Stochastic Nash Equilibrium Problems: Sample Average Approximation and Applications

Huifu Xu[1] and Dali Zhang
School of Mathematics
University of Southampton
Highfield Southampton, UK

December 19, 2008

Abstract

This paper presents a Nash equilibrium model where the underlying objective functions involve uncertainties and nonsmoothness. The well known sample average approximation method is applied to solve the problem and the first order equilibrium conditions are characterized in terms of Clarke generalized gradients. Under some moderate conditions, it is shown that with probability one, a statistical estimator obtained from sample average approximate equilibrium problem converges to its true counterpart. Moreover, under some calmness conditions of the generalized gradients and metric regularity of the set-valued mappings which characterize the first order equilibrium conditions, it is shown that with probability approaching one exponentially fast with the increase of sample size, the statistical estimator converge to its true counterparts. Finally, the model is applied to an equilibrium problem in electricity market.

Key words. Stochastic Nash equilibrium, exponential convergence, H-calmness, Clarke generalized gradients, metric regularity.

1 Introduction

Let $X_i \subset \mathbb{R}^{n_i}$, $i = 1, \cdots, \hat{i}$, be a closed convex subset of $\mathbb{R}^{n_i}$, where $\hat{i}$ and n_i are positive integers. Let $X_{-i} = X_1 \times \cdots \times X_{i-1} \times X_{i+1} \times \cdots \times X_{\hat{i}}$. We consider the following *stochastic Nash equilibrium* problem: find $(x_1^*, \cdots, x_{\hat{i}}^*) \in X_1 \times \cdots \times X_{\hat{i}}$ such that

$$\vartheta_i(x_i^*, x_{-i}^*) = \min_{x_i \in X_i} \mathbb{E}[v_i(x_i, x_{-i}^*, \xi)], \quad \text{for } i = 1, \cdots, \hat{i}, \tag{1.1}$$

where $x_{-i} = (x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_{\hat{i}}) \in X_{-i}$, $v_i(\cdot, x_{-i}, \xi) : \mathbb{R}^{n_i} \to \mathbb{R}$ is Lipschitz continuous, $\xi : \Omega \to \Xi \subset \mathbb{R}^k$ is a random vector defined on probability space $(\Omega, \mathcal{F}, P)$, $\mathbb{E}$ denotes the mathematical expectation. We make a blanket assumption that $\mathbb{E}[v_i(x_i, x_{-i}, \xi)]$ is well defined for all $x_i \in X_i$ and $x_{-i} \in X_{-i}$, $i = 1, \cdots, \hat{i}$ and an equilibrium of (1.1) exists.

Nash equilibrium models have been well studied and have found many interesting applications in economics and engineering. Our Nash equilibrium model has two specific features: one is that the underlying functions involve some random variables, the other is that these functions are not necessarily continuously differentiable with respect to the decision variables. The model reflects the stochastic nature and/or possible nonsmoothness in some practical Nash equilibrium

[1]Email: h.xu@soton.ac.uk.

problems such as multiple leader stochastic Stackelberg Nash-Cournot models for future market competition [10], stochastic equilibrium program with equilibrium (SEPEC) models for electricity markets [15, 51], Nash equilibrium model in transportation [45] and signal transmission in wireless networks [25]. Note that the SEPEC models in [10, 15, 51] are *two* stage Nash equilibrium problems and they can be reformulated as (1.1) only when the second stage equilibrium can be explicitly or implicitly represented by the variables at the first stage. In general circumstances, such reformulation may not be possible or desirable particularly when the second stage problem has multiple equilibria. A promising approach proposed by Henrion and Römisch [15] is to look into the M-stationary point of the two stage SEPEC model, see [15] for details.

In this paper, we are concerned with the numerical methods for solving (1.1). In particular, we deal with the complications resulting from the randomness and nonsmoothness. Note that if one knows the distribution of ξ and can integrate out the expected value $\mathbb{E}[v_i(x_i, x_{-i}, \xi)]$ explicitly, then the problem becomes a deterministic minimization problem. Throughout this paper, we assume that $\mathbb{E}[v_i(x_i, x_{-i}, \xi)]$ cannot be calculated in a closed form so that we will have to approximate it through discretization.

One of the best known discretization approaches is Monte Carlo simulation based method. The basic idea of the method is to generate an independent identically distributed (i.i.d.) sample $\xi^1, \cdots, \xi^N$ of ξ and then approximate the expected value with sample average. In this context, the objective function of (1.1) is approximated by

$$\vartheta_i^N(x_i, x_{-i}) := \frac{1}{N}\sum_{k=1}^{N} v_i(x_i, x_{-i}, \xi^k)$$

for $i = 1, \cdots, \hat{\imath}$, and consequently we may consider the following sample average approximate Nash equilibrium problem: find $x^N := (x_1^N, \cdots, x_{\hat{\imath}}^N) \in X_1 \times \cdots \times X_{\hat{\imath}}$ such that

$$\vartheta_i^N(x_i^N, x_{-i}^N) = \min_{x_i \in X_i} \vartheta_i^N(x_i, x_{-i}^N), \text{ for } i = 1, \cdots, \hat{\imath}. \tag{1.2}$$

We refer to (1.1) as *true problem* and (1.2) as *Sample Average Approximation (SAA) problem*. Naturally we will use x^N as a statistical estimator of its true counterpart. SAA is a very popular method in stochastic programming. It is also known as Sample Path Optimization (SPO) method [26, 30]. There has been extensive literature on SAA and SPO. See recent work [3, 20, 26, 30, 32, 41, 8, 23, 48] and a comprehensive review by Shapiro in [40].

Our focus here is on the convergence (also known as asymptotic consistency in some references) of x^N to its true counterpart as the sample size N increases. There are essentially two ways to carry out the analysis: one is through convergence of function values, that is, the convergence of ϑ_i^N to $\mathbb{E}[v_i(x_i, x_{-i}, \xi)]$ as N tends to infinity. This approach has been widely used in SAA method for stochastic optimization problems. See [40] and references therein. The other is through the convergence of derivatives of ϑ_i^N, that is, by considering the first order equilibrium condition of (1.2). This approach has also been used recently in stochastic programming for analyzing convergence of stationary points of sample average optimization problems, see for instance [43, 48, 50]. There are also other ways such as epi-convergence where convergence of optimal values and solutions are investigated through the asymptotic consistency of epi-graphs of objective functions. See [20].

In this paper, we investigate the convergence of $\{x^N\}$ through the first order equilibrium condition of (1.1) rather than (1.1) itself because (1.1) involves $\hat{\imath}$ stochastic optimization problems where a decision variable of one problem becomes a parameter of another problem while the first order equilibrium conditions of (1.1) can be put under a unified framework of generalized equations where x_i, $i = 1, \cdots, \hat{i}$ are all treated as variables. The main disadvantage of this approach is that the first order equilibrium conditions may involve set-valued mappings when $v_i(x_i, x_{-i}, \xi)$ is not continuously differentiable in x_i. Note that when $v_i(x_i, x_{-i}, \xi)$, $i = 1, \cdots, \hat{i}$, is continuously differentiable with respect to x_i, the first order equilibrium condition of (1.2) reduces to a variational inequality problem or a nonlinear complementarity problem [10].

There are two types of convergence one may consider: almost sure convergence and exponential convergence. The former concerns whether or not the statistical estimator of an SAA problem converges to its true counterpart. This type of convergence is usually obtained by applying classical uniform strong law of large numbers (SLLN) ([34, Lemma A1]) to the underlying functions which define the statistical estimator. The uniform SLLN requires the random functions to be Lipschitz continuous. More recently, the classical uniform SLLN has been extended to random upper semi-continuous random compact set-valued mappings [43, Theorem 1]. The extension allows one to analyze statistical estimators defined by set-valued mappings, e.g. stationary points characterized by Clarke generalized gradients in stochastic nonsmooth optimization. See [43, 50].

Almost sure convergence does not tell us how fast the convergence is and exponential convergence addresses this. To obtain a rate of convergence, one may use the well known Cramer's theorem in large deviation theory [9] to investigate the probability of the deviation of a statistical estimator from its true counterpart as sample size increases and show that the probability goes to zero at exponential rate of sample size. Over the past few years, various exponential convergence results have been established for sample average approximate optimization problems and the focuses are largely on optimal solutions and/or optimal values. See for instances [21, 32, 41, 40, 42] and the references therein. Similar to almost sure convergence, exponential convergence of a statistical estimator in stochastic programming is usually obtained by the uniform exponential convergence of the underlying functions which define the estimator. It also requires some additional sensitivity conditions which ensure the deviation of a statistical estimator is bounded by that of the underlying functions defining it, see for instance, second order growth condition in [38, 32].

In this paper, the underlying functions of (1.1) and (1.2) are not necessarily continuously differentiable and consequently their first order equilibrium conditions have to be characterized in terms of generalized gradients. We investigate both almost sure convergence and exponential convergence of $\{x^N\}$ through the first order equilibrium conditions. The former can be readily obtained by applying the uniform SLLN [43]. The latter is more challenging and it is indeed what this paper is mainly focused on. The main challenges and complications arise from the necessity to establish exponential convergence of sample average of the generalized gradients which characterizes $\{x^N\}$ in the first order equilibrium conditions.

The key steps we take to tackle the challenges and complications are as follows: we use metric regularity to obtain a bound of deviation of x^N to its true counterpart by the deviation of the underlying sample average set-valued mappings from its expectation, and then transform the latter into the maximum difference of its support functions over a unit ball; finally we extend the uniform exponential convergence of Hölder continuous random functions established by Shapiro and Xu [42, Theorem 5.1] to H-calm (from above or below) functions [31] which accommodate discontinuity at some points and apply the extended results to the support functions of sample average set-valued mappings which characterize the first order Nash equilibrium condition.

As far as we are concerned, the main contributions of this paper can be summarized as follows: we propose a general nonsmooth stochastic Nash equilibrium model and apply the well known sample average approximation method to solve it; we establish almost sure convergence and exponential convergence of Nash equilibrium estimator obtained from the sample average Nash equilibrium problem. Finally we model the competition in electricity spot market as a stochastic nonsmooth Nash equilibrium problem and use a smoothing SAA method to solve it.

2 Preliminaries

Throughout this paper, we use the following notation. $x^T y$ denotes the scalar products of two vectors x and y, $\|\cdot\|$ denotes the Euclidean norm of a vector and a compact set of vectors. If D is a compact set of vectors, then

$$\|D\| := \max_{x\in D} \|x\|.$$

$d(x, D) := \inf_{x'\in D} \|x - x'\|$ denotes the distance from point x to set D. For two compact sets D_1 and D_2,

$$\mathbb{D}(D_1, D_2) := \sup_{x\in D_1} d(x, D_2)$$

denotes the deviation from set D_1 to set D_2 (in some references [14] it is also called *excess* of D_1 over D_2), and $\mathbb{H}(D_1, D_2)$ denotes the Hausdorff distance between the two sets, that is,

$$\mathbb{H}(D_1, D_2) := \max\left(\mathbb{D}(D_1, D_2), \mathbb{D}(D_1, D_2)\right).$$

We use $D_1 + D_2$ to denote the Minkowski addition of D_1 and D_2, that is, $D_1 + D_2 = \{x + y : x \in D_1, y \in D_2\}$. We use $B(x, \delta)$ to denote the closed ball with radius δ and center x, that is $B(x, \delta) := \{x' : \|x' - x\| \le \delta\}$. When δ is dropped, $B(x)$ represents a neighborhood of point x. Finally we use $\mathcal{B}$ to denote the unit ball in a finite dimensional space. Finally, for a closed convex set D, we use $\mathcal{N}_C(x)$ to denote the normal cone of D at x, that is,

$$\mathcal{N}_D(x) := \left\{z \in \mathbb{R}^m : z^T(x' - x) \le 0, \ \ \forall x' \in D\right\}, \ \text{ if } x \in D.$$

For a closed set D in $\mathbb{R}^m$, the *support function* of D is defined as

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

for every $u \in \mathbb{R}^m$. The following results are known as Hömander's formulae.

Proposition 2.1 *([6, Theorem II-18]) Let D_1, D_2 be two 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*

$$\mathbb{D}(D_1, D_2) = \max_{\|u\|\le 1} \left(\sigma(D_1, u) - \sigma(D_2, u)\right)$$

and

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

2.1 Set-valued mappings

Let $\mathcal{X}$ be a closed subset of $\mathbb{R}^n$. Recall that a set-valued mapping $F : \mathcal{X} \to 2^{\mathbb{R}^m}$ is said to be *closed* at $x \in \mathcal{X}$ if for $x_k \subset \mathcal{X}$, $x_k \to x$, $y_k \in F(x_k)$ and $y_k \to \bar{y}$ implies $\bar{y} \in F(\bar{y})$. F is said to be *uniformly compact* near $\bar{x} \in \mathcal{X}$ if there is a neighborhood $B(\bar{x})$ of $\bar{x}$ such that the closure of $\bigcup_{x \in B(\bar{x})} F(x)$ is compact. F is said to be *upper semi-continuous* at $\bar{x} \in \mathcal{X}$ if for every $\epsilon > 0$, there exists a $\delta > 0$ such that

$$F(\bar{x} + \delta\mathcal{B}) \subset F(\bar{x}) + \epsilon\mathcal{B}.$$

The following result was established by Hogan [16].

Proposition 2.2 *Let $F : \mathcal{X} \to 2^{\mathbb{R}^m}$ be uniformly compact near $\bar{x}$. Then F is upper semi-continuous at $\bar{x}$ if and only if F is closed.*

The proposition is useful because in some cases we need to establish upper semi-continuity of a set-valued mapping which is either closed or uniformly compact.

Definition 2.1 *Let $F : \mathcal{X} \to 2^{\mathbb{R}^m}$ be a closed set valued mapping. For $\bar{x} \in \mathcal{X}$ and $\bar{y} \in F(\bar{x})$, F is said to be* metrically regular *at $\bar{x}$ for $\bar{y}$ if there exists a constant $\alpha > 0$ such that*

$$d(x, F^{-1}(y)) \leq \alpha d(y, F(x)) \text{ for all } (x, y) \text{ close to} (\bar{x}, \bar{y}).$$

Here the inverse mapping F^{-1} is defined as $F^{-1}(y) = \{x \in \mathcal{X} : y \in F(x)\}$ and the minimal constant $\alpha < \infty$ which makes the above inequality hold is called regularity modulus.

Metric regularity is a generalization of Jacobian nonsingularity of a (vector valued) function to set-valued mappings [29]. The property is equivalent to nonsingularity of the Mordukhovich coderivative of F at $\bar{x}$ for $\bar{y}$ and to Aubin's property of F^{-1}. For a comprehensive discussion of the history and recent development of the notion, see [12, 31] and references therein. Using the notion of metric regularity, we can analyze the sensitivity of generalized equations.

Proposition 2.3 *Let $F, G : \mathcal{X} \to 2^{\mathbb{R}^m}$ be two set valued mappings. Let $\bar{x} \in \mathcal{X}$ and $0 \in F(\bar{x})$. Let $0 \in G(x)$ with x being close to $\bar{x}$. Suppose that F is metrically regular at $\bar{x}$ for 0. Then*

$$d\left(x, F^{-1}(0)\right) \leq \alpha \mathbb{D}(G(x), F(x)), \tag{2.3}$$

where α is the regularity modulus of F at $\bar{x}$ for 0.

