Stochastic Nash Equilibrium Problems: Sample Average Approximation and Applications

Huifu Xu[1] and Dali Zhang[1]

(May 2008, revised January 2010)

Abstract

This paper presents a Nash equilibrium model where the underlying objective functions involve uncertainty 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 (a Nash equilibrium or a Nash-C-stationary point) obtained from sample average approximate equilibrium problem converges to its true counterpart. Moreover, under some calmness conditions of the Clarke generalized derivatives, it is shown that with probability approaching one exponentially fast by increasing sample size, the Nash-C-stationary point converges to a weak Nash-C-stationary point of the true problem. Finally, the model is applied to stochastic Nash equilibrium problem in the wholesale electricity market.

Key words. Stochastic Nash equilibrium, exponential convergence, H-calmness, Nash-C-stationary point.

1 Introduction

Let X_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 $X_{-i} = X_1 \times \cdots \times X_{i-1} \times X_{i+1} \times \cdots \times X_{\hat{i}}$ where "$\times$" denotes Cartesian product. 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

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

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 with respect to the distribution of the random vector ξ. 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. To ease notation, we will use ξ to denote either the random vector $\xi(\omega)$ or an element of $\mathbb{R}^k$, depending on the context.

Nash equilibrium models have been well studied and have found many applications in economics and engineering, see for instances [16, 40, 20] for recent developments on the topic. Our Nash equilibrium model (1.1) 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 nonsmoothness of v_i allows us to broaden the scope of the model (1.1) to include the following *two stage stochastic Nash equilibrium* problem

$$\begin{aligned} \min_{x_i \in X_i, y(\cdot)} \quad & \mathbb{E}[f_i(x_i, x_{-i}, y(\omega), \xi(\omega))] \\ \text{s.t.} \quad & y(\omega) \in \mathcal{F}(x, \xi(\omega)), \quad \text{a.e.} \quad \omega \in \Omega, \end{aligned} \tag{1.2}$$

where f_i is a continuously differentiable function of x_i, x_{-i}, y and ξ, and $\mathcal{F}(x, \xi(\omega))$ is a closed subset of $\mathbb{R}^m$. To see this, let $v_i(x_i, x_{-i}, \xi)$ denote the optimal value function of the following second stage problem:

$$\begin{aligned} \min_y \quad & f_i(x_i, x_{-i}, y, \xi) \\ \text{s.t.} \quad & y \in \mathcal{F}(x, \xi). \end{aligned} \tag{1.3}$$

---

[1]School of Mathematics, University of Southampton, Highfield Southampton, UK. Email: h.xu@soton.ac.uk, dali.zhang.ustc@gmail.com.

It is well known (see e.g. [30, 35]) that under some metric regularity conditions of $\mathcal{F}(x, \xi)$, $v_i(x_i, x_{-i}, \xi)$ is locally Lipschitz continuous. Moreover that under some moderate conditions (1.2) is equivalent to (1.1) and (1.3), see [49, 56]. A particular interesting case is when $\mathcal{F}(x, \xi(\omega))$ is the solution set of a parametric generalized equation, that is,

$$\mathcal{F}(x, \xi) = \{y \in \mathbb{R}^m : 0 \in H(x, y, \xi) + \mathcal{N}_Q(y)\}, \tag{1.4}$$

where Q is a closed subset in $\mathbb{R}^m$ and $\mathcal{N}_Q(y)$ denotes the Clarke normal cone to set Q at point y. In such a case, we call (1.2) the *two stage stochastic equilibrium program with equilibrium constraints* (SEPEC) in that the generalized equation in (1.4) often represents an equilibrium in practical applications. Another interesting case is when $\mathcal{F}(x, \xi)$ is presented by a system of equalities and inequalities, that is,

$$\mathcal{F}(x, \xi) = \{y \in \mathbb{R}^m : h(x, y, \xi) = 0, g(x, y, \xi) \leq 0\}, \tag{1.5}$$

where h and g are some continuous vector-valued functions. In such a case we call the corresponding Nash equilibrium problem (1.2) the *two stage stochastic generalized Nash equilibrium* (SGNE) problem. Here "generalized" is used to distinguish the Nash model where players strategy spaces are independent. Mathematically, equalities and inequalities can be recovered from the generalized equation in (1.4) by considering a specific set Q, which means SEPEC subsumes SGNE. In Section 6, we will discuss applications of the stochastic Nash equilibrium model (1.1) and the two stage SEPEC model (defined by (1.2) and (1.4)) in electricity markets.

Note that stochastic Nash equilibrium/game is not a new concept. Over the past few decades, various stochastic Nash equilibrium models have been proposed to study specific practical decision making problems which involve random data as well as multiple decision makers who are in a competitive relationship. For instance, Watling [58] proposed a stochastic Nash equilibrium model in transportation and Ngo and Krishnamurthy [37] proposed a stochastic Nash equilibrium model for signal transmission in wireless networks [37]. Haurie, Zaccour and Smeers [20] introduced an S-adapted open-loop equilibrium model which is essentially a two stage SGNE. In a more recent development, DeMiguel and Xu proposed a two stage multiple-leader stochastic Stackelberg Nash-Cournot models for future market competition [14], Henrion and Römisch [22], and Zhang, Xu and Wu [65] proposed two stage stochastic equilibrium program with equilibrium constraints (SEPEC) models for electricity markets [22, 65]. A broader literature review of the subject of stochastic equilibrium/game may also include Harsanyi's Bayesian equilibrium model [19], Aumann's correlated equilibrium model [7], Shubik and Sobel's Markov perfect equilibrium model and Klemperer and Meyer's supply function equilibrium model [32] although they may have different mathematical formulations.

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 the Monte Carlo simulation based method. The basic idea of the method is to generate an independent and 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 $i = 1, \cdots, \hat{i}$, and consequently we may consider the following sample average approximate Nash equilibrium problem: find $x^N := (x_1^N, \cdots, x_{\hat{i}}^N) \in X_1 \times \cdots \times X_{\hat{i}}$ such that

$$\vartheta_i^N(x_i^N, x_{-i}^N) = \min_{x_i \in X_i} \vartheta_i^N(x_i, x_{-i}^N), \text{ for } i = 1, \cdots, \hat{i}. \tag{1.6}$$

We refer to (1.1) as the *true problem* and (1.6) as the *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 optimization and it is also known as Sample Path Optimization (SPO) method [43]. There has been extensive literature on SAA and SPO. See recent work [3, 29, 45, 54, 11, 34, 60, 56] and a comprehensive review by Shapiro in [53].

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. This includes two cases: (a) x^N is Nash equilibrium of (1.6), (b) x^N is a Nash stationary point (to be defined in Section 3) of (1.6). The notion of Nash stationary point is proposed by Hu and Ralph [26] (they called Nash-stationary equilibrium) for modeling a bilevel game in an electricity market.

There are essentially two ways to carry out the convergence 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 [53] 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 (1.6). This approach has also been used recently in stochastic programming for analyzing convergence of stationary points of sample average optimization problems, see for instance [55, 60, 61]. 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 [29].

In this paper, we investigate the convergence of $\{x^N\}$ through the first order equilibrium condition of the true problem (1.1) rather than through (1.1) itself because the latter involves $\hat{i}$ stochastic optimization problems where a decision variable of one problem becomes a parameter of another problem while the first order equilibrium conditions can be put under a unified framework of generalized equations with x_i, $i = 1, \cdots, \hat{i}$ being treated equally 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 (1.6) reduces to a variational inequality problem or a nonlinear complementarity problem [14].

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) ([47, 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 outer semicontinuous random compact set-valued mappings [55, 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 [55, 61].

Almost sure convergence does not address the rate of convergence and exponential convergence does. A popular way for the latter is to use the well-known Cramer's theorem in large deviation theory [13] 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 focus has been largely on optimal solutions and/or optimal values. See for instance [31, 45, 54, 53, 11, 56] 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 [51, 45]. More recently, Homem-de-Mello [27] presented a strong exponential convergence of the SAA optimal solution to its true counterpart without a second order growth condition.

In this paper, the underlying functions of (1.1) and (1.6) 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. Since generalized gradients are usually set-valued mappings, the convergence results in the literature we have reviewed cannot be used in our setting. Consequently our analysis will be carried out through some kind of uniform semi-convergence of sample average random mapping. 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 derive a uniform exponential convergence for sample average H-calm random functions, Proposition 4.1, and then apply it to

the support function of random set-valued mappings and establish one sided uniform exponential convergence of sample average random set-valued mappings through Hömander's formula (Theorem 4.3). By using a sensitivity result (Lemma 4.1), we translate the uniform set-valued convergence (Theorem 4.3) into to that of the SAA Nash-stationary points (Theorem 4.4).

As far as we are concerned, the main contributions of this paper can be summarized as follows:

(a) We propose a general stochastic Nash equilibrium model where the underlying function may be nonsmooth and nonconvex and apply the well-known sample average approximation method to solve it. We establish two types of convergence: almost sure convergence and exponential convergence for the Nash equilibrium estimator obtained from sample average Nash equilibrium problem. The a.s. convergence consists of two parts: a.s. convergence of Nash-C-stationary points (Theorem 4.1) and a.s. convergence of Nash equilibrium (Theorem 4.2) both of which are established under moderate conditions.

The exponential convergence (Theorem 4.4) addresses the rate of convergence of Nash-C-stationary points. As far as we are concerned, this is the first result which shows the exponential convergence of the SAA estimators of the Nash stationary points of nonsmooth stochastic Nash equilibrium problems. This result is established through the uniform exponential semi-convergence results, Proposition 4.1, under H-calmness. The latter is an extension of [42, Lemma 3.5], a core result that Ralph and Xu used to establish exponential convergence of SAA generalized equations ([42, Theorem 3.6]). Note that [42, Theorem 3.6] is obtained under some metric regularity type condition while our main convergence result, Theorem 4.4, is obtained without such a condition.

There are a couple of other papers that may be related to the exponential convergence results in this paper. Shapiro and Xu [56, Theorem 5.1] established a uniform exponential convergence of a class of Hölder continuous random functions. This result has been widely used to analyze SAA optimal solutions, stationary points and equilibria where the underlying functions are single valued and Hölder continuous. Obviously [56, Theorem 5.1] cannot be used to analyze Clarke stationary points/equilibria because the underlying functions of the latter are *not* Hölder continuous in general. Homem-De-Mello [27] also presented an exponential result for SAA optimal solutions under Lipschitz continuity, see e.g. [27, Theorem 2.3]. His result makes two interesting contributions: (i) the sampling is not necessarily i.i.d.; (ii) the metric regularity is not needed to obtain the exponential convergence of SAA optimal solution to its true counterpart. Our Theorem 4.4 enjoys the advantage of [27, Theorem 2.3] in the sense of (ii) but it has strengthened or complemented the latter in the sense that it deals with stationary points characterized by Clarke generalized gradients which are not necessarily continuous.

Note that DeMiguel and Xu [14] investigated a stochastic multiple-leaders Stackelberg game which can be essentially reformulated as (1.1). The objective functions in the decision problems there are convex under some moderate conditions and the analysis concerns the exponential convergence of SAA equilibrium by exploiting the Hausdorff continuity of epsilon-convex subdifferentials of a convex function. The same analysis cannot be used in this paper where the underlying functions are not necessarily convex.

To summarize, all of our convergence results are new in the sense that no existing convergence result of SAA for optimization problems can be applied to the stochastic Nash equilibrium problem although some of the tools used to establish these convergence might have some links to our previous work. The analysis tool for a.s. convergence (Theorem 4.1) is not new. This was first proposed in Xu and Meng's paper ([60, Lemma 3.2]). The tool for exponential convergence (Theorem 4.4) is Proposition 3.1. As indicated, Proposition 3.1 (i) is from Ralph and Xu's paper ([42, Lemma 3.5]) and Proposition 3.1 (iii) is Shapiro and Xu's [56, Theorem 5.1]. However Proposition 3.1 gives a unified treatment of uniform exponential convergence of sample average random functions under H-calmness. In other words the proposition is not entirely new but it has consolidated the previous results. The main convergence result Theorem 4.4 is novel. This is the strongest convergence result as far as we know: it gives exponential rate for the convergence of SAA stationary sequence $\{x^N\}$ to its true counterpart x^* without metric regularity type condition.

(b) We propose a smoothing model for solving the SAA Nash equilibrium problem coupled with convergence analysis. The smoothing scheme is not new but the proposed smoothing SAA scheme is very relevant in the sense that the complicated nonsmooth stochastic Nash equilibrium problem can now be approximated by an ordinary variational inequality for which many powerful numerical methods have been proposed over the past couple of decades [17].

