Available online at www.sciencedirect.com
ScienceDirect
J. Math. Anal. Appl. 325 (2007) 1390–1399
Journal of MATHEMATICAL ANALYSIS AND APPLICATIONS
www.elsevier.com/locate/jmaa

Uniform laws of large numbers for set-valued mappings and subdifferentials of random functions

Alexander Shapiro [a,*,1], Huifu Xu [b,2]

[a] School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0205, USA
[b] School of Mathematics, University of Southampton, Highfield Southampton, SO17 1BJ, UK

Received 19 December 2005

Available online 29 March 2006

Submitted by B.S. Mordukhovich

Abstract

We derive a uniform (strong) Law of Large Numbers (LLN) for random set-valued mappings. The result can be viewed as an extension of both, a uniform LLN for random functions and LLN for random sets. We apply the established results to a consistency analysis of stationary points of sample average approximations of nonsmooth stochastic programs.

Keywords: Random sets; Artstein–Vitale law of large numbers; Set-valued mappings; Uniform LLN; Generalized gradients; Stochastic programming; Sample average approximation

1. Introduction

A uniform version of the (strong) Law of Large Numbers (LLN) for considered (real-valued) random functions has been instrumental in consistency analysis of optimal values and solutions in stochastic optimization (see, e.g., [6, Sections 2.6 and 6.2]). However, those results cannot be applied to a similar analysis of stationary points, in nonsmooth nonconvex stochastic optimization, formulated in terms of subdifferentials (generalized gradients) of the corresponding

* Corresponding author.
E-mail addresses: ashapiro@isye.gatech.edu (A. Shapiro), h.xu@maths.soton.ac.uk (H. Xu).
1 Research of this author was partly supported by the NSF grant DMS-0510324.
2 The work of this author is supported by the United Kingdom Engineering and Physical Sciences Research Council grant GR/S90850/01.

0022-247X/$ – see front matter
doi:10.1016/j.jmaa.2006.02.078

objective functions. Consequently a uniform LLN for random set-valued mappings (multifunctions) is needed. In this paper we present such a result. An LLN for random sets was originally derived by Artstein and Vitale [1] and extended to a Banach space setting in [2]. This has a natural application to establishing LLN for subdifferentials of random functions at a given (fixed) point. In studying convergence properties of stationary points of sample average approximations of stochastic programs, one needs a *uniform* type of LLN for random multifunctions given by subdifferentials (generalized gradients) of considered random functions. In this paper we extend the Artstein–Vitale LLN to a uniform setting and show how it can be applied to establishing convergence of stationary points of sample average approximations of stochastic programs.

We use the following notation throughout the paper. For a point x in a metric space $(\mathcal{X}, \rho)$ and $r \geqslant 0$, we denote by $B_r(x) := \{x' \in \mathcal{X}\colon\ \rho(x', x) \leqslant r\}$ the ball of radius r centered at x. For a set $A \subset \mathbb{R}^n$ and $y \in \mathbb{R}^n$, we denote by $\operatorname{dist}(y, A) := \inf_{z \in A} \|y - z\|$ the distance from y to A with respect to the Euclidean norm $\|\cdot\|$. For two sets $A, C \subset \mathbb{R}^n$ we denote by

$$\mathbb{D}(A, C) := \sup_{x \in A} \operatorname{dist}(x, C)$$

the deviation of the set A from the set C, by

$$\mathbb{H}(A, C) := \max\bigl\{\mathbb{D}(A, C), \mathbb{D}(C, A)\bigr\}$$

the Hausdorff distance between A and C, and $\|A\| := \sup_{y \in A} \|y\|$. Note that both $\mathbb{D}(\cdot,\cdot)$ and $\mathbb{H}(\cdot,\cdot)$ satisfy the triangle inequality, i.e., for sets $A, B, C \subset \mathbb{R}^n$, the following inequality holds:

$$\mathbb{D}(A, C) \leqslant \mathbb{D}(A, B) + \mathbb{D}(B, C), \tag{1.1}$$

and similarly for the Hausdorff distance. Note also that for sets $A, B, A', B' \subset \mathbb{R}^n$, the following inequality holds:

$$\mathbb{D}(A + B, A' + B') \leqslant \mathbb{D}(A, A') + \mathbb{D}(B, B'). \tag{1.2}$$

Both inequalities (1.1) and (1.2) follow easily from the formula

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

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

2. Uniform LLN for set-valued mappings

Let $(\mathcal{X}, \rho)$ be a metric space, $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{A}: \mathcal{X} \times \Omega \rightrightarrows \mathbb{R}^n$ be a multifunction (set-valued mapping), which maps $(x, \omega) \in \mathcal{X} \times \Omega$ into a subset of $\mathbb{R}^n$. We assume that $\mathcal{A}(x, \omega)$ is *compact valued*, i.e., for every $(x, \omega) \in \operatorname{dom}(\mathcal{A})$ we have that $\mathcal{A}(x, \omega) \in \mathfrak{C}_n$, where $\mathfrak{C}_n$ denotes the space of nonempty compact subsets of $\mathbb{R}^n$. Equipped with the Hausdorff metric $\mathbb{H}(\cdot,\cdot)$, the space $\mathfrak{C}_n$ becomes a metric space. We can view the multifunction $\mathcal{A}$ as a single-valued mapping from $\operatorname{dom}(\mathcal{A})$ into $\mathfrak{C}_n$. With some abuse of notation we denote this single-valued mapping also by $\mathcal{A}$. Also for $x \in \mathcal{X}$, we sometimes use the notation $\mathcal{A}_x(\omega) := \mathcal{A}(x, \omega)$.