Proof. Since F is metrically regular at $\bar{x}$ for 0, there exists a constant $\alpha > 0$ such that $d(x, F^{-1}(0)) \leq \alpha d(0, F(x))$. Since $0 \in G(x)$, then by the definition of $\mathbb{D}$, $d(0, F(x)) \leq \mathbb{D}(G(x), F(x))$. ■

The above result can be explained as follows. Suppose we want to solve generalized equation $0 \in F(x)$. We do so by solving an approximate equation $0 \in G(x)$ where G is an approximation of F, and obtaining a solution x for the approximate equation. Suppose also that x is close to a true solution $\bar{x} \in F^{-1}(0)$ and F is metrically regular at $\bar{x}$, then the deviation of x from $F^{-1}(0)$ is bounded by the deviation of $G(x)$ from $F(x)$. This type of error bound is numerically useful because we may use $\mathbb{D}(G(x), F(x))$ to estimate $d(x, F^{-1}(0))$.

2.2 Expectation of random set-valued mappings

Consider now a random set-valued mapping $F(\cdot, \xi(\cdot)) : \mathcal{X} \times \Omega \to 2^{\mathbb{R}^n}$ (we are slightly abusing the notation F) where $\mathcal{X}$ is a closed subset of $\mathbb{R}^n$ and ξ is a random vector defined on probability space $(\Omega, \mathcal{F}, P)$. Let $x \in \mathcal{X}$ be fixed and consider the measurability of set-valued mapping $F(x, \xi(\cdot)) : \Omega \to 2^{\mathbb{R}^n}$. Let $\mathfrak{B}$ denote the space of nonempty, closed subsets of $\mathbb{R}^n$. Then $F(x, \xi(\cdot))$ can be viewed as a single valued mapping from Ω to $\mathfrak{B}$. Using [31, Theorem 14.4], we know that $F(x, \xi(\cdot))$ is measurable if and only if for every $B \in \mathfrak{B}$, $F(x, \xi(\cdot))^{-1}B$ is $\mathcal{F}$-measurable.

Recall that $A(x, \xi(\omega)) \in F(x, \xi(\omega))$ is said to be a *measurable selection* of the random set $\mathcal{A}(x, \xi(\omega))$, if $A(x, \xi(\omega))$ is measurable. Measurable selections exist, see [2] and references therein. The *expectation of* $F(x, \xi(\omega))$, denoted by $\mathbb{E}[F(x, \xi(\omega))]$, is defined as the collection of $\mathbb{E}[A(x, \xi(\omega))]$, where $A(x, \xi(\omega))$ is an integrable measurable selection. The expected value is also known as Aumann's integral [14] as it was first studied comprehensively by Aumann in [5]. $\mathbb{E}[F(x, \xi(\omega))]$ is regarded as well defined if $\mathbb{E}[F(x, \xi(\omega))] \in \mathfrak{B}$ is nonempty. A sufficient condition of this is $\mathbb{E}[\|F(x, \xi(\omega))\|] := \mathbb{E}[\mathbb{H}(0, F(x, \xi(\omega)))] < \infty$, see [2]. In such a case, F is said to be *integrably bounded* [5, 14].

2.3 Clarke generalized gradients of a random function

We are interested in the cases when the integrand functions in (1.1) Lipschitz continuous. Let $f(x, \xi(\cdot)) : \mathbb{R}^n \times \Omega \to \mathbb{R}$ be a random function that is locally Lipschitz continuous with respect to x, let ξ be a realization of $\xi(\omega)$. The Clarke generalized gradient [7] of $f(x, \xi)$ with respect to x at point $x \in \mathbb{R}^n$ is defined as

$$\partial_x f(x, \xi) := \text{conv} \left\{ \lim_{\substack{y \in D_{f(\cdot, \xi)} \\ y \to x}} \nabla_x f(y, \xi) \right\},$$

where $D_{f(\cdot, \xi)}$ denotes the set of points near x where $f(x, \xi)$ is Frechét differentiable with respect to x, $\nabla_x f(y, \xi)$ denotes the usual gradient of $f(x, \xi)$ in x and 'conv' denotes the convex hull of a set. It is well known that the Clarke generalized gradient $\partial_x f(x, \xi)$ is a convex compact set and it is upper semicontinuous [7, Proposition 2.1.2 and 2.1.5]. When $f(\cdot, \xi)$ is continuously differentiable at x, $\partial_x f(x, \xi)$ coincides with $\nabla_x f(x, \xi)$.

3 First order equilibrium conditions

In this section, we discuss the first order equilibrium conditions of stochastic Nash equilibrium problem (1.1) and its sample average approximation (1.2) in terms of Clarke generalized gradient. For $i = 1, \cdots, \hat{i}$, assume that the Lipschitz modulus of $v_i(x_i, x_{-i}, \xi)$ with x_i is integrable for every $x_{-i} \in X_{-i}$. It is well known [36] that $\mathbb{E}[v_i(x_i, x_{-i}, \xi)]$ is also Lipschitz continuous in x_i and hence the Clarke generalized gradient of $\mathbb{E}[v_i(x_i, x_{-i}, \xi)]$ in x_i, denoted by $\partial_{x_i} \mathbb{E}[v_i(x_i, x_{-i}, \xi)]$, is well defined. Consequently we can characterize the first order equilibrium condition of (1.1) at a Nash equilibrium in terms of the Clarke generalized gradients as follows:

$$0 \in \partial_{x_i} \mathbb{E}[v_i(x_i, x_{-i}, \xi)] + \mathcal{N}_{X_i}(x_i), \ i = 1, \cdots, \hat{i}. \tag{3.4}$$

Here and later on, the addition of sets is in the sense of Minkowski. We call a point x^* satisfying (3.4) a *stochastic C-Nash stationary point*. Obviously if x^* is stochastic Nash equilibrium, then it must satisfy (3.4) and hence it is a stochastic C-Nash stationary point, but not vice versa. However, if $\mathbb{E}[v_i(x_i, x_{-i}, \xi)]$ is convex in x_i for each i, then a C-Nash stationary point is a Nash equilibrium. Note that

$$\partial_{x_i}\mathbb{E}[v_i(x_i, x_{-i}, \xi)] \subset \mathbb{E}[\partial_{x_i} v_i(x_i, x_{-i}, \xi)].$$

The equality holds when v_i is Clarke regular [7] in x_i. See for instance [17, 46, 28]. Consequently, we may consider a weaker condition than (3.4)

$$0 \in \mathbb{E}[\partial_{x_i} v_i(x_i, x_{-i}, \xi)] + \mathcal{N}_{X_i}(x_i), \ i = 1, \cdots, \hat{\imath}. \tag{3.5}$$

We call (3.5) *weak* Clarke first order equilibrium condition of (1.1) and a point satisfying (3.5) a *weak stochastic C-Nash stationary point*. Here "weak" is in the sense that a stochastic C-stationary point is a weak stochastic C-Nash stationary point but not vice versa. When v_i, $i = 1, \cdots, \hat{\imath}$, is convex in x_i, these stationary points coincide with Nash equilibrium points.

Using the Clarke generalized gradient, we can also characterize the first order equilibrium condition of the sample average approximation equilibrium (1.2) as follows:

$$0 \in \frac{1}{N}\sum_{k=1}^{N} \partial_{x_i} v_i(x_i, x_{-i}, \xi^k) + \mathcal{N}_{X_i}(x_i), \ i = 1, \cdots, \hat{\imath}. \tag{3.6}$$

We call a point x^N satisfying (3.6) *SAA C-Nash stationary point*. In Section 4, we will investigate the convergence of x^N as sample size N increases and show under some appropriate conditions that w.p.1 an accumulation point of $\{x^N\}$ is a weak stochastic C-Nash stationary point. This is why we consider condition (3.5).

Next, we look into the upper semi-continuity of the set-valued mapping $\partial_{x_i} v_i(x_i, x_{-i}, \xi)$ with respect to (x_i, x_{-i}).

Assumption 3.1 *Let $v_i(x_i, x_{-i}, \xi)$, $i = 1, \cdots, \hat{\imath}$, be defined as in (1.1) and $\partial_{x_i} v_i(x_i, x_{-i}, \xi)$ be its Clarke generalized gradient with respect to x_i.*

(a) $v_i(\cdot, x_{-i}, \xi)$ is Lipchitz continuous on X_i with modulus $\kappa_i(\xi)$, where $\mathbb{E}[\kappa_i(\xi)] < \infty$.

(b) $\partial_{x_i} v_i(x_i, x_{-i}, \xi)$ is closed with respect to x_{-i} on space X_{-i}.

Corollary 3.1 *Under Assumption 3.1, $\partial_{x_i} v_i(x_i, x_{-i}, \xi)$ is upper semi-continuous with respect to both x_i and x_{-i}.*

Proof. Upper semi-continuity with respect to x_i follows from the property of the Clarke generalized gradient [7]. It suffices to show the upper semi-continuity with respect to x_{-i}. Under Assumption 3.1 (a), $\partial_{x_i} v_i(x_i, x_{-i}, \xi)$ is bounded by $\kappa_i(\xi)$ which gives rises to local compactness of the set valued mapping, and under Assumption 3.1 (b), $\partial_{x_i} v_i(x_i, x_{-i}, \xi)$ is closed in x_{-i}. Applying Proposition 2.2, we obtain the upper semi-continuity of $\partial_{x_i} v_i(x_i, x_{-i}, \xi)$ in x_{-i} as desired. ■

Finally, we look into the measurability and integrability of the set-valued mapping $\partial_{x_i} v_i(x_i, x_{-i}, \xi(\cdot)) : \Omega \to 2^{\mathbb{R}^m}$.

Proposition 3.1 *Let $\partial_{x_i} v_i(x_i, x_{-i}, \xi)$ be the Clarke generalized gradient of v_i with respect to x_i. Then*

(i) $\partial_{x_i} v_i(x_i, x_{-i}, \xi(\cdot)) : \Omega \to 2^{\mathbb{R}^m}$ is measurable;

(ii) under Assumption 3.1 (a), $\mathbb{E}[\partial_{x_i} v_i(x_i, x_{-i}, \xi)]$ is well defined.

Proof. Part (i). Let $d_i \in \mathbb{R}^{n_i}$ be fixed. By definition, the Clarke generalized derivative [7] of $v_i(x_i, x_{-i}, \xi)$ with respect to x_i at a point x_i in direction d_i is defined as

$$(v_i)^o_{x_i}(x_i, x_{-i}, \xi; d_i) := \limsup_{\substack{y_i \to x_i \\ t \to 0}} [v_i(y_i + td_i, x_{-i}, \xi) - v_i(y_i, x_{-i}, \xi)]/t.$$

Since v_i is continuous in ξ and $\xi(\omega)$ is a random vector, then v_i is measurable, and by [4, Lemma 8.2.12], $(v_i)^o_{x_i}(x_i, x_{-i}, \xi; d_i)$ is also measurable. Since $(v_i)^o_{x_i}(x_i, x_{-i}, \xi; d_i)$ is the support function of $\partial_{x_i} v_i(x_i, x_{-i}, \xi)$, by [4, Theorem 8.2.14], $\partial_{x_i} v_i(x_i, x_{-i}, \xi)$ is measurable. Part (ii). Assumption 3.1 (a) indicates that $\partial_{x_i} v_i(x_i, x_{-i}, \xi)$ is integrably bounded. Together with the measurability as proved in Part (i), this gives the well definedness of $\mathbb{E}[\partial_{x_i} v_i(x_i, x_{-i}, \xi)]$. ■

4 Convergence analysis

In this section, we analyze convergence of a sequence of SAA Nash stationary points $\{x^N\}$ defined by (3.6). The analysis is carried out in two steps. First, we show almost sure convergence, that is, w.p.1, an accumulation point of $\{x^N\}$ satisfies (3.4). Second, under additional condition, namely H-calmness, of the generalized gradient $\partial_{x_i} v_i(x_i, x_{-i}, \xi)$, we show that with probability approaching one exponentially fast with the increase of sample size N, $\{x^N\}$ converges to a weak Nash stationary point.

For the simplicity of notation, we denote throughout this section $\partial_{x_i} v_i(x_i, x_{-i}, \xi)$ by $\partial_{x_i} v_i(x, \xi)$. This will not cause a confusion because both x_i and x_{-i} are treated as variables in the analysis. Let

$$\mathcal{A}v(x, \xi) := \partial_{x_1} v_1(x, \xi) \times \cdots \times \partial_{x_{\hat{i}}} v_{\hat{i}}(x, \xi) \tag{4.7}$$

and

$$G_X(x) := \mathcal{N}_{X_1}(x_1) \times \cdots \times \mathcal{N}_{X_{\hat{i}}}(x_{\hat{i}}). \tag{4.8}$$

The first order equilibrium condition (3.4) can be written as

$$0 \in \mathbb{E}[\mathcal{A}v(x, \xi)] + G_X(x), \tag{4.9}$$

where

$$\mathbb{E}[\mathcal{A}v(x, \xi)] := \mathbb{E}[\partial_{x_1} v_1(x, \xi)] \times \cdots \times \mathbb{E}[\partial_{x_{\hat{i}}} v_{\hat{i}}(x, \xi)].$$

By Proposition 3.1, $\mathbb{E}[\mathcal{A}v(x, \xi)]$ is well defined. Likewise, the first order equilibrium condition (3.6) can be written as

$$0 \in \mathcal{A}\vartheta^N(x) + G_X(x), \tag{4.10}$$

where

$$\mathcal{A}\vartheta^N(x) := \frac{1}{N}\sum_{k=1}^{N}\partial_{x_1}v_1(x,\xi^k)\times\cdots\times\frac{1}{N}\sum_{k=1}^{N}\partial_{x_{\hat{i}}}v_{\hat{i}}(x,\xi^k). \tag{4.11}$$

Obviously, $\mathbb{E}[\mathcal{A}\vartheta^N(x)] = \mathbb{E}[\mathcal{A}v(x,\xi)]$. For the simplicity of discussion, we make a blanket assumption that (3.6) has a solution for all N. Further discussion on the existence issue requires sensitivity analysis of generalized equation (4.9). Note that in deterministic case, King and Rockafellar [18] investigated the existence of solution to a perturbed generalized equation when the original equation has a solution under subinvertibility of set-valued mappings. We refer the interested readers to [18] for details.

4.1 Almost sure convergence

Our idea to obtain almost sure convergence is to apply the uniform SLLN for sample average random set-valued mapping which is recently established by Shapiro and Xu [43] to set-valued mapping $\mathcal{A}v(x,\xi)$. This approach has been used in nonsmooth stochastic optimization in [43, 50]. Here we use the approach to an equilibrium problem.

Theorem 4.1 *Let x^N be a solution of (3.6) and Assumption 3.1 hold. Assume that w.p.1 the sequence $\{x^N\}$ is contained in a compact subset $\mathcal{X}$ of X. Then w.p.1, an accumulation point of $\{x^N\}$ satisfies (3.4).*

Proof. Under Assumption 3.1, $\mathcal{A}v(\cdot,\xi)$ is upper semi-continuous on X and

$$\|\mathcal{A}v(x,\xi)\| \leq \sum_{i=1}^{\hat{i}}\kappa_i(\xi),$$

where $\mathbb{E}[\sum_{i=1}^{\hat{i}}\kappa_i(\xi)] < \infty$. Let x^* be a stationary point of $\{x^N\}$. By applying [43, Theorem 2] and [50, Theorem 4.3] on a compact set $\mathcal{X}$, we know that w.p.1 x^* satisfies (4.9). ■

Theorem 4.1 states that if $\{x^N\}$ is a sequence of SAA C-stationary points of problem (1.2), then w.p.1 its accumulation point is a weak C-Nash stationary point of the true problem. In some cases, one may be able to obtain a Nash equilibrium in solving SAA problem, that is, x^N is a Nash equilibrium of (1.2). Consequently we may want to know whether an accumulation point of $\{x^N\}$ is a Nash equilibrium of the true problem (1.1). The following theorem addresses this.