(c) Finally we model the competition in the 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 [21] 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, $\mathbb{H}(D_1, D_2) := \max\left(\mathbb{D}(D_1, D_2), \mathbb{D}(D_2, D_1)\right)$. 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 $\mathcal{N}_D(x)$ 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 *([8, Theorem II-18]) Let D_1, D_2 be two convex and compact subsets of $\mathbb{R}^m$. Let $\sigma(D_1, u)$ and $\sigma(D_2, u)$ denote the support functions of D_1 and D_2 respectively. Then*

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

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 $\bar{y} \in F(x)$. 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 *outer semicontinuous* ([44, Definition 5.4]) at $\bar{x} \in \mathcal{X}$ if

$$\overline{\lim_{x \to \bar{x}}} F(x) \subset F(\bar{x})$$

where

$$\overline{\lim_{x \to \bar{x}}} F(x) = \{u : \exists x^\nu \to \bar{x}, \exists u^\nu \to \bar{u} \text{ with } u^\nu \in S(x^\nu)\}.$$

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 [44, 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. It is well-known that 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 [21] as it was first studied comprehensively by Aumann in [6]. $\mathbb{E}[F(x, \xi(\omega))]$ 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* [6, 21].

2.3 Clarke generalized gradients of a random function

We are interested in the cases when the integrand functions in (1.1) are Lipschitz continuous. Let $f(\cdot, \xi(\cdot))$: $\mathbb{R}^n \times \Omega \to \mathbb{R}$ 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 [9] of $f(x, \xi)$ with respect to x at point $x \in \mathbb{R}^n$ is defined as

$$\partial_x f(x, \xi) := \mathrm{conv}\big\{ \lim_{y \in D_{f(\cdot,\xi)}, y \to x} \nabla_x f(y, \xi) \big\},$$

where $D_{f(\cdot,\xi)}$ denotes the set of points near x where $f(x, \xi)$ is Fréchet 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 outer semicontinuous [9, 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 (1.6) 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 [49] 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:

$$0 \in \partial_{x_i} \mathbb{E}[v_i(x_i, x_{-i}, \xi)] + \mathcal{N}_{X_i}(x_i), \ i = 1, \cdots, \hat{i}. \tag{3.7}$$

Here and later on, the addition of sets is in the sense of Minkowski. We call a point x^* satisfying (3.7) a *stochastic Nash-C-stationary point*. As we discussed in the introduction, the notion of Nash-stationary point/equilibrium is introduced by Hu and Ralph [26]. Obviously if x^* is stochastic Nash equilibrium, then it must satisfy (3.7) and hence it is a stochastic Nash-C-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 Nash-C-stationary point is a Nash equilibrium. Note that

$$\partial_{x_i} \mathbb{E}[v_i(x_i, x_{-i}, \xi)] \subset \mathbb{E}[\partial_{x_i} v_i(x_i, x_{-i}, \xi)]. \tag{3.8}$$

The equality holds when v_i is Clarke regular [9] in x_i, see for instance [9, Theorem 2.7.2]. Consequently, we may consider a weaker condition than (3.7)

$$0 \in \mathbb{E}[\partial_{x_i} v_i(x_i, x_{-i}, \xi)] + \mathcal{N}_{X_i}(x_i), \ i = 1, \cdots, \hat{i}. \tag{3.9}$$

We call (3.9) *weak* Clarke first order equilibrium condition of (1.1) and a point satisfying (3.9) a *weak stochastic Nash-C-stationary point*. Here "weak" is in the sense that a stochastic Nash-C-stationary point is a weak stochastic Nash-C-stationary point but not vice versa.

Using the Clarke generalized gradient, we can also characterize the first order equilibrium condition of the sample average approximation equilibrium (1.6) as follows:

$$0 \in \frac{1}{N} \sum_{k=1}^{N} \partial_{x_i} v_i(x_i, x_{-i}, \xi^k) + \mathcal{N}_{X_i}(x_i), \ i = 1, \cdots, \hat{i}. \tag{3.10}$$

We call a point x^N satisfying (3.10) an *SAA Nash-C-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 Nash-C-stationary point. Our general idea to investigate some kind of uniform almost sure and exponential convergence of $\frac{1}{N} \sum_{k=1}^{N} \partial_{x_i} v_i(x_i, x_{-i}, \xi^k)$ to $\mathbb{E}[\partial_{x_i} v_i(x_i, x_{-i}, \xi)]$ over a compact subset of X. The main challenges arise from the set-valued nature of the SAA mappings.

Remark 3.1 *When $v_i(x_i, x_{-i}, \xi)$ is Clarke regular [9] with respect to x_i, $i = 1, \cdots, \hat{i}$, equality in (3.8) holds and hence weak stochastic Nash-C-stationary point is a stochastic Nash-C-stationary point. This includes*

two important cases: (a) v_i is continuously differentiable in x_i; (b) v_i is convex in x_i (when v_i, $i = 1, \cdots, \hat{i}$, is convex in x_i, these stationary points coincide with Nash equilibria). In this paper, our discussion is not restricted to the two cases. Note that without convexity, the existence of Nash equilibrium might be a problem. This is why we consider Nash-stationary point, a notion which is used by Hu and Ralph [26] in a deterministic Nash game, see also Remark 4.1 regarding the existence of equilibrium.

Let us now comment on the relevance of weak Nash equilibrium condition (without C-regularity) and the relationship between weak Nash-C-stationary point and a Nash-C-stationary point. Note first that the SAA Nash-C-stationary points converge to a weak Nash-C-stationary point, see Theorem 4.1. Note also that in many practical instances, v_i is smooth (continuously differentiable) for almost every ξ at "almost every" point, see Remark 4.3. At a "smooth" point, one has $\mathbb{E}[\partial_{x_i} v_i] = \partial_{x_i} \mathbb{E}[v_i] = \nabla_{x_i} \mathbb{E}[v_i]$. Therefore, there is no difference between Nash-C-stationarity and weak Nash-C-stationarity at such a point. At a "nonsmooth" point where $\mathbb{E}[\partial_{x_i} v_i]$ is larger than $\partial_{x_i} \mathbb{E}[v_i]$, the difference between the two stationary points relies on the local sensitivity at the set of weak Nash-C-stationary points with respect to the difference of the two set-valued mappings. Metric regularity [44, 15] is perhaps a proper concept to address this issue.

Note also that many practical problems are "piecewise convex". If we may obtain a weak Nash-C-stationary point where $\mathbb{E}[v_i(\cdot, x_{-i}, \xi)]$ is locally convex at x_i, then x is a local Nash equilibrium.

In conclusion, a weak Nash-C-stationary condition is used as a convenient and unified mathematical framework to describe the optimality conditions of our problem at both "smooth" and "nonsmooth" points.

Next, we look into the outer semicontinuity 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 $v_i(x_i, x_{-i}, \xi)$, $i = 1, \cdots, \hat{i}$, 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 Lipschitz 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, x_{-i}) on space $X_i \times X_{-i}$.

Corollary 3.1 *Under Assumption 3.1, $\partial_{x_i} v_i(x_i, x_{-i}, \xi)$ is outer semicontinuous with respect to (x_i, x_{-i}).*

Proof. Recall that a set-valued mapping is closed on its domain if and only if its graph is closed, see Section 1.3 in [4]. On the other hand, [44, Theorem 5.7] (a) states that a set-valued mapping is outer semicontinuous if and only if its graph is closed, in other words, the set-valued mapping is closed on its domain. Using these arguments, the conclusion then follows straightforwardly under Assumption 3.1 (b). ∎

Finally, we state the measurability and integrability of the set-valued mapping $\partial_{x_i} v_i(x_i, x_{-i}, \xi(\cdot)) : \Omega \to 2^{\mathbb{R}^m}$ in the following proposition. The proof is standard, we move it to the appendix.

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.*

4   Convergence analysis

In this section, we analyze the convergence of a sequence of SAA Nash stationary points $\{x^N\}$ defined by (3.10). 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 (3.9). Second, under additional conditions, 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 by increasing the sample size N, $\{x^N\}$ converges to a weak Nash-C-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 confusion because both x_i and x_{-i} are treated as variables in the analysis. Let

$$\mathcal{A}v(x, \xi) := \partial_{x_1} v_1(x, \xi) \times \cdots \times \partial_{x_{\hat{i}}} v_{\hat{i}}(x, \xi) \tag{4.11}$$

and

$$G_X(x) := \mathcal{N}_{X_1}(x_1) \times \cdots \times \mathcal{N}_{X_{\hat{i}}}(x_{\hat{i}}).\tag{4.12}$$

The first order equilibrium condition (3.7) can be written as

$$0 \in \mathbb{E}[\mathcal{A}v(x,\xi)] + G_X(x),\tag{4.13}$$

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 (3.10) can be written as

$$0 \in \mathcal{A}\vartheta^N(x) + G_X(x),\tag{4.14}$$

where

$$\mathcal{A}\vartheta^N(x) := \frac{1}{N}\sum_{k=1}^{N}\partial_{x_1}v_1(x,\xi^k) \times \cdots \times \frac{1}{N}\sum_{k=1}^{N}\partial_{x_{\hat{i}}}v_{\hat{i}}(x,\xi^k).\tag{4.15}$$

Obviously, $\mathbb{E}[\mathcal{A}\vartheta^N(x)] = \mathbb{E}[\mathcal{A}v(x,\xi)]$.

Remark 4.1 *For the simplicity of discussion, we make a blanket assumption that (3.10) has a solution for all N. In doing so, we do not restrict the integrand $v_i(x_i, x_{-i}, \xi)$ to be convex or strategy space X_i to be bounded. Further discussion on the existence issue requires sensitivity analysis of generalized equation (4.13). For instance, Shapiro [53, Section 7] investigated a similar issue under the context of stochastic generalized equation (SGE). Under the condition that the underlying function in SGE is continuously differentiable (corresponding to our case that v_i is twice continuously differentiable) and the true SGE satisfies a strong metric regularity condition, he showed the existence of a solution to the SAA-SGE, see [53, Theorem 22] for details.*

Note that in deterministic cases, King and Rockafellar [28] 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 [28] for details.

Note finally that if $X = X_1 \times \cdots \times X_{\hat{i}}$ is compact, then the set of solutions to the generalized equations (4.13) and (4.15) are nonempty. This follows easily from [5, Theorem 9.9] in that the set-valued mappings $\mathbb{E}[\mathcal{A}v(\cdot,\xi)]$, $\mathcal{A}\vartheta^N(\cdot)$ are compact and convex valued; X is a convex set, and $G_X(\cdot)$ is outer semicontinuous on X. The same argument is used by Outrata to show the existence of Clarke stationary points in deterministic equilibrium programs with equilibrium constraints (EPEC), see [39, Theorem 3.3].

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, recently established by Shapiro and Xu [55] to set-valued mapping $\mathcal{A}v(x,\xi)$. This approach has been used in nonsmooth stochastic optimization in [55, 61]. Here we use the approach to an equilibrium problem. Note that a key condition we need to impose is that the sequence of SAA Nash-C-stationary points $\{x^N\}$ is bounded w.p.1. Various conditions may lead to this, e.g., X is a compact set; v_i is coercive in x_i uniformly.

Theorem 4.1 *Let x^N be a solution of (3.10) 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 (3.9).*

Proof. Under Assumption 3.1, $\mathcal{A}v(\cdot,\xi)$ is outer semicontinuous on X and

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

where $\mathbb{E}[\sum_{i=1}^{\hat{i}} \kappa_i(\xi)] < \infty$. Let $\delta > 0$ be any fixed positive number. Applying [55, Theorem 2] to $\mathcal{A}\vartheta^N(x)$ on the compact set $\mathcal{X}$,

$$\varlimsup_{N\to\infty} \mathcal{A}\vartheta^N(x) \subset \mathbb{E}\left[\text{conv } \mathcal{A}^\delta v(x,\xi)\right], \text{ w.p.1} \tag{4.16}$$

uniformly for $x \in X$, where $\mathcal{A}^\delta v(x,\xi) := \bigcup_{x' \in B(x,\delta)} \mathcal{A}v(x',\xi)$. Let x^* be an accumulation point of $\{x^N\}$ and assume (by taking a subsequence if necessary) that $\{x^N\} \to x^*$. Then (4.16) implies

$$\varlimsup_{N\to\infty} \mathcal{A}\vartheta^N(x^N) \subset \mathbb{E}\left[\text{conv } \mathcal{A}^\delta v(x^*,\xi)\right], \text{ w.p.1}.$$

This and the outer semicontinuity of $G_X(\cdot)$ further imply

$$0 \in \mathbb{E}\left[\text{conv } \mathcal{A}^\delta v(x^*,\xi)\right] + G_X(x^*).$$

Note first that $\mathcal{A}^\delta v(x^*,\xi)$ is integrably bounded by $\sum_{i=1}^{\hat{i}} \kappa_i(\xi)$ and $\|\mathcal{A}^\delta v(x^*,\xi)\|$ is decreasing on δ, therefore $\|\mathcal{A}^\delta v(x^*,\xi)\|$ is uniformly integrable. Moreover, $\mathcal{A}^\delta v(x^*,\xi)$ is closed. To see this, let $\{\eta^k\}$ be a sequence such that $\eta^k \in \mathcal{A}^\delta v(x^*,\xi)$ and $\eta^k \to \bar{\eta}$. By definition, there exists $x^k \in \mathcal{X}$ such that $x^k \in B(x^*,\delta)$ and $\eta^k \in \mathcal{A}v(x^k,\xi)$. Suppose without loss of generality that $x^k \to \bar{x}$. Then by [9, Proposition 2.1.5(b)], $\bar{\eta} \in \mathcal{A}v(x^*,\xi) \subset \mathcal{A}^\delta v(x^*,\xi)$. Furthermore, for every ξ, it follows from [9, Proposition 2.1.5(b)] that

$$\lim_{\delta\to 0} \mathcal{A}^\delta v(x^*,\xi) = \bigcap_{\delta>0} \bigcup_{x \in B(x^*,\delta)} \mathcal{A}v(x,\xi) = \mathcal{A}v(x^*,\xi).$$

