. The *Clarke generalized gradient* [7] of $h(x)$ at x, denoted by $\partial h(x)$, is a subset of the dual space X^* of continuous linear functionals on X defined as

$$\partial h(x) := \left\{\eta \in X^*:\ h^\circ(x;d) \geqslant \langle \eta, d\rangle\right\},$$

see [7, p. 27]. Throughout this paper, we denote by $\partial_x f(x,\xi)$ the Clarke generalized gradient of function $f(x,\xi)$ with respect to x. It follows from Assumption 1(c) that $\partial_x f(x,\xi)$ is well defined and $\|\partial_x f(x,\xi)\|$ is bounded by $\kappa(\xi)$.

We now consider the SAA problem (2). Let $\hat{f}_N(x)$ be defined as in (2), that is,

$$\hat{f}_N(x) = \frac{1}{N}\sum_{i=1}^{N} f\left(x,\xi^i\right).$$

We need to consider the Clarke generalized gradient of $\hat{f}_N(x)$. For this purpose, we define the following two sets both of which may serve as an upper bound of $\partial \hat{f}_N(x)$:

$$\mathcal{S}_N(x) := \frac{1}{N}\sum_{i=1}^{N} \partial_x f\left(x,\xi^i\right)$$

and

$$\mathcal{S}^r_N(x) := \frac{1}{N}\sum_{i=1}^{N} \partial^r_x f\left(x,\xi^i\right),$$

where $r > 0$ is a constant and

$$\partial_x^r f(x,\xi) := \bigcup_{x' \in B_r(x) \cap \mathcal{X}} \partial_x f(x',\xi).$$

Proposition 1. *The following hold.*

(i) $\partial \hat{f}_N(x) \subset \mathcal{S}_N(x)$;
(ii) $\partial \mathbb{E}[f(x,\xi)] \subset \mathbb{E}[\partial_x f(x,\xi)]$.

Proof. The first inclusion follows from [7] and second from [15, Lemma 6.18]. Notice that $\mathbb{E}[\partial_x f(x,\xi)]$ is well defined. To see this, we note that because f is measurable, by [4, Lemma 8.2.12], we can easily see that the Clarke generalized derivative $f^\circ(x,\xi;d)$ is also measurable. Since $f^\circ(x,\xi;d)$ is the support function of $\partial_x f(x,\xi)$, we know from [4, Theorem 8.2.14] that $\partial_x f(x,\xi)$ is also measurable. The well-definedness follows from the measurability and the integrably boundedness. □

Let $x \in \mathcal{X}$. A vector v in X is tangent to $\mathcal{X}$ at x if the Clarke generalized derivative $d^\circ_{\mathcal{X}}(x;v)$ of distance function $d(x,\mathcal{X})$ is zero. Let $T_{\mathcal{X}}(x)$ denote the set of all tangents and $\mathcal{N}_{\mathcal{X}}(x)$ denote the *normal cone* to $\mathcal{X}$ at x. In the case when $\mathcal{X}$ is convex,

$$\mathcal{N}_{\mathcal{X}}(x) := \big\{\eta \in X^*\colon\ \langle \eta, v\rangle \leqslant 0,\ \forall v \in T_{\mathcal{X}}(x)\big\},$$

where X^* is the dual space of X. A point $x^* \in \mathcal{X}$ is said to be a *Clarke stationary point* of (1) if

$$0 \in \partial \mathbb{E}\big[f(x^*,\xi)\big] + \mathcal{N}_{\mathcal{X}}(x^*).$$

A point $x^* \in \mathcal{X}$ is said to be a *weak Clarke stationary point* of (1) if

$$0 \in \mathbb{E}\big[\partial_x f(x^*,\xi)\big] + \mathcal{N}_{\mathcal{X}}(x^*).$$

Because $\partial \mathbb{E}[f(x,\xi)] \subset \mathbb{E}[\partial_x f(x,\xi)]$, it is obvious that a Clarke stationary point is a weak Clarke stationary point but not vice versa unless $f(x,\xi)$ is Clarke regular at x^* for all ξ.