Theorem 4.2 *Assume the setting and conditions of Theorem 4.1. Let $\{x^N\}$ be a sequence of Nash equilibria of the SAA problem (1.2). Then w.p.1 an accumulation point of $\{x^N\}$ is a Nash equilibrium of the true problem (1.1) if one of the following conditions hold:*

(a) $v_i(x_i, x_{-i}, \xi)$, $i = 1, \cdots, \hat{i}$, *is convex w.r.t.* x_i;

(b) $v_i(x_i, x_{-i}, \xi)$, $i = 1, \cdots, \hat{i}$, *is Lipschitz continuous w.r.t.* x_{-i} *on* X_i *with Lipschitz modulus* $\kappa_{-i}(\xi)$ *where* $\mathbb{E}[\kappa_{-i}(\xi)] < \infty$.

Proof. Under condition (a), the set of weak C-Nash stationary points of the true problem coincides with the set of its Nash equilibria. In what follows, we prove the conclusion under condition (b). Let

$$\rho(y, x) := \sum_{i=1}^{\hat{i}} \vartheta_i(y_i, x_{-i})$$

and

$$\hat{\rho}^N(y, x) := \sum_{i=1}^{\hat{i}} \hat{\vartheta}_i^N(y_i, x_{-i}).$$

It is well known (see e.g. [33]) that $x^* \in X$ is a Nash equilibrium of the true problem (1.1) if and only if x^* solves the following minimization problem

$$\min_{y \in X} \rho(y, x^*).$$

Similarly $x^N \in X$ is a Nash equilibrium of the SAA problem (1.2) if and only if x^N solves the following minimization problem

$$\min_{y \in X} \hat{\rho}^N(y, x^N).$$

Assume without loss of generality (by taking a subsequence if necessary) that $\{x^N\}$ converges to x^* w.p.1. In what follows, we show that w.p.1 $\hat{\rho}^N(y, x^N)$ converges to $\rho(y, x^*)$ uniformly w.r.t y. Let us consider

$$\hat{\rho}^N(y, x^N) - \rho(y, x^*) = \hat{\rho}^N(y, x^N) - \hat{\rho}^N(y, x^*) + \hat{\rho}^N(y, x^*) - \rho(y, x^*).$$

Since $v_i(x_i, x_{-i}, \xi)$ is Lipschitz w.r.t. x_{-i} with modulus $\kappa_{-i}(\xi)$, we have

$$\begin{aligned} |\hat{\rho}^N(y, x^N) - \hat{\rho}^N(y, x^*)| &\leq \sum_{i=1}^{\hat{i}} \frac{1}{N} \sum_{j=1}^{N} |v_i(y_i, x_{-i}^N, \xi^j) - v_i(y_i, x_{-i}^*, \xi^j)| \\ &\leq \sum_{i=1}^{\hat{i}} \frac{1}{N} \sum_{j=1}^{N} \kappa_{-i}(\xi^j) \|x^N - x^*\|. \end{aligned}$$

The last term tends to 0 uniformly w.r.t. y when $N \to \infty$ because $\frac{1}{N}\sum_{j=1}^{N} \kappa_{-i}(\xi^j) \to \mathbb{E}[\kappa_{-i}(\xi)] < \infty$. In the same manner, we can show that $\hat{\rho}^N(y, x^*) - \rho(y, x^*) \to 0$ uniformly w.r.t. y w.p.1 as $N \to \infty$. This shows w.p.1 $\hat{\rho}^N(y, x^N)$ converges to $\rho(y, x^*)$ uniformly w.r.t. y. It is well known that the uniform convergence implies that the limit of the global minimizer of $\hat{\rho}^N(y, x^N)$ over compact set X is a global minimizer of $\rho(y, x^*)$ over X (hence a Nash equilibrium of the true problem), see for instance [34, Theorem A1][2]. ■

4.2 Exponential convergence

Next we discuss the exponential convergence of $\{x^N\}$ as N goes to infinity. We do so in three steps. First, we extend Shapiro and Xu's uniform exponential convergence results ([42, Theorem 5.1]) to a class of random semi-continuous functions that are H-calm (from above or below).

[2] In the theorem, the convergence of $\bar{v}_N$ to v^* was proved under the condition that v^* is a unique global minimizer of $l(v)$ but the conclusion can be easily extended to the case when $l(v)$ has multiple minimizers in which case one can prove that $d(\bar{v}^N, V^*) \to 0$ where V^* denotes the set of global minimizers of $l(v)$.

Second, we show uniform exponential convergence of $\mathbb{D}(\mathcal{A}\vartheta^N(x), \mathbb{E}[\mathcal{A}v(x,\xi)])$ when $\mathcal{A}v$ is upper hemi-continuous and H-calm from above. We do so by reformulating $\mathbb{D}(\mathcal{A}\vartheta^N(x), \mathbb{E}[\mathcal{A}v(x,\xi)])$ as the difference of the support functions of $\mathcal{A}\vartheta^N(x)$ and $\mathbb{E}[\mathcal{A}v(x,\xi)]$ using the well known Hörmander's formulae, Lemma 2.1, and then applying uniform exponential convergence established in the first step to the sample average of the support functions. Finally, we obtain an error bound for $\|x^N - x^*\|$ in terms of $\mathbb{D}(\mathcal{A}\vartheta^N(x), \mathbb{E}[\mathcal{A}v(x,\xi)])$ under some metric regularity of $\mathbb{E}[\mathcal{A}v]$ at x^*, where x^* is a weak KKT equilibrium of the true problem, and subsequently the exponential convergence of $\|x^N - x^*\|$.

Definition 4.1 *Let $\phi : \mathbb{R}^n \times \Xi \to \mathbb{R}$ be a real valued function and $\xi : \Omega \to \Xi \subset \mathbb{R}^k$ a random vector defined on probability space $(\Omega, \mathcal{F}, P)$, let $\mathcal{X} \subset \mathbb{R}^n$ be a subset of $\mathbb{R}^n$ and $x \in \mathcal{X}$. ϕ is said to be*

- H-calm at x from above *with modulus $\kappa(\xi)$ and order γ if $\phi(x,\xi)$ is finite and there exists a (measurable) function $\kappa : \Xi \to \mathbb{R}_+$, positive numbers γ and δ such that*

$$\phi(x',\xi) - \phi(x,\xi) \leq \kappa(\xi)\|x' - x\|^\gamma \tag{4.12}$$

  for all $x' \in \mathcal{X}$ with $\|x' - x\| \leq \delta$ and $\xi \in \Xi$;

- H-calm at x from below *with modulus $\kappa(\xi)$ and order γ if $\phi(x,\xi)$ is finite and there exists a (measurable) function $\kappa : \Xi \to \mathbb{R}_+$, positive numbers γ and δ such that*

$$\phi(x',\xi) - \phi(x,\xi) \geq -\kappa(\xi)\|x' - x\|^\gamma \tag{4.13}$$

  for all $x' \in \mathcal{X}$ with $\|x' - x\| \leq \delta$ and $\xi \in \Xi$;

- H-calm at x *with modulus $\kappa(\xi)$ and order γ if $\phi(x,\xi)$ is finite and there exists a (measurable) function $\kappa : \Xi \to \mathbb{R}_+$, positive numbers γ and δ such that*

$$|\phi(x',\xi) - \phi(x,\xi)| \leq \kappa(\xi)\|x' - x\|^\gamma \tag{4.14}$$

  for all $x' \in \mathcal{X}$ with $\|x' - x\| \leq \delta$ and $\xi \in \Xi$.

ϕ is said to be H-calm from above, calm from below, calm on set $\mathcal{X}$ if the respective properties stated above hold at every point of $\mathcal{X}$.

Calmness of a deterministic real valued function is well known. See for instance [31, Page 322]. The property is a generalization of Lipschitz continuity, that is, a locally Lipschitz continuous function is calm but the converse is not necessarily true, see discussions in [31, Page 350-352]. Our definition is slightly different the calmness in [31] in that we allow a nonlinear growth bound and therefore we use term "H-calmness" to indicate that the property is a generalization of Hölder continuity. Note that γ is *not* restricted to positive values between 0 and 1, instead, it may take any positive values.

In what follows, we discuss uniform exponential convergence of sample average random function $\phi(x,\xi)$ under H-calmness. Let $\xi^1, ..., \xi^N$ be an iid sample of the random vector $\xi(\omega)$. We consider the sample average function

$$\psi_N(x) := \frac{1}{N}\sum_{k=1}^{N} \phi(x, \xi^k).$$

Let $\psi(x) = \mathbb{E}[\phi(x,\xi)]$. We use the large deviation theorem to investigate the probability of $\psi_N(x)$ deviating from $\psi(x)$ over a compact $\mathcal{X} \subset \mathbb{R}^n$ as sample size N increases. Let

$$M_x(t) := \mathbb{E}\left\{e^{t[\phi(x,\xi(\omega))-\psi(x)]}\right\}$$

denote the moment generating function of the random variable $\phi(x,\xi(\omega)) - \psi(x)$. We make the following assumption.

Assumption 4.1 *Let $\phi : \mathbb{R}^n \times \Xi \to \mathbb{R}$ be a random function and ξ be a random vector, let $\mathcal{X} \subset \mathbb{R}^n$ be a compact subset of $\mathbb{R}^n$.*

(a) For every $x \in \mathcal{X}$, the moment generating function $M_x(t)$ is finite valued for all t in a neighborhood of zero.

(b) $\psi(x)$ is continuous on $\mathcal{X}$.

Assumption 4.1 (a) implies that the probability distribution of random variable $\phi(x,\xi)$ dies exponentially fast in the tails. In particular, it holds if this random variable has a distribution supported on a bounded subset of $\mathbb{R}$. See similar assumptions in [42]. Assumption 4.1 (b) holds when $\phi(x,\xi)$ is continuous w.p.1 and bounded by an integrable function. Comprehensive discussions on the continuity of the expectation of piecewise continuous random set-valued mappings (real-valued random function is just a special case) can be found in [28, Section 4].

Proposition 4.1 *Let $\phi : \mathbb{R}^n \times \Xi \to \mathbb{R}$ be a real valued function and ξ be a random vector, let $\mathcal{X} \subset \mathbb{R}^n$ be a compact subset of $\mathbb{R}^n$ and Assumption 4.1 hold.*

(i) If $\phi(\cdot,\xi)$ is H-calm from above on $\mathcal{X}$ with modulus $\kappa(\xi)$ and order γ and the moment generating function $\mathbb{E}\left[e^{\kappa(\xi)t}\right]$ is finite valued for t close to 0, then for every $\epsilon > 0$, there exist positive constants $c(\epsilon)$ and $\beta(\epsilon)$, independent of N, such that

$$\text{Prob}\left\{\sup_{x\in\mathcal{X}}(\psi_N(x) - \psi(x)) \geq \epsilon\right\} \leq c(\epsilon)e^{-N\beta(\epsilon)}. \tag{4.15}$$

(ii) If $\phi(\cdot,\xi)$ is H-calm from below on $\mathcal{X}$ with modulus $\kappa(\xi)$ and order γ and the moment generating function $\mathbb{E}\left[e^{\kappa(\xi)t}\right]$ is finite valued for t close to 0, then for every $\epsilon > 0$, there exist positive constants $c(\epsilon)$ and $\beta(\epsilon)$, independent of N, such that

$$\text{Prob}\left\{\inf_{x\in\mathcal{X}}(\psi_N(x) - \psi(x)) \leq -\epsilon\right\} \leq c(\epsilon)e^{-N\beta(\epsilon)}. \tag{4.16}$$

(iii) If $\phi(\cdot,\xi)$ is H-calm on $\mathcal{X}$ with modulus $\kappa(\xi)$ and order γ and the moment generating function $\mathbb{E}\left[e^{\kappa(\xi)t}\right]$ is finite valued for t close to 0, then for every $\epsilon > 0$, there exist positive constants $c(\epsilon)$ and $\beta(\epsilon)$, independent of N, such that

$$\text{Prob}\left\{\sup_{x\in\mathcal{X}}|\psi_N(x) - \psi(x)| \geq \epsilon\right\} \leq c(\epsilon)e^{-N\beta(\epsilon)}. \tag{4.17}$$

Part (i) is proved in a recent paper [28]. We include the proof in the appendix for complete as the paper has not been published yet. Part (ii) can be proved in a similar way to Part (i) and Part (iii) is a combination of Parts (i) and (ii).

Proposition 4.1 is a generalization of [42, Theorem 5.1] where an exponential convergence of the sample average of a random function is obtained under Assumption 4.1 and uniform Hölder continuity of $\psi(x)$ in x. The significance of the results here is that we extend the exponential convergence to a class of random functions which may be discontinuous at some points. The results can be easily used to establish exponential convergence of sample average approximation of stochastic optimization problems where the underlying functions are lower or upper semi-continuous. Our main purpose here, however, is to use Proposition 4.1 to establish the exponential convergence of random set-valued mappings $\mathcal{A}\vartheta^N(x)$ over a compact subset of X.

Let $F : \mathbb{R}^n \times \Xi \to 2^{\mathbb{R}^m}$ be a random set-valued mapping, $\mathcal{X} \subset \mathbb{R}^n$ be a subset of $\mathbb{R}^n$ and $x \in \mathcal{X}$. Recall that F is said to be *upper hemi-continuous* at x if for any $u \in \mathbb{R}^n$, the support function $\sigma(F(\cdot,\xi),u)$ is upper semi-continuous at x. F is upper hemi-continuous on $\mathcal{X}$ if and only if the support function $\sigma(F(\cdot,\xi),u)$ is upper semi-continuous on $\mathcal{X}$ for every $u \in \mathbb{R}^n$. For a more detailed discussion of upper hemi-continuity, see [4, Section 2.6].

Definition 4.2 *Let $F : \mathbb{R}^n \times \Xi \to 2^{\mathbb{R}^m}$ be a random set-valued mapping which is upper hemi-continuous at $x \in \mathcal{X}$, F is said to be*

- H-calm from above *at x, if for any $u \in \mathbb{R}^m$, the support function $\sigma(F(\cdot,\xi),u)$ is H-calm from above at x;*
- H-calm *at x, if for any $u \in \mathbb{R}^m$, the support function $\sigma(F(\cdot,\xi),u)$ is H-calm at x.*

F is said to be H-calm (from above) if respective properties above hold at every point of $\mathcal{X}$.