We assume that for every $x \in \mathcal{X}$, the multifunction $\mathcal{A}_x(\cdot)$ is *measurable*. That is, for any closed set $C \subset \mathbb{R}^n$, the set

$$\mathcal{A}_x^{-1}(C) = \bigl\{\omega \in \Omega\colon\ \mathcal{A}_x(\omega) \cap C \neq \emptyset\bigr\}$$

is $\mathcal{F}$-measurable. We have that $\mathcal{A}_x$ is measurable iff the corresponding single-valued mapping $\mathcal{A}_x : \operatorname{dom}(\mathcal{A}_x) \to \mathfrak{C}_n$ is measurable in the usual sense, i.e., the inverse of every Borel subset of the metric space $(\mathfrak{C}_n, \mathbb{H})$, belongs to the sigma algebra $\mathcal{F}$ (e.g., [5, Theorem 14.4]).

We say that $\mathcal{A}_i : \mathcal{X} \times \Omega \to \mathfrak{C}_n$, $i = 1, \ldots$, is an independent identically distributed (iid) sequence of realizations of the multifunction $\mathcal{A}$ if for every $x \in \mathcal{X}$, the multifunction $\mathcal{A}_i(x) = \mathcal{A}_i(x, \cdot)$ is measurable and has the same probability distribution as $\mathcal{A}(x, \omega)$, and each $\mathcal{A}_i(x)$ is independent of $\{\mathcal{A}_j(x)\}_{j \neq i}$. We have the following (strong) law of large numbers for the iid sequence $A_i = \mathcal{A}_i(x)$ of random sets (for fixed $x \in \mathbb{R}^n$), due to Artstein and Vitale [1].

Theorem 1. *Let A_i, $i = 1, \ldots$, be an iid sequence of realizations of a (measurable) mapping $\mathcal{A} : \Omega \to \mathfrak{C}_n$ such that $\mathbb{E}\|\mathcal{A}\| < \infty$. Then*

$$N^{-1}(A_1 + \cdots + A_N) \to \mathbb{E}\big[\mathrm{conv}(\mathcal{A})\big] \quad \text{w.p. } 1 \text{ as } N \to \infty. \tag{2.1}$$

Here $\mathrm{conv}(\mathcal{A})$ denotes the convex hull of $\mathcal{A}$. The convergence in (2.1) is taken in the (Hausdorff) metric of the space $\mathfrak{C}_n$. The expectation $\mathbb{E}[\mathrm{conv}(\mathcal{A})]$ is defined as the set of integrals $\int_\Omega a(\omega)\, dP(\omega)$ taken over all measurable selections $a(\omega) \in \mathrm{conv}(\mathcal{A})(\omega)$. Recall that by a theorem due to Aumann [3],

$$\mathbb{E}\big[\mathrm{conv}(\mathcal{A})\big] = \mathbb{E}[\mathcal{A}]$$

if the probability space $(\Omega, \mathcal{F}, P)$ is nonatomic.

We say that a multifunction $\mathcal{A} : \mathcal{X} \to \mathfrak{C}_n$ is upper semicontinuous at a point $\bar{x} \in \mathcal{X}$ if for any neighborhood $\mathcal{V}$ of the set $\mathcal{A}(\bar{x})$ there exists a neighborhood $\mathcal{N}$ of $\bar{x}$ such that for every $x \in \mathcal{N}$ the inclusion $\mathcal{A}(x) \subset \mathcal{V}$ holds. Since here the multifunction $\mathcal{A}$ is assumed to be compact valued, this is equivalent to the condition

$$\lim_{x \to \bar{x}} \mathbb{D}\big(\mathcal{A}(x), \mathcal{A}(\bar{x})\big) = 0. \tag{2.2}$$

It is said that $\mathcal{A}$ is upper semicontinuous if it is upper semicontinuous at every point of $\mathcal{X}$. For $r \geqslant 0$ we denote

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

Note that for any $t \geqslant 0$ and $A \subset \mathbb{R}^n$ it holds that $\mathrm{conv}(A + t\mathcal{B}) = \mathrm{conv}(A) + t\mathcal{B}$, where $\mathcal{B}$ is the unit ball in $\mathbb{R}^n$. It follows that if the multifunction $\mathcal{A}$ is upper semicontinuous, then the multifunction $\mathrm{conv}(\mathcal{A})$ is also upper semicontinuous.

We prove now the main result of this paper which can be viewed as an extension of Theorem 1 to a uniform setting.