A point $x_N \in \mathcal{X}$ is said to be a *Clarke stationary point* of the SAA problem (2) if

$$0 \in \partial_x \hat{f}_N(x_N) + \mathcal{N}_{\mathcal{X}}(x_N). \tag{17}$$

In what follows, we use Theorem 1 to analyze the convergence of a sequence of Clarke stationary points $\{x_N\}$ as $N \to \infty$.

Theorem 3. *Let $\mathcal{X} \subset X$ be a compact set and X^* be a separable Banach space. Let $\{x_N\} \subset X$ be a sequence of Clarke stationary points which satisfies* (17)*. Assume*: *Assumption* 1 *holds, $\partial_x f(x,\xi)$ is compact valued for every $x \in \mathcal{X}$ and $\xi \in \Xi$ and upper semicontinuous with respect to x for every ξ. Then every accumulation point of $\{x_N\}$ is a weak stationary point of* (1) *almost surely.*

Proof. Let x^* be an accumulation point of $\{x_N\}$, and $\epsilon, r > 0$ be small numbers. By triangle inequality,

$$\mathbb{D}\big(\mathcal{S}_N(x_N), \mathbb{E}\big[\partial_x^{2r} f(x^*,\xi)\big] + \epsilon\mathcal{B}^*\big) \leqslant \mathbb{D}\big(\mathcal{S}_N(x_N), \mathcal{S}_N^r(x^*)\big) + \mathbb{D}\big(\mathcal{S}_N^r(x^*), \mathbb{E}\big[\partial_x^{2r} f(x^*,\xi)\big] + \epsilon\mathcal{B}^*\big) \tag{18}$$

for all x_N, where $\mathcal{B}^*$ denotes the unit ball in X^*. By considering a subsequence if necessary, we assume for the simplicity of notation that $\{x_N\} \to x^*$ almost surely as $N \to \infty$. Let $\bar{N}(\omega)$ be sufficiently large such that for all $N \geqslant \bar{N}(\omega)$, $x_N \in B_r(x^*)$. Then $\mathcal{S}_N(x_N) \subset \mathcal{S}_N^r(x^*)$ and hence

$$\mathbb{D}\big(\mathcal{S}_N(x_N), \mathcal{S}_N^r(x^*)\big) = 0.$$

In what follows, we use Theorem 1 to show that the second term of (18) tends to zero almost surely as $N \to \infty$. To this end, we need to verify the conditions of Theorem 1 for the Clarke generalized gradient mapping $\partial_x f(\cdot,\xi(\cdot)) : X \times \Omega \to 2^{X^*}$. By assumption, $\partial_x f(\cdot,\xi(\omega))$ is upper semicontinuous for P-almost every ω and it is convex compact set-valued. Moreover, by Assumption 1, $\|\partial_x f(x,\xi(\omega))\|_* \leqslant \kappa(\xi)$, where $\|c\|_*$ denotes the norm of $c \in X^*$ and for a set $\mathcal{C} \in X^*$,

$$\|\mathcal{C}\|_* = \sup\big\{\|c\|_*\colon\ c \in \mathcal{C}\big\}. \tag{19}$$

Therefore all conditions of Theorem 1 are satisfied by $\mathcal{A} := \partial_x f$. We apply the theorem. Let $\epsilon > 0$ be given and $\bar{N}(\omega)$ be sufficiently large such that for $N > \bar{N}(\omega)$

$$\mathcal{S}_N(x) \subset \mathbb{E}\big[\partial_x^r f(x,\xi)\big] + \epsilon\mathcal{B}^*$$

for all $x \in \mathcal{X}$. Then

$$\mathcal{S}_N^r(x^*) = \bigcup_{x' \in B_r(x^*)} \mathcal{S}_N(x') \subset \bigcup_{x' \in B_r(x^*)} \big\{\mathbb{E}\big[\partial_x^r f(x',\xi)\big] + \epsilon\mathcal{B}^*\big\} \subset \mathbb{E}\big[\partial_x^{2r} f(x^*,\xi)\big] + \epsilon\mathcal{B}^*.$$