By [23, Theorems 2.5] (or [23, Theorem 2.8] and the following remark, or [36, Theorem 1.43 (iii)]),

$$\lim_{\delta\to 0} \mathbb{E}\left[\mathcal{A}^\delta v(x^*,\xi)\right] = \mathbb{E}\left[\lim_{\delta\to 0} \mathcal{A}^\delta v(x^*,\xi)\right] = \mathbb{E}[\mathcal{A}v(x^*,\xi)].$$

which means that x^* satisfies (4.13). ∎

Theorem 4.1 states that if $\{x^N\}$ is a sequence of SAA Nash-C-stationary points of problem (1.6), then w.p.1 its accumulation point is a weak Nash-C-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 (1.6). 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 *Let $\{x^N\}$ be a sequence of Nash equilibria of obtained from solving the SAA problem (1.6) 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 $\{x^N\}$ is a Nash equilibrium of the true problem (1.1) if one of the following conditions holds:*

(a) *$v_i(x_i, x_{-i}, \xi)$, $i = 1, \cdots, \hat{i}$, is convex with respect to x_i for almost every $\xi \in \Xi$;*

(b) *$v_i(x_i, x_{-i}, \xi)$, $i = 1, \cdots, \hat{i}$, is Lipschitz continuous with respect to 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 Nash-C-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

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

9

It is well-known (see e.g. [46]) 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 (1.6) 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. We show that w.p.1 $\hat{\rho}^N(y, x^N)$ converges to $\rho(y, x^*)$ uniformly with respect to 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 with respect to 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 with respect to y when $N \to \infty$ because $\frac{1}{N} \sum_{j=1}^{N} \kappa_{-i}(\xi^j) \to \mathbb{E}[\kappa_{-i}(\xi)] < \infty$. On the other hand, we can show, through the application of the classical uniform law of large numbers [47, Lemma A1], that $\hat{\rho}^N(y, x^*) - \rho(y, x^*) \to 0$ on the compact set $\mathcal{X}$ uniformly with respect to y w.p.1 as $N \to \infty$. This shows that w.p.1 $\hat{\rho}^N(y, x^N)$ converges to $\rho(y, x^*)$ uniformly with respect to 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 set X is a global minimizer of $\rho(y, x^*)$ over X (hence a Nash equilibrium of the true problem), see for instance [47, Theorem A1][2]. ∎

4.2 Exponential convergence

Next we discuss the exponential convergence of $\{x^N\}$ as N goes to infinity. As we discussed in Section 1, there are many discussions in the literature on the exponential convergence of SAA optimization problems. However, as far as we are concerned, our discussion on the topic here is new in two fold: (a) this is first work which deals with the exponential convergence of SAA Nash-C-stationary point of a stochastic Nash equilibrium problem; (b) the underlying functions in first order equilibrium conditions (3.9) and (3.10) (or equivalently (4.13) and (4.14)) are set-valued.

We carry out our analysis in three steps. First, we extend Shapiro and Xu's uniform exponential convergence results ([56, Theorem 5.1]) to a class of random semicontinuous functions that are H-calm (from above or below). Second, we show uniform exponential convergence of $\mathbb{D}(\mathcal{A}\vartheta^N(x), \mathbb{E}[\mathcal{A}v(x, \xi)])$ under some moderate conditions of the Clarke generalized derivative $(v_i)^o$. 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 $d(x^N, X^*)$ 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 the set of weak Nash-C-stationary points of the true problem, and subsequently the exponential convergence of $d(x^N, X^*)$.

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 $\mathcal{X}$ 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*

$$\phi(x', \xi) - \phi(x, \xi) \leq \kappa(\xi)\|x' - x\|^{\gamma} \tag{4.17}$$

---

[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 $l(v)$.

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

- 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*

$$\phi(x',\xi) - \phi(x,\xi) \geq -\kappa(\xi)\|x' - x\|^{\gamma} \tag{4.18}$$

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

- 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*

$$|\phi(x',\xi) - \phi(x,\xi)| \leq \kappa(\xi)\|x' - x\|^{\gamma} \tag{4.19}$$

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

Note that the constants δ, γ and $\kappa(\xi)$ may depend on point x in the above definition. 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 [44, 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 [44, page 350-352]. Our definition is slightly different from the calmness in [44] 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.

Remark 4.2 *From (4.17) and (4.18), it is easy to observe that H-calmness from above implies upper semi-continuity while H-calmness from below implies lower semicontinuity.*

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 i.i.d. 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 set $\mathcal{X} \subset \mathbb{R}^n$ 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 [56]. 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 [42, 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 $\mathcal{X}$ be a compact subset of $\mathbb{R}^n$ and Assumption 4.1 hold.*

(i) *(Ralph and Xu [42, Lemma 3.6]) 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*

$$\text{Prob}\left\{\sup_{x \in \mathcal{X}}(\psi_N(x) - \psi(x)) \geq \epsilon\right\} \leq c(\epsilon)e^{-N\beta(\epsilon)}. \tag{4.20}$$

(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*

$$\text{Prob}\left\{\inf_{x \in \mathcal{X}}(\psi_N(x) - \psi(x)) \leq -\epsilon\right\} \leq c(\epsilon)e^{-N\beta(\epsilon)}. \tag{4.21}$$

(iii) *(Shapiro and Xu [56, Theorem 5.1]) 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*

$$\text{Prob}\left\{\sup_{x \in \mathcal{X}}|\psi_N(x) - \psi(x)| \geq \epsilon\right\} \leq c(\epsilon)e^{-N\beta(\epsilon)}. \tag{4.22}$$

Part (i) is proved in [42]. We include the proof in the appendix for completeness as the paper has not been published yet. Part (ii) can be proved in a similar way to Part (i) and Part (iii) is [56, Theorem 5.1] which is a combination of Parts (i) and (ii).

Proposition 4.1 is a generalization of [56, 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 the exponential convergence of sample average approximation of stochastic optimization problems where the underlying functions are lower or upper semicontinuous and satisfy certain calmness conditions. Our main purpose here, however, is to use Proposition 4.1 to establish the exponential rate of convergence of random set-valued mappings $\mathcal{A}\vartheta^N(x)$ over a compact subset of X.

We are now ready to present one of the main results of this section. For $i = 1, \cdots, \hat{i}$, let $(v_i)^o_{x_i}(x, \xi, u_i)$ denote the Clarke generalized derivative of v_i with respect to x_i in direction u_i, where $u_i \in \mathbb{R}^{n_i}$ and $\|u_i\| \leq 1$. Then $(v_i)^o_{x_i}(x, \xi, u_i) = \sigma(\partial_{x_i}v_i(x, \xi), u_i)$.

Theorem 4.3 *Let $\mathcal{A}v(x, \xi)$ be defined by (4.11) and $\mathcal{X}$ be a nonempty compact subset of X. Assume: (a) part (a) of Assumption 3.1 holds, (b) $\mathbb{E}[(v_i)^o_{x_i}(x, \xi, u_i)]$ is continuous on $\mathcal{X}$, (c) $(v_i)^o_{x_i}(x, \xi, u_i)$ 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*

$$\text{Prob}\left\{\sup_{x \in \mathcal{X}}\mathbb{D}(\mathcal{A}\vartheta^N(x), \mathbb{E}[\mathcal{A}v(x, \xi)]) \geq \epsilon\right\} \leq \hat{c}(\epsilon)e^{-\hat{\beta}(\epsilon)N}. \tag{4.23}$$

Proof. We use Proposition 4.1 to prove the result. First, for any $u := (u_1, \cdots, u_{\hat{i}})$, it follows from [38, Proposition 3.4]

$$\mathbb{E}[\sigma(\mathcal{A}v(x, \xi), u)] = \sigma(\mathbb{E}[\mathcal{A}v(x, \xi)], u). \tag{4.24}$$

Observe next that

$$\mathbb{D}(\mathcal{A}\vartheta^N(x), \mathbb{E}[\mathcal{A}v(x, \xi)]) \leq \sum_{i=1}^{\hat{i}} \mathbb{D}(\partial_{x_i}v_i^N(x, \xi), \mathbb{E}[\partial_{x_i}v_i(x, \xi)]), \tag{4.25}$$

12

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,

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

Using this, Proposition 2.1 and (4.24), we obtain

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

Consequently

$$\sup_{x\in\mathcal{X}}\mathbb{D}(\partial_{x_i}\vartheta_i^N(x,\xi),\mathbb{E}[\partial_{x_i}v_i(x,\xi)]) \leq \sup_{\|u_i\|\leq 1, x\in\mathcal{X}}\left(\frac{1}{N}\sum_{k=1}^N (v_i)^o_{x_i}(x,\xi^k,u_i)-\mathbb{E}[(v_i)^o_{x_i}(x,\xi,u_i)]\right). \qquad (4.26)$$

Let $Z_i := \{u_i \in \mathbb{R}^{n_i} : \|u_i\| \leq 1\}\times\mathcal{X}$. 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

$$\|(v_i)^o_{x_i}(x,\xi,u_i)\| \leq \|\partial_{x_i}v_i(x,\xi)\| \leq \kappa_i(\xi),$$

and by assumption $(v_i)^o_{x_i}(x,\xi,u_i)$ is H-calm from above in x with modulus $a_i(\xi)$ and order γ. Thus

$$(v_i)^o_{x_i}(x',\xi,u_i')-(v_i)^o_{x_i}(x,\xi,u_i) \leq a_i(\xi)\|x'-x\|^\gamma+\kappa_i(\xi)\|u_i'-u_i\| \leq p_i(\xi)\|z_i'-z_i\|^{\min(\gamma,1)},$$

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 $\|z_i'-z_i\| \leq 1$. This shows the H-calmness from above of $(v_i)^o_{x_i}(x,\xi,u_i)$ with respect to (x,u_i) set Z_i. Notice that $\mathbb{E}[(v_i)^o_{x_i}(x,\xi,u_i)]$ is continuous in x by assumption and because $(v_i)^o_{x_i}(x,\xi,u_i)$ is Lipschitz continuous in u_i with integrable modulus $\kappa_i(\xi)$, $\mathbb{E}[(v_i)^o_{x_i}(x,\xi,u_i)]$ is also continuous in u_i. This shows the continuity of $\mathbb{E}[(v_i)^o_{x_i}(x,\xi,u_i)]$ 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

$$\text{Prob}\left\{\sup_{(x,u_i)\in Z_i}\left(\frac{1}{N}\sum_{k=1}^N (v_i)^o_{x_i}(x,\xi^k,u_i)-\mathbb{E}[(v_i)^o_{x_i}(x,\xi,u_i)]\right) \geq \epsilon_i\right\} \leq \hat{c}_i(\epsilon_i)e^{-N\hat{\beta}_i(\epsilon_i)}, \qquad (4.27)$$

for $i=1,\cdots,\hat{i}$. For any $\epsilon > 0$, let $\epsilon_i > 0$ be such that $\sum_{i=1}^{\hat{i}}\epsilon_i \leq \epsilon$. Then by combining (4.25)-(4.27), we obtain

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

This shows (4.23) with $\hat{c}(\epsilon)=\hat{i}\max_{i=1}^{\hat{i}}\hat{c}_i(\epsilon_i)$ and $\hat{\beta}(\epsilon)=\min_{i=1}^{\hat{i}}\hat{\beta}_i(\epsilon_i)$. $\blacksquare$

Remark 4.3 *The H-calmness from above of $(v_i)^o_{x_i}(x,\xi,u_i)$ and continuity of $\mathbb{E}[\partial_{x_i}v_i]$ in a certain compact set $\mathcal{X}$ play an important role in Theorem 4.3. It is therefore natural to ask when the properties hold.*

(i) *(H-calmness from above) Except in some pathological examples, v_i is often piecewise smooth in many practical instances, that is, for almost every $\xi \in \Xi$, v_i can be expressed as a finite number of smooth functions of x, see [50] for detailed discussion of piecewise smooth functions. Let us consider the case that for almost every $\xi \in \Xi$, v_i is piecewise twice continuously differentiable with respect to x. Then $(v_i)^o_{x_i}(x, \xi, u_i)$ is H-calm in x. The conclusion follows from a number of claims which can be verified without much difficulties: (a) $\partial_{x_i} v_i(x, \xi, u_i)$ is outer semicontinuous with respect to x; (b) by [4, Proposition 2.6.4], claim (a) implies that $(v_i)^o_{x_i}(x, \xi, u_i)$ is upper semicontinuous with respect to x; (c) $(v_i)^o_{x_i}(x, \xi, u_i)$ is upper bounded by $\max_{j=1}^J \|\nabla_{x_i} v_i^j(x, \xi, u_i)\|$ where v_i^j, $j = 1, \cdots, J$ denote active pieces of v_i at point x. Therefore,*

$$(v_i)^o_{x_i}(x', \xi, u_i) \le (v_i)^o_{x_i}(x, \xi, u_i) + \max_{j=1}^J \{\|\nabla^2_{x_i x} v_i^j(x, \xi, u_i)\|\|x' - x\| + o(\|x' - x\|)\}$$

for all x' in a small neighborhood of x relative to X (here we assume that the existence of such neighborhood is independent of ξ). Let $\kappa_i(\xi) := \max_{x \in \mathcal{X}} \max_{j=1}^J \|\nabla^2_{x_i x} v_i^j(x, \xi, u_i)\|$. The H-calmness of $(v_i)^o_{x_i}$ follows.