The definition is a generalization of clamness of deterministic set-valued mappings in Rockafellar and Wets' book [31, Chapter 9, Section I], where a set-valued mapping $F(\cdot) : \mathbb{R}^n \to 2^{\mathbb{R}^m}$ is said to be *calm* at a point $x \in \mathbb{R}^n$ if

$$F(x') \subset F(x) + \varrho\|x' - x\|\mathcal{B}$$

for all x' close to x, where ϱ is a positive constant. Obviously, the calmness implies the H-calmness from above with modulus ϱ and order 1 in that the inclusion above implies

$$\sigma(F(x),u) \le \sigma(F(\bar{x}),u) + \varrho\|x - \bar{x}\|$$

for all $u \in \mathbb{R}^m$. Rockafellar and Wets observed that if $F(x)$ is a piecewise polyhedral mapping, that is, the graph of S is piecewise polyhedral (expressible as the union of finitely many polyhedral sets), then $F(x)$ is calm on its domain. See [31, Example 9.57] for details.

When the calmness of Rockafellar and Wets is generalized to random set-valued mappings in our context, it requires the existence of a (measurable) function $\varrho : \Xi \to \mathbb{R}_+$ and positive constant $\delta > 0$ such that

$$F(x',\xi) \subset F(x,\xi) + \varrho(\xi)\|x' - x\|B \tag{4.18}$$

for all $\xi \in \Xi$ and x' with $\|x'-x\| \le \delta$. This implies

$$\sigma(F(x',\xi),u) \le \sigma(F(x,\xi),u) + \varrho(\xi)\|x'-x\|$$

for all $u \in {\rm I\!R}^m$, which is the H-calmness from above at x. In other words, (4.18) is sufficient for the H-calmness of F.

We are now ready to present one of the main results of this section.

Theorem 4.3 *For $i = 1, \cdots, \hat{i}$, let $\phi_i(x,\xi,u_i) = \sigma(\partial_{x_i} v_i(x,\xi), u_i)$ where $u_i \in {\rm I\!R}^{n_i}$ and $\|u_i\| \le 1$. Let $\mathcal{A}v(x,\xi)$ be defined by (4.7) and $\mathcal{X}$ be a nonempty compact subset of X. Assume: (a) Assumption 3.1 hold, (b) $\mathbb{E}[\phi_i(x,\xi,u_i)]$ is continuous on $\mathcal{X}$, (c) $\partial_{x_i} v_i(\cdot,\cdot,\xi)$ is H-calm from above on $\mathcal{X}$ with modulus $a_i(\xi)$ and order γ, (d) for $p_i(\xi) \equiv \kappa_i(\xi) + a_i(\xi)$, where κ_i is defined as in Assumption 3.1, the moment generating function $\mathbb{E}\left[e^{tp_i(\xi)}\right]$ of $p_i(\xi)$, is finite valued for t close to 0. Then for any small positive number $\epsilon > 0$, there exist $\hat{c}(\epsilon) > 0$ and $\hat{\beta}(\epsilon) > 0$, independent of N, such that for N sufficiently large*

$$\text{Prob}\left\{\sup_{x\in\mathcal{X}} \mathbb{D}(\mathcal{A}\vartheta^N(x), \mathbb{E}[\mathcal{A}v(x,\xi)]) \ge \epsilon\right\} \;\le\; \hat{c}(\epsilon)e^{-\hat{\beta}(\epsilon)N}. \tag{4.19}$$

Proof. We use Proposition 4.1 to prove the result. First, we show that for any $u := (u_1, \cdots, u_{\hat{i}})$,

$$\mathbb{E}[\sigma(\mathcal{A}v(x,\xi),u)] = \sigma(\mathbb{E}[\mathcal{A}v(x,\xi)],u). \tag{4.20}$$

Let $\eta \in \mathcal{A}v(x,\xi)$ be such that $\sigma(\mathcal{A}v(x,\xi),u) = x^T\eta$. By [4, Theorem 8.2.11], η is measurable and hence $\mathbb{E}[\eta] \in \mathbb{E}[\mathcal{A}v(x,\xi)]$. Therefore

$$\mathbb{E}[\sigma(\mathcal{A}v(x,\xi),u)] = \mathbb{E}[x^T\eta] \le \sigma(u\mathbb{E}[\mathcal{A}v(x,\xi)],u).$$

Conversely, let $\hat{\eta} \in \mathbb{E}[\mathcal{A}v(x,\xi)]$ such that $\sigma(\mathbb{E}[\mathcal{A}v(x,\xi)],u) = x^T\hat{\eta}$. Then there exists $\eta \in \mathcal{A}v(x,\xi)$ such that $\hat{\eta} = \mathbb{E}[\eta]$. Hence

$$\mathbb{E}[\sigma(\mathcal{A}v(x,\xi),u)] \ge \mathbb{E}[x^T\eta] = \sigma(\mathbb{E}[\mathcal{A}v(x,\xi)],u).$$

This shows (4.20). Observe next that

$$\mathbb{D}(\mathcal{A}\vartheta^N(x), \mathbb{E}[\mathcal{A}v(x,\xi)]) \le \sum_{i=1}^{\hat{i}} \mathbb{D}(\partial_{x_i} v_i^N(x,\xi), \mathbb{E}[\partial_{x_i} v_i(x,\xi)]), \tag{4.21}$$

where

$$\partial_{x_i}\vartheta_i^N(x,\xi) := \frac{1}{N}\sum_{k=1}^N \partial_{x_i} v_i(x,\xi^k), \quad i = 1,\cdots,\hat{i}.$$

Since $\partial_{x_i} v_i(x,\xi)$ is a convex set,

$$\sigma(\partial_{x_i}\vartheta_i^N(x,\xi),u_i) = \frac{1}{N}\sum_{k=1}^N \sigma(\partial_{x_i} v_i(x,\xi^k),u_i), \quad i = 1,\cdots,\hat{i}.$$

Using this, Proposition 2.1 and (4.20), we obtain

$$\begin{aligned}\mathbb{D}(\partial_{x_i}\vartheta_i^N(x,\xi),\mathbb{E}[\partial_{x_i}v_i(x,\xi)]) &= \sup_{\|u_i\|\le 1}\left(\sigma(\partial_{x_i}\vartheta_i^N(x,\xi),u_i)-\sigma(\mathbb{E}[\partial_{x_i}v_i(x,\xi)],u_i)\right)\\ &= \sup_{\|u_i\|\le 1}\left(\frac{1}{N}\sum_{k=1}^N \phi_i(x,\xi^k,u_i)-\sigma(\mathbb{E}[\partial_{x_i}v_i(x,\xi)],u_i)\right)\\ &= \sup_{\|u_i\|\le 1}\left(\frac{1}{N}\sum_{k=1}^N \phi_i(x,\xi^k,u_i)-\mathbb{E}[\phi_i(x,\xi,u_i)]\right).\end{aligned}$$

Consequently

$$\sup_{x\in\mathcal{X}}\mathbb{D}(\partial_{x_i}\vartheta_i^N(x,\xi),\mathbb{E}[\partial_{x_i}v_i(x,\xi)]) \le \sup_{\|u_i\|\le 1, x\in\mathcal{X}}\left(\frac{1}{N}\sum_{k=1}^N \phi_i(x,\xi^k,u_i)-\mathbb{E}[\phi_i(x,\xi,u_i)]\right). \quad (4.22)$$

Let $Z_i := \{u_i \in \mathbb{R}^{n_i} : \|u_i\| \le 1\} \times \mathcal{X}$. In what follows, we show the uniform exponential convergence of the right hand side of the above inequality by applying Proposition 4.1 (i) to $\phi(x,\xi,u_i)$ with variable (x,u_i). Observe that

$$\|\phi_i(x,\xi,u_i)\| \le \|\partial_{x_i}v_i(x,\xi)\| \le \kappa_i(\xi),$$

and by assumption $\phi_i(x,\xi,u_i)$ is H-calm from above in x with modulus $a_i(\xi)$ and order γ. Thus

$$\phi_i(x',\xi,u_i')-\phi_i(x,\xi,u_i) \le a_i(\xi)\|x'-x\|^\gamma + \kappa_i(\xi)\|u_i'-u_i\| \le p_i(\xi)\|z_i'-z_i\|^{\min(\gamma,1)},$$

where $z_i = (x,u_i)$ and the last inequality is due to the fact that we only use the inequality for z_i' close to z and hence may assume without lost of generality that $\|z_i'-z_i\| \le 1$. This shows the H-calmness from above of $\phi(x,\xi,u_i)$ with respect to (x,u_i) set Z_i. Notice that $\mathbb{E}[\phi(x,\xi,u_i)]$ is continuous in x by assumption and because $\phi(x,\xi,u_i)$ is Lipschitz continuous in u_i with integrable modulus $\kappa_i(\xi)$, $\mathbb{E}[\phi(x,\xi,u_i)]$ is also continuous in u_i. This shows the continuity of $\mathbb{E}[\phi_i(x,\xi,u_i)]$ with respect to (x,u_i) on set Z_i.

By Proposition 4.1, for any $\epsilon_i > 0$, there exist positive constants $\hat{c}_i(\epsilon_i)$ and $\hat{\beta}_i(\epsilon_i)$, independent of N, such that

$$\text{Prob}\left\{\sup_{(x,u_i)\in Z_i}\left(\frac{1}{N}\sum_{k=1}^N \phi_i(x,\xi^k,u_i)-\mathbb{E}[\phi_i(x,\xi,u_i)]\right)\ge \epsilon_i\right\} \le \hat{c}_i(\epsilon_i)e^{-N\hat{\beta}_i(\epsilon_i)}, \quad (4.23)$$

for $i = 1,\cdots,\hat{i}$. For any $\epsilon > 0$, let $\epsilon_i > 0$ be such that $\sum_{i=1}^{\hat{i}}\epsilon_i \le \epsilon$. Then by combining (4.21)-(4.23), we obtain

$$\begin{aligned}\text{Prob}\left\{\sup_{x\in\mathcal{X}}\mathbb{D}(\mathcal{A}\vartheta^N(x),\mathbb{E}[\mathcal{A}v(x,\xi)])\ge\epsilon\right\} &\le \sum_{i=1}^{\hat{i}}\text{Prob}\left\{\sup_{x\in\mathcal{X}}\mathbb{D}(\partial_{x_i}\vartheta_i^N(x,\xi),\mathbb{E}[\partial_{x_i}v_i(x,\xi)])\ge\epsilon_i\right\}\\ &\le \sum_{i=1}^{\hat{i}}\hat{c}_i(\epsilon_i)e^{-N\hat{\beta}_i(\epsilon_i)} \le \hat{c}(\epsilon)e^{-N\hat{\beta}(\epsilon)}.\end{aligned}$$

This shows (4.19) with $\hat{c}(\epsilon) = \hat{i}\max_{i=1}^{\hat{i}}\hat{c}_i(\epsilon_i)$ and $\hat{\beta}(\epsilon) = \min_{i=1}^{\hat{i}}\hat{\beta}_i(\epsilon_i)$. ■

Remark 4.1 *The H-calmness from above of $\partial_{x_i}v_i$ and continuity of $\mathbb{E}[\partial_{x_i}v_i]$ play an important role in Theorem 4.3. It is therefore natural to ask when the properties hold.*

(i) Except in some pathological examples, v_i is often piecewise smooth in many practical instances, that is, it can be expressed as a finite number of smooth functions. See [37] for detailed discussion of piecewise function. When v_i is piecewise twice continuously differentiable, the gradient of each smooth piece is H-calm in its domain. Since the Clarke generalized gradient at a point is the convex hull of the gradient of some active pieces at the point, this means the Clarke generalized gradient of v_i is H-calm in its domain. We omit technical details.

(ii) Shapiro [39, Proposition 4.1] showed that if a random function is continuously differentiable w.p.1 and is Lipschitz continuous with integrable Lipschitz modulus, then the expected value of the function is continuously differentiable. In the context of this paper, the proposition implies that if v_i is continuously differentiable with respect to x_i w.p.1 and Assumption 3.1 holds, then $\mathbb{E}[v_i]$ is continuously differentiable in x_i and

$$\partial_{x_i}\mathbb{E}[v_i] = \nabla_{x_i}\mathbb{E}[v_i] = \mathbb{E}[\nabla_{x_i} v_i] = \mathbb{E}[\partial_{x_i} v_i].$$

Moreover, since $\nabla_{x_i} v_i$ is single valued, the upper semi-continuity of $\partial_{x_i} v_i$ with respect to x_{-i} established in Proposition 3.1 implies the continuity. This gives us the desired continuity of $\mathbb{E}[\partial_{x_i} v_i]$ with respect to x.

When $\partial_{x_i} v_i$, $i = 1, \cdots, \hat{\imath}$, is H-calm, the exponential convergence of Theorem 4.3 can be strengthened by replacing $\mathbb{D}$ with $\mathbb{H}$ in (4.19). We state this in the following corollary.

Corollary 4.1 *If $\partial_{x_i} v_i(\cdot,\cdot,\xi)$ is H-calm on $\mathcal{X}$ with modulus $a_i(\xi)$ and order γ, and for $p_i(\xi) \equiv \kappa_i(\xi)+c_i(\xi)$, where κ_i is defined as in Assumption 3.1, the moment generating function $\mathbb{E}\left[e^{tp_i(\xi)}\right]$ of $p_i(\xi)$, is finite valued for t close to 0, then for any small positive number $\epsilon > 0$, there exist $\hat{c}(\epsilon) > 0$ and $\hat{\beta}(\epsilon) > 0$, independent of N, such that for N sufficiently large*

$$\text{Prob}\left\{\sup_{x\in\mathcal{X}} \mathbb{H}(\mathcal{A}\vartheta^N(x), \mathbb{E}[\mathcal{A}v(x,\xi)]) \geq \epsilon\right\} \;\leq\; \hat{c}(\epsilon)e^{-\hat{\beta}(\epsilon)N}. \tag{4.24}$$

Proof. Under the H-calmness, we can show, similar to the proof of Theorem 4.3, that

$$\mathbb{D}(\mathbb{E}[\partial_{x_i} v_i(x,\xi)], \partial_{x_i}\vartheta_i^N(x,\xi)) \;=\; \sup_{\|u_i\|\leq 1}\left(\mathbb{E}[\phi_i(x,\xi,u_i)] - \frac{1}{N}\sum_{k=1}^N \phi_i(x,\xi^k,u_i)\right).$$

Moreover,

$$\begin{aligned}
&\text{Prob}\left\{\sup_{(x,u_i)\in Z_i}\left(\mathbb{E}[\phi_i(x,\xi,u_i)] - \frac{1}{N}\sum_{k=1}^N \phi_i(x,\xi^k,u_i)\right) \geq \epsilon_i\right\}\\
&= \text{Prob}\left\{\inf_{(x,u_i)\in Z_i}\left(\frac{1}{N}\sum_{k=1}^N \phi_i(x,\xi^k,u_i) - \mathbb{E}[\phi_i(x,\xi,u_i)]\right) \leq -\epsilon_i\right\}.
\end{aligned}$$

Notice that H-calmness of $\partial_{x_i} v_i(x,\xi)$ implies that $\phi_i(x,\xi,u_i)$ is H-calm from below for every fixed u_i. By applying Proposition 4.1 (iii) to the right hand side of the above equation, we obtain that for any small positive number $\epsilon > 0$, there exist $\hat{c}(\epsilon) > 0$ and $\hat{\beta}(\epsilon) > 0$, independent of N, such that for N sufficiently large

$$\text{Prob}\left\{\sup_{x\in\mathcal{X}} \mathbb{D}(\mathbb{E}[\mathcal{A}v(x,\xi)], \mathcal{A}\vartheta^N(x)) \geq \epsilon\right\} \;\leq\; \hat{c}(\epsilon)e^{-\hat{\beta}(\epsilon)N}. \tag{4.25}$$

The conclusion follows by combining (4.25) and (4.19). ■

Theorem 4.4 *Let $\mathcal{X} \subset X$ be a nonempty compact subset of X and $x^* \in \mathcal{X}$ be a weak stochastic Nash stationary point of the true problem (1.1). Let $\Psi(x) = \mathbb{E}[\mathcal{A}v(x,\xi)] + G_X(x)$, where $G_X(x)$ is defined by (4.8). Assume that Ψ is metric regular at x^* for 0 and for $\bar{N}$ sufficiently, the sequence $\{x^N\}_{N>\bar{N}}$ is located in $\mathcal{X}$ w.p.1. Then under conditions (a)-(d) of Theorem 4.3,*

(i) $\{x^N\}$ converges to x^* at exponential rate, that is, for any small positive number $\epsilon > 0$, there exist $\hat{c}(\epsilon) > 0$ and $\hat{\beta}(\epsilon) > 0$, independent of N, such that for N sufficiently large

$$\text{Prob}\big(\|x^N - x^*\| \geq \epsilon\big) \;\leq\; \hat{c}(\epsilon)e^{-\hat{\beta}(\epsilon)N}. \tag{4.26}$$

(ii) if , in addition, one of the conditions of Theorem 4.2 holds, and $\{x^N\}$ is a sequence of Nash equilibria of SAA problem (1.2), then x^* is a Nash equilibrium of the true problem (1.1).