Theorem 2. *Let $\mathcal{A}_i(x)$, $i = 1, \ldots$, be an iid sequence of realizations of a (measurable) mapping $\mathcal{A} : \mathcal{X} \times \Omega \to \mathfrak{C}_n$ and $\mathcal{S}_N(x) := N^{-1} \sum_{i=1}^N \mathcal{A}_i(x)$. Suppose that the metric space $(\mathcal{X}, \rho)$ is compact, there exists a P-integrable function $\kappa : \Omega \to \mathbb{R}_+$ such that*

$$\big\|\mathcal{A}(x, \omega)\big\| \leqslant \kappa(\omega), \quad \forall (x, \omega) \in \mathcal{X} \times \Omega, \tag{2.3}$$

and that for any $x \in \mathcal{X}$ the multifunction $\mathcal{A}(\cdot, \omega)$ is upper semicontinuous at x for P-almost every $\omega \in \Omega$. Then the expected value $\mathcal{E}(x) := \mathbb{E}[\mathrm{conv}(\mathcal{A})(x)]$ is well defined, the multifunction $\mathcal{E} : \mathcal{X} \to \mathfrak{C}_n$ is upper semicontinuous and for any $r > 0$, the following limits hold:

$$\sup_{x \in \mathcal{X}} \mathbb{D}\big(\mathcal{S}_N(x), \mathcal{E}^r(x)\big) \to 0 \quad \text{w.p. } 1 \text{ as } N \to \infty, \tag{2.4}$$

$$\sup_{x \in \mathcal{X}} \mathbb{D}\big(\mathcal{E}(x), \mathcal{S}_N^r(x)\big) \to 0 \quad \text{w.p. } 1 \text{ as } N \to \infty. \tag{2.5}$$

Proof. First note that by assumption (2.3), we have

$$\mathbb{E}\|\mathrm{conv}(\mathcal{A})(x)\| \leqslant \mathbb{E}[\kappa] < \infty, \quad \forall x \in \mathcal{X}.$$

It follows that the expected value $\mathcal{E}(x)$ is well defined and $\|\mathcal{E}(x)\| \leqslant \mathbb{E}[\kappa]$ for all $x \in \mathcal{X}$. We also have by (1.2) and (2.3) that for any $x, x' \in \mathcal{X}$ and $\omega \in \Omega$,

$$\mathbb{D}\big(\mathrm{conv}(\mathcal{A})(x',\omega), \mathrm{conv}(\mathcal{A})(x,\omega)\big) \leqslant 2\kappa(\omega). \tag{2.6}$$

Consider a sequence $\gamma_k \downarrow 0$ and define $V_k(x) := B_{\gamma_k}(x)$, i.e., $V_k(x)$ is a ball of radius γ_k centered at x. Because of (2.6), we have by the Lebesgue Dominated Convergence Theorem that for any (fixed) point $x \in \mathcal{X}$,

$$\begin{aligned}\lim_{k\to\infty} \int_\Omega \sup_{x'\in V_k(x)} \mathbb{D}\big(\mathrm{conv}(\mathcal{A})(x',\omega), \mathrm{conv}(\mathcal{A})(x,\omega)\big)\, dP(\omega)\\ = \int_\Omega \lim_{k\to\infty} \sup_{x'\in V_k(x)} \mathbb{D}\big(\mathrm{conv}(\mathcal{A})(x',\omega), \mathrm{conv}(\mathcal{A})(x,\omega)\big)\, dP(\omega).\end{aligned} \tag{2.7}$$

Since $\mathcal{A}(\cdot,\omega)$, and hence $\mathrm{conv}(\mathcal{A})(\cdot,\omega)$, is upper semicontinuous at x w.p. 1, we have that the supremum inside the integral in the right-hand side of (2.7) tends to zero for a.e. $\omega \in \Omega$. We obtain that

$$\lim_{k\to\infty} \int_\Omega \sup_{x'\in V_k(x)} \mathbb{D}\big(\mathrm{conv}(\mathcal{A})(x',\omega), \mathrm{conv}(\mathcal{A})(x,\omega)\big)\, dP(\omega) = 0. \tag{2.8}$$

We also have that

$$\mathbb{D}\big(\mathcal{E}(x'), \mathcal{E}(x)\big) \leqslant \int_\Omega \mathbb{D}\big(\mathrm{conv}(\mathcal{A})(x',\omega), \mathrm{conv}(\mathcal{A})(x,\omega)\big)\, dP(\omega), \quad \forall x, x' \in \mathcal{X}. \tag{2.9}$$

It follows that $\mathbb{D}(\mathcal{E}(x_k), \mathcal{E}(x)) \to 0$ for any $x_k \to x$, and hence $\mathcal{E}(\cdot)$ is upper semicontinuous at x.