Therefore the second term of (18) must converge to zero w.p. 1. Since ϵ can be any positive number, the discussion above shows that for any given $\epsilon > 0$, we can choose $\bar{N}(\omega)$ sufficiently large such that for $N > \bar{N}(\omega)$, we have

$$\mathcal{S}_N(x_N) \subset \mathbb{E}\big[\partial_x^{2r} f(x^*, \xi)\big] + \epsilon\mathcal{B}^*.$$

This implies that

$$\limsup_{N\to\infty} \mathcal{S}_N(x_N) \subset \mathbb{E}\big[\partial_x^{2r} f(x^*, \xi)\big] + 2\epsilon\mathcal{B}^*,$$

almost surely, where "lim sup" denotes the upper limit of a set-valued mapping (see [4]). On the other hand, from the definition of the normal cone, it is easy to show that

$$\mathcal{N}_{\mathcal{X}}(x_N) \subset \mathcal{N}_{\mathcal{X}}(x^*)$$

for large N. Taking the limit on both sides of (17) and using Proposition 1(i), we have

$$\begin{aligned} 0 &\in \limsup_{N\to\infty} \big\{\partial_x \hat{f}_N(x_N) + \mathcal{N}_{\mathcal{X}}(x_N)\big\} \\ &\subset \limsup_{N\to\infty} \big\{\mathcal{S}_N(x_N) + \mathcal{N}_{\mathcal{X}}(x_N)\big\} \\ &\subset \mathbb{E}\big[\partial_x^{2r} f(x^*, \xi)\big] + \mathcal{N}_{\mathcal{X}}(x^*) + 2\epsilon\mathcal{B}^* \end{aligned}$$

almost surely. Next we use the dominated convergence theorem to show that for any monotonically decreasing sequence of positive numbers $\{r_n\} \to 0$,

$$\lim_{r_n\to 0} \mathbb{E}\big[\partial^{2r_n} f(x^*, \xi)\big] = \mathbb{E}\Big[\lim_{r_n\to 0} \partial^{2r_n} f(x^*, \xi)\Big] = \mathbb{E}\big[\partial f(x^*, \xi)\big]. \tag{20}$$

To this end, we make the following observations. First, by assumption, $\partial_x^{2r_n} f(x^*, \xi)$ is integrably bounded by $\kappa(\xi)$ and $\|\partial^{2r_n} f(x^*), \xi)\|$ is decreasing on r_n, therefore $\|\partial^{2r_n} f(x^*), \xi)\|$ is uniformly integrable. Second, $\partial_x^{2r_n} f(x^*, \xi)$ is closed. To see this, let η^i be a sequence of X^* such that $\eta^i \in \partial_x^{2r_n} f(x^*)$ and $\eta^i \to \bar{\eta}$. By definition, there exists $x^i \in X$ such that $x^i \in x^* + 2r_n\mathcal{B}$ and $\eta^i \in \partial f(x^i, \xi)$. Suppose without loss of generality that $x^i \to \bar{x}$. Then by [7, Proposition 2.1.5(b)], $\bar{\eta} \in \partial f(x^*, \xi) \subset \partial_x^{2r_n} f(x^*, \xi)$. Third, for every ξ, it follows from [7, Proposition 2.1.5(b)] that

$$\lim_{r_n\to 0} \partial_x^{2r_n} f(x^*, \xi) = \bigcap_{r_n>0} \bigcup_{x\in x^*+2r_n\mathcal{B}} \partial f(x, \xi) = \partial f(x^*, \xi).$$