Proof. Part (i). By Proposition 2.3, the metric regularity of $\Psi(x)$ at x^* implies that there exists a positive scalar $\alpha > 0$ such that

$$\begin{aligned} \|x^N - x^*\| &\leq \alpha\mathbb{D}(\mathcal{A}\vartheta^N(x^N) + G_X(x^N), \Psi(x^N)) = \alpha\mathbb{D}(\mathcal{A}\vartheta^N(x^N), \mathbb{E}[\mathcal{A}v(x^N,\xi)]) \\ &\leq \alpha \sup_{x\in\mathcal{X}} \mathbb{D}(\mathcal{A}\vartheta^N(x), \mathbb{E}[\mathcal{A}v(x,\xi)]) \end{aligned}$$

for $x^N \in \mathcal{X}$, where $\mathcal{A}\vartheta^N$ is defined as in (4.11). The rest follows from Theorem 4.3.

Part (ii) follows from Theorem 4.2. ■

The metric regularity of Ψ at x^* is the main condition in this theorem. From the proof of the theorem, we can see that this condition can be weakened to

$$\|x^N - x^*\| \leq \alpha \sup_{x\in\mathcal{X}} \mathbb{D}(\mathcal{A}\vartheta^N(x), \mathbb{E}[\mathcal{A}v(x,\xi)]). \tag{4.27}$$

In what follows, we discuss inequality (4.27) under the circumstance that $v_i(x_i, x_{-i}, \xi)$ is continuously differentiable with respect to x_i for every x_{-i} and ξ. For the simplicity of notation, let $\mathcal{A}\vartheta(x) := \mathbb{E}[\mathcal{A}v(x,\xi)]$. The weak first equilibrium condition (4.9) can be written as

$$0 \in \Psi(x) := \mathcal{A}\vartheta(x) + G_X(x). \tag{4.28}$$

Since $\mathcal{A}\vartheta(x)$ and $\mathcal{A}\vartheta^N(x)$ are single valued, both (4.28) and (4.10) are essentially a variational inequality and $\mathbb{D}(\mathcal{A}\vartheta^N(x), \mathbb{E}[\mathcal{A}v(x,\xi)])$ reduces to $\|\mathcal{A}\vartheta^N(x) - \mathcal{A}\vartheta(x)\|$. Facchinei and Pang [13, Chapter 5] presented an intensive discussion on stability and sensitivity of variational inequalities. If $\mathcal{A}\vartheta(x)$ is B-differentiable (directionally differentiable and Lipschitz continuous) and for every vector u in the critical cone of (4.28) $u^T\mathcal{A}\vartheta'(x,u) > 0$, where ϑ' denotes the directional derivative, then from [13, Proposition 5.1.6 and Corollary 5.1.8], one can obtain the following:

- (4.28) has a solution;
- (4.10) has a solution provided that the quantity $\sup_{x\in\mathcal{X}} \mathbb{D}\|\mathcal{A}\vartheta^N(x) - \mathcal{A}\vartheta(x)\|$ is sufficiently small;
- (4.27) holds.

In Section 6, we will show that $\nabla\mathcal{A}\vartheta(x)$ is nonsingular in a practical example. See Lemma 6.1.

5 A smoothing approach

Having established the convergence results in the preceding section, we now turn to discuss the numerical solution of the sample average Nash equilibrium problem (1.2). For fixed sample, this is a deterministic nonsmooth equilibrium problem and one may use well known bundle methods [22, 35] to solve it.

In this section, we consider the case when the underlying function has simple nonsmoothness structure. Our idea here is to approximate v_i by a parameterized smooth function and then solve the smoothed sample average approximation problem. The approach is known as smoothing and has been used to deal with nonsmooth stochastic optimization problem in [50]. It is shown that the approach is very effective when the nonsmoothness of underlying functions is caused by a few simple operations such as max (min)-function. The approach is even more attractive here because once the function is smoothed, the first order equilibrium conditions are reduced to variational inequalities or nonlinear complementarity problems for which many numerical methods are available [13]. Let us first describe the smoothing method.

Definition 5.1 *Let $f : \mathbb{R}^m \to \mathbb{R}$ be a locally Lipschitz continuous function and $\epsilon \in \mathbb{R}$ be a parameter. $\hat{f}(x, \epsilon) : \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}$ is a* smoothing *of f if it satisfies the following:*

(a) for every $x \in \mathbb{R}^m$, $\hat{f}(x, 0) = f(x)$;

(b) for every $x \in \mathbb{R}^m$, $\hat{f}$ is locally Lipschitz continuous at $(x, 0)$;

(c) $\hat{f}$ is continuously differentiable on $\mathbb{R}^m \times \mathbb{R}\backslash\{0\}$;

(d) $\hat{f}$ is convex if f is convex.

A smoothing scheme, excluding (d), was first proposed by Ralph and Xu [27] for obtaining a smooth approximation of an implicit function defined by a nonsmooth system of equations and was applied to tackle nonsmooth in nonsmooth stochastic minimization problem in [50]. Parts (a) and (c) in the definition require that the smoothing function match the original function when the smoothing parameter is zero and be continuously differentiable when the smoothing parameter is nonzero. The Lipschitz continuity in part (b) implies that the Clarke generalized gradient $\partial_{(x,\epsilon)}\hat{f}(x, 0)$ is well defined and this allows us to compare the generalized gradient of the smoothed function at point $(x, 0)$ with that of the original function. If

$$\pi_x \partial_{(x,\epsilon)}\hat{f}(x, 0) \subset \partial_x f(x),$$

where $\pi_x \partial_{(x,\epsilon)}\hat{f}(x, 0)$ denotes the set of all m-dimensional vectors a such that, for some scalar c, the $(m + 1)$-dimensional vector (a, c) belongs to $\partial_{(x,\epsilon)}\hat{f}(x, 0)$, then $\hat{f}$ is said to satisfy *gradient consistency* (which is known as Jacobian consistency when f is vector valued, see [27] and references therein). This is a key property that will be used in the analysis of the first order optimality condition later on. Property (d) requires the smoothing function preserve the convexity. This is particularly relevant in this paper because Nash equilibria are closed related to convexity.

Using the smoothing function, we may consider the smoothed true problem: find $x^*(\epsilon)$ such that

$$\hat{\vartheta}_i(x_i^*(\epsilon), x_{-i}^*(\epsilon), \epsilon) = \min_{x_i \in X_i} E[\hat{v}_i(x_i, x_{-i}^*(\epsilon), \xi, \epsilon)], \quad \text{for } i = 1, \cdots, \hat{i}, \tag{5.29}$$

and its sample average approximation: find $(x_1^N(\epsilon), \cdots, x_{\hat{i}}^N(\epsilon)) \in X_1 \times \cdots \times X_{\hat{i}}$ such that

$$\hat{\vartheta}_i^N(x_i^N(\epsilon), x_{-i}^N(\epsilon), \epsilon) = \min_{x_i \in X_i} \hat{\vartheta}_i^N(x_i, x_{-i}^N(\epsilon), \epsilon) \text{ for } i = 1, \cdots, \hat{i}, \tag{5.30}$$

where

$$\hat{\vartheta}_i^N(x_i, x_{-i}^N(\epsilon), \epsilon) = \frac{1}{N} \sum_{k=1}^N \hat{v}_i(x_i, x_{-i}^N(\epsilon), \xi^k, \epsilon).$$

The first order equilibrium conditions of (5.29) and (5.30) can be written respectively as

$$0 \in E[\nabla_{x_i} \hat{v}_i(x_i, x_{-i}, \xi, \epsilon)] + \mathcal{N}_{X_i}(x_i), \ i = 1, \cdots, \hat{i}, \tag{5.31}$$

and

$$0 \in \frac{1}{N} \sum_{k=1}^N \nabla_{x_i} \hat{v}_i(x_i, x_{-i}, \xi^k, \epsilon) + \mathcal{N}_{X_i}(x_i), \ i = 1, \cdots, \hat{i}. \tag{5.32}$$

Note that in numerical implementation, we may solve (5.30) by fixing sample and driving the smoothing parameter to zero or fixing the smoothing parameter and increasing the sample size. The former might be more preferable because it is numerically cheaper to reduce the smoothing parameter than increasing sample size. In what follows, we give a statement of convergence for both cases. Let $S(\epsilon)$ denote the set of solutions to (5.31) and $S^N(\epsilon)$ the set of solutions to (5.32).

Theorem 5.1 *For $i = 1, \cdots, \hat{i}$, let $\hat{v}_i(x_i, x_{-i}, \xi, \epsilon)$ be a smoothing of $v_i(x_i, x_{-i}, \xi)$. Suppose: (a) there exists an integrable function $\kappa_i(\xi)$ such that the Lipschitz modulus of $\hat{v}_i(x, \xi, \epsilon)$ with respect to x_i is bounded by $\kappa_i(\xi)$, (b) for almost every ξ,*

$$\overline{\lim_{x' \to x, \epsilon \to 0}} \{\nabla_{x_i} \hat{v}_i(x'_i, x'_{-i}, \xi, \epsilon)\} \subset \partial_{x_i} v_i(x_i, x_{-i}, \xi). \tag{5.33}$$

(i) If $S(\epsilon)$ is nonempty for all ϵ close to 0, then $\overline{\lim}_{\epsilon \to 0} S(\epsilon) \subset S$, w.p.1., where S denotes the set of solutions to (3.4).

(ii) If $S^N(\epsilon)$ is nonempty for all N sufficiently large, then $\overline{\lim}_{\epsilon \to 0} S^N(\epsilon) \subset S^N$, w.p.1, where S^N denotes the solution set of (3.6);

(iii) If $S^N(\epsilon)$ is nonempty for all N sufficiently large, then $\overline{\lim}_{N \to \infty} S^N(\epsilon) \subset S(\epsilon)$, w.p.1.

Proof. The proof of parts (i) and (ii) is similar to the proof of [50, Theorem 3.1]. We omit details.

Part (iii). Let $\mathcal{X}$ be a compact subset of X such that $\overline{\lim}_{\epsilon \to 0} S^N(\epsilon) \subset \mathcal{X}$. Under condition (a), $\nabla_{x_i} \hat{v}_i(x_i, x_{-i}, \xi, \epsilon)$ is bounded by $\kappa_i(\xi)$. Applying the uniform strong law of large numbers [34, Lemma A1] to $\frac{1}{N} \sum_{k=1}^N \nabla_{x_i} \hat{v}_i(x_i, x_{-i}, \xi^k, \epsilon)$ over set $\mathcal{X}$, we have

$$\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^N \nabla_{x_i} \hat{v}_i(x_i, x_{-i}, \xi^k, \epsilon) = \mathbb{E}[\nabla_{x_i} \hat{v}_i(x_i, x_{-i}, \xi, \epsilon)], \text{ for } i = 1, \cdots, \hat{i},$$

w.p.1. The rest is straightforward given the upper semi-continuity of normal cones $\mathcal{N}_{X_i}(\cdot)$. ∎

Analogous to Theorem 4.4, it is possible to analyze the rate of convergence in part (ii) of Theorem 5.1. That is, if: (a) $x^N(\epsilon) \in S^N(\epsilon)$ and for N sufficiently large, it is located within a neighborhood of $x(\epsilon) \in S(\epsilon)$, (b) the moment generating functions $e^{(\hat{v}_i(x_i,x_{-i},\xi,\epsilon)-\mathbb{E}[\hat{v}_i(x_i,x_{-i},\xi,\epsilon)])t}$ and $e^{\kappa_i(x)t}$ are finite valued for t close to 0, $i = 1, \cdots, \hat{i}$ and (c) the set valued mapping

$$\Phi(x) = \mathcal{A}v(x, \epsilon) + G_X(x)$$

is metric regular at $x(\epsilon)$ for 0, where

$$\mathcal{A}v(x, \epsilon) = \mathbb{E}[\nabla_{x_1}\hat{v}_1(x_1, x_{-1}, \xi, \epsilon)] \times \cdots \times \mathbb{E}[\nabla_{x_{\hat{i}}}\hat{v}_{\hat{i}}(x_{\hat{i}}, x_{-\hat{i}}, \xi, \epsilon)]$$

and $G_X(x)$ is defined as in (4.8), then one can obtain exponential rate of convergence for $x^N(\epsilon)$ to $x(\epsilon)$. We omit the technical details.

6 Applications

In this section, we discuss a stochastic Nash equilibrium problem arising from competition of generators in a wholesale electricity market and use the sample average approximation method to solve the problem.

6.1 A stochastic Nash equilibrium model for competition in electricity markets

Consider an electricity spot market with M generators competing in a non-collaborative manner to bid for dispatch of electricity before market demand is realized. The market demand is characterized by an inverse demand function $p(q, \xi(\omega))$, where $p(q, \xi(\omega))$ is the market price, q is the total supply to the market, and $\xi : \Omega \to \mathbb{R}$ is a continuous random variable with support set Ξ. Demand uncertainty is thus characterized by the distribution of the random variable ξ.

Before market demand is realized, generator i, $i = 1, \cdots, M$, chooses its quantity for dispatch, denoted by q_i. The generator's expected profit can then be formulated as

$$R_i(q_i, Q_{-i}) = \mathbb{E}[q_i p(Q, \xi) - C_i(q_i) + H_i(p(Q, \xi))]. \tag{6.34}$$

Here Q_{-i} denotes the total bids by i's competitors, $Q = q_i + Q_{-i}$ is the total bids by all generators to the market, $q_i p(Q, \xi)$ is the total revenue for generator by selling amount q_i of electricity if the market demand scenario turns out to be $p(Q, \xi)$, $C_i(q_i)$ denotes the total cost for producing q_i amount of electricity, $H_i(p)$ denotes the payments related to forward contracts that generators signed with retailers before entering spot market.

Forward contracts are financial instruments which are typically used to hedge risks arising from uncertainties in spot market. There are essentially two types of contracts: two-way contracts and one-way contracts. A two-way contract is a contract for differences. Assuming generator i signs a two-way contract at a strike price f_2 for a quantity s_i, and the market clearing price in the spot market is $p(Q, \xi)$, then the generator will pay $s_i\,(p(Q, \xi) - f_2)$ to the contract holder if $p \geq f_2$. Conversely, if $p < f_2$, then the generator will get paid from the contract holder an amount of $s_i\,(f_2 - p)$.

A one-way contract is a call or put option. For the simplicity of notation, here we only consider the call option. By selling a call option at a strike price f_1, generator i will pay $w_i(p - f_1)$ to the contract holder if $p > f_1$ and the option is exercised, but no payment is made if $p \leq f_1$, where w_i is the quantity signed by generator i on one-way contract at strike price f_1. Assuming that generator i signs two way contract of quantity s_i at strike price f_2 and one way contract of w_i at strike price f_1, we can formulate $H_i(p(Q,\xi))$ as follows:

$$H_i(p(Q,\xi)) = -w_i \max\left(p(Q,\xi) - f_1, 0\right) - s_i\left(p(Q,\xi) - f_2\right).$$

Generator i's decision making problem is to choose an optimal quantity q_i such that its expected profit defined in (6.34) is maximized. Assuming that every generator is a profit maximizer, we can formulate the competition as a stochastic Nash equilibrium problem. For the simplicity of notation, let

$$r_i(q_i, Q_{-i}, \xi) := q_i p(Q,\xi) - C_i(q_i) - w_i \max\left(p(Q,\xi) - f_1, 0\right) - s_i\left(p(Q,\xi) - f_2\right).$$

Then $R_i(q_i, Q_{-i}) = \mathbb{E}[r_i(q_i, Q_{-i}, \xi)]$.

Definition 6.1 *A stochastic Nash Equilibrium is an M-tuple* $\mathbf{q}^* = (q_1^*, \ldots, q_M^*)$ *such that*

$$-R_i(q_i^*, Q_{-i}^*) = \min_{q_i \in \mathcal{Q}_i} -R_i(q_i, Q_{-i}^*). \tag{6.35}$$

for $i = 1, \cdots, M$, *where* $\mathcal{Q}_i := [0, q_i^u]$, *and* q_i^u *is the capacity limit of generator* i.

Obviously the above Nash equilibrium problem is an example of (1.1). In what follows, we apply the SAA method to this problem and investigate the convergence of SAA equilibrium as sample size increases using our established results in the preceding sections. Note that the existence and uniqueness of equilibrium of problem (6.35) can be obtained under the following assumptions.

Assumption 6.1 *The inverse demand function* $p(q,\xi)$ *and the cost function* $C_i(q_i)$ *satisfy the following conditions:*

(a) $p(q,\xi)$ *is twice continuously differentiable and strictly decreasing in* q *for any fixed* $\xi \in \Xi$;

(b) there exists an integrable function $\kappa(\xi)$ *such that*

$$\max(|p(q,\xi)|, |p_q'(q,\xi)|, |p_{qq}''(q,\xi)|) \leq \kappa(\xi)$$

for all $\xi \in \Xi$;

(c) $p_q'(q,\xi) + q p_q''(q,\xi) \leq 0$, *for all* $q \geq 0$ *and* $\xi \in \Xi$;

(d) the cost function $C_i(q_i)$, $i = 1, 2, \ldots, M$, *is twice continuously differentiable and* $C_i'(q_i) \geq 0$ *and* $C_i''(q_i) \geq 0$ *for all* $q_i > 0$.

The assumptions are fairly standard, see similar ones in [10, 11, 44].

Proposition 6.1 *Under Assumption 6.1,*

(i) $R_i(q_i, Q_{-i})$ *is continuously differentiable in* q_i *and its derivative is continuously differentiable with respect to* $\mathbf{q}$*;*

(ii) $r_i(q_i, Q_{-i}, \xi)$ *and* $R_i(q_i, Q_{-i})$ *are strictly concave in* q_i*.*

Proof. Part (i). Observe first that $r_i(q_i, Q_{-i}, \xi)$ is piecewise continuously differentiable, that is, if $p(Q, \xi) > f_1$, then

$$(r_i)'_{q_i}(q_i, Q_{-i}, \xi) = p(Q, \xi) - C'_i(q_i) + (q_i - w_i - s_i)p'_q(Q, \xi)$$

and if $p(Q, \xi) < f_1$,

$$(r_i)'_{q_i}(q_i, Q_{-i}, \xi) = p(Q, \xi) - C'_i(q_i) + (q_i - s_i)p'_q(Q, \xi).$$

The function is not differentiable at point q_i where $p(Q, \xi) = f_1$. Under Assumption 6.1 (a), $\nabla_{q_i} r_i(q_i, Q_{-i}, \xi)$ is bounded by

$$L(q_i, \xi) := \max(\kappa(\xi) + |C'_i(q_i)| + |q_i - w_i - s_i|\kappa(\xi), \kappa(\xi) + |C'_i(q_i)| + |q_i - s_i|\kappa(\xi)),$$

which implies that $r_i(q_i, Q_{-i}, \xi)$ is globally Lipschitz continuous in q_i with an integrable modulus $\max_{q_i \in \mathcal{Q}_i} L(q_i, \xi)$. Notice that for every fixed Q, the strict monotonic decreasing property of $p(\cdot, \xi)$ implies that there exists at most one ξ value such that $p(Q, \xi) = f_1$. This means that $r_i(q_i, Q_{-i}, \xi)$ is continuously differentiable in q_i w.p.1. By [36, Chapter 2, Proposition 2], $\mathbb{E}[r_i(q_i, Q_{-i}, \xi)]$ is also continuously differentiable in q_i. In a similar manner, we can show that the derivative of $\mathbb{E}[r_i(q_i, Q_{-i}, \xi)]$ in q_i is also continuously differentiable with respect to $\mathbf{q}$ by verifying that the function $(r_i)'_{q_i}(q_i, Q_{-i}, \xi)$ is Lipschitz continuous in $\mathbf{q}$ with some integrable modulus and it is continuously differentiable in $\mathbf{q}$ w.p.1. We omit the details.

Part (ii). Under Assumption 6.1, one can easily show the strict concavity of $r_i(q_i, Q_{-i}, \xi)$ and hence $R_i(q_i, Q_{-i})$. Again we omit the details. ■

With Proposition 6.1 and [33, Theorems 1 and 2], we can show the existence and uniqueness of equilibrium of (6.35). We omit the details because it is not the main focus of this paper.

6.2 Sample average approximation

Stochastic Nash equilibrium problem (6.35) makes a good case for SAA method in that: (a) the distribution of ξ is not necessarily known but it may be obtained by sampling from past data or computer simulation; (b) the presence of max-operator in $r_i(q_i, Q_{-i}, \xi)$ makes it difficult to obtain a closed form of $R_i(q_i, Q_{-i})$ even when the distribution of ξ is known.

Let $\xi^1, \cdots, \xi^N$ be an i.i.d. sample of $\xi(\omega)$. The sample average approximation of Nash equilibrium problem (6.35) is: find $\mathbf{q}^N := (q_1^N, q_2^N, \ldots, q_M^N) \in \mathcal{Q}_1 \times \mathcal{Q}_2 \times \cdots \times \mathcal{Q}_M$ such that

$$-R_i^N(q_i^N, Q_{-i}^N) = \min_{q_i \in \mathcal{Q}_i} -R_i^N(q_i, Q_{-i}^N) := \frac{1}{N}\sum_{k=1}^{N} -r_i(q_i, Q_{-i}, \xi^k) \quad (6.36)$$

for $i = 1, \cdots, M$. In order to study the convergence of $\mathbf{q}^N$, we need first order equilibrium conditions of both the true problem and its sample average approximation. Observe first that from Proposition 6.1 (i), $\mathbb{E}[r_i(q_i, Q_{-i}, \xi)]$ is continuously differentiable. Therefore the weak first order equilibrium condition of the true problem (6.35) coincides with the first order equilibrium condition which can be written as:

$$0 \in -\mathbb{E}\left[\nabla_{q_i} r_i(q_i, Q_{-i}, \xi)\right] + \mathcal{N}_{\mathcal{Q}_i}(q_i), \ i = 1, 2, \ldots, M. \tag{6.37}$$

For the SAA problem, the underlying functions are piecewise continuously differentiable. Therefore we use the Clarke generalized gradient to characterize the first order equilibrium condition as follows:

$$0 \in -\frac{1}{N} \sum_{k=1}^{N} \partial_{q_i} r_i(q_i, Q_{-i}, \xi^k) + \mathcal{N}_{\mathcal{Q}_i}(q_i), \ i = 1, 2, \ldots, M. \tag{6.38}$$

Let

$$F(\mathbf{q}, \xi) := \left((r_1)'_{q_1}(q_1, Q_{-1}, \xi), \cdots, (r_M)'_{q_M}(q_M, Q_{-M}, \xi)\right)^T.$$

and $G(\mathbf{q}) = \mathcal{N}_{\mathcal{Q}_1}(q_1) \times \cdots \mathcal{N}_{\mathcal{Q}_M}(q_M)$, let $\Psi(\mathbf{q}) = -\mathbb{E}[F(\mathbf{q}, \xi)] + G(\mathbf{q})$. Then the first order equilibrium condition (6.37) can be written as

$$0 \in \Psi(\mathbf{q}).$$

Lemma 6.1 *Under Assumption 6.1, $\nabla_{\mathbf{q}} \mathbb{E}[F(\mathbf{q}, \xi)] = \mathbb{E}[\nabla_{\mathbf{q}} F(\mathbf{q}, \xi)]$ and $\mathbb{E}[\nabla_{\mathbf{q}} F(\mathbf{q}, \xi)]$ is a nonsingular for all $\mathbf{q} \in \mathcal{Q}$.*

The proof can be obtained by a detailed calculation of the determinant of matrix $\mathbb{E}[\nabla_{\mathbf{q}} F(\mathbf{q}, \xi)]$. We include it in the Appendix.

Theorem 6.1 *Let $\mathbf{q}^*$ be a Nash stationary point defined by (6.37)and $\mathbf{q}^N$ be a Nash stationary point defined by (6.38).Under Assumption 6.1, the sequence $\{\mathbf{q}^N\}$ converges to $\mathbf{q}^*$ at exponential rate, with the increase of sample size N, that is, for any small positive number $\epsilon > 0$, there exists constants $\hat{c}_1(\epsilon) > 0$ and $\hat{c}_2(\epsilon) > 0$, independent of N, such that for N sufficiently large*

$$\text{Prob}\left(\|\mathbf{q}^N - \mathbf{q}^*\| \geq \epsilon\right) \leq \hat{c}_1(\epsilon) e^{-\hat{c}_2(\epsilon) N}. \tag{6.39}$$

Proof. We use Theorem 4.4 to prove the results. To this end, we verify conditions (a)-(e) of Theorem 4.4 in this context. Conditions (a) and (b) follow from Proposition 6.1 which shows that $\mathbb{E}[r_i(q_i, Q_{-i}, \xi)]$ is continuously differentiable in q_i, and its derivative is continuous in $\mathbf{q}$ and bounded by an integrable function. Conditions (c) follows from Lemma 6.1. This is because the nonsingularity of matrix $\nabla_{\mathbf{q}} \mathbb{E}[F(\mathbf{q}, \xi)]$ implies that $\Psi(\mathbf{q})$ is metric regular for all $\mathbf{q} \in \mathcal{Q}$, see a discussion in [31, Page 388]. Condition (d) is not needed because the metric regularity condition holds in the whole region $\mathcal{Q}$. Condition (e) is satisfied following Remark 4.1. ■

6.3 Smoothing approximation and convergence

In Proposition 6.1, we have shown that $\mathbb{E}[r_i(q_i, Q_{-i}, \xi)]$ is continuously differentiable in q_i. However, $r_i(q_i, Q_{-i}, \xi^k)$ in the sample average Nash equilibrium problem (6.36) is not necessarily continuously differentiable for every ξ^k. The possible nonsmoothness results from the max-function. In what follows, we consider a simple smoothing scheme used in [50] to smooth the max-function. Let $\epsilon \in \mathbb{R}_+$ and $\hat{a}(z,\epsilon)$ be such that for every $\epsilon > 0$, let

$$\hat{a}(z,\epsilon) := \epsilon \ln(1 + e^{z/\epsilon}) \tag{6.40}$$

and for $\epsilon = 0$, $\hat{a}(z,0) := \max(z,0)$. It is proved in [50, Example 3.1] that $\hat{a}(z,\epsilon)$ satisfies properties (a)-(c) specified in Definition 5.1. Moreover, $\hat{a}(z,\epsilon)$ is convex in z and it satisfies derivative consistency

$$\overline{\lim_{(z',\epsilon)\to(z,0)}} \frac{d\hat{a}(z,\epsilon)}{dz} = [0,1] = \partial_z \max(z,0). \tag{6.41}$$

Let $\hat{h}_i(q_i, Q_{-i}, \xi, \epsilon) := \hat{a}(p(q_i + Q_{-i}, \xi) - f_1, \epsilon)$ and

$$\hat{r}_i(q_i, Q_{-i}, \xi, \epsilon) := q_i p(Q,\xi) - C_i(q_i) - w_i \hat{h}_i(q_i, Q_{-i}, \xi, \epsilon) - s_i\left(p(Q,\xi) - f_2\right).$$

Then we may solve the following smoothed SAA problem instead of (6.36): find an M-tuple $\mathbf{q}^N(\epsilon) := \left(q_1^N(\epsilon), \ldots, q_M^N(\epsilon)\right)$ such that

$$-\hat{R}_i^N(q_i^N(\epsilon), Q_{-i}^N(\epsilon), \epsilon) = \min_{q_i \in \mathcal{Q}_i} -\frac{1}{N}\sum_{k=1}^N \hat{r}_i(q_i, Q_{-i}^N(\epsilon), \xi^k, \epsilon),\ i = 1, 2, \ldots, M. \tag{6.42}$$

The above problem is the sample average approximation of the following smoothed true problem: find an M-tuple $\mathbf{q}(\epsilon) := (q_1(\epsilon), \ldots, q_M(\epsilon))$ such that

$$-\hat{R}_i(q_i(\epsilon), Q_{-i}, \epsilon) = \max_{q_i \in \mathcal{Q}_i} -\hat{R}_i(q_i, Q_{-i}(\epsilon), \epsilon),\ i = 1, 2, \ldots, M. \tag{6.43}$$

Note that since the smoothing preserves convexity, $\hat{R}_i(q_i, Q_{-i}, \epsilon)$ is convex in q_i. Therefore both (6.42) and (6.43) have a unique equilibrium. Moreover, we can solve (6.42) by solving

$$0 \in -\frac{1}{N}\sum_{k=1}^N (\hat{r}_i)'_{q_i}(q_i, Q_{-i}^N(\epsilon), \xi^k, \epsilon) + \mathcal{N}_{\mathcal{Q}_i}(q_i),\ i = 1, 2, \ldots, M. \tag{6.44}$$

The following proposition states the convergence of $\mathbf{q}^N(\epsilon)$ and $\mathbf{q}(\epsilon)$ as $N \to \infty$ and $\epsilon \to 0$.

Proposition 6.2 *Under Assumption 6.1, $\mathbf{q}^N(\epsilon)$ converges to $\mathbf{q}(\epsilon)$ w.p.1 as $N \to \infty$ for fixed $\epsilon > 0$ and $\mathbf{q}(\epsilon)$ converges to a stochastic Nash equilibrium of (6.35) w.p.1 as ϵ tends to* 0.

Proof. We use Theorem 5.1 to prove the results. Since $\hat{a}(z,\epsilon)$ is a smoothing of $\max(0,z)$, it is easy to verify that $\hat{r}_i(q_i, Q_{-i}, \xi, \epsilon)$ is a smoothing of r_i. Moreover, since $\hat{a}_z(z,\epsilon)$ is bounded by a constant ([50, Example 3.1]), then under Assumption 6.1, there exists an integrable function $\kappa_i(\xi)$ such that the Lipschitz modulus of $\hat{r}_i(q_i, Q_{-i}, \xi, \epsilon)$ with respect to q_i is bounded by $\kappa_i(\xi)$. Moreover, (6.41) implies that for almost every ξ,

$$\overline{\lim_{\mathbf{q}'\to\mathbf{q},\epsilon\to 0}} \{\nabla_{q_i}\hat{r}_i(q_i', q_{-i}', \xi, \epsilon)\} \subset \partial_{q_i} r_i(q_i, q_{-i}, \xi).$$

The conclusion follows. ■

6.4 Numerical tests

We carry out numerical tests on the proposed smoothing SAA scheme for solving the stochastic Nash equilibrium problem (6.35) with three generator signing both two-way and one-way contracts. We use mathematical programming codes in GAMS installed in a PC with Windows XP operating system and the built-in solver *path* for solving (6.44). Our tests are focused on different values of the smoothing parameter ϵ and sample size N.

Example 6.1 *Consider problem (6.35) with three generators which compete with each other in dispatching electricity. The inverse demand function is*

$$p(q, \xi) = \alpha(\xi) - \beta q,$$

where ξ is a random variable with uniform distribution on $[0, 1]$, β is deterministic (we set $\beta = 1$), and $\alpha(\xi)$ takes a form of $\alpha\xi + \alpha_0$ with $\alpha = 20$, $\alpha_0 = 30$ and $\beta = 1$. Table 1 displays generator's i the quantities of one-way contract w_i and two-way contract s_i, as well as generator i's cost function $C_i(q_i)$. The strike prices are: $f_1 = 22$ and $f_2 = \mathbb{E}[p(q, \xi)]$.

Table 1: Quantities of contracts and cost functions

Generator	w_i	s_i	$C_i(q_i)$
1	10	0	$q_1^2 + 2q_1$
2	8	0	$2q_2^2 + 2q_2$
3	0	6	$2q_3^2 + 3q_3$

We can solve the true problem analytically and obtain the exact equilibrium and other related quantities as displayed in Table 2.

Table 2: Exact result of Example 6.1.

True problem	(q_1^*, q_2^*, q_3^*)	Q^*	$\mathbb{E}[p(Q^*, \xi)]$	(R_1^*, R_2^*, R_3^*)
	$(8.300, 4.781, 4.987)$	18.067	21.933	$(71.886, 29.851, 44.677)$

We carry out tests for different parameter values $\epsilon = 2, 0.2, 0.02$ and sample sizes $N = 500, 1000, 5000$. The test results are displayed in Table 3. In the table, we use the following notation: $q_i^N(\epsilon)$ denotes generator i's decision on dispatch in the smoothed SAA Nash equilibrium with sample size N, $R_i^N(\epsilon)$ denotes generator i's profit in the smoothed SAA Nash equilibrium, and finally $Q^N(\epsilon)$ denotes the aggregate dispatch in the smoothed SAA Nash equilibrium; $\bar{p}(Q^N(\epsilon))$ denotes the average price in the smoothed SAA Nash equilibrium.

The results show that the convergence is not very sensitive to changes of the value of ϵ so long as ϵ is sufficiently small. This is consistent with the observations obtained in the literature. See [23, 50].

Example 6.2 *Consider Example 6.1. Assume now ξ follows a truncated normal distribution with mean value* 0.5 *and standard deviation* 1*, and truncated* 0.5 *above and below the mean value.*

Table 3: Numerical results of Example 6.1.

ϵ	N	$(q_1^N(\epsilon), q_2^N(\epsilon), q_3^N(\epsilon))$	$Q^N(\epsilon)$	$\bar{p}(Q^N(\epsilon))$	$(R_1^N(\epsilon), R_2^N(\epsilon), R_3^N(\epsilon))$
2	500	$(8.114, 4.680, 4.926)$	17.719	21.629	$(71.043, 30.209, 43.539)$
	1000	$(8.222, 4.739, 4.963)$	17.924	21.816	$(71.291, 29.788, 44.238)$
	5000	$(8.281, 4.771, 4.980)$	18.033	21.901	$(71.900, 29.969, 44.556)$
0.2	500	$(8.163, 4.706, 4.937)$	17.806	21.683	$(71.296, 30.191, 43.741)$
	1000	$(8.242, 4.751, 4.975)$	17.969	21.877	$(71.516, 29.800, 44.466)$
	5000	$(8.297, 4.779, 4.983)$	18.059	21.914	$(71.823, 29.845, 44.608)$
0.02	500	$(8.168, 4.708, 4.936)$	17.811	21.678	$(71.299, 30.164, 43.720)$
	1000	$(8.250, 4.755, 4.973)$	17.979	21.867	$(71.521, 29.805, 44.429)$
	5000	$(8.299, 4.780, 4.982)$	18.061	21.912	$(71.824, 29.846, 44.598)$

Assume also the strike price of the one-way contract is 27 *while all other parameters are the same.*

We carry out numerical tests for this example with fixed the smoothing parameter $\epsilon = 0.2$ and varying sample size $N = 500, 1000, 5000$. We do so for the study of convergence of the approximation with respect to the sample size. The results are displayed in Table 4.

Table 4: Numerical results of Example 6.2.

N	$(q_1^N(\epsilon), q_2^N(\epsilon), q_3^N(\epsilon))$	$Q^N(\epsilon)$	$\bar{p}(Q^N(\epsilon))$	$(R_1^N(\epsilon), R_2^N(\epsilon), R_3^N(\epsilon))$
500	$(7.648, 4.487, 5.079)$	17.214	22.397	$(89.612, 44.942, 46.923)$
1000	$(7.745, 4.539, 5.108)$	17.393	22.541	$(91.659, 46.073, 47.631)$
5000	$(7.755, 4.544, 5.108)$	17.407	22.542	$(91.739, 46.109, 47.638)$
10000	$(7.760, 4.547, 5.112)$	17.409	22.550	$(91.872, 46.166, 47.662)$

The results show that there is no significant improvement when the sample size in changed from 500 to 1000, 5000, or even 10000. This reflects the fast convergence of the sample average approximation.

7 Concluding remarks

In this paper, we present a comprehensive convergence analysis of sample average approximation method for a class of stochastic Nash equilibrium problems where the underlying functions are not necessarily continuously differentiable. The analysis is carried through the first order equilibrium conditions characterized in terms of Clarke generalized gradients. Almost sure convergence is established under the condition that the underlying functions are Lipschitz continuous with integrable modulus. Under the additional conditions that the expectation of the Clarke generalized gradients continuous, H-calm and metric regular, we show that with probability approaching one exponentially fast with the increase of sample size, an estimator of a Nash stationary point from SAA problem converges to its true counterpart.

While both almost convergence and exponential convergence cover a large class of practically important problems, we note that it may be interesting from theoretical point of view to extend the discussions to a more general classes of problems. For instance, we may consider the case when the underlying functions are lower semi-continuous as opposed to Lipschitz continuous. In such a case, we may use so-called general subgradient [31] rather than Clarke generalized gradient to characterize the first order equilibrium condition. As the former is generally unbounded, one cannot use the uniform SLLN for set-valued mapping to establish almost convergence. Wets and Xu [47] are recently developing graphical convergence of SAA random set-valued mapping and we expect that the new tool will effectively deal with this challenge.

It is also possible to strengthen the exponential convergence to the case where the expected value of the Clarke generalized gradient is merely piecewise continuous. We avoid this as it incurs a lot of very sophisticated technical details. This type of work is carried out in a separate work by Ralph and Xu in [28] for stochastic generalized equations. The convergence result can also be obtained by replacing the Clarke generalized gradients with its continuous approximation [49].

Acknowledgements: The first author would like to thank Professor Alexander Shapiro for his valuable comments on Proposition 4.1 (i). He is also grateful to Professor Werner Römisch for bringing to his attention of reference [15] and to an anonymous associate editor of SPEPS for valuable comments.

References

[1] Z. Artstein, On the calculus of closed set-valued functions, *Indiana University Mathematical Journal*, Vol. 24, pp. 433-441, 1974.

[2] Z. Artstein and R. A. Vitale, A strong law of large numbers for random compact sets, *The Annals of Probability*, Vol. 3, pp. 879-882, 1975.

[3] Z. Artstein and R. J-B Wets, Consistency of minimizers and the SLLN for stochastic programs, *Journal of Convex Analysis*, Vol. 2, pp. 1–17, 1995.

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

[5] R.J. Aumann, Integrals of set-valued functions, *Journal of Mathematical Analysis and Applications*, Vol. 12, PP. 1-12, 1965.

[6] C. Castaing and M. Valadier, convex analysis and measurable multifunctions, Lecture Notes in Mathematics, Vol 580, Springer, Berlin, 1977.

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

[8] L. Dai, C. H. Chen and J. R. Birge, Convergence properties of two stage stochastic programming, *Journal of Optimization Theory and Applications*, Vol. 106, pp. 489-509, 2000.

[9] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Springer-Verlag, New York, 1998.

[10] V. DeMiguel and H. Xu, A Stochastic Multiple Leader Stackelberg Model: Analysis, Computation, and Application, manuscripts, 2006.

[11] D. DeWolf and Y. Smeers, A stochastic version of a Stackelberg-Nash-Cournot equilibrium model, *Management Sciences*, Vol. 43, pp. 190-197, 1997.

[12] A. L. Dontchev, A. S. Lewis and R. T. Rockafellar, The radius of metric regularity, *Transactions of the American Mathematics Society*, Vol. 355, pp. 493-517, 2004.

[13] F. Facchinei and J-S Pang, *Finite-dimensional Variational Inequalities and Complementarity Problems*, Springer, 2003.

[14] 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.

[15] R. Henerion and W. Römisch, On M-stationary point for a stochastic equilibrium problem under equilibrium constraints in electricity spot market modeling, *Applications of Mathematics*, Vol. 52, pp.473-494, 2007.

[16] W.W. Hogan, Point-to-set maps in mathematical programming, *SIAM Review*, Vol. 15, pp.591-603, 1973.

[17] T. Homen-De-Mello, Estimation of derivatives of nonsmooth performance measures in regenerative systems, *Mathematics of Operations Research*, Vol. 26, pp. 741-768, 2001.

[18] A. J. King and R. T. Rockafellar, Sensitivity analysis for nonsmooth generalized equations, *Mathematical Programming*, Vol. 55, pp. 191-212, 1992.

[19] A. J. King and R. T. Rockafellar, Asymptotic theory for solutions in statistical estimation and stochastic programming, *Mathematics of Operations Research*, Vol. 18, pp. 148-162, 1993.

[20] A. J. King and R. J.-B. Wets, Epi-consistency of convex stochastic programs, *Stochastics Stochastics Reports*, Vol. 34, pp. 83–92, 1991.

[21] Y. M. Kaniovski, A. J. King and R. J.-B. Wets, Probabilistic bounds (via large deviation) for solutions of stochastic programming problems, *Annals of Operations Research*, Vol. 56, pp. 189-208, 1995.

[22] C. Lemaréchal, Bundle methods in nonsmooth optimization. In *Nonsmooth Optimization*, C. Lemaréchal and R. Mifflin (Eds), Pergamon Press, Oxford, pp. 79-102, 1978.

[23] J. Linderoth, A. Shapiro and S. Wright, The empirical behavior of sampling methods for stochastic programming, *Annals of Operations Research*, Vol. 142, pp. 215-241, 2006.

[24] B.S. Mordukhovich, *Variational Analysis and Generalized Differentiation, I: Basic Theory; II Applications*. Grundlehren Series (Fundamental Principles of Mathematical Sciences), Vol. 330 and 331, Springer, 2006.

[25] M. H. Ngo and V. Krishnamurthy, Game theoretic cross-layer transmission policies in multipacket reception wireless networks, *IEEE Transaction on Signal Processing*, Vol. 55, pp. 1911-1926, 2007.

[26] E. L. Plambeck, B. R. Fu, S. M. Robinson and R. Suri, Sample-path optimization of convex stochastic performances functions, *Mathematical Programming*, Vol. 75, pp. 137–176, 1996.

[27] D. Ralph and H. Xu, Implicit smoothing and its application to optimization with piecewise smooth equality constraints, *Journal of Optimization Theory and Applications*, Vol. 124, pp. 673-699, 2005.

[28] D. Ralph and H. Xu, Asympototic analysis of Stationary Points of Sample Average Two Stage Stochastic Programs: A Generalized Equation Approach, manuscript, 2008.

[29] S. M. Robinson, Regularity and stability for convex multivalued functions, *Mathematics of Operations Research*, Vol. 1, pp. 131-143, 1976.

[30] S. M. Robinson, Analysis of sample-path optimization, *Mathematics of Operations Research*, Vol. 21, pp. 513–528, 1996.

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

[32] W. Römisch and R. Schultz, Lipschitz stability for stochastic programs with complete recourse, *SIAM Journal on Optimization*, Vol. 6, pp. 531–547, 1996.

[33] J. B. Rosen, Existence and uniqueness of equilibrium points for concave N-person games, *Econometrica*, Vol. 33, pp. 520-534, 1965.

[34] R. Y. Rubinstein and A. Shapiro, *Discrete Events Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Methods*, John Wiley and Sons, New York, 1993.

[35] A. Rusczyński, A linearization method for nonsmooth stochastic programming problems, *Mathematics of Operations Research*, Vol. 12, pp. 32 - 49, 1987.

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

[37] S. Scholtes, *Introduction to piecewise smooth equations*, Habilitation Thesis, University of Karlsruhe, 1994.

[38] A. Shapiro, Quantitative stability in stochastic programming, *Mathematical Programming*, Vol. 67, pp. 99-108, 1994.

[39] A. Shapiro, Asymptotic properties of statistical estimators in stochastic programming. *Annals of Statistics*, Vol. 17, pp. 841-858, 1989.

[40] 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.

[41] A. Shapiro and T. Homem-de-Mello, On rate of convergence of Monte Carlo approximations of stochastic programs, *SIAM Journal on Optimization*, Vol. 11, pp. 70-86, 2000.

[42] A. Shapiro and H. Xu, Stochastic Mathematical Programs with Equilibrium Constraints, Modeling and Sample Average Approximation, *Optimization*, Vol. 57, pp. 395-418, 2008.

[43] A. Shapiro and H. Xu, Uniform Laws of Large Numbers for Set-Valued Mappings and Subdifferentials of Random Functions, *Journal of Mathematical Analysis and Applications*, Vol. 325, pp. 1390-1399, 2006.

[44] H. D. Sherali, A. L. Soyster and F. H. Murphy, Stackelberg-Nash-Cournot equilibria: characterizations and computations, *Operations Research*, Vol.31, pp. 253-276, 1983.