Let us observe that it follows from (1.2) that for any $x, x' \in \mathcal{X}$, it holds that

$$\mathbb{D}\big(\mathcal{S}_N(x'), \mathcal{S}_N(x)\big) \leqslant N^{-1} \sum_{i=1}^N \mathbb{D}\big(\mathcal{A}_i(x'), \mathcal{A}_i(x)\big). \tag{2.10}$$

Consequently

$$\sup_{x'\in V_k(x)} \mathbb{D}\big(\mathcal{S}_N(x'), \mathcal{S}_N(x)\big) \leqslant N^{-1} \sum_{i=1}^N \sup_{x'\in V_k(x)} \mathbb{D}\big(\mathcal{A}_i(x'), \mathcal{A}_i(x)\big). \tag{2.11}$$

By the (strong) LLN we have that w.p. 1,

$$N^{-1} \sum_{i=1}^N \sup_{x'\in V_k(x)} \mathbb{D}\big(\mathcal{A}_i(x'), \mathcal{A}_i(x)\big) \to \mathbb{E}\Big[\sup_{x'\in V_k(x)} \mathbb{D}\big(\mathcal{A}(x'), \mathcal{A}(x)\big)\Big]. \tag{2.12}$$

Consider an arbitrary point $x \in \mathcal{X}$. Of course, the limit of form (2.8) holds for the multifunction $\mathcal{A}$ as well. Consequently, we have that for any given $\varepsilon > 0$ the right-hand side of (2.12) is less than ε for all k large enough. Together with (2.11) and (2.12) this implies that there exists $\delta > 0$ such that for a.e. $\omega \in \Omega$ there exists $\bar{N}(\omega)$ such that

$$\sup_{x'\in B_\delta(x)} \mathbb{D}\big(\mathcal{S}_N(x'), \mathcal{S}_N(x)\big) \leqslant \varepsilon \tag{2.13}$$

for all $N \geqslant \bar{N}(\omega)$. Consequently, by compactness of $\mathcal{X}$, given $\delta > 0$ there exist a finite set of points $x_j \in \mathcal{X}$, $j = 1, \ldots, \ell$, with respective neighborhoods $W_j := B_\delta(x_j)$, such that $\mathcal{X} \subset \bigcup_{j=1}^{\ell} W_j$ and for a.e. $\omega \in \Omega$ there is $\bar{N}(\omega)$ such that

$$\sup_{x' \in W_j} \mathbb{D}\big(\mathcal{S}_N(x'), \mathcal{S}_N(x_j)\big) \leqslant \varepsilon, \quad j = 1, \ldots, \ell, \tag{2.14}$$

for all $N \geqslant \bar{N}(\omega)$. Of course, we can take $\delta \leqslant r$.

Now we can apply the pointwise LLN of Theorem 1 at every point x_j to conclude (probably by taking a larger value of $\bar{N}(\omega)$) that

$$\mathbb{D}\big(\mathcal{S}_N(x_j), \mathcal{E}(x_j)\big) \leqslant \varepsilon, \quad j = 1, \ldots, \ell, \tag{2.15}$$

for all $N \geqslant \bar{N}(\omega)$. For a point $x \in \mathcal{X}$ we have that $x \in W_j$ for some $j \in \{1, \ldots, \ell\}$. Then

$$\mathbb{D}\big(\mathcal{S}_N(x), \mathcal{E}^r(x)\big) \leqslant \mathbb{D}\big(\mathcal{S}_N(x), \mathcal{S}_N(x_j)\big) + \mathbb{D}\big(\mathcal{S}_N(x_j), \mathcal{E}(x_j)\big) + \mathbb{D}\big(\mathcal{E}(x_j), \mathcal{E}^r(x)\big). \tag{2.16}$$

It follows from (2.14) and (2.15) that the first and second terms in the right-hand side of (2.16) are less than or equal to ε for $N \geqslant \bar{N}(\omega)$, and the last term is zero since $\rho(x_j, x) \leqslant r$. We obtain that for $N \geqslant \bar{N}(\omega)$,

$$\sup_{x \in \mathcal{X}} \mathbb{D}\big(\mathcal{S}_N(x), \mathcal{E}^r(x)\big) \leqslant 2\varepsilon, \tag{2.17}$$

and hence (2.4) follows.

The other limit (2.5) can be proved in a similar way. That is, for a given $\varepsilon > 0$, we can choose a finite number of points $x_j \in \mathcal{X}$, $j = 1, \ldots, \ell$, and $\delta \in (0, r)$, such that the neighborhoods $W_j := B_\delta(x_j)$, $j = 1, \ldots, \ell$, cover $\mathcal{X}$, and

$$\sup_{x \in W_j} \mathbb{D}\big(\mathcal{E}(x), \mathcal{E}(x_j)\big) \leqslant \varepsilon, \quad j = 1, \ldots, \ell, \tag{2.18}$$

and (by Theorem 1) for a.e. $\omega \in \Omega$ there exists $\bar{N}(\omega)$ such that

$$\mathbb{D}\big(\mathcal{E}(x_j), \mathcal{S}_N(x_j)\big) \leqslant \varepsilon, \quad j = 1, \ldots, \ell, \tag{2.19}$$

for all $N \geqslant \bar{N}(\omega)$. Consequently

$$\mathbb{D}\big(\mathcal{E}(x), \mathcal{S}_N^r(x)\big) \leqslant \mathbb{D}\big(\mathcal{E}(x), \mathcal{E}(x_j)\big) + \mathbb{D}\big(\mathcal{E}(x_j), \mathcal{S}_N(x_j)\big) + \mathbb{D}\big(\mathcal{S}_N(x_j), \mathcal{S}_N^r(x)\big) \leqslant 2\varepsilon,$$

and hence (2.5) follows. □

Remark 1. If $\mathcal{A}(\cdot, \omega)$ is compact valued and upper semicontinuous, then $\mathcal{A}^r(\cdot, \omega)$ is also compact valued and upper semicontinuous. By replacing $\mathcal{A}$ with $\mathcal{A}^r$ we obtain that under the assumptions of Theorem 2, limits (2.4) and (2.5) imply that

$$\sup_{x \in \mathcal{X}} \mathbb{D}\big(\mathcal{S}_N^{r'}(x), \mathcal{E}^r(x)\big) \to 0 \quad \text{w.p. 1 as } N \to \infty, \tag{2.20}$$

$$\sup_{x \in \mathcal{X}} \mathbb{D}\big(\mathcal{E}^{r'}(x), \mathcal{S}_N^r(x)\big) \to 0 \quad \text{w.p. 1 as } N \to \infty, \tag{2.21}$$

for any $r > r' \geqslant 0$.

Remark 2. The natural question is whether limit (2.4) holds for $r = 0$. It looks that in the absence of continuity of $\mathcal{A}(\cdot, \omega)$ this could be not true. However, we do not have such a counterexample.

Of course, the same question can be asked about limit (2.5) for $r = 0$, and ultimately whether the uniform LLN (2.23) holds under the assumptions of Theorem 2. As far as we know these questions are open.

If the expected value multifunction $\mathcal{E}(x)$ is continuous (in the Hausdorff metric), then it is possible to strengthen the convergence result (2.4) to $r = 0$. That is, the following result holds.

Theorem 3. *If in addition to the assumptions of Theorem* 2 *the expected value multifunction* $\mathcal{E}(x)$ *is continuous, then*

$$\sup_{x\in\mathcal{X}} \mathbb{D}\big(\mathcal{S}_N(x), \mathcal{E}(x)\big) \to 0 \quad \textit{w.p. } 1 \textit{ as } N \to \infty. \tag{2.22}$$

Proof. Convergence (2.22) follows from assertion (2.4) of Theorem 2 and the following fact:

($\star$) Let $(\mathcal{X}, \rho)$ be a compact metric space and $\mathcal{E} : \mathcal{X} \to \mathfrak{C}_n$ be a continuous mapping. Then for any $\varepsilon > 0$ there exists $\delta > 0$ such that $\mathbb{H}(\mathcal{E}^\delta(x), \mathcal{E}(x)) \leqslant \varepsilon$ for all $x \in \mathcal{X}$.

Proof of the above assertion ($\star$) is rather standard. Let us argue by a contradiction. Suppose that the assertion is false. Then there exist $\varepsilon > 0$ and sequences $x_k \in \mathcal{X}$ and $r_k \downarrow 0$ such that $\mathbb{H}(\mathcal{E}^{r_k}(x_k), \mathcal{E}(x_k)) > \varepsilon$ for all k. Since $\mathcal{X}$ is compact, we can assume that x_k converges to a point $x^* \in \mathcal{X}$. Moreover, since $\mathcal{E}$ is continuous at x^*, we have that there is $\delta > 0$ such that $\mathbb{H}(\mathcal{E}(x), \mathcal{E}(x')) < \varepsilon$ for all $x, x' \in B_\delta(x^*)$. This, however, contradicts the assumption that $\mathbb{H}(\mathcal{E}^{r_k}(x_k), \mathcal{E}(x_k)) > \varepsilon$ for $\rho(x_k, x^*) < \delta/2$ and $r_k < \delta/2$. □

Moreover, if for every $x \in \mathcal{X}$ the multifunction $\mathcal{A}(x, \omega)$ is a singleton w.p. 1, i.e., $\mathcal{A}(x, \omega) = \{a(x, \omega)\}$ for a.e. $\omega \in \Omega$, then the expected value $\mathcal{E}(x) = \{e(x)\}$ is also single valued. In that case the upper semicontinuity of $\mathcal{E}(x)$ is equivalent to continuity of the corresponding mapping $e : \mathcal{X} \to \mathbb{R}^n$. The following result then is a consequence of Theorem 3.

Corollary 1. *Suppose that, in addition to the assumptions of Theorem* 2, *for every* $x \in \mathcal{X}$ *the set* $\mathcal{A}(x, \omega)$ *is a singleton for a.e.* $\omega \in \Omega$. *Then the expected value* $\mathcal{E}(x) = \{e(x)\}$ *is single valued, the mapping* $e : \mathcal{X} \to \mathbb{R}^n$ *is continuous and convergence* (2.22) *follows.*

If we assume that the multifunction $\mathcal{A}(\cdot, \omega)$ is continuous (rather than just upper semicontinuous) we obtain the following result.

Theorem 4. *Suppose that, in addition to the assumptions of Theorem* 2, *for every* $x \in \mathcal{X}$ *the multifunction* $\mathcal{A}(\cdot, \omega)$ *is continuous at* x *for a.e.* $\omega \in \Omega$. *Then the expected value multifunction* $\mathcal{E}(x)$ *is continuous and the following uniform LLN holds*:

$$\sup_{x\in\mathcal{X}} \mathbb{H}\big(\mathcal{S}_N(x), \mathcal{E}(x)\big) \to 0 \quad \textit{w.p. } 1 \textit{ as } N \to \infty. \tag{2.23}$$

Proof. One can proceed in proving continuity of $\mathcal{E}(x)$ exactly in the same way as the proof of the upper semicontinuity in Theorem 2, by replacing $\mathbb{D}$ in Eqs. (2.6)–(2.9) with $\mathbb{H}$. Also in order to prove (2.23) one can use the same arguments as in the proof of Theorem 2. The crucial difference here is that by continuity of $\mathcal{E}(x)$ we can use inequality (2.16) with $\mathbb{D}$ replaced by $\mathbb{H}$ and $r = 0$. □

3. Uniform LLN for subdifferentials of random functions

In this section we discuss applications of the results of the previous section to studying convergence of subdifferentials of sample averages of random functions. Let $\xi_i : \Omega \to \mathbb{R}^d$, $i = 1, \dots,$ be an iid sequence of random vectors supported on a set Ξ, i.e., Ξ is a closed subset of $\mathbb{R}^d$ and $\xi_i \in \Xi$ w.p. 1. Consider a function $F : \mathbb{R}^n \times \Xi \to \mathbb{R}$. We assume that $F(x, \xi)$ is a *Carathéodory* function, i.e., $F(x, \cdot)$ is (Borel) measurable for every $x \in \mathbb{R}^n$ and $F(\cdot, \xi)$ is continuous for a.e. $\xi \in \Xi$ (in fact we assume later that $F(\cdot, \xi)$ is Lipschitz continuous). We assume that the expected value function $f(x) := \mathbb{E}[F(x, \xi)]$ is *well defined* (finite valued). Consider the sample average function:

$$\hat{f}_N(x) := N^{-1} \sum_{i=1}^{N} F(x, \xi_i).$$

Let $\mathcal{X}$ be a nonempty compact subset of $\mathbb{R}^n$. Suppose that for a.e. $\xi \in \Xi$ the function $F_\xi(\cdot) = F(\cdot, \xi)$ is Lipschitz continuous on a neighborhood of $\mathcal{X}$, and let $\partial F_\xi(x)$ be its Clarke's generalized gradient at $x \in \mathcal{X}$ (cf., [4]). There are several equivalent ways to define $\partial F_\xi(x)$. For example, consider the so-called generalized directional derivative

$$F_\xi^\circ(x, d) := \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{F_\xi(y + td) - F_\xi(y)}{t}.$$

We have that $F_\xi^\circ(x, \cdot)$ is convex positively homogeneous and is the support function of the closed compact set $\partial F_\xi(x)$ [4, Proposition 2.1.2]. For a given $x \in \mathcal{X}$, we can view $\xi \mapsto F_\xi^\circ(x, \cdot)$ as a mapping from $\mathcal{X} \times \Xi$ into the space $C(S^{n-1})$ of continuous functions, on the unit sphere S^{n-1} of $\mathbb{R}^n$, equipped with the sup-norm. Since there is an isometric correspondence between the space of convex compact subsets of $\mathbb{R}^n$, equipped with the Hausdorff metric, and the corresponding subset of $C(S^{n-1})$, it follows that the multifunction $\xi \mapsto \partial F_\xi(x)$ is measurable (cf., [5, Theorem 14.4]).

Consider the multifunction $\mathcal{A}(x, \xi) := \partial F_\xi(x)$. By the above, we have that for every $x \in \mathcal{X}$ the multifunction $\mathcal{A}(x, \cdot) : \Xi \rightrightarrows \mathbb{R}^n$ is convex and compact valued and measurable. It is a property of Clarke's generalized gradient that this multifunction is upper semicontinuous [4, Proposition 2.1.5]. Suppose, further, that:

(i) there is a measurable function $\kappa : \Xi \to \mathbb{R}_+$ such that $\mathbb{E}[\kappa(\xi)] < \infty$ and

$$\left| F(x', \xi) - F(x, \xi) \right| \leqslant \kappa(\xi) \|x' - x\|, \quad \text{for all } x', x \in \mathcal{X} \text{ and } \xi \in \Xi. \tag{3.1}$$

The above assumption (i) implies condition (2.3) of Theorem 2. We can apply now Theorem 2 to the multifunction $\mathcal{A}(x, \xi)$. Let

$$\mathcal{S}_N(x) := N^{-1} \sum_{i=1}^{N} \mathcal{A}(x, \xi_i) = N^{-1} \sum_{i=1}^{N} \partial F_{\xi_i}(x)$$

and $\mathcal{E}(x) := \mathbb{E}[\mathcal{A}(x)]$ be the corresponding sample average and the expected value multifunctions, respectively. Then the following inclusions hold $\partial \hat{f}_N(x) \subset \mathcal{S}_N(x)$ and $\partial f(x) \subset \mathcal{E}(x)$ [4, Theorem 2.7.2]. Unfortunately, these inclusions can be strict. Therefore, in order to get a meaningful uniform LLN for Clarke's generalized gradients, we need some additional assumptions.

Recall that a locally Lipschitz function $g:\mathbb{R}^n \to \mathbb{R}$ is said to be *regular* at a point $x \in \mathbb{R}^n$, in the sense of Clarke, if $g(\cdot)$ is directionally differentiable at x and the directional derivative $g'(x,\cdot)$ coincides with the generalized directional derivative $g^\circ(x,\cdot)$ (cf., [4, Definition 2.3.4]). Then $g'(x,\cdot)$ coincides with the support function of the generalized gradient $\partial g(x)$, and hence $g'(x,\cdot)$ is convex. Denote

$$\partial^r g(x) := \bigcup_{x' \in B_r(x)} \partial g(x'), \tag{3.2}$$

where the ball $B_r(x)$ is taken with respect to the Euclidean norm of $\mathbb{R}^n$. As a consequence of Theorem 2 we have the following result.

Theorem 5. *Let ξ_i, $i = 1,\dots,N$, be an iid sequence of random vectors supported on a set $\Xi \subset \mathbb{R}^d$, $F:\mathbb{R}^n \times \Xi \to \mathbb{R}$ be a Carathéodory function, and $\mathcal{X}$ be a nonempty compact subset of $\mathbb{R}^n$. Suppose that the above condition* (i) *and the following condition are satisfied*: (ii) *for any $x \in \mathcal{X}$ the function $F(\cdot,\xi)$ is regular at x for a.e. $\xi \in \Xi$. Then for any $r > r' \geqslant 0$ the following limits hold*:

$$\sup_{x \in \mathcal{X}} \mathbb{D}\big(\partial^{r'} \hat{f}_N(x), \partial^r f(x)\big) \to 0 \quad \textit{w.p. } 1 \textit{ as } N \to \infty, \tag{3.3}$$

$$\sup_{x \in \mathcal{X}} \mathbb{D}\big(\partial^{r'} f(x), \partial^r \hat{f}_N(x)\big) \to 0 \quad \textit{w.p. } 1 \textit{ as } N \to \infty. \tag{3.4}$$

Proof. It follows from assumption (i) that for all $\xi \in \Xi$ the function $F(\cdot,\xi)$ is Lipschitz continuous on a neighborhood of $\mathcal{X}$, and hence the generalized gradient $\partial F_\xi(x)$, of $F(\cdot,\xi)$ at x is well defined convex and compact for any $x \in \mathcal{X}$. Moreover, it follows from (2.23) that $\|\partial F_\xi(x)\| \leqslant \kappa(\xi)$ for all $(x,\xi) \in \mathcal{X} \times \Xi$. Now because of the regularity assumption (ii), we have that the sample average function $\hat{f}_N(x)$ is regular (cf., [4, Proposition 2.3.6]) and $\partial \hat{f}_N(x) = N^{-1}\sum_{i=1}^N \partial F_{\xi_i}(x)$. Also, because of the regularity assumption (ii), we have that $f(x)$ is regular at every $x \in \mathcal{X}$ and $\partial f(x) = \mathbb{E}[\partial F_\xi(x)]$ [4, Theorem 2.7.2]. The uniform convergence properties (3.3) and (3.4) follow now by Theorem 2 (see Remark 1). □

By Corollary 1 we have the following result (cf., [7, Proposition 2.2]).

Theorem 6. *Let ξ_i, $i = 1,\dots,N$, be an iid sequence of random vectors supported on a set $\Xi \subset \mathbb{R}^d$, $F:\mathbb{R}^n \times \Xi \to \mathbb{R}$ be a Carathéodory function, and $\mathcal{X}$ be a nonempty compact subset of $\mathbb{R}^n$. Suppose that condition* (i) *and the following condition are satisfied*: (iii) *for every $x \in \mathcal{X}$ the set $\partial F_\xi(x)$ is a singleton for a.e. $\xi \in \Xi$. Then the expected value function $f(x)$ is continuously differentiable, $\mathbb{E}[\partial F_\xi(x)] = \{\nabla f(x)\}$ for any $x \in \mathcal{X}$ and*

$$\sup_{x \in \mathcal{X}} \mathbb{D}\big(\partial \hat{f}_N(x), \{\nabla f(x)\}\big) \to 0 \quad \textit{w.p. } 1 \textit{ as } N \to \infty. \tag{3.5}$$

4. Applications to stochastic optimization

In this section, we discuss an application of the established results to stochastic programming. Consider the following stochastic optimization problem:

$$\mathrm{Min}_{x \in \mathcal{X}} \big\{ f(x) := \mathbb{E}\big[F(x,\xi)\big] \big\}, \tag{4.1}$$

where $\mathcal{X}$ is a nonempty convex compact subset of $\mathbb{R}^n$, ξ is a random vector supported on a set $\Xi \subset \mathbb{R}^d$, and $F(x,\xi)$ is a random function satisfying the assumptions of Theorem 5.