(ii) *(Continuity of $\mathbb{E}[\partial_{x_i} v_i]$) Shapiro [52, Proposition 4.1] showed that if a random function is continuously differentiable w.p.1 and 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 outer semicontinuity 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 $\partial_{x_i} v_i$, $i = 1, \cdots, \hat{i}$, is H-calm, the exponential convergence of Theorem 4.3 can be strengthened by replacing $\mathbb{D}$ with $\mathbb{H}$ in (4.23). We state this in the following corollary.

Corollary 4.1 *If $(v_i)^o_{x_i}(x, \xi, u_i)$ is H-calm with respect to x 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*

$$\text{Prob}\left\{\sup_{x \in \mathcal{X}} \mathbb{H}(\mathcal{A}\vartheta^N(x), \mathbb{E}[\mathcal{A}v(x, \xi)]) \ge \epsilon\right\} \le \hat{c}(\epsilon) e^{-\hat{\beta}(\epsilon)N}. \tag{4.28}$$

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

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

Moreover,

$$\text{Prob}\left\{\sup_{(x, u_i) \in Z_i} \left(\mathbb{E}[(v_i)^o_{x_i}(x, \xi, u_i)] - \frac{1}{N} \sum_{k=1}^N (v_i)^o_{x_i}(x, \xi^k, u_i)\right) \ge \epsilon_i\right\}$$

$$= \text{Prob}\left\{\inf_{(x, u_i) \in Z_i} \left(\frac{1}{N} \sum_{k=1}^N (v_i)^o_{x_i}(x, \xi^k, u_i) - \mathbb{E}[(v_i)^o_{x_i}(x, \xi, u_i)]\right) \le -\epsilon_i\right\}.$$

Notice that H-calmness of $\partial_{x_i} v_i(x, \xi)$ implies that $(v_i)^o_{x_i}(x, \xi, u_i)$ is H-calm from below for every fixed u_i. By applying Proposition 4.1 (ii) 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

$$\text{Prob}\left\{\sup_{x \in \mathcal{X}} \mathbb{D}(\mathbb{E}[\mathcal{A}v(x, \xi)], \mathcal{A}\vartheta^N(x)) \ge \epsilon\right\} \le \hat{c}(\epsilon) e^{-\hat{\beta}(\epsilon)N}. \tag{4.29}$$

The conclusion follows by combining (4.29) and (4.23). ∎

Theorem 4.4 *Let $\mathcal{X} \subset X$ be a nonempty compact subset of X and X^* be a set of weak stochastic Nash-C-stationary points of the true problem (1.1) within $\mathcal{X}$. Let $\Psi(x) = \mathbb{E}[\mathcal{A}v(x,\xi)] + G_X(x)$, where $G_X(x)$ is defined by (4.12). Assume that for $\bar{N}$ sufficiently large, the sequence $\{x^N\}_{N>\bar{N}}$ is located in $\mathcal{X}$ w.p.1. Then under conditions (a)-(d) of Theorem 4.3, $\{x^N\}$ converges to X^* at an exponential rate, that is, for any small positive number $\epsilon > 0$, there exist $\tilde{c}(\epsilon) > 0$ and $\tilde{\beta}(\epsilon) > 0$, independent of N, such that for N sufficiently large*

$$\text{Prob}\big(d(x^N, X^*) \geq \epsilon\big) \; \leq \; \tilde{c}(\epsilon)e^{-\tilde{\beta}(\epsilon)N}. \tag{4.30}$$

We need an intermediate result on sensitivity analysis to prove the result. Consider the following generalized equation

$$0 \in H(x) + \mathcal{N}_{\mathcal{C}}(x), \tag{4.31}$$

where $H(x) : \mathcal{C} \to 2^{\mathbb{R}^m}$ is a closed set-valued mapping, $\mathcal{C}$ is a closed convex subset of $\mathbb{R}^m$. Let $\tilde{H}(x)$ be a perturbation of $H(x)$ and we consider the perturbed equation

$$0 \in \tilde{H}(x) + \mathcal{N}_{\mathcal{C}}(x). \tag{4.32}$$

The following lemma states that when $\mathbb{D}(\tilde{H}(x), H(x))$ is sufficiently small uniformly with respect to x, then the solution set of (4.32) is close to the solution set of (4.31).

Lemma 4.1 *([62, Lemma 4.2]) Let $\mathcal{X}$ be a compact subset of $\mathbb{R}^m$. Let X^* denote the set of solutions to (4.31) restricted to $\mathcal{X}$ and Y^* denote the set of solutions to (4.32) restricted to $\mathcal{X}$. Suppose that both X^* and Y^* are nonempty. Then*

(i) for any $\epsilon > 0$ there exists a $\delta > 0$ such that if H is outer semicontinuous in $\mathcal{X}$ and $\sup_{x \in \mathcal{X}} \mathbb{D}(\tilde{H}(x), H(x)) < \delta$, then $\mathbb{D}(Y^*, X^*) < \epsilon$.

(ii) for any $\epsilon > 0$ there exists a $\delta > 0$ such that if $\tilde{H}$ is outer semicontinuous in $\mathcal{X}$ and $\sup_{x \in \mathcal{X}} \mathbb{H}(\tilde{H}(x), H(x)) < \delta$, then $\mathbb{H}(Y^*, X^*) < \epsilon$.

Proof of Theorem 4.4. The conclusion follows from Theorem 4.3 and Lemma 4.1 (i). Indeed, for any $\epsilon > 0$, it follows from Lemma 4.1 (i) that there exists a $\delta > 0$ such that $d(x^N, X^*) \leq \epsilon$ so long as $\sup_{x \in \mathcal{X}} \mathbb{D}(\mathcal{A}\vartheta^N(x), \mathbb{E}[\mathcal{A}v(x,\xi)]) \leq \epsilon$. This is equivalent to

$$\text{Prob}\big(d(x^N, X^*) \geq \epsilon\big) \leq \text{Prob}\left\{\sup_{x \in \mathcal{X}} \mathbb{D}(\mathcal{A}\vartheta^N(x), \mathbb{E}[\mathcal{A}v(x,\xi)]) \geq \epsilon\right\}.$$

The rest follows from (4.23). ∎

Remark 4.4 *It is possible to express the constants $\tilde{c}(\epsilon)$ and $\tilde{\beta}(\epsilon)$ in Theorem 4.4 in terms their counterparts in Theorem 4.3 under some stronger conditions. For instance, if there exists a strictly monotonically increasing function $\Theta : \mathbb{R}^+ \to \mathbb{R}^+$ such that*

$$d(x, X^*) \leq \Theta(\sup_{x \in \mathcal{X}} d(0, \mathbb{E}[\mathcal{A}\vartheta(x,\xi)] + G_X(x))), \tag{4.33}$$

then we can easily obtain that

$$d(x^N, X^*) \leq \Theta(\sup_{x \in \mathcal{X}} (\mathcal{A}\vartheta^N(x), \mathbb{E}[\mathcal{A}\vartheta(x,\xi)])).$$

Consequently we have $\tilde{c}(\epsilon) = \hat{c}(\Theta^{-1}(\epsilon))$ and $\tilde{\beta}(\epsilon) = \hat{\beta}(\Theta^{-1}(\epsilon))$. In a particular case when $\Theta(t) = \alpha t$, inequality (4.33) is implied by the metric regularity, see recent discussions about this by Ralph and Xu in [42]. A stronger regularity condition is also considered for SAA-SGE by Shapiro in [53, Section 7], see Remark 4.1.

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 (1.6). For fixed sample, this is a deterministic nonsmooth equilibrium problem and one may use well-known bundle methods [33, 48] 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 [61]. 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 [17]. 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 [41] for obtaining a smooth approximation of an implicit function defined by a nonsmooth system of equations and was applied to tackle nonsmoothness in nonsmooth stochastic minimization problem in [61]. 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 [41] 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

$$\hat{\vartheta}_i(x_i^*(\epsilon), x_{-i}^*(\epsilon), \epsilon) = \min_{x_i \in X_i} \mathbb{E}[\hat{v}_i(x_i, x_{-i}^*(\epsilon), \xi, \epsilon)], \text{ for } i = 1, \cdots, \hat{i}, \tag{5.34}$$

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

$$\hat{\vartheta}_i^N(x_i^N(\epsilon), x_{-i}^N(\epsilon), \epsilon) = \min_{x_i \in X_i} \hat{\vartheta}_i^N(x_i, x_{-i}^N(\epsilon), \epsilon) \text{ for } i = 1, \cdots, \hat{i}, \tag{5.35}$$

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 (5.34) and (5.35) can be written respectively as

$$0 \in \mathbb{E}[\nabla_{x_i}\hat{v}_i(x_i, x_{-i}, \xi, \epsilon)] + \mathcal{N}_{X_i}(x_i), \ i = 1, \cdots, \hat{i}, \tag{5.36}$$

and

$$0 \in \frac{1}{N} \sum_{k=1}^{N} \nabla_{x_i} \hat{v}_i(x_i, x_{-i}, \xi^k, \epsilon) + \mathcal{N}_{X_i}(x_i), \ i = 1, \cdots, \hat{i}. \tag{5.37}$$

Note that in numerical implementation, we may solve (5.35) 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 (5.36) and $S^N(\epsilon)$ the set of solutions to (5.37).

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,*

$$\varlimsup_{x' \to x, \epsilon \to 0} \{\nabla_{x_i} \hat{v}_i(x'_i, x'_{-i}, \xi, \epsilon)\} \subset \partial_{x_i} v_i(x_i, x_{-i}, \xi). \tag{5.38}$$

(i) *If $S(\epsilon)$ is nonempty for all ϵ close to 0, then $\varlimsup_{\epsilon \to 0} S(\epsilon) \subset S$, w.p.1., where S denotes the set of solutions to (3.7).*

(ii) *If $\epsilon_N \to 0$ as $N \to \infty$, $S^N(\epsilon_N)$ is nonempty for all N sufficiently large, then $\varlimsup_{N \to \infty} S^N(\epsilon_N) \subset S$, w.p.1, where S^N denotes the solution set of (3.10);*

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

Proof. The proof of parts (i) and (ii) is similar to the proof of [61, Theorem 3.1]. Here we give a brief proof sketch.

Part (i). By assumption, $\nabla_{x_i} \hat{v}_i(x'_i, x'_{-i}, \xi, \epsilon)$ is integrably bounded by $\kappa_i(\xi)$. Then similar to the proof of [61, Theorem 3.1],

$$\varlimsup_{x' \to x, \epsilon \to 0} \mathbb{E}\left[\nabla_{x_i} \hat{v}_i(x'_i, x'_{-i}, \xi, \epsilon)\right] = \mathbb{E}\left[\varlimsup_{x' \to x, \epsilon \to 0} \nabla_{x_i} \hat{v}_i(x'_i, x'_{-i}, \xi, \epsilon)\right] \subset \mathbb{E}\left[\partial_{x_i} v_i(x_i, x_{-i}, \xi)\right].$$

The inclusion follows from (5.38). The rest is straightforward.

Part (ii). Let

$$\mathcal{G}(x, \xi, \epsilon) = \begin{cases} \nabla_{x_i} \hat{v}_i(x, \xi, \epsilon), & \text{if } \epsilon > 0, \\ \varlimsup_{(x', \epsilon') \to (x, 0)} \nabla_{x_i} \hat{v}_i(x', \xi, \epsilon'), & \text{if } \epsilon = 0, \end{cases}$$

and

$$\mathcal{F}^\delta(x, \xi, 0) = \bigcup_{(x', \epsilon') \in B((x,0), \delta)} \mathcal{G}(x', \xi, \epsilon').$$

By [55, Theorem 2],

$$\varlimsup_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \nabla_{x_i} \hat{v}_i(x_i, x_{-i}, \xi^k, \epsilon_N) \subset \mathbb{E}[\mathcal{F}^\delta(x, \xi, 0)]$$

w.p.1 uniformly with respect to x. Thus for any sequence $\{x^N\} \subset S(\epsilon_N)$ with accumulation point x^*

$$\varlimsup_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} \nabla_{x_i} \hat{v}_i(x^N, \xi^k, \epsilon_N) \subset \mathbb{E}[\mathcal{F}^\delta(x^*, \xi, 0)].$$

By [23, Theorems 2.5] (or [23, Theorem 2.8] and the following remark, or [36, Theorem 1.43 (iii)]),

$$\lim_{\delta \to 0} \mathbb{E}[\mathcal{F}^\delta(x^*, \xi, 0)] = \mathbb{E}[\lim_{\delta \to 0} \mathcal{F}^\delta(x^*, \xi, 0)] \subset \partial_{x_i} v_i(x_i, x_{-i}, \xi).$$

The last inclusion follows from (5.38). The rest is obvious.

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 [47, 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 outer semicontinuity 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 (4.12), 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 the impact of options in electricity markets

6.1.1 The model

Consider an electricity spot market with M generators competing in a non-cooperative 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

$$R_i(q_i, Q_{-i}) = \mathbb{E}[q_i p(Q, \xi) - C_i(q_i) + H_i(p(Q, \xi))], \tag{6.39}$$

where $Q = q_i + Q_{-i}$ is the total bids by all generators to the market, Q_{-i} denotes the total bids by i's competitors, $q_i p(Q, \xi)$ is the total revenue for generator i from 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 and finally $H_i(p)$ denotes the payments related to contracts which generators sign with retailers before entering the spot market.

Contracts are typically used to hedge risks arising from uncertainties in the spot market. They are financial instruments which do not involve the actual generation of electricity, but the money paid under the contract is tied to the pool price. There are essentially two types of contracts: a *one-way contract* such as a put or a call option where only one party of the contract commits to pay the difference between the strike price and the spot price for the contracted quantity, and a *two-way contract* where both parties of the contracts commit to the difference as opposed to the one-way contract. We assume that both generators

and retailers are risk neutral, that no retailer will sign a two way contract with strike price greater than the expected spot price and generators, in equilibrium, will not find it advantageous to offer lower contract price than expected spot price. Under this kind of assumption which is also known as no-arbitrage condition in the literature [1, 18], the strike price (denoted by f) of a two way contract is equal to the expected value of the spot market price $\mathbb{E}[p(Q,\xi)]$, consequently the payment related to the two way contract does not appear in the expected profit in that for any contacted quantity q, $\mathbb{E}[q(f - p(Q,\xi))] = 0$. Based on this discussion, here we only consider one way contracts, specifically call option and put option can be dealt with in the same manner. By selling a call option at a strike price f, generator i will pay $w_i\,(p - f)$ to the contract holder if $p > f$, but no payment is made if $p \leq f$, where w_i is the quantity signed by generator i on one-way contract at strike price f. Assuming that generator i enters a call option of quantity w_i at strike price f, we can formulate $H_i(p(Q,\xi))$ as follows:

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

Generator i's decision making problem is to choose an optimal quantity q_i such that its expected profit defined in (6.39) is maximized. Let

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

Then generator's total expected profit can be written as $R_i(q_i, Q_{-i}) = \mathbb{E}\left[r_i(q_i, Q_{-i}, \xi)\right]$. Assuming that every generator is a profit maximizer, we can formulate the competition as a stochastic Nash equilibrium problem.

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

$$-R_i(q_i^*, Q_{-i}^*) = \min_{q_i \in \mathcal{Q}_i} -R_i(q_i, Q_{-i}^*). \tag{6.40}$$

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

The Nash equilibrium problem above is a practical example of the general model (1.1). In what follows, we apply the SAA method to this problem and investigate the convergence of sample average approximate equilibrium as sample size increases using our established results in the preceding sections. We need 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 [14, 57].

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.

19

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

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

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

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

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

$$L(q_i, \xi) := \max\{\kappa(\xi) + |C'_i(q_i)| + |q_i - w_i|\kappa(\xi), \kappa(\xi) + |C'_i(q_i)| + q_i\kappa(\xi)\}, \tag{6.41}$$

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 $p(Q, \xi) = f$. This means that $r_i(q_i, Q_{-i}, \xi)$ is continuously differentiable in q_i w.p.1. By [49, 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 [46, Theorems 1 and 2], we can show the existence and uniqueness of the equilibrium of (6.40). We omit the details because they are not the main focus of this paper.

6.1.2   Sample average approximation

Stochastic Nash equilibrium problem (6.40) 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 the Nash equilibrium problem (6.40) 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

$$-R_i^N(q_i^N, Q_{-i}^N) = \min_{q_i \in \mathcal{Q}_i} -R_i^N(q_i, Q_{-i}^N) := \frac{1}{N} \sum_{k=1}^{N} -r_i(q_i, Q_{-i}, \xi^k) \tag{6.42}$$

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 (6.40) coincides with the first order equilibrium condition which can be written as:

$$0 \in -\mathbb{E}[\nabla_{q_i} r_i(q_i, Q_{-i}, \xi)] + \mathcal{N}_{\mathcal{Q}_i}(q_i), \ i = 1, 2, \ldots, M. \tag{6.43}$$

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

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

For simplicity of notation, let

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

$G(\mathbf{q}) = \mathcal{N}_{\mathcal{Q}_1}(q_1) \times \cdots \mathcal{N}_{\mathcal{Q}_M}(q_M)$ and $\Psi(\mathbf{q}) = -\mathbb{E}[F(\mathbf{q}, \xi)] + G(\mathbf{q})$. Then the first order equilibrium condition (6.43) 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 of this lemma 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 $\mathbf{q}^N$ be a Nash-C-stationary point defined by (6.44). Under Assumption 6.1, there exists a unique Nash equilibrium $\mathbf{q}^*$ for problem (6.43) and the sequence $\{\mathbf{q}^N\}$ converges to $\mathbf{q}^*$ at an exponential rate, with the increase of sample size N, that is, for any small positive number $\epsilon > 0$, there exist constants $\hat{c}_1(\epsilon) > 0$ and $\hat{c}_2(\epsilon) > 0$, independent of N, such that for N sufficiently large*

$$\text{Prob}\left(\|\mathbf{q}^N - \mathbf{q}^*\| \geq \epsilon\right) \leq \hat{c}_1(\epsilon) e^{-\hat{c}_2(\epsilon)N}. \tag{6.45}$$

Proof. Under Assumption 6.1, we can show through Proposition 6.1 the existence and uniqueness of the Nash equilibrium $\mathbf{q}^*$. We use Theorem 4.4 to prove the rest. To this end, we verify conditions (a)-(d) of Theorem 4.3 in this context. Condition (a) is part (a) of Assumption 3.1. Obviously $r_i(q_i, Q_{-i}, \xi)$ is Lipschitz continuous with respect to q_i with modulus $L(q_i, \xi)$ defined as in (6.41). Condition (b) follows from Proposition 6.1 which shows that $\mathbb{E}[r_i(q_i, Q_{-i}, \xi)]$ is continuous. Conditions (c) follows from Remark 4.3. To see this, observe that r_i is piecewise twice continuously differentiable. It is easy to verify that the Clarke generalized derivative $(r_i)_{q_i}^o$ is upper semicontinuous with respect to $\mathbf{q}$. Condition (d) is automatically satisfied because the support set Ξ is bounded, see comments by Shapiro and Xu about conditions (C1) and (C3) in [56, page 408]. ∎

Note that 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 by Rockafellar and Wets in [44, page 388]. Therefore Remark 4.4 applies in this case, that is, (4.33) holds with $\Theta(t) = \alpha t$. We omit the details.

6.1.3 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 (6.42) 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 [61] to smooth the max-function. Let $\epsilon \in \mathbb{R}_+$ and $\hat{a}(z, \epsilon)$ be such that for every $\epsilon > 0$, let

$$\hat{a}(z, \epsilon) := \epsilon \ln(1 + e^{z/\epsilon}) \tag{6.46}$$

and for $\epsilon = 0$, $\hat{a}(z, 0) := \max(z, 0)$. It is proved in [61, 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 the gradient (subdifferential) consistency, that is,

$$\overline{\lim_{(z', \epsilon) \to (z, 0)}} \frac{d\hat{a}(z, \epsilon)}{dz} = [0, 1] = \partial_z \max(z, 0). \tag{6.47}$$

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

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

Then we may solve the following smoothed SAA problem instead of (6.42): find an M-tuple $\mathbf{q}^N(\epsilon) := \left(q_1^N(\epsilon), \ldots, q_M^N(\epsilon)\right)$ such that

$$-\hat{R}_i^N(q_i^N(\epsilon), Q_{-i}^N(\epsilon), \epsilon) = \min_{q_i \in \mathcal{Q}_i} -\frac{1}{N} \sum_{k=1}^N \hat{r}_i(q_i, Q_{-i}^N(\epsilon), \xi^k, \epsilon), \ i = 1, 2, \ldots, M. \tag{6.48}$$

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

$$-\hat{R}_i(q_i(\epsilon), Q_{-i}, \epsilon) = \max_{q_i \in \mathcal{Q}_i} -\hat{R}_i(q_i, Q_{-i}(\epsilon), \epsilon), \ i = 1, 2, \ldots, M. \tag{6.49}$$

Since the smoothing preserves the convexity, $\hat{R}_i(q_i, Q_{-i}, \epsilon)$ is also convex in q_i. Therefore both (6.48) and (6.49) have a unique equilibrium under Assumption 6.1. Moreover, the convexity of the problem means that

we can we can solve (6.48) by solving the following KKT conditions (which form a variational inequality):

$$0 \in -\frac{1}{N} \sum_{k=1}^{N} (\hat{r}_i)'_{q_i} (q_i, Q_{-i}^N(\epsilon), \xi^k, \epsilon) + \mathcal{N}_{\mathcal{Q}_i}(q_i), \ i = 1, 2, \ldots, M. \tag{6.50}$$

The following proposition summarizes 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 (6.40) 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 ([61, 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, (6.47) 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 from Theorem 5.1. ∎

6.1.4 Numerical tests

We carry out numerical tests on the proposed smoothing SAA scheme for solving the stochastic Nash equilibrium problem (6.40) with three generators signing call options. We use mathematical programming codes in GAMS installed in a PC with Windows XP operating system and the solver *PATH* for solving (6.50). Our tests are focused on different values of the smoothing parameter ϵ and sample size N.

Example 6.1 *Consider problem (6.40) 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 $\alpha = 20$, $\alpha_0 = 30$. Table 1 lists the three generators' quantities of call option w_i and cost functions. The strike price is $f = 22$.*

Table 1: Quantities of the contracts and cost functions

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

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 the numerical tests with different smoothing parameter values and sample sizes. The results are displayed in Table 3 where $q_i^N(\epsilon)$ denotes generator i's approximate dispatch, $R_i^N(\epsilon)$ denotes i's approximate profit, $Q^N(\epsilon)$ denotes the approximate aggregate dispatch, $\bar{p}(Q^N(\epsilon))$ denotes the approximate average price.

The results show that the convergence is not very sensitive to changes of the value of ϵ when it is in the range of $[0.02, 2]$. This is consistent with the observations obtained in the literature, see [34, 61].

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	$(71.824, 29.846, 44.598)$

To investigate the smoothing parameter issue further, we fix the sample size to $N = 5000$, and look into the computing time, main iterations (defined by GAMS) and function evaluations as ϵ is reduced. It turns out that there is no further improvement of approximate solutions, and yet the computing time for getting a solution significantly increases, see our report in Table 4. Let us explain this phenomena. Our smoothing function in this example is $\hat{a}(z, \epsilon) = \epsilon \ln(1 + e^{z/\epsilon})$. The second order derivative of the function is $\hat{a}''_{zz}(z, \epsilon) = \frac{e^{z/\epsilon}}{\epsilon(1 + e^{z/\epsilon})^2}$. It is easy to verify that $\hat{a}''_{zz}(z, \epsilon) \to +\infty$ as $\epsilon \to 0$ for $z < 0$. This means the smoothed SAA problem (6.49) becomes ill-conditioned as $\epsilon \to 0$. Consequently more computational time and function evaluations are needed to get a solution although we solve it through its first order equilibrium conditions (6.50).

Table 4: Numerical results of Example 6.2.

ϵ	Major Iteration	Minor Iteration	Function Evaluation	Computation Time
2	1	1	4	0.140 sec
0.6	1	1	4	0.140 sec
0.2	2	2	5	0.156 sec
0.06	2	2	5	0.172 sec
0.02	2	2	5	0.188 sec
0.006	190	194	3184	232.304 sec
0.002	203	207	3364	274.399 sec

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. 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 in order to study the convergence of the approximation with respect to the sample size. The results are displayed in Table 5.

Table 5: 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 is changed from 500 to 10000. This reflects the fast convergence of the sample average approximation.

Our conclusion from the preliminary numerical tests is that the smoothed SAA method through (5.37) might provide a convenient approach to obtain an approximate solution with relatively low precision (due

to the limitation in reducing smoothing parameter). To obtain a more accurate solution, one may apply some iterative method such as the well-known bundle method [33] to solve the SAA problem (1.6) directly (without smoothing) with the obtained approximate solution as a warmstart (initial point).

6.2 A stochastic Nash equilibrium model for electricity markets with network

The stochastic Nash equilibrium model in Section 6.1 is convex and focuses on a single node electricity pool market. In this section, we consider a nonconvex problem which models generators' competition for dispatch in a centrally dispatched wholesale spot market with demand uncertainty and transmission network constraints. The aim is to study electricity market competition from a different angle and show the applicability of model (1.1).

6.2.1 Problem description

Let us consider a market which consists of a set of nodal spot markets connected through transmission grids. There exists a pool operated by an Independent System Operator (ISO), which serves as a broker, and makes decisions on the market clearing price and power transactions. The ISO leases the transmission system from the network owners and controls the power flows in order to maintain the feasibility of the transmission network.

Over the past few years, game-theoretic models have been extensively used to investigate strategic behavior in deregulated electricity markets with network constraints. Hobbs, Metzler and Pang [24] proposed a strategic gaming model for analyzing an oligopolistic electricity market economy with several dominant firms in an electric power network, where individual generators' decision problem was formulated as a mathematical program with equilibrium constraints (MPEC). Hu and Ralph [26] apparently made the first attempt to investigate a bilevel game-theoretic model for electricity markets with transmission networks, where the corresponding bilevel game is recast as an equilibrium problem with equilibrium constraints (EPEC). Moreover, due to nonconvexity of the transmission constraints, they introduced some new concepts such as local Nash equilibria and Nash stationary points for characterizing the equilibrium conditions of EPEC which complement the standard concept of (global) Nash equilibria. In the literature, there exist a volume of game-theoretic models or equilibrium models employed to investigate strategic behavior in electricity markets with transmission constraints, see Yao, Oren and Adler [63] for using an SEPEC to model an equilibrium in the spatial market with the demand uncertainties at each node, Wei and Smeers [59] for a generalized Nash equilibrium model for an oligopolistic electricity market with spatially dispersed generators and consumers, and Day, Hobbs and Pang [12] for a conjectured supply function equilibrium model of competition among generators on a linearized DC network.

The stochastic equilibrium model to be discussed in this section follows primarily from Yao, Oren and Alder's model in [63] and [64], where the generators anticipate the impact of the transmission line congestion and take the effect into account in their production decisions. Here we discuss how the model can be fit to our stochastic Nash equilibrium framework and how the proposed numerical schemes could possibly be applied to it.

Consider a lossless direct current (DC) network with a set of transmission lines denoted by $L = \{1, \cdots, L\}$. The network underlying the spot market consists of a set of nodes denoted by $\mathcal{G} = \{1, \cdots, G\}$ and a set of generators indexed by $i = 1, 2, \cdots, M$. In the spot market, we assume that generator i operates the units at a subset of locations (nodes) $\mathcal{G}_i \subset \mathcal{G}$ for $i = 1, 2, \cdots, M$. We also assume that at most one generator operates at a node, and if necessary, we can introduce artificial nodes to meet this assumption [64]. Moreover, we use a vector of continuous random variables $\xi(\omega) = (\xi_1(\omega), \cdots, \xi_G(\omega))$, with support set Ξ to characterize the demand uncertainty in the wholesale electricity market where $\xi_g(\omega)$ denotes the uncertainty at node $g \in \mathcal{G}$. As in Section 6.1, we use $p_g(\tau_g, \xi(\omega))$ to describe the inverse demand function at node g. That is, if the total supply to the node is τ_g, then the market price is $p_g(\tau_g, \xi)$ at scenario $\xi(\omega) = \xi$. A significant difference from the previous model in Section 6.1 is that here the market may have a number of different nodal prices over the network.

In this game-theoretic model, the generators compete in a noncollaborative manner before the uncertainties are realized, each of which wishes to maximize its individual expected profit by taking the ISO's decision

24

making process into account. After knowing a particular realization of the demand uncertainty ($\xi = \xi(\omega)$) and receiving individual generator's bids for supply, the ISO makes a decision on each generator's dispatch and the import or export quantity $r_g(\xi)$ at each node $g \in \mathcal{G}$, and hence the flow on every transmission line $l \in \mathrm{L}$.

Let us first formulate ISO's decision problem. Given that the realization of the demand uncertainty ξ and generator i's production quantity q_g at node $g \in \mathcal{G}_i$, the ISO determines generators' dispatch and the import/export $r_g(\xi)$ at every node $g \in \mathcal{G}$. These quantities are required to satisfy the network feasibility constraints, that is, the resulting power flows should not exceed the thermal limits K_l of the transmission line l in both directions. In this model, we consider a lossless DC approximation of Kirchhoff's laws. Specifically, flows on lines can be calculated using power transfer distribution factor (PTDF) $D_{l,g}$, which specifies the proportion of flow on a line $l \in \mathrm{L}$ resulting from an injection of one-unit electricity at node $g \in \mathcal{G}$ and a corresponding one-unit withdrawal at some fixed reference node [10]. The ISO's objective is to maximize the total social welfare of the entire network which is defined as the total consumer willingness-to-pay, less the sum of all generation costs. Mathematically, this is the aggregated area under the curve of the nodal inverse demand function $p_g(\cdot, \xi)$. The ISO solves the following problem parametric on the generators' production decision q_g, $g \in \mathcal{G}$, and $\xi \in \Xi$:

$$
\begin{aligned}
\max_{r_g(\xi),\, g \in \mathcal{G}} \quad & \sum_{g \in \mathcal{G}} \left(\int_0^{r_g(\xi) + q_g} p_g(\tau_g, \xi) d\tau_g - C_g(q_g) \right) \\
\text{s.t.} \quad & \sum_{g \in \mathcal{G}} r_g(\xi) = 0, \\
& r_g(\xi) + q_g \geq 0,\ g \in \mathcal{G}, \\
& \sum_{g \in \mathcal{G}} D_{l,g} r_g(\xi) \geq -K_l,\ l \in \mathrm{L}, \\
& \sum_{g \in \mathcal{G}} D_{l,g} r_g(\xi) \leq K_l,\ l \in \mathrm{L}.
\end{aligned}
\tag{6.51}
$$

Here $C_g(q_q)$ is a generator's total generation cost in producing a quantity q_g of electricity at node g, the first constraint indicates the load and the generation must be balanced at all times which means that all import and export quantities must add up to zero; the second constraint means that the net consumption of electricity at node g is nonnegative; the last two constraints mean that the quantities flowing through a line in both directions should not exceed the thermal limits, see [64] for details.

Let $\varrho(\xi)$, $\eta_g(\xi)$, $\lambda_l^-(\xi)$ and $\lambda_l^+(\xi)$ denote the Lagrange multipliers corresponding to the constraints in (6.51). The first order necessary conditions (the Karush-Kuhn-Tucker (KKT) conditions) of the ISO's problem can be written as:

$$
\begin{aligned}
0 &= p_g\left(r_g(\xi) + q_g, \xi\right) - \varrho(\xi) + \eta_g(\xi) - \sum_{l \in \mathrm{L}} D_{l,g}\left(\lambda_l^+(\xi) - \lambda_l^-(\xi)\right), & g &\in \mathcal{G}, \\
0 &\leq \eta_g(\xi) \perp r_g(\xi) + q_g \geq 0, & g &\in \mathcal{G}, \\
0 &= \sum_{g \in \mathcal{G}} r_g(\xi), & & \\
0 &\leq \lambda_l^-(\xi) \perp K_l + \sum_{g \in \mathcal{G}} D_{l,g} r_g(\xi) \geq 0, & l &\in \mathrm{L}, \\
0 &\leq \lambda_l^+(\xi) \perp K_l - \sum_{g \in \mathcal{G}} D_{l,g} r_g(\xi) \geq 0, & l &\in \mathrm{L}.
\end{aligned}
\tag{6.52}
$$

Assume that the inverse demand function $p_g(q, \xi)$ is decreasing with respect to q for every fixed ξ. It is easy to verify that the objective function in the ISO's problem (6.51) is concave with respect to $r_g, g \in \mathcal{G}$. Since all of the constraint functions are linear, then the ISO's problem is a concave program. This implies that problem (6.51) is *equivalent* to the KKT system (6.52). In particular, when $p_g(q, \xi)$ is is monotonically decreasing, both (6.51) and (6.52) have a unique solution.

Let us now consider an individual generator's optimal decision making problem. In the pool, generator i, $i = 1, 2, \ldots, M$, determines the output of its generation unit at each node $g \in \mathcal{G}_i$ before the actual market demand is known. Assume that each generator aims at maximizing its expected profit and in doing so the generator anticipates the impact of its production on ISO's decision making (e.g. on power dispatch, flows and market clearing price) and its competitors' response. Consequently we can formulate the generators' optimal decision making problems as follows: for $i = 1, \cdots, M$

$$\max_{q_g, r_g(\cdot), g \in \mathcal{G}_i} \quad \mathbb{E}\left[\sum_{g \in \mathcal{G}_i} p_g(r_g(\xi(\omega)) + q_g, \xi(\omega)) q_g \right] - \sum_{g \in \mathcal{G}_i} C_g(q_g)$$

$$\text{s.t.} \quad \text{for a.e. } \omega \in \Omega$$
$$0 \le q_g \le \bar{q}_g, g \in \mathcal{G}_i,$$
$$0 = p_g \left(r_g(\xi(\omega)) + q_g, \xi(\omega) \right) - \varrho(\xi(\omega)) + \eta_g(\xi(\omega))$$
$$\quad - \sum_{l \in \mathrm{L}} D_{l,g} \left(\lambda_l^+(\xi(\omega)) - \lambda_l^-(\xi(\omega)) \right), g \in \mathcal{G},$$
$$0 = \sum_{g \in \mathcal{G}} r_g(\xi(\omega)),$$
$$0 \le \eta_g(\xi(\omega)) \perp r_g(\xi(\omega)) + q_g \ge 0, g \in \mathcal{G},$$
$$0 \le \lambda_l^-(\xi(\omega)) \perp K_l + \sum_{g \in \mathcal{G}} D_{l,g} r_g(\xi(\omega)) \ge 0, l \in \mathrm{L},$$
$$0 \le \lambda_l^+(\xi(\omega)) \perp K_l - \sum_{g \in \mathcal{G}} D_{l,g} r_g(\xi(\omega)) \ge 0, l \in \mathrm{L}.$$

(6.53)

This is a two stage stochastic Nash equilibrium problem with equilibrium constraints: in the first stage, generators make a decision on their production quantities and in doing so they anticipate the ISO to decide power flows r_g, $g \in \mathcal{G}$ in the second stage by solving (6.51) which becomes an equilibrium constraint in (6.53) (through the equivalent formulation (6.52)). Theoretically speaking, the ISO's problem may have multiple optimal solutions or equivalently the second stage equilibrium constraints in (6.53) may define multiple equilibria. By considering the maximization with respect to $r_g(\cdot)$ in the formulation we implicitly assume that each generator is optimistic in anticipating the ISO's decision outcome. For simplicity of discussion, we assume that ISO's decision problem has a unique optimal solution and consequently maximization with respect to $r_g(\cdot)$ can be dropped.

Note also that the equilibrium constraints form a complementarity problem which defines a nonconvex feasible set and hence problem (6.53) is *nonconvex* in general. Moreover following our discussion in the introduction, the two stage stochastic Nash equilibrium problem can be reformulated as a one stage stochastic Nash equilibrium problem. The objective function of the reformulated problem is the expected value of the optimal value function of the second stage problem which is usually nonsmooth and nonconvex.

We apply the sample average approximation method to problem (6.53). Let $\xi^1, \ldots, \xi^N$ be an i.i.d. sample of $\xi(\omega)$. The sample average approximation of Nash equilibrium problem (6.51) is: find $\mathbf{q}^N := \left(q_1^N, \cdots, q_G^N \right) \in \mathcal{Q}_1 \times \mathcal{Q}_2 \times \cdots \times \mathcal{Q}_G$ such that for $i = 1, 2, \cdots, M$, $\mathbf{q}_i^N = \{q_g^N\}_{g \in \mathcal{G}_i}$ solves the following optimization problem:

$$\max_{q_g, g \in \mathcal{G}_i} \quad \frac{1}{N} \sum_{k=1}^{N} \left[\sum_{g \in \mathcal{G}_i} p_g \left(r_g \left(\xi^k \right) + q_g, \xi^k \right) q_g \right] - \sum_{g \in \mathcal{G}_i} C_g(q_g)$$

$$\text{s.t.} \quad \text{for } k = 1, 2, \ldots, N,$$
$$0 = p_g \left(r_g \left(\xi^k \right) + q_g, \xi^k \right) - \varrho \left(\xi^k \right) + \eta_g \left(\xi^k \right) - \sum_{l \in \mathrm{L}} D_{l,g} \left(\lambda_l^+ \left(\xi^k \right) - \lambda_l^- \left(\xi^k \right) \right), \quad g \in \mathcal{G},$$
$$0 = \sum_{g \in \mathcal{G}} r_g \left(\xi^k \right),$$
$$0 \le \eta_g \left(\xi^k \right) \perp r_g \left(\xi^k \right) + q_g \ge 0, \quad g \in \mathcal{G},$$
$$0 \le \lambda_l^- \left(\xi^k \right) \perp K_l + \sum_{g \in \mathcal{G}} D_{l,g} r_g \left(\xi^k \right) \ge 0, \quad l \in \mathrm{L},$$
$$0 \le \lambda_l^+ \left(\xi^k \right) \perp K_l - \sum_{g \in \mathcal{G}} D_{l,g} r_g \left(\xi^k \right) \ge 0, \quad l \in \mathrm{L},$$
$$0 \le q_g, \quad g \in \mathcal{G}_i, \ i = 1, 2, \cdots, M,$$
$$0 \le \bar{q}_g - q_g, \quad g \in \mathcal{G}_i, \ i = 1, 2, \cdots, M,$$

(6.54)

where the feasible set of the generation quantity of generator i at node $g \in \mathcal{G}_i \subset \mathcal{G}$ is $\mathcal{Q}_g = [0, \bar{q}_g]$.

6.2.2   A three-node network

To explain the numerical schemes outlined in the preceding subsection in detail, we consider a three-node network connected with three transmission lines as indicated in Figure 1. Demand occurs at each node but there are only two generators located in node 1 and node 2. Following the notation in the preceding
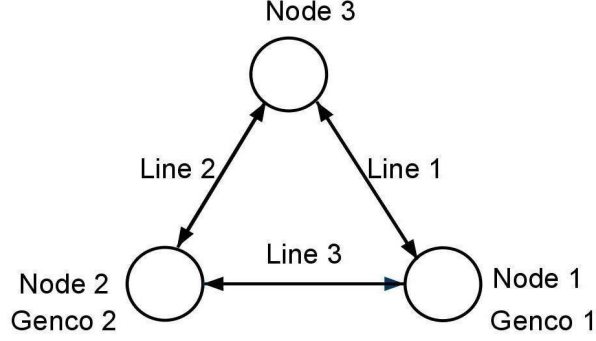
Figure 1: A three-node example with two generators.

subsection, $\mathcal{G}_1 = \{1\}$ and $\mathcal{G}_2 = \{2\}$, $\mathcal{G} = \{1, 2, 3\}$ and $M = 2$, $\text{L} = \{1, 2, 3\}$ with the total number of lines $L = 3$. To avoid confusion, we denote the transmission lines by $\text{L} = \{l_1, l_2, l_3\}$. Other parameters are summarized in the following tables:

Table 6: Price functions and cost functions at each nodal market

node g	generator i	$p_g(\tau_g, \xi)$	$C_i(q_i)$
$g = 1$	1	$a_1(\xi) - b_1(\xi)\tau_1$	$\alpha_1 q_1 + \beta_1 q_1^2$
$g = 2$	2	$a_2(\xi) - b_2(\xi)\tau_2$	$\alpha_2 q_2 + \beta_2 q_2^2$
$g = 3$	-	$a_3(\xi) - b_3(\xi)\tau_3$	-

Table 7: Power transfer distribution factors (PTDFs) at the market

node g	$D_{l_1,g}$	$D_{l_2,g}$	$D_{l_3,g}$
$g = 1$	$D_{1,1}$	$D_{2,1}$	$D_{3,1}$
$g = 2$	$D_{1,2}$	$D_{2,2}$	$D_{3,2}$
$g = 3$	$D_{1,3}$	$D_{2,3}$	$D_{3,3}$
K_l	K_1	K_2	K_3

In Table 7, the power transfer distribution factor $D_{l,g}$ specifies the proportion of flow on a line $l \in \text{L}$ resulting from an injection of one-unit electricity at node $g \in \mathcal{G}$ and corresponding one-unit withdrawal at some fixed reference node. In this model, we set node 3 as a reference node, which means $D_{l,3} = 0$ for all $l = 1, 2$ and 3. The choice of the reference node only affects the matrix of the power transfer distribution factors, and the power flows to a certain reference node is not of importance for the global system solution, but is used in the calculation of the transmission flows and is necessary for a proper understanding of PTDFs, see [10] for details.

Based on the structure of the system as indicated in Figure 1 and the parameters listed in Tables 6 and 7, the ISO's decision problem for a particular realization $\xi = \xi(\omega)$ can be written as:

$$
\begin{aligned}
\max_{r_1(\xi), r_2(\xi), r_3(\xi)} \quad & \sum_{g=1}^{3} \left(\int_0^{r_g(\xi) + q_g} (a_g(\xi) - b_g(\xi)\tau_g)\, d\tau_g \right) - \left(\alpha_1 q_1 + \beta_1 q_1^2 \right) - \left(\alpha_2 + \beta_2 q_2^2 \right) \\
\text{s.t.} \quad & r_1(\xi) + r_2(\xi) + r_3(\xi) = 0, \\
& r_1(\xi) + q_1 \geq 0, \\
& r_2(\xi) + q_2 \geq 0, \\
& r_3(\xi) \geq 0, \\
& -K_l \leq D_{l,1} r_1(\xi) + D_{l,2} r_2(\xi) + D_{l,3} r_3(\xi) \leq K_l, \quad \text{for } l = 1, 2, 3.
\end{aligned}
\tag{6.55}
$$

Consequently, competition between the two generators can be mathematically modeled as a stochastic Nash

27

equilibrium problem: find (q_1^*, q_2^*) which solves the following problem:

$$
\begin{aligned}
\max_{q_g, g \in \mathcal{G}_i} \quad & \mathbb{E}\left[(a_g(\xi(\omega)) - b_g(\xi(\omega))(r_g(\xi(\omega)) + q_g))q_g\right] - \left(\alpha_g q_g + \beta_g q_g^2\right) \\
\text{s.t.} \quad & \text{for a.e. } \omega \in \Omega \\
& 0 = a_g(\xi(\omega)) - b_g(\xi(\omega))(r_g(\xi(\omega)) + q_g) - \varrho(\xi(\omega)) + \eta_g(\xi(\omega)) \\
& \quad - \sum_{l=1}^{3} D_{l,g}\left(\lambda_l^+(\xi(\omega)) - \lambda_l^-(\xi(\omega))\right), \; g = 1, 2, \\
& 0 = a_g(\xi(\omega)) - b_g(\xi(\omega))r_g(\xi(\omega)) - \varrho(\xi(\omega)) + \eta_g(\xi(\omega)) \\
& \quad - \sum_{l=1}^{3} D_{l,g}\left(\lambda_l^+(\xi(\omega)) - \lambda_l^-(\xi(\omega))\right), \; g = 3, \\
& 0 = r_1(\xi(\omega)) + r_2(\xi(\omega)) + r_3(\xi(\omega)), \\
& 0 \le \eta_g(\xi(\omega)) \perp r_g(\xi(\omega)) + q_g \ge 0, g = 1, 2, \\
& 0 \le \eta_g(\xi(\omega)) \perp r_g(\xi(\omega)) \ge 0, g = 3, \\
& 0 \le \lambda_l^-(\xi(\omega)) \perp K_l + \sum_{g=1}^{3} D_{l,g}r_g(\xi(\omega)) \ge 0, l = 1, 2, 3, \\
& 0 \le \lambda_l^+(\xi(\omega)) \perp K_l - \sum_{g=1}^{3} D_{l,g}r_g(\xi(\omega)) \ge 0, l = 1, 2, 3, \\
& 0 \le q_g, g = 1, 2, \\
& 0 \le \bar{q}_g - q_g, g = 1, 2.
\end{aligned}
\tag{6.56}
$$

We carry out a number of numerical tests for solving sample average approximation of (6.56) with the data specified in Examples 6.3-6.4: In Example 6.3, we test the convergence as the sample size N increases; in Example 6.4, we perform comparative static analysis (with fixed sample size) on the sensitivity of various economic variables such as generator's dispatch, expected profits and the expected market price with respect to the change of generator 1's marginal cost. We use mathematical programming codes in GAMS installed in a PC with Windows XP operating system and solver *PATH* for solving sample average approximation of (6.56).

Example 6.3 *Consider the inverse demand function $p_g(q_g, \xi_g) = a_g + \xi_g - b_g q_g$, at node $g = 1, 2, 3$, where ξ_g follows a truncated normal distribution with support $[-5, 5]$, mean value 0 and standard deviation 1. Let $\xi = (\xi_1, \xi_2, \xi_3)$ (In this case, ξ_g describes the demand uncertainty at node g while ξ represents all of the uncertainties. The power flow $\tau_g(\xi)$ depends ξ in general). The other parameters are given in Tables 8 and 9.*

Table 8: Price functions and cost functions at each nodal market

node g	generator i	$p_g(\tau_g, \xi)$	$C_i(q_i)$
$g = 1$	$i = 1$	$25 + \xi_1 - 1.2(q_1 + \tau_1(\xi))$	$5q_1 + q_1^2$
$g = 2$	$i = 2$	$26.5 + \xi_2 - 1.3(q_2 + \tau_2(\xi))$	$2q_2 + 1.5q_2^2$
$g = 3$	-	$27.9 + \xi_3 - 1.5\tau_3$	-

Table 9: Power transfer distribution factors (PTDFs) at the network

node g	$D_{l_1,g}$	$D_{l_2,g}$	$D_{l_3,g}$
$g = 1$	0.5	-0.5	-0.5
$g = 2$	0.5	-0.5	0.5
K_l	1.8	1.8	1.8

We carry out a number of numerical tests for sample average approximation of (6.56) with sample sizes $500, 1000$ and 2000. The results are displayed in Table 10 where $\bar{r}_g(q_g^N) = \frac{1}{N}\sum_{k=1}^{N}\tau_g(\xi^k)$, $\bar{p}_g(q_g^N) = \frac{1}{N}\sum_{k=1}^{N}p_g(q_g + \tau_g(\xi^k), \xi^k)$ for $g = 1, 2, 3$ and R_i^N, $i = 1, 2$, is the sample average of generator i's profit. The results show that there is no significant improvement as the sample size increases from 500 to 2000. This is consistent with the observation in Subsection 6.1.

It is important to note that (q_1^N, q_2^N) displayed in Table 10 are obtained from solving sample average approximation of (6.56) and they are (approximate) weak Nash-C-stationary points of the true problem (6.56). In order to verify whether they are Nash equilibria or not, we may look into the quasiconvexity of each generator's objective function by fixing its rival's dispatch. It is very difficult to do this in this case because each generator's decision problem is a two-stage mathematical program with equilibrium constraints. Similar comments apply to Example 6.4. In the case when $r_g(\xi)$, $g = 1, 2$, is piecewise linear function of q_1, q_2, we can prove that the objective function in (6.56) are piecewise concave provided that $\beta_g \geq b_g(\xi)$ w.p.1 for $g = 1, 2$ and subsequently (q_1^N, q_2^N) is an (approximate) local Nash equilibrium.

Table 10: Numerical results of Example 6.3.

N	q_1^N, q_2^N	$\bar{r}_1(q_g^N), \bar{r}_2(q_g^N), \bar{r}_3(q_g^N)$	$\bar{p}_1(q_1^N), \bar{p}_2(q_2^N), \bar{p}_3(q_3^N)$	R_1^N, R_2^N
500	$5.407, 4.915$	$-2.573, -1.003, 3.575$	$21.560, 21.415, 22.638$	$60.516, 59.186$
1000	$5.413, 4.920$	$-2.575, -1.004, 3.579$	$21.594, 21.409, 22.632$	$60.524, 59.182$
2000	$5.416, 4.922$	$-2.576, -1.004, 3.580$	$21.591, 21.406, 22.630$	$60.528, 59.180$

Example 6.4 *We now move on to investigate the impact of the change of a single generator's marginal cost on the other generator's dispatch and nodal prices etc. Specifically, we consider generator 1's cost function $C_1(q_1) = 5q_1 + \beta_1 q_1^2$ and look into the impact when parameter β_1 increases from 0.7 to 1.6. Observe that the marginal cost is $5 + 2\beta_1 q_1$ and it increases as parameter β_1 increases. We list details of other data in Tables 6.4 and 6.4 and leave distribution of ξ as in Example 6.3. The sample size is 1000.*

Table 11: Price functions and cost functions at each nodal market

node market g	generator i	$p_g(\tau_g, \xi)$	$C_i(q_i)$
$g = 1$	1	$26.5 + \xi_1 - 1.2(q_1 + \tau_1(\xi))$	$5q_1 + \beta_1 q_1^2$
$g = 2$	2	$26.5 + \xi_2 - 1.2(q_2 + \tau_2(\xi))$	$5q_2 + 1.2q_2^2$
$g = 3$	-	$28.0 + \xi_3 - 1.5\tau_3(\xi)$	-

Table 12: Power transfer distribution factors (PTDFs) at the market

node g	$D_{l_1,g}$	$D_{l_2,g}$	$D_{l_3,g}$
$g = 1$	0.5	-0.5	-0.5
$g = 2$	0.5	-0.5	0.5
K_l	1.8	1.8	1.8

Figures 2-4 show how the change of parameter value β_1 from 0.7 to 1.6 affects the dispatch and profits of the two generators and market prices at the three nodes. Figure 2 shows that when $\beta_1 = 1.2$ the dispatches of the two generators are equal as the two generators are in a symmetric position. Figure 3 depicts the change of average nodal prices as β_1 increases. Note that when $\beta_1 \leq 1.1$, power flow on each line reaches the line capacity limit ($K_l = 1.8$) therefore the price at node 3 remains constant. The price at node 2 is slightly higher than the price at node 1 because generator 2 produces less power at node 2. When $\beta_1 > 1.1$, power flow on each line drops below the line capacity limit. Consequently the three nodal prices coincide. Figure 4 depicts the change of expected profits of the two generators.
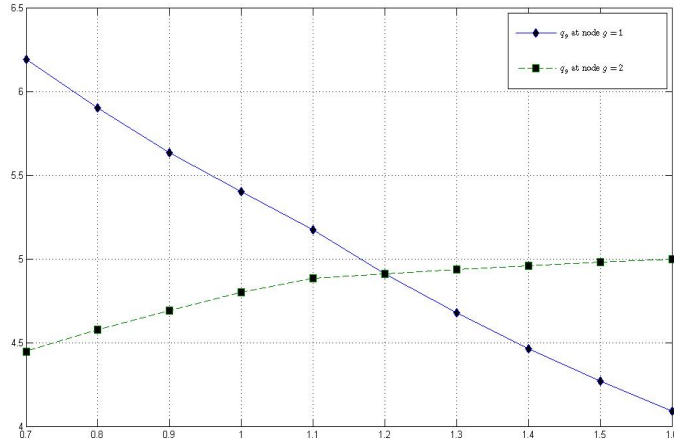
Figure 2: Comparative static analysis on the dispatch quantities with respect to the change of β_1.
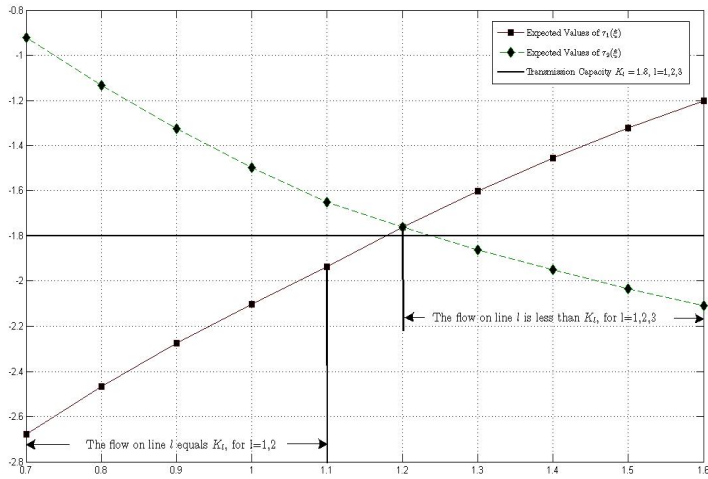

Figure 3: Comparative static analysis on the expected transmission flows with respect to the change of β_1.
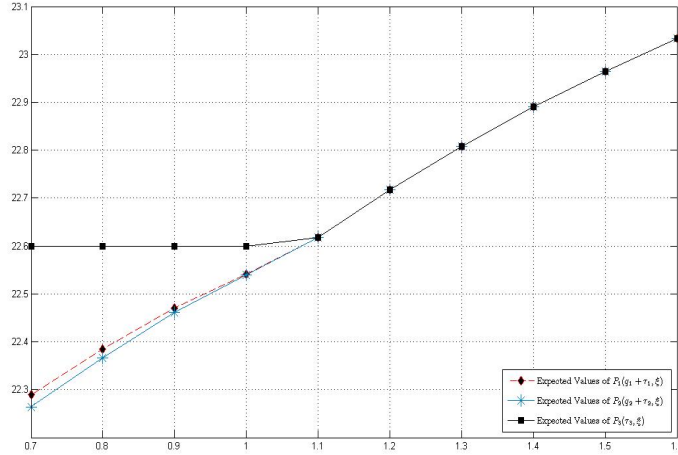
Figure 4: Comparative static analysis on the expected node prices with respect to the change of β_1.
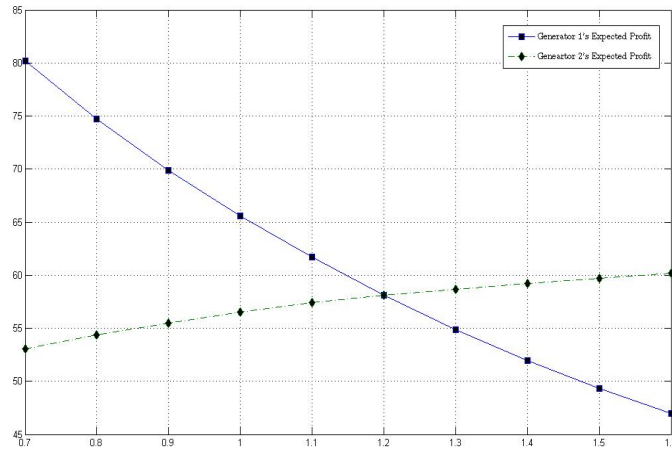

Figure 5: Comparative static analysis on the expected profits with respect to the change of β_1.

31

References

[1] B. Allaz and J-L. Vila, Cournot competition, futures markets and efficiency, *Journal of Economic Theory*, Vol. 59, pp. 1-16, 1993.

[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] J.-P. Aubin, *Set-Valued Analysis*, Birkhauser, Boston, 1990. *Optima and Equilibria: An Introduction to Nonlinear Analysis*, Springer-Verlag, 1993.

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

[7] R. J. Aumann, Correlated equilibrium as an expression of Baysesian rationality, *Economica*, Vol. 55, pp. 1-18, 1987.

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

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

[10] E. Delarue, D. Bekaert, R. Belmans, and W. D'haeseleer, Development of a comprehensive electricity generation simulation model using a mixed integer programming approach, *International Journal of Electrical, Computer, and Systems Engineering*, Vol. 1, pp. 92-97, 2007.

[11] 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.

[12] C. J. Day, B. F. Hobbs and J-S. Pang, Oligopolistic competition in power networks: A conjectured supply function approach, *IEEE Transactions on Power Systems*, Vol. 17, pp. 597-607, 2002.

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

[14] V. DeMiguel and H. Xu, A stochastic multiple leader Stackelberg model: analysis, computation, and application, *Operations Research*, Vol. 57, pp. 12201235, 2009.

[15] 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.

[16] F. Facchinei, A. Fischer and V. Piccialli, Generalized Nash equilibrium problems and Newton methods, *Mathematical Programming*, Vol. 117, pp. 163-194, 2009.

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

[18] J. S. Gans, D. Price and K. and Woods, Contracts and electricity pool prices, *Australian Journal of Management*, Vol. 23, pp. 8396, 1998.

[19] J. C. Harsanyi, Games with incomplete information played by "Bayesian" players, Part I. The basic model, *Management Science*, Vol. 14, pp. 159-182, 1967.

[20] A. Haurie, G. Zaccour and Y. Smeers, Stochastic equilibrium programming for dynamic oligopolistic markets, *Journal of Optimization Theory and Applications*, Vol. 66, pp. 243-253, 1990.

[21] 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.

[22] R. Henrion 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.

[23] F. Hiai, Convergence of conditional expectations and strong law of large numbers for multivalued random variables, *Transaction of The American Mathematical Society*, Vol. 291, pp. 613-627, 1985.

[24] B. F. Hobbs, C. B. Metzler and J.-S. Pang, Strategic gaming analysis for electric power systems: An MPEC approach, *IEEE Transaction on Power Systems*, Vol. 15, pp. 637-645, 2000.

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

[26] X. Hu and D. Ralph, Using EPECs to model bilevel games in restructured electricity markets with locational prices, *Operations Research*, Vol. 55, pp. 809-827, 2007.

[27] T. Homen-De-Mello, On rates of convergence for stochastic optimization problems under non-independent and identically distributed sampling, *SIAM Journal on Optimization*, Vol. 19, pp. 524-551, 2008.

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

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

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

[31] 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.

[32] P. D. Klemperer and M. A. Meyer, Supply function equilibria in oligopoly under uncertainty, *Econometrica*, Vol. 57, pp. 1243-1277, 1989.

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

[34] 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.

[35] Y. Lucet and J.J. Ye, Sensitivity analysis of the value function for optimization problems with variational inequality constraints, *SIAM Journal on Control and Optimization*, Vol. 40, pp. 699-723, 2001.

[36] I. Molchanov, *Theory of Random Sets*, in *Probability and Its Applications*, J. Gani et al eds. Springer, 2005.

[37] 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.

[38] N. Papageorgiou, On the theory of Banach space valued multifunctions 1. Integration and conditional expectation, *Journal of Multivariate Analysis*, Vol. 17, pp. 185-206, 1985.

[39] J. Outrata, A note on a class of equilibrium problems with equilibrium constraints, Vol. 40, pp. 585-594, 2004.

[40] J. S. Pang and M. Fukushima, Quasi-variational inequalities, generalized Nash equilibria, and multi-leader-follower games, *Computational Management Science*, Vol. 2, pp. 21-56, 2005.

[41] 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.

[42] D. Ralph and H. Xu, Asympototic analysis of Stationary Points of Sample Average Two Stage Stochastic Programs: A Generalized Equation Approach, manuscript, 2008. Submitted to *Mathematics of Operations Research*, under revision.

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

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

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

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

[47] 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.

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

[49] 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.

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

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

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

[53] 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.

[54] 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.

[55] 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, 2007.

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

[57] 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.

[58] 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.

[59] J-Y Wei and Y. Smeers, Spatial oligopolistic electricity models with cournot generators and regulated transmission prices, *Operations Research*, Vol. 47, pp. 102-112, 1999.

[60] 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.

[61] H. Xu and D. Zhang, Smooth sample average approximation of nonsmooth stochastic problems and applications, *Mathematical Programming Series A*, Vol. 119, pp. 371-401, 2009.

[62] H. Xu, Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming, School of Mathematics, University of Southampton, 2008.

[63] J. Yao, S. S. Oren and I. Adler, Two-settlement electricity markets with price caps and Cournot generation firms, *European Journal of Operational Research,* Vol. 181, pp. 1279-1296, 2007.

[64] J. Yao, I. Adler and S. S. Oren, Modeling and computing two-settlement oligopolistic equilibrium in a congested electricity network, *Operations Research,* Vol. 56, pp. 34-47, 2008.

[65] D. Zhang, H. Xu and Y. Wu, A stochastic two stage equilibrium model for electricity markets with two way contracts, To appear in *Mathematical Methods of Operations Research.*

Appendix

Proof of Proposition 3.1. Part (i). Let $u_i \in \mathbb{R}^{n_i}$ be fixed. By definition, the Clarke generalized derivative [9] of $v_i(x_i, x_{-i}, \xi)$ with respect to x_i at a point x_i in direction u_i is defined as

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

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], $(v_i)^o_{x_i}(x_i, x_{-i}, \xi; u_i)$ is also measurable. Since $(v_i)^o_{x_i}(x_i, x_{-i}, \xi; u_i)$ 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 $\mathbb{E}[\partial_{x_i} v_i(x_i, x_{-i}, \xi)]$. ∎

Proof of Proposition 4.1 part (i). The proof is similar to that of [56, Theorem 5.1]. In [56, Theorem 5.1] it is assumed that (4.19) holds for all $x, x' \in \mathcal{X}$ and all $\xi \in \Xi$. However in the proof of the theorem, this condition is used only at a finite number of points $\bar{x}_1, ..., \bar{x}_M \in \mathcal{X}$ which form a δ-net of $\mathcal{X}$ (that is, for every $x \in \mathcal{X}$, there exists $\bar{x}_i$, $i \in \{1, ..., M\}$ such that $\|x - \bar{x}_i\| \leq \nu$). Also, the Hölder condition holds "for all $\xi \in \Xi$ can obviously be weakened to "for almost every $\xi \in \Xi$". In part (i), this condition corresponds to (4.17). Under condition (C2), we can find a unified integrable function $\kappa(\xi)$ and constant γ such that (4.17) holds at the points $\bar{x}_1, \cdots, \bar{x}_M$. This can be achieved by setting $\kappa(\xi) = \max_{i=1}^M \kappa_{\bar{x}_i}(\xi)$ and $\gamma = \min_{i=1}^M \gamma_{\bar{x}_i}$ where $\kappa_{\bar{x}_i}(\xi)$ and $\gamma_{\bar{x}_i}$ denote the κ and γ at point $\bar{x}_i$, $i = 1, \cdots, M$. The rest of the proof is similar to that of [56, Theorem 5.1]. ∎

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 $q_i, Q_{-i} \geq 0$. From the definition of $r_i(q_i, Q_{-i}, \xi)$, we have for $i = 1, 2, \ldots, M$,

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

Hence, for $j \neq i$,

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

and $j = i$,

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

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

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

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

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

Let

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

and

$$B_i(q_i, Q_{-i}) = \quad \mathbb{E}[2p_q'(Q,\xi) - C_i''(q_i) + q_i p_{qq}''(Q,\xi)]\text{Prob}\,(p(Q,\xi) < f)$$
$$+\mathbb{E}[2p_q'(Q,\xi) - C_i''(q_i) + (q_i - w_i)p_{qq}''(Q,\xi)]\text{Prob}\,(p(Q,\xi) > f)\,.$$

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

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

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. ∎