[45] D. Watling, User equilibrium traffic network assignment with stochastic travel times and late arrival penalty, *European Journal of Operational Research*, Vol. 175, pp. 1539-1556, 2006.

[46] R. Wets, Stochastic Programming, Handbooks in OR & MS, Vol. 1, G. L. Nemhauser et al eds, pp. 573-629, 1989.

[47] R. Wets and H. Xu, Graphical convergence of sample average random set-valued mappings, manuscript, 2008.

[48] H. Xu and F. Meng, Convergence analysis of sample average approximation methods for a class of stochastic mathematical programs with equality constraints, *Mathematics of Operations Research*, Vol. 32, pp. 648-668, 2007.

[49] H. Xu, A.M. Rubinov and B.M. Glover, Continuous approximations to generalized Jacobians, *Optimization*, Vol. 46, pp. 221–246, 1999.

[50] H. Xu and D. Zhang, Smooth Sample Average Approximation of Nonsmooth Stochastic Problems and Applications, *Mathematical Programming Series A*, 2008.

[51] D. Zhang, H. Xu and Y. Wu, A Stochastic EPEC Model for Electricity Markets with Two Way Contracts, To appear in *Mathematical Methods of Operations Research*.

Appendix

Proof of Proposition 4.1 part (i). By Cramér's Large Deviation (LD) Theorem [9], we have that for any $x \in \mathcal{X}$ and $\epsilon > 0$ it holds that

$$\text{Prob}\left\{\psi_N(x) - \psi(x) \geq \epsilon\right\} \leq e^{-NI_x(\epsilon)}, \tag{7.45}$$

where

$$I_x(z) := \sup_{t\in\mathbb{R}} \left\{zt - \log M_x(t)\right\}$$

is the LD rate function of random variable $\phi(x, \xi(\omega)) - \psi(x)$. By Assumption 4.1 (a), we have that $I_x(\epsilon)$ is positive for every $x \in \mathcal{X}$. For given $\epsilon > 0$, let $\nu > 0$ and $\bar{x}_1, ..., \bar{x}_K \in \mathcal{X}$ be such that for every $x \in \mathcal{X}$, there exists $\bar{x}_i$, $i \in \{1, ..., K\}$ such that

$$\|x - \bar{x}_i\| \leq \nu,$$

i.e., $\{\bar{x}_1, ..., \bar{x}_K\}$ is a ν-net in $\mathcal{X}$,

$$\phi(x, \xi) - \phi(\bar{x}_i, \xi) \leq \kappa(\xi)\|x - \bar{x}_i\|^\gamma \tag{7.46}$$

and

$$|\psi(x) - \psi(\bar{x}_i)| \leq \epsilon/4. \tag{7.47}$$

This can be achieved by applying the finite cover theorem on compact set $\mathcal{X}$ and the continuity of $\psi(x)$ on $\mathcal{X}$. Note that we can choose this net in such a way that $K \leq O(1)(D/\nu)^n$, where

$D := \sup_{x',x\in\mathcal{X}} \|x'-x\|$ is the diameter of $\mathcal{X}$ and $O(1)$ is a generic constant. By (7.46) we have that

$$\psi_N(x) - \psi_N(\bar{x}_i) \le \kappa_N \nu^\gamma,$$

for $\|x-\bar{x}_i\| \le \nu$, where $\kappa_N := N^{-1}\sum_{k=1}^N \kappa(\xi^k)$. Because the moment generating function $\mathbb{E}\left[e^{\kappa(\xi)t}\right]$ is finite valued for t close to 0, by Cramér's LD Theorem [9], we have that for any $L' > \mathbb{E}[\kappa(\xi)]$, there is a positive constant λ such that

$$\text{Prob}\left\{\kappa_N \ge L'\right\} \le e^{-N\lambda}. \tag{7.48}$$

Consider $Z_i := \psi_N(\bar{x}_i) - \psi(\bar{x}_i)$, $i = 1, ..., K$. We have that the event $\{\max_{1\le i\le K} Z_i \ge \epsilon\}$ is equal to the union of the events $\{Z_i \ge \epsilon\}$, $i = 1, ..., M$, and hence

$$\text{Prob}\left\{\max_{1\le i\le K} Z_i \ge \epsilon\right\} \le \sum_{i=1}^K \text{Prob}\left\{Z_i \ge \epsilon\right\}.$$

Together with (7.45) this implies that

$$\text{Prob}\left\{\max_{1\le i\le K} (\psi_N(\bar{x}_i) - \psi(\bar{x}_i)) \ge \epsilon\right\} \le \sum_{i=1}^K e^{-NI_{\bar{x}_i}(\epsilon)}.$$

For an $x \in \mathcal{X}$ let $i(x) \in \arg\min_{1\le i\le M} \|x-\bar{x}_i\|$. By construction of the ν-net we have that $\|x-\bar{x}_{i(x)}\| \le \nu$ for every $x \in \mathcal{X}$. Then

$$\begin{aligned}\psi_N(x) - \psi(x) &\le \psi_N(x) - \psi_N(\bar{x}_{i(x)}) + \psi_N(\bar{x}_{i(x)}) - \psi(\bar{x}_{i(x)}) + \psi(\bar{x}_{i(x)}) - \psi(x)\\ &\le \kappa_N\nu^\gamma + \psi_N(\bar{x}_{i(x)}) - \psi(\bar{x}_{i(x)}) + \epsilon/4.\end{aligned}$$

Then

$$\text{Prob}\left\{\sup_{x\in\mathcal{X}}(\psi_N(x) - \psi(x)) \ge \epsilon\right\} \le \text{Prob}\left\{\kappa_N\nu^\gamma + \max_{1\le i\le K}(\psi_N(\bar{x}_i) - \psi(\bar{x}_i)) \ge 3\epsilon/4\right\}.$$

Moreover, by (7.48) we have that

$$\text{Prob}\left\{\kappa_N\nu^\gamma \ge \epsilon/2\right\} \le e^{-N\lambda}$$

for some $\lambda > 0$ (by setting $\epsilon/(2\nu^\gamma) > \mathbb{E}[\kappa(\xi)]$), and hence

$$\begin{aligned}\text{Prob}\left\{\sup_{x\in\mathcal{X}}(\psi_N(x) - \psi(x)) \ge \epsilon\right\} &\le e^{-N\lambda} + \text{Prob}\left\{\max_{1\le i\le K}(\psi_N(\bar{x}_i) - \psi(\bar{x}_i)) \ge \epsilon/4\right\}\\ &\le e^{-N\lambda} + \sum_{i=1}^K e^{-NI_{\bar{x}_i}(\epsilon/4)}.\end{aligned} \tag{7.49}$$

Since the above choice of the ν-net does not depend on the sample (although it depends on ϵ), and $I_{\bar{x}_i}(\epsilon/4)$ is positive, $i = 1, ..., K$, we obtain that (7.49) implies (4.17), and hence complete the proof. ■

Proof of Lemma 6.1. The first part of the conclusion follows from Proposition 6.1 (i). In what follows, we show that $\mathbb{E}[\nabla_{\mathbf{q}} F(\mathbf{q},\xi)]$ is a nonsingular matrix under Assumption 6.1 for all $q_i, Q_{-i} \ge 0$. From the definition of $r_i(q_i, Q_{-i}, \xi)$, we have for $i = 1, 2, \dots, M$,

$$(r_i)'_{q_i}(q_i, Q_{-i}, \xi) = \begin{cases} p(Q,\xi) - C_i'(q_i) + (q_i - s_i)p_q'(Q,\xi), & \text{if } p(Q,\xi) < f_1;\\ p(Q,\xi) - C_i'(q_i) + (q_i - w_i - s_i)p_q'(Q,\xi), & \text{if } p(Q,\xi) > f_1.\end{cases}$$

Hence, for $j \neq i$,

$$(r_i)''_{q_iq_j}(q_i, Q_{-i}, \xi) = \begin{cases} p'_q(Q,\xi) + (q_i - s_i)p''_{qq}(Q,\xi), & \text{if } p(Q,\xi) < f_1; \\ p'_q(Q,\xi) + (q_i - w_i - s_i)p'_{qq}(Q,\xi), & \text{if } p(Q,\xi) > f_1, \end{cases}$$

and $j = i$,

$$(r_i)''_{q_iq_i}(q_i, Q_{-i}, \xi) = \begin{cases} 2p'_q(Q,\xi) - C''_i(q_i) + (q_i - s_i)p''_{qq}(Q,\xi), & \text{if } p(Q,\xi) < f_1; \\ 2p'_q(Q,\xi) - C''_i(q_i) + (q_i - w_i - s_i)p'_{qq}(Q,\xi), & \text{if } p(Q,\xi) > f_1. \end{cases}$$

Therefore, we can write the matrix $\mathbb{E}[\nabla_{\mathbf{q}} F(\mathbf{q},\xi)]$ as the sum of two conditional expectations as

$$\begin{aligned} \mathbb{E}[\nabla_{\mathbf{q}} F(\mathbf{q},\xi)] &= \mathbb{E}[\nabla_{\mathbf{q}} F(\mathbf{q},\xi)|p(Q,\xi) < f_1]\text{Prob}\,(p(Q,\xi) < f_1) \\ &\quad + \mathbb{E}[\nabla_{\mathbf{q}} F(\mathbf{q},\xi)|p(Q,\xi) \geq f_1]\text{Prob}\,(p(Q,\xi) \geq f_1)\,. \end{aligned}$$

Since the set $\{\xi|\text{Prob}\,(p(Q,\xi) = f_1)\}$ has measure zero, we have

$$\begin{aligned} \mathbb{E}[\nabla_{\mathbf{q}} F(\mathbf{q},\xi)] &= \mathbb{E}[\nabla_{\mathbf{q}} F(\mathbf{q},\xi)|p(Q,\xi) < f_1]\text{Prob}\,(p(Q,\xi) < f_1) \\ &\quad + \mathbb{E}[\nabla_{\mathbf{q}} F(\mathbf{q},\xi)|p(Q,\xi) > f_1]\text{Prob}\,(p(Q,\xi) > f_1)\,. \end{aligned}$$

Let

$$\begin{aligned} A_i(q_i, Q_{-i}) = &\ \mathbb{E}[p'_q(Q,\xi) + (q_i - s_i)p''_{qq}(Q,\xi)]\text{Prob}\,(p(Q,\xi) < f_1) \\ &+ \mathbb{E}[p'_q(Q,\xi) + (q_i - w_i - s_i)p''_{qq}(Q,\xi)]\text{Prob}\,(p(Q,\xi) > f_1) \end{aligned}$$

and

$$\begin{aligned} B_i(q_i, Q_{-i}) = &\ \mathbb{E}[2p'_q(Q,\xi) - C''_i(q_i) + (q_i - s_i)p''_{qq}(Q,\xi)]\text{Prob}\,(p(Q,\xi) < f_1) \\ &+ \mathbb{E}[2p'_q(Q,\xi) - C''_i(q_i) + (q_i - w_i - s_i)p''_{qq}(Q,\xi)]\text{Prob}\,(p(Q,\xi) > f_1)\,. \end{aligned}$$

From Assumption 6.1 and the convexity of $p(\cdot,\xi)$, we have

$$A_i(q_i, Q_{-i}) \leq \mathbb{E}[p'_q + q_i p''_{qq}]\text{Prob}\,(p(Q,\xi) < f_1) + \mathbb{E}[p'_q + q_i p''_{qq}]\text{Prob}\,(p(Q,\xi) > f_1) \leq 0$$

and

$$B_i(q_i, Q_{-i}) = A_i(q_i, Q_{-i}) + \mathbb{E}[p'_q(Q,\xi)] - C''_i(q_i) < 0.$$

Hence, the matrix $\mathbb{E}[\nabla_{\mathbf{q}} F(\mathbf{q},\xi)]$ can be formulated as

$$\begin{pmatrix} B_1(q_1,Q_{-1}) & A_2(q_2,Q_{-2}) & \cdots & A_{M-1}(q_{M-1},Q_{-(M-1)}) & A_M(q_M,Q_{-M}) \\ A_1(q_1,Q_{-1}) & B_2(q_2,Q_{-2}) & \cdots & A_{M-1}(q_{M-1},Q_{-(M-1)}) & A_M(q_M,Q_{-M}) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ A_1(q_1,Q_{-1}) & A_2(q_2,Q_{-2}) & \cdots & B_{M-1}(q_{M-1},Q_{-(M-1)}) & A_M(q_M,Q_{-M}) \\ A_1(q_1,Q_{-1}) & A_2(q_2,Q_{-2}) & \cdots & A_{M-1}(q_{M-1},Q_{-(M-1)}) & B_M(q_M,Q_{-M}) \end{pmatrix}$$

and its determinant is the same as the following matrix

$$\begin{pmatrix} \mathbb{E}[p'_q] - C''_1(q_1) & -\mathbb{E}[p'_q] + C''_2(q_2) & 0 & \cdots & 0 & 0 \\ 0 & \mathbb{E}[p'_q] - C''_2(q_2) & -\mathbb{E}[p'_q] + C''_3(q_3) & \cdots & 0 & 0 \\ 0 & 0 & \mathbb{E}[p'_q] - C''_3(q_3) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \mathbb{E}[p'_q] - C''_{M-1}(q_{M-1}) & -\mathbb{E}[p'_q] + C''_M(q_M) \\ A_1(q_1,Q_{-1}) & A_2(q_2,Q_{-2}) & A_3(q_3,Q_{-3}) & \cdots & A_{M-1}(q_{M-1},Q_{-(M-1)}) & B_M(q_M,Q_{-M}) \end{pmatrix}$$

Hence, the determinant equals to

$$|\mathbb{E}[\nabla_{\mathbf{q}} F(\mathbf{q},\xi)]| = \sum_{i=1}^{M} \left[A_i(q_i, Q_{-i}) \prod_{j \neq i} \left(\mathbb{E}[p_q'(Q,\xi)] - C_j''(q_j) \right) \right] + \prod_{i=1}^{M} \left(\mathbb{E}[p_q'(Q,\xi)] - C_i''(q_i) \right).$$

From Assumption 6.1, we have, for every $i = 1, 2, \ldots, M$, $p_q'(Q,\xi) < 0$ and $C_i''(q) \geq 0$ for any fixed $Q, q \geq 0$ and hence $\mathbb{E}[p_q'] - C_i'' < 0$. Therefore we can rewrite $|\mathbb{E}[\nabla_{\mathbf{q}} F(\mathbf{q},\xi)]|$ as

$$|\mathbb{E}[\nabla_{\mathbf{q}} F(\mathbf{q},\xi)]| = (-1)^M \sum_{i=1}^{M} |A_i(q_i, Q_{-i})| \prod_{j \neq i} \left| \mathbb{E}[p_q'(Q,\xi)] - C_j''(q_j) \right| + (-1)^M \prod_{i=1}^{M} \left| \mathbb{E}[p_q'(Q,\xi)] - C_i''(q_i) \right|.$$

Moreover, since for $i = 1, 2, \ldots, M$, $\mathbb{E}[p_q'] - C_i'' < 0$, we have $|\mathbb{E}[\nabla_{\mathbf{q}} F(\mathbf{q},\xi)]| \neq 0$ and hence $\mathbb{E}[\nabla_{\mathbf{q}} F(\mathbf{q},\xi)]$ is nonsingular. ■