By [11, Theorem 2.5] (or [11, Theorem 2.8] and the following remark, or [14, Theorem 1.43(iii)]), (20) holds. Consequently

$$\begin{aligned} 0 &\in \lim_{r\to 0} \mathbb{E}\big[\partial^{2r} f(x^*, \xi)\big] + \mathcal{N}_{\mathcal{X}}(x^*) + 2\epsilon\mathcal{B}^* \\ &= \lim_{r_n\to 0} \mathbb{E}\big[\partial^{2r_n} f(x^*, \xi)\big] + \mathcal{N}_{\mathcal{X}}(x^*) + 2\epsilon\mathcal{B}^* \\ &= \mathbb{E}\Big[\lim_{r_n\to 0} \partial^{2r_n} f(x^*, \xi)\Big] + \mathcal{N}_{\mathcal{X}}(x^*) + 2\epsilon\mathcal{B}^* \\ &= \mathbb{E}\big[\partial f(x^*, \xi)\big] + \mathcal{N}_{\mathcal{X}}(x^*) + 2\epsilon\mathcal{B}^*. \end{aligned}$$

Since $\epsilon > 0$ can be arbitrary, we conclude from the above equation that x^* is a weak Clarke stationary point of (1) almost surely. □

In some practical instances, $f(x, \xi)$ may take specific forms. The following corollary addresses the case when f is a composition of a smooth vector valued function and a nonsmooth function.

Corollary 2. *Let $h : X \times \Omega \to R^n$ be a smooth function and $g : R^n \to R$ be a nonsmooth function which is locally Lipschitz continuous. Then the sequence of stationary points $\{x_k\}$ of (2) for the random function $f := g \circ h$ converges to the stationary point of the true problem.*

The result is immediate from Theorem 3 in that by [7, Theorem 2.3.10], we have

$$\partial_x f(x, \xi) \subset \partial g\big(h(x, \xi)\big) D_x h(x, \xi),$$

where D_x denote the Fréchet derivative, and the set at the right-hand side of the above equation is a compact set.

To conclude this section, we note that it is possible to consider generalized gradients other than Clarke's such as Michel–Penot subdifferential and Mordukhovich subdifferential in the analysis presented in this section so long as the generalized gradient mapping satisfies the properties required by Theorem 1.

5. Exponential convergence

The convergence result established in Theorem 3 does not provide us with any information about the rate of convergence. In this section, we investigate this issue. To this end, we assume throughout this section that $f(x,\xi)$ is continuously differentiable with respect to x. Consequently the first order optimality conditions of the true and SAA problems can be written respectively as

$$0 \in \mathbb{E}\big[\{D_x f(x,\xi)\}\big] + \mathcal{N}_{\mathcal{X}}(x) \tag{21}$$

and

$$0 \in \{D_x \hat{f}_N(x)\} + \mathcal{N}_{\mathcal{X}}(x), \tag{22}$$

where D_x denotes the Fréchet derivative and

$$\{D_x \hat{f}_N(x)\} = \frac{1}{N}\sum_{i=1}^{N}\{D_x f(x,\xi^i)\}.$$

Note that both $\{D_x f(x,\xi)\}$ and $\{D_x \hat{f}_N(x)\}$ are single valued. We write them this way rather than $D_x f(x,\xi)$ and $D_x \hat{f}_N(x)$ to indicate that they are specific set valued mappings so that we can apply Theorem 2 readily in the later discussion (in the proof of Theorem 5). In what follows, we analyze the convergence rate of the sequence of the stationary point defined by (22) as sample size $N \to \infty$. We need the theory of metric regularity.

Let $\Gamma : \mathcal{X} \to 2^X$ be a set valued mapping. Γ is said to be *closed* at x if for $x_k \subset \mathcal{X}$, $x_k \to x$, $y_k \in \Gamma(x_k)$ and $y_k \to \bar{y}$ implies $\bar{y} \in \Gamma(\bar{y})$. For $\bar{x} \in \mathcal{X}$ and $\bar{y} \in \Gamma(\bar{x})$, a closed set-valued mapping Γ is said to be *metrically regular* at $\bar{x}$ for $\bar{y}$ if there exists a constant $\alpha > 0$ such that

$$d\big(x, \Gamma^{-1}(y)\big) \leqslant \alpha d\big(y, \Gamma(x)\big) \quad \text{for all } (x,y) \text{ close to } (\bar{x},\bar{y}).$$

Here the inverse mapping Γ^{-1} is defined as $\Gamma^{-1}(y) = \{x \in \mathcal{X}:\ y \in \Gamma(x)\}$ and the minimal α which makes the above inequality hold is called *regularity modulus* [9]. The metric regularity is equivalent to the surjectivity of coderivative of Γ at $\bar{x}$ for $\bar{y}$ or Aubin's property of Γ^{-1} at $\bar{y}$. In particular, it holds under the graphic convexity of Γ [9]. For a comprehensive discussion of the history and recent development of the notion, see [9], [18, Chapter 9] and references therein.

Theorem 4. *Let*

$$\Gamma(x) := \mathbb{E}\big[\{D_x f(x,\xi)\}\big] + \mathcal{N}_{\mathcal{X}}(x),$$

x_N be a stationary point which satisfies (22) *and sequence $\{x_N\}$ converge to x^* w.p.* 1 *as $N \to \infty$. Suppose that*:

(a) *Assumption* 1 *holds*;
(b) *$f(x,\xi)$ is continuously differentiable with respect to x*;
(c) *Γ is metrically regular at x^* for* 0.

Then for N sufficiently large

$$d\big(x_N, \Gamma^{-1}(0)\big) \leqslant \alpha \mathbb{H}\big(\mathbb{E}\big[\{D_x f(x_N,\xi)\}\big], \{D_x \hat{f}_N(x_N)\}\big), \tag{23}$$

where α is the regularity modulus of Γ at x^* for 0.

Proof. Let $\bar{N}$ be sufficiently large such that for all $N > \bar{N}$, w.p. 1 x_N falls into a neighborhood of x^* where the metric regularity applies. By the metric regularity of Γ at x^* for 0, there exists a constant $\alpha > 0$ such that

$$d\big(x_N, \Gamma^{-1}(0)\big) \leqslant \alpha d\big(0, \Gamma(x_N)\big). \tag{24}$$

By the definition of $\mathbb{D}$,

$$\begin{aligned} d\big(x_N, \Gamma^{-1}(0)\big) &\leqslant \alpha d\big(0, \Gamma(x_N)\big) \\ &\leqslant \alpha \mathbb{D}\big(\{D_x \hat{f}_N(x_N)\} + \mathcal{N}_{\mathcal{X}}(x_N), \mathbb{E}\big[\{D_x \hat{f}(x_N,\xi)\}\big] + \mathcal{N}_{\mathcal{X}}(x_N)\big) \\ &\leqslant \alpha \mathbb{D}\big(\{D_x \hat{f}_N(x_N)\}, \mathbb{E}\big[\{D_x \hat{f}(x_N,\xi)\}\big]\big). \end{aligned}$$

The last inequality is due to (4). The proof is complete. □

A note on the regularity assumption. The optimality condition of the true problem $0 \in \Gamma(x)$ is essentially an infinite dimensional stochastic variational inequality problem. The metric regularity of the latter is investigated by Dontchev, Lewis and Rockafellar [9]. See [9, Theorem 5.1].

Theorem 4 is interesting from numerical perspective in that it gives an error bound for x_N in terms of $\mathbb{H}(\mathbb{E}[D_x f(x_N,\xi)], \{D_x\hat{f}_N(x_N)\})$. The latter may be used as a stopping criterion. In what follows we use the result to derive the exponential convergence rate.

Let $M_x(t) := \mathbb{E}[e^{\mathbb{H}(\{D_x f(x,\xi)\},\mathbb{E}[\{D_x f(x,\xi)\}])t}]$ denote the moment generating function of random variable $\mathbb{H}(\{D_x f(x,\xi)\}, \mathbb{E}[\{D_x f(x,\xi)\}])$.

Theorem 5. *Assume conditions in Theorem* 4*. Assume, in addition, that*:

(a) *the moment generating function $M_x(t)$ is finite valued for t close to zero*;
(b) *there exists a positive integrable function $\kappa(\xi) > 0$ such that*

$$\mathbb{H}\big(\{D_x f(x',\xi)\}, \{D_x f(x'',\xi)\}\big) \leqslant \kappa(\xi)\|x' - x''\|, \quad \forall x', x'' \in \mathcal{X};$$

(c) *X^* is a separable Banach space.*

Then for any small positive number $\epsilon > 0$, there exist positive constants $C(\epsilon) > 0$, $\beta(\epsilon) > 0$ independent of N such that for N sufficiently large

$$\text{Prob}\big(d\big(x_N, \Gamma^{-1}(0)\big) \geqslant \alpha\epsilon\big) \leqslant C(\epsilon)e^{-\beta(\epsilon)N}, \tag{25}$$

w.p. 1*, where α is the regularity modulus of Γ at x^*.*

Proof. Let $\epsilon > 0$ be a small number. Under conditions (a), (b) and (c), it follows by Theorem 2 and Remark 1 that there exist positive constants $C(\epsilon) > 0$ and $\beta(\epsilon) > 0$ independent of N such that for N sufficiently large

$$\text{Prob}\Big(\sup_{x\in\mathcal{X}} \mathbb{H}\big(\{D_x\hat{f}_N(x)\}, \mathbb{E}\big[\{D_x f(x,\xi)\}\big]\big) \geqslant \epsilon\Big) \leqslant C(\epsilon)e^{-\beta(\epsilon)N}.$$

Let $\bar{N}$ be such that for all $N > \bar{N}$, w.p. 1, x_N satisfies (23). Combining the above equation with (23), we have that

$$\begin{aligned}
\text{Prob}\big(d\big(x_N, \Gamma^{-1}(0)\big) \geqslant \alpha\epsilon\big) &\leqslant \text{Prob}\big(\alpha\mathbb{H}\big(\mathbb{E}\big[\{D_x f(x_N,\xi)\}\big], \{D_x\hat{f}_N(x_N)\}\big) \geqslant \alpha\epsilon\big) \\
&\leqslant \text{Prob}\Big(\sup_{x\in\mathcal{X}} \mathbb{H}\big(\{D_x\hat{f}_N(x)\}, \mathbb{E}\big[\{D_x f(x,\xi)\}\big]\big) \geqslant \epsilon\Big) \\
&\leqslant C(\epsilon)e^{-\beta(\epsilon)N}.
\end{aligned}$$

The proof is complete. □

The established exponential convergence should be distinguished from those in the literature [21,22] where it is often shown that with probability approaching one exponentially fast with the increase of sample size, an optimal solution of the true problem becomes an ϵ-optimal solution of its sample average approximation. The latter is numerically useful only when the ϵ-optimal solution set is small. Our result here is stronger in the sense we measure the distance of a computed approximate stationary point to the set of stationary points of the true problem.

Note that our rate is obtained for $d(x_N, \Gamma^{-1}(0))$ rather than $\|x_N - x^*\|$. The latter is larger than the former. It is possible to obtain an estimate for the latter under some stronger regularity conditions. We omit details.

Note also that our results in this section can be easily extended to stochastic variational inequality problems in a Banach space. To see this, we only need to replace $D_x f(x,\xi)$ with a general smooth integrable vector valued function $F(x,\xi)$ in (21) and $\{D_x\hat{f}_N(x)\}$ with its sample average in (22). Stochastic variational inequality model in finite dimensional space has been proposed by Gürkan, Özge and Robinson [10]. The model has interesting applications in economics and engineering where equilibrium problems can be modeled as a variational inequality.

Finally, we note that our results established in this section are based on a key assumption that $f(x,\xi)$ is continuously differentiable with respect to x. It is possible to extend the results to the case when $f(x,\xi)$ is nonsmooth but has some specific structure such as a composite function as discussed in Corollary 2. In such a case, we may replace D_x with some approximate subdifferential which is Lipschitz continuous with respect to x. Consider for instance a composite function $f := g \circ h$ as defined in Corollary 2. If g is convex, then we can construct an approximate subdifferential of f by using the ϵ-convex subdifferential. It is well known [12, Theorem 4.1.3] that the latter is Lipschitz continuous for fixed $\epsilon > 0$. Consequently, we can establish exponential convergence rates of a sequence of stationary points of SAA characterized by such an approximate subdifferential. We omit the details as they are purely technical.

Acknowledgments

We would like to thank Pedro Terán for helpful comments on an earlier version of this paper. We would also like to thank an anonymous referee for insightful comments which leads to a significant improvement of the paper.

References

[1] Z. Artstein, On the calculus of closed set-valued functions, Indiana Univ. Math. J. 24 (1974) 433–441.
[2] Z. Artstein, J.C. Hansen, Convexification in limit law of random sets in Banach spaces, Ann. Probab. 13 (1985) 307–309.
[3] Z. Artstein, R.A. Vitale, A strong law of large numbers for random compact sets, Ann. Probab. 3 (1975) 879–882.
[4] J.-P. Aubin, H. Frankowska, Set-Valued Analysis, Birkhäuser, Boston, 1990.
[5] R.J. Aumann, Integrals of set-valued functions, J. Math. Anal. Appl. 12 (1965) 1–12.
[6] C. Castaing, M. Valadier, Convex Analysis and Measurable Multifunctions, Lecture Notes in Math., vol. 580, Springer, Berlin, 1977.
[7] F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley, New York, 1983.
[8] J. Diestel, J.J. Uhl, Vector Measures, Math. Surveys, vol. 15, American Mathematical Society, 1977.
[9] A.L. Dontchev, A.S. Lewis, R.T. Rockafellar, The radius of metric regularity, Trans. Amer. Math. Soc. 355 (2004) 493–517.
[10] G. Gürkan, A.Y. Özge, S.M. Robinson, Sample-path solution of stochastic variational inequalities, Math. Program. 84 (1999) 313–333.
[11] F. Hiai, Convergence of conditional expectations and strong law of large numbers for multivalued random variables, Trans. Amer. Math. Soc. 291 (1985) 613–627.
[12] J.B. Hiriart-Urruty, C. Lemaréchal, Convex Analysis and Minimization Algorithms. I, Springer-Verlag, Berlin, 1993.
[13] C. Hess, Set-valued integration and set-valued probability theory: an overview, in: Handbook of Measure Theory, vols. I, II, North-Holland, Amsterdam, 2002, pp. 617–673.
[14] I. Molchanov, Theory of random sets, in: J. Gani, et al. (Eds.), Probability and Its Applications, Springer, 2005.
[15] B.S. Mordukhovich, Variational Analysis and Generalized Differentiation II. Applications, Springer, Berlin, 2006.
[16] N. Papageorgiou, On the theory of Banach space valued multifunctions 1. Integration and conditional expectation, J. Multivariate Anal. 17 (1985) 185–206.
[17] S.M. Robinson, Analysis of sample-path optimization, Math. Oper. Res. 21 (1996) 513–528.
[18] R.T. Rockafellar, R.J.-B. Wets, Variational Analysis, Springer-Verlag, Berlin, 1998.
[19] R.Y. Rubinstein, A. Shapiro, Discrete Events Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Methods, John Wiley and Sons, New York, 1993.
[20] A. Rusczyński, A. Shapiro (Eds.), Stochastic Programming, Handbooks Oper. Res. Management Sci., vol. 10, North-Holland, Amsterdam, 2003.
[21] A. Shapiro, T. Homem-de-Mello, On rate of convergence of Monte Carlo approximations of stochastic programs, SIAM J. Optim. 11 (2000) 70–86.
[22] A. Shapiro, Monte Carlo sampling methods, in: A. Rusczyński, A. Shapiro (Eds.), Stochastic Programming, in: Handbooks Oper. Res. Management Sci., vol. 10, North-Holland, Amsterdam, 2003.
[23] A. Shapiro, H. Xu, Stochastic mathematical programs with equilibrium constraints, modeling and sample average approximation, Optimization 57 (2008) 395–418.
[24] A. Shapiro, H. Xu, Uniform Laws of large numbers for set-valued mappings and subdifferentials of random functions, J. Math. Anal. Appl. 325 (2007) 1390–1399.
[25] Pedro Terán, On a uniform law of large numbers for random sets and subdifferentials of random functions, Statist. Probab. Lett. 78 (2008) 42–49.