For $\varepsilon \geqslant 0$ we say that a point $\bar{x} \in \mathcal{X}$ is an ε-stationary point of problem (4.1) if it satisfies the following equation:

$$0 \in \partial^{\varepsilon} f(x) + \mathcal{N}_{\mathcal{X}}(x). \tag{4.2}$$

Here

$$\mathcal{N}_{\mathcal{X}}(x) := \{ z \in \mathbb{R}^n \colon \langle z, x' - x \rangle \leqslant 0,\ \forall x' \in \mathcal{X} \}$$

denotes the normal cone of the set $\mathcal{X}$ at $x \in \mathcal{X}$ and $\partial^{\varepsilon} f(x)$ is defined in (3.2). (Note that, by the definition, $\mathcal{N}_{\mathcal{X}}(x) = \emptyset$ if $x \neq \mathcal{X}$.) In particular, for $\varepsilon = 0$ we refer to a point $\bar{x} \in \mathcal{X}$ satisfying (4.2) as a stationary point.

Let $\xi_1, \dots,$ be an iid sample of the random vector ξ. Consider the following sample average program:

$$\operatorname*{Min}_{x \in \mathcal{X}} \left\{ \hat{f}_N(x) := \frac{1}{N} \sum_{i=1}^{N} F(x, \xi_i) \right\}. \tag{4.3}$$

As above, we say that a point $\bar{x} \in \mathcal{X}$ is an ε-stationary point of problem (4.3) if it satisfies the equation

$$0 \in \partial^{\varepsilon} \hat{f}_N(x) + \mathcal{N}_{\mathcal{X}}(x). \tag{4.4}$$

We denote by Σ^{ε} and $\hat{\Sigma}_N^{\varepsilon}$ the sets of ε-stationary points of problems (4.1) and (4.3), respectively. Note that (4.2) is a necessary condition for a point $x \in \mathcal{X}$ to be a locally optimal solution of problem (4.1). Therefore the set Σ^{ε} is nonempty provided that $f(x)$ is continuous and $\mathcal{X}$ is nonempty and compact. Similar remark applies to the set $\hat{\Sigma}_N^{\varepsilon}$ as well.

Theorem 7. *Let ξ_i be an iid sequence of random vectors supported on a set $\Xi \subset \mathbb{R}^d$, $F : \mathbb{R}^n \times \Xi \to \mathbb{R}$ be a Carathéodory function, $\mathcal{X}$ be a nonempty convex compact subset of $\mathbb{R}^n$ and ε_N be a sequence of nonnegative numbers converging to zero. Suppose that assumptions* (i) *and* (ii) *of Theorem* 5 *hold. Then*

$$\mathbb{D}\big(\hat{\Sigma}_N^{\varepsilon_N}, \Sigma^0\big) \to 0 \quad \text{w.p. } 1 \text{ as } N \to \infty. \tag{4.5}$$

Proof. By using compactness of $\mathcal{X}$ and the uniform convergence result (3.3), it is quite straightforward to prove that for any $\epsilon > 0$, it holds that $\mathbb{D}(\hat{\Sigma}_N^{\varepsilon_N}, \Sigma^{\epsilon}) \to 0$ w.p. 1 as $N \to \infty$ (see, e.g., [8, Section 7.1]). Also by the upper semicontinuity of the multifunction $x \mapsto \partial f(x)$ we have that

$$\bigcap_{\epsilon > 0} \big(\partial^{\epsilon} f(x) + \mathcal{N}_{\mathcal{X}}(x) \big) = \partial f(x) + \mathcal{N}_{\mathcal{X}}(x).$$

Because of the compactness of $\mathcal{X}$, these two properties imply (4.5). □

The above theorem shows that w.p. 1 stationary points of the sample average approximation problem (4.3) converge to the set of stationary points of the true problem (4.1) as sample size tends to infinity. This result is particularly interesting in the context of nonsmooth nonconvex stochastic programming where a stationary point, rather than a local minimizer, is more likely to be obtained. Of course, in case $F(\cdot, \xi)$ is convex w.p. 1, problems (4.1) and (4.3) are convex and the sets $\hat{\Sigma}_N^0$ and Σ^0 do coincide with the sets of optimal solutions of the respective problems.

References

[1] Z. Artstein, R.A. Vitale, A strong law of large numbers for random compact sets, Ann. Probab. 3 (1975) 879–882.

[2] Z. Artstein, J.C. Hansen, Convexification in limit law of random sets in Banach spaces, Ann. Probab. 13 (1985) 307–309.

[3] R.J. Aumann, Integrals of set-valued functions, J. Math. Anal. Appl. 12 (1965) 1–12.

[4] F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley, New York, 1983.

[5] R.T. Rockafellar, Variational Analysis, Springer-Verlag, New York, 1998.

[6] 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.

[7] A. Shapiro, Asymptotic properties of statistical estimators in stochastic programming, Ann. Statist. 17 (1989) 841–858.

[8] A. Shapiro, Monte Carlo sampling methods, in: A. Rusczyński, A. Shapiro (Eds.), Stochastic Programming, in: Handbooks Oper. Res. Management Sci., vol. 10, North-Holland, 2003.