Math. Program., Ser. A
DOI 10.1007/s10107-008-0214-0

FULL LENGTH PAPER

Smooth sample average approximation of stationary points in nonsmooth stochastic optimization and applications

Huifu Xu · Dali Zhang

Received: 14 January 2007 / Accepted: 12 November 2007

Abstract Inspired by a recent work by Alexander et al. (J Bank Finance 30:583–605, 2006) which proposes a smoothing method to deal with nonsmoothness in a conditional value-at-risk problem, we consider a smoothing scheme for a general class of nonsmooth stochastic problems. Assuming that a smoothed problem is solved by a sample average approximation method, we investigate the convergence of stationary points of the smoothed sample average approximation problem as sample size increases and show that w.p.1 accumulation points of the stationary points of the approximation problem are weak stationary points of their counterparts of the true problem. Moreover, under some metric regularity conditions, we obtain an error bound on approximate stationary points. The convergence result is applied to a conditional value-at-risk problem and an inventory control problem.

Keywords Smoothing method · Sample average approximation · Stationary points · Error bound

Mathematics Subject Classification (2000) 90C15 · 65K05 · 91B28

1 Introduction

In this paper, we consider the following nonsmooth stochastic minimization problem

$$\begin{array}{l} \min \ \mathbb{E}\left[f(x, \xi(\omega))\right] \\ \text{s.t.} \ \ x \in \mathcal{X}, \end{array} \tag{1.1}$$

H. Xu (✉) · D. Zhang
School of Mathematics, University of Southampton, Highfield Southampton, UK
e-mail: h.xu@soton.ac.uk

where $f : \mathbb{R}^m \times \mathbb{R}^k \to \mathbb{R}$ is locally Lipschitz continuous but not necessarily continuously differentiable, $\xi : \Omega \to \Xi \subset \mathbb{R}^k$ is a random vector defined on probability space $(\Omega, \mathcal{F}, P)$, $\mathbb{E}$ denotes the mathematical expectation, $x \in \mathcal{X}$ is a decision vector with $\mathcal{X}$ being a nonempty subset of $\mathbb{R}^m$. Throughout this paper, we assume that $\mathbb{E}[f(x, \xi(\omega))]$ is well defined for every $x \in \mathcal{X}$. To ease the notation, we will write $\xi(\omega)$ as ξ and this should be distinguished from ξ being a deterministic vector of Ξ in a context.

Nonsmooth stochastic programming model (1.1) covers a number of interesting problems such as stochastic programs with recourse and stochastic min-max problems, see [11,32,38] and references therein.

This paper is concerned with numerical methods for solving (1.1). We deal with two main issues: one is the nonsmoothness of $f(x, \xi)$ and the other is the mathematical expectation operator in the objective function. Over the past few decades, a number of effective numerical methods such as bundle methods [18] and aggregate subgradient methods [17] have been proposed for solving general nonsmooth deterministic optimization problems. Rusczyński [32] proposes a stochastic bundle-like method for solving nonsmooth stochastic optimization problems and shows the convergence of stationary points. Another well known method for solving nonsmooth stochastic optimization problem is stochastic quasi-subgradient method [10]. The method is simple to implement in that it only requires calculation of a quasi-subgradient of the objective function at each iteration albeit the convergence of the method is relatively slow.

In practical applications, nonsmooth problems are often well structured. For instance, Rockafelar and Uryasev [27] reformulate the minimization of conditional value-at-risk (CVaR for short) in finance as a nonsmooth stochastic minimization problem where the nonsmoothness is essentially caused by a max-function. Similar cases can also be found in inventory control problems in supply chain. Undoubtedly these nonsmooth stochastic minimization problems can be solved by the stochastic bundle-like methods [32], however they can be also treated by simple smoothing techniques. For instance, Alexander et al. [1] use an elementary smoothing function to deal with the nonsmoothness of stochastic program in CVaR where the nonsmoothness is caused by a max-function. They show how the smoothing method can save significant amount of calculations from a linear programming approach. This kind of elementary smoothing techniques is easy to handle and numerically effective in many cases, indeed they have been extensively exploited to deal with nonlinear complementarity problems and mathematical problems with equilibrium constraints (MPEC) [12,20]. More recently the smoothing methods have been used for two stage stochastic programs with nonsmooth equality constraints and stochastic MPECs [21,39,40].

In this paper we propose a general smoothing scheme which uses a smoothing function $\hat{f}(x, \xi, \epsilon)$ parameterized by a number ϵ with small absolute value to approximate $f(x, \xi)$. This type of smoothing is considered in [24] and covers many useful elementary smoothing functions. Consequently we consider a smoothed stochastic program

$$\begin{aligned} &\min\ \mathbb{E}\left[\hat{f}(x, \xi, \epsilon)\right] \\ &\text{s.t.}\ \ x \in \mathcal{X}, \end{aligned} \tag{1.2}$$

and investigate the approximation of (1.2)–(1.1) as the smoothing parameter ϵ is driven to zero. Here $\mathcal{X}$ is defined as in (1.1).

We next deal with the mathematical expectation in the objective function. The way to tackle this issue depends on the availability of the information of ξ and the property of f. If we know the distribution of ξ and can integrate out the expected value explicitly, then the problem becomes a deterministic minimization problem, no discretization procedures are required. Throughout this paper, we assume that $\mathbb{E}[f(x,\xi)]$ cannot be calculated in a closed form so that we will have to approximate it through discretization.

One of the most well known discretization approaches is Monte Carlo simulation based method. The basic idea of the method is to generate an independent identically distributed (i.i.d.) sample $\xi^1,\ldots,\xi^N$ of ξ and then approximate the expected value with sample average, that is,

$$\begin{array}{ll}\min\limits_{x} & \frac{1}{N}\sum_{i=1}^{N}\hat{f}(x,\xi^i,\epsilon)\\ \text{s.t.} & x\in\mathcal{X}.\end{array}\tag{1.3}$$

We refer to (1.1) as true problem and (1.3) as sample average approximation (SAA) problem. SAA methods have been extensively investigated in stochastic optimization. This type of methods are also known as sample path optimization (SPO) methods. There has been extensive literature on SAA and SPO. See recent work [3,16,19,23,26,30,35] and a comprehensive review by Shapiro [34].

Most convergence analysis of SAA problems in the literature concerns the convergence of optimal solutions and optimal values [34], that is, if we solve (1.3) and obtain an optimal solution, what is the convergence of the optimal solution sequence as sample size N increases? Our interest here, however, is on the convergence of stationary points, that is, if we obtain a stationary point of (1.3) which is not necessarily an optimal solution, then what is the accumulation point of the SAA stationary sequence? The rational behind this is that in some practical instances, $f(x,\xi)$ is non-convex and consequently the smoothed SAA problem (1.3) is also non-convex. Under these circumstances, it is more likely to obtain a stationary point rather than an optimal solution in solving (1.3). To investigate whether or not an accumulation point of a stationary sequence of SAA is an optimal solution of the true problem, we will need more information about the properties of the true problem such as convexity or quasi-convexity of the objective function and the structure of feasible set $\mathcal{X}$.

The main contributions of this paper as far as we are concerned can be summarized as follows: we propose a smoothing SAA method for solving a general class of one stage nonsmooth stochastic problems. We generalize a convergence theorem established by Shapiro [34] on SAA method for a stochastic generalized equation and use it to show that under moderate conditions w.p.1 the stationary points of smoothed sample average approximation problem converge to the weak stationary points of the true problem and, when the underlying functions are convex, to optimal solutions. When the smoothing parameter is fixed, we obtain an error bound for the SAA stationary points under some metric regularity condition. Finally we apply the convergence results to a CVaR problem and an inventory control problem in supply chain.

The rest of the paper is organized as follows. In Sect. 2, we discuss the measurability of Clarke generalized gradient of locally Lipschitz continuous random functions. In

Sect. 3, we discuss smoothing techniques and smooth approximation of stationary points. In Sect. 4, we discuss sample average approximation of the smoothed problem and analyze convergence of stationary point of the smoothed SAA problem. In Sect. 5, we apply the convergence results established in Sect. 4 to a CVaR problem and an inventory control problem in supply chain.

2 Preliminaries

Throughout this paper, we use the following notation. $\|\cdot\|$ denotes the Euclidean norm of a vector and a compact set of vectors. When $\mathcal{M}$ is a compact set of vectors, $\|\mathcal{M}\| := \max_{M\in\mathcal{M}} \|M\|$. $d(x,\mathcal{D}) := \inf_{x'\in\mathcal{D}} \|x-x'\|$ denotes the distance from point x to set $\mathcal{D}$. For two compact sets $\mathcal{C}$ and $\mathcal{D}$,

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

denotes the deviation from $\mathcal{C}$ to $\mathcal{D}$ (in some references [13] also called *excess* of $\mathcal{C}$ over $\mathcal{D}$) , and $\mathbb{H}(\mathcal{C},\mathcal{D})$ denotes the Hausdorff distance between the two sets, that is,

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

We use $B(x,\delta)$ to denote the closed ball in $\mathbb{R}^m$ with radius δ and center x, that is $B(x,\delta) := \left\{x' \in \mathbb{R}^m : \|x'-x\| \le \delta\right\}$. When δ is dropped, $B(x)$ represents a neighborhood of point x. Finally, for a set-valued mapping $\mathcal{A} : \mathbb{R}^m \to 2^{\mathbb{R}^m}$, we use $\overline{\lim}_{y\to x} \mathcal{A}(y)$ to denote the outer limit of the mapping at point x.

2.1 Clarke generalized gradient and measurability

Let $\xi \in \Xi$ be fixed. The Clarke generalized gradient [8] of $f(x,\xi)$ with respect to x at $x \in \mathbb{R}^m$ is defined as

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

where $D_{f(\cdot,\xi)}$ denotes the set of points near x where $f(x,\xi)$ is Frechét differentiable with respect to x, $\nabla_x f(y,\xi)$ denotes the usual gradient of $f(x,\xi)$ in x and 'conv' denotes the convex hull of a set. It is well known that the Clarke generalized gradient $\partial_x f(x,\xi)$ is a convex compact set and it is upper semicontinuous [8, Proposition 2.1.2, 2.1.5]. In this paper, we assume that for every $\xi \in \Xi$, $f(x,\xi)$ is locally Lipschitz continuous, hence $\partial_x f(x,\xi)$ is well defined.

In what follows, we fix x and discuss the measurability of the set-valued mapping $\partial_x f(x,\xi(\cdot)) : \Omega \to 2^{\mathbb{R}^m}$.

Let $\mathfrak{B}$ denote the space of nonempty, compact subsets of $\mathbb{R}^m$ equipped with the Hausdorff distance. Then $\partial_x f(x,\xi(\cdot))$ can be viewed as a single valued mapping from

Ω to $\mathfrak{B}$. By [29, Theorem 14.4], we know that $\partial_x f(x, \xi(\cdot))$ is measurable if and only if for every $B \in \mathfrak{B}$, $\partial_x f(x, \xi(\cdot))^{-1}B$ is $\mathcal{F}$-measurable.

Proposition 2.1 *Let $f(x, \xi)$ be a locally Lipschitz continuous function in both x and ξ. The Clarke generalized gradient $\partial_x f(x, \xi)$ is measurable.*

Proof Let $d \in \mathbb{R}^m$ be fixed. By definition, the Clarke generalized derivative [8] of $f(x, \xi)$ with respect to x at a point x in direction d is defined as

$$f^o(x, \xi; d) := \limsup_{\substack{y \to x \\ t \to 0}} [f(y + td, \xi) - f(y, \xi)]/t.$$

Since f is continuous in ξ and ξ is a random vector, then f is measurable and so is $f^o(x, \xi; d)$ [5, Lemma 8.2.12]. Since $f^o(x, \xi; d)$ is the support function of $\partial_x f(x, \xi)$, by [5, Theorem 8.2.14], $\partial_x f(x, \xi)$ is measurable. □

Remark 2.1 The conclusion also holds when $f(x, \xi)$ is a vector valued function, in which case $\partial_x f(x, \xi)$ is the Clarke generalized Jacobian [8, Definition 2.6.1]. To see this, notice that the Clarke generalized Jacobian at a point x is defined as the convex hull of the limiting classical Jacobians of f at points near x where f is Frechét differentiable while the classical Jacobians are measurable, by [29, Theorem 14.20] and [5, Theorem 8.2.2], both the limit operation and the convex hull operation preserve the measurability.

In some cases, one may consider the Clarke generalized gradient with respect to ξ, that is, $\partial_\xi f(x, \xi)$. In such a case, the Clarke generalized gradient is also measurable in that $\partial_\xi f(x, \xi)$ is upper semicontinuous in ξ, hence it is measurable in ξ by [5, Proposition 8.2.1]. The composition of measurable mappings is measurable.

Proposition 2.1 ensures that $\partial_x f(x, \xi)$ is a random set-valued mapping. We now define the expectation of $\partial_x f(x, \xi)$. A *selection* from a random set $\mathcal{A}(x, \xi(\omega))$ is a random vector $A(x, \xi(\omega)) \in \mathcal{A}(x, \xi(\omega))$, which means $A(x, \xi(\omega))$ is measurable. Note that such selections exist, see [4]. The *expectation of* $\mathcal{A}(x, \xi(\omega))$ which is widely known as Aumann's integral [4,13], is defined as the collection of $\mathbb{E}[A(x, \xi(\omega))]$ where $A(x, \xi(\omega))$ is a selection. We denote the expected value by $\mathbb{E}[\mathcal{A}(x, \xi(\omega))]$. We regard $\mathbb{E}[\mathcal{A}(x, \xi(\omega))]$ as well defined if $\mathbb{E}[\mathcal{A}(x, \xi(\omega))] \in \mathfrak{B}$. A sufficient condition of the well definedness of the expectation is

$$\mathbb{E}[\|\mathcal{A}(x, \xi(\omega))\|] := \mathbb{E}[\mathbb{H}(0, \mathcal{A}(x, \xi(\omega)))] < \infty,$$

see [2] and a comprehensive review by Hess [13] for the Aumann's integral of a random set-valued mapping. From discussions above, we immediately have the following.

Proposition 2.2 *Let $f(x, \xi)$ be a locally Lipschitz continuous function in both x and ξ. Suppose that there exists a measurable function $\kappa(\xi)$ such that $\mathbb{E}[\kappa(\xi)] < \infty$ and*

$$\|\partial_x f(x,\xi)\| \le \kappa(\xi)$$

for all $x \in \mathcal{X}$ *and* $\xi \in \Xi$. *Then* $\mathbb{E}[\partial_x f(x,\xi)]$ *is well defined.*

This result can be generalized to the case when f is vector valued.

3 Smooth approximations

In this section, we introduce a smooth approximation to the nonsmooth stochastic minimization problem (1.1). First we give a general definition of the smoothing.

Definition 3.1 Let $\phi : \mathbb{R}^m \to \mathbb{R}$ be a locally Lipschitz continuous function and $\epsilon \in \mathbb{R}$ be a parameter. $\hat{\phi}(x,\epsilon) : \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}$ is a *smoothing* of ϕ if it satisfies the following:

(a) for every $x \in \mathbb{R}^m$, $\hat{\phi}(x,0) = \phi(x)$;
(b) for every $x \in \mathbb{R}^m$, $\hat{\phi}$ is locally Lipschitz continuous at $(x,0)$;
(c) $\hat{\phi}$ is continuously differentiable on $\mathbb{R}^m \times \mathbb{R}\backslash\{0\}$.

The properties specified in parts (a) and (c) are of common sense, that is, the smoothing function must match the original function when the smoothing parameter is zero and when the smoothing parameter is nonzero, the smoothing function is continuously differentiable. The Lipschitz continuity in part (b) needs a bit specific explanation: it implies that the Clarke generalized gradient $\partial_{(x,\epsilon)}\hat{\phi}(x,0)$ 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

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

where $\pi_x \partial_{(x,\epsilon)}\hat{\phi}(x,0)$ denotes the set of all m-dimensional vectors a such that, for some scalar c, the $(m+1)$-dimensional vector (a,c) belongs to $\partial_{(x,\epsilon)}\hat{\phi}(x,0)$, then $\hat{\phi}$ is said to satisfy *gradient consistency* (which is known as Jacobian consistency when f is vector valued, see [24] and references therein). This is a key property that will be used in the analysis of the first order optimality condition later on.

The definition was first introduced in [24] for smoothing deterministic Lipschitz continuous functions and it is shown that this type of smoothing covers a range of interesting elementary smoothing functions in the literature. In practical applications, some nonsmooth functions have specific structures. The proposition below addresses the case when $f(x,\xi)$ is a composition of a nonsmooth function and a smooth function.

Proposition 3.1 *Let* $f(x,\xi) = \psi(g(x,\xi))$*, where* $\psi : \mathbb{R}^n \to \mathbb{R}$ *is a locally Lipschitz continuous function and* $g : \mathcal{X} \times \Xi \to \mathbb{R}^n$ *be a continuously differentiable function. Let* $\hat{\psi}(z,\epsilon)$ *be a smoothing of* $\psi(z)$ *and* $\hat{f}(x,\xi,\epsilon) := \hat{\psi}(g(x,\xi),\epsilon)$*. Then* $\hat{f}(x,\xi,\epsilon)$ *is a smoothing of* $f(x,\xi)$*. Moreover, if* $\hat{f}_i(x,\xi,\epsilon)$ *is a smoothing of* $f_i(x,\xi)$*,* $i = 1,\ldots,m$*, then* $\sum_{i=1}^m \hat{f}_i(x,\xi,\epsilon)$ *is a smoothing of* $\sum_{i=1}^m f_i(x,\xi)$.

We omit the proof as it is straightforward. In what follows, we use a simple example to illustrate how it can be applied to a nonsmooth random function.

Example 3.1 Consider a composite function $f(x,\xi) = p(g(x,\xi))$ where $p(z) = \max(0,z)$ is the max-function and $g(x,\xi)$ is a continuously differentiable function. Obviously the nonsmoothness is caused by the max-function. To smooth $f(x,\xi)$, it suffices to smooth the max-function. Here we consider two well known smoothing techniques for the max-function and demonstrate that they fit in the smoothing scheme specified by Definition 3.1.

First we consider a smoothing technique used by Alexander et al. [1]. Let $\epsilon \in \mathbb{R}_+$ and $\hat{p}_1(z,\epsilon)$ be such that for every $\epsilon > 0$,

$$\hat{p}_1(z,\epsilon) := \begin{cases} z, & z > \epsilon; \\ \frac{1}{4\epsilon}(z^2 + 2z\epsilon + \epsilon^2), & -\epsilon \le z \le \epsilon; \\ 0, & z < -\epsilon; \end{cases} \tag{3.4}$$

and for $\epsilon = 0$, $\hat{p}_1(z,0) := p(z)$. It is easy to check that $\lim_{\epsilon\to 0} \hat{p}_1(z,\epsilon) = p(z)$, which implies that $\hat{p}_1(z,\epsilon)$ is continuous in ϵ at $\epsilon = 0$ for every z. Note that the continuity of $\hat{p}_1$ in ϵ on $(0,\infty)$ is obvious.

In what follows, we examine the Lipschitz continuity. To do this, we investigate the Lipschitz continuity with respect to z and ϵ separately. First, let us fix z and discuss $|\hat{p}_1(z,\epsilon') - \hat{p}_1(z,\epsilon)|$ for $\epsilon',\epsilon \in [0,\infty)$. Observe that if $\epsilon',\epsilon > 0$, then $\hat{p}_1(z,\cdot)$ is piecewise smooth, that is,

$$d\hat{p}_1(z,\epsilon)/d\epsilon = \begin{cases} 0, & z > \epsilon; \\ \frac{1}{4}(1 - \frac{z^2}{\epsilon^2}), & -\epsilon \le z \le \epsilon; \\ 0, & z < -\epsilon. \end{cases} \tag{3.5}$$

This shows $d\hat{p}_1(z,\epsilon)/d\epsilon \le \frac{1}{4}$ for all z and $\epsilon > 0$. In the case when ϵ' is 0, we have that

$$\begin{aligned} |\hat{p}_1(z,\epsilon) - \hat{p}_1(z,0)|/\epsilon &= \begin{cases} 0, & z > \epsilon, \\ \frac{1}{4\epsilon^2}(z^2 - 2z\epsilon + \epsilon^2), & 0 \le z \le \epsilon, \\ \frac{1}{4\epsilon^2}(z^2 + 2z\epsilon + \epsilon^2), & -\epsilon \le z < 0, \\ 0, & z < -\epsilon, \end{cases} \\ &\le \frac{1}{4}. \end{aligned} \tag{3.6}$$

This shows that $\hat{p}_1(z,\cdot)$ is uniformly globally Lipschit continuous with respect to ϵ with module $\frac{1}{4}$. Similarly, we can show that $\hat{p}_1(\cdot,\epsilon)$ is uniformly globally Lipschitz continuous with module 1. This verifies part (a) and (b) of Definition 3.1.

The continuous differentiability of $\hat{p}_1(z,\epsilon)$ on $\mathbb{R} \times (0,\infty)$ is obvious. Thus (c) is satisfied and hence $\hat{p}_1(z,\epsilon)$ is a smoothing in the sense of Definition 3.1.

Note that $\hat{p}_1(z,\epsilon)$ is not necessarily differentiable at $(z,0)$. In the convergence analysis later on, we will consider the outer limit of $\nabla_z \hat{p}_1(z',\epsilon)$ as$(z',\epsilon) \to (z,0)$

for any $z \in \mathbb{R}$. For this purpose, let us calculate the partial derivative of $\hat{p}_1$ in z when $\epsilon > 0$:

$$\frac{d\hat{p}_1(z,\epsilon)}{dz} = \begin{cases} 1, & z > \epsilon, \\ \frac{1}{2\epsilon}(z+\epsilon), & -\epsilon \le z \le \epsilon, \\ 0, & z < -\epsilon. \end{cases} \tag{3.7}$$

It is easy to obtain that

$$\overline{\lim_{(z',\epsilon)\to(z,0)}} \frac{d\hat{p}_1(z,\epsilon)}{dz} = [0,1] = \partial_z p(z), \tag{3.8}$$

which means p_1 satisfies the gradient consistency [24] at z.

Peng [22] proposes another way to smooth a general max function $\max(z_1,\ldots,z_n)$. We apply the method to the max function in this example. For $\epsilon > 0$, let

$$\hat{p}_2(z,\epsilon) := \epsilon \ln(1 + e^{z/\epsilon}) \tag{3.9}$$

and $\hat{p}_2(z,0) := p(z)$. Then $\hat{p}_2(z,\epsilon)$ is continuous in $z \in \mathbb{R}$ and $\epsilon \in \mathbb{R}_+\backslash\{0\}$. Moreover,

$$\lim_{\epsilon\to 0} \hat{p}_2(z,\epsilon) = \lim_{\epsilon\to 0} z\ln(1+e^{\frac{z}{\epsilon}})^{\frac{\epsilon}{z}} = \begin{cases} z, & z \ge 0, \\ 0, & z < 0, \end{cases}$$

which coincides with $p(z)$ for any z. Furthermore, it follows from [22, Lemma 2.1] that $\hat{p}_2(x,\epsilon)$ is continuously differentiable for $\epsilon > 0$ and

$$\frac{d\hat{p}_2(z,\epsilon)}{dz} = \frac{d\epsilon \ln(1+e^{z/\epsilon})}{dz} = \frac{e^{z/\epsilon}}{1+e^{z/\epsilon}} \in (0,1). \tag{3.10}$$

This and the fact that $\hat{p}_2(z,0)$ is globally Lipschitz continuous with a modulus of 1 imply that $\hat{p}_2(z,\epsilon)$ is uniformly globally Lipschtz continuous in z with a modulus 1.

In what follows, we show the uniform local Lipschitzness of $\hat{p}_2(z,\cdot)$ near $\epsilon = 0$. By a simple calculation, we have

$$\frac{d\hat{p}_2(z,\epsilon)}{d\epsilon} = \ln\left(1+e^{-|z|/\epsilon}\right) + \frac{|z|}{\epsilon(1+e^{|z|/\epsilon})} \le \ln 2 + \frac{e^{|z|/\epsilon}}{1+e^{|z|/\epsilon}} \le 1 + \ln 2$$

for all $z \in \mathbb{R}$. Therefore $\hat{p}_2(z,\epsilon)$ is locally Lipschitz continuous at point $(z,0)$ for any $z \in \mathbb{R}$ and hence $\hat{p}_2$ is a smoothing in the sense of Definition 3.1.

Note that like $\hat{p}_1$, $\hat{p}_2$ is not necessarily differentiable at $(z,0)$. Peng proved in [22, Lemma 2.1] the following gradient consistency

$$\overline{\lim_{(z',\epsilon)\to(z,0)}} \frac{d\hat{p}_2(z,\epsilon)}{dz} \subset [0,1] = \partial_z p(z). \tag{3.11}$$

In what follows, we investigate the smooth approximation of (1.2)–(1.1) as the smoothing parameter ϵ is driven to zero.

Suppose that $\mathbb{E}[f(x,\xi)]$ is well defined and it is locally Lipschitz continuous. Then the first order necessary condition of (1.1) in terms of Clarke generalized gradient can be written as

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

where $\mathcal{N}_{\mathcal{X}}(x)$ denotes the normal cone (see e.g. [6]) of $\mathcal{X}$ at x, that is,

$$\mathcal{N}_{\mathcal{X}}(x) := [T_{\mathcal{X}}(x)]^{-} = \left\{\varsigma \in \mathbb{R}^m \mid \langle \varsigma, z\rangle \le 0 \text{ for all } z \in T_{\mathcal{X}}(x)\right\},$$

where $T_{\mathcal{X}}(x) := \overline{\lim}_{t\downarrow 0}(\mathcal{X} - x)/t$ and $x^T y$ denotes the scalar product of two vectors. It is well known that when $\mathcal{X}$ is convex, the normal cone reduces to

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

A point satisfying the above equation is called a *stationary point* of the true problem. We make a blanket assumption that the set of stationary points of the true problem is non-empty. This may be satisfied when $\mathcal{X}$ is compact or $\mathbb{E}[f(x,\xi)]$ tends to ∞ as $\|x\| \to \infty$. Suppose now that $\mathbb{E}[\partial_x f(x,\xi)]$ is well defined, then under some boundedness conditions on $\partial_x f(x,\xi)$, we have that $\partial\mathbb{E}[f(x,\xi)] \subset \mathbb{E}[\partial_x f(x,\xi)]$ and consequently we may consider a weaker first order necessary condition of (1.1)

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

A point satisfying the above equation is called a *weak stationary point* of the true problem. It is well known [15,38] that when f is Clarke regular [8, Definition 2.3.4] on $\mathcal{X}$, then $\mathbb{E}[\partial_x f(x,\xi)] = \partial\mathbb{E}[f(x,\xi)]$ and hence the set of weak stationary points coincides with the set of stationary points. In particular, when f is convex in x, then these points are optimal solutions of the true problem.

In correspondence to the weak first order necessary condition, we consider a weak first order necessary condition of the smoothed problem (1.2)

$$0 \in \mathbb{E}\left[\nabla_x \hat{f}(x,\xi,\epsilon)\right] + \mathcal{N}_{\mathcal{X}}(x). \tag{3.12}$$

A point satisfying the above equation is called a *weak stationary point* of the smoothed problem (1.2). By [33, Proposition 2, Chap. 2], if $\nabla_x \hat{f}(x,\xi,\epsilon)$ is integrably bounded w.p.1, then $\mathbb{E}[\nabla_x \hat{f}(x,\xi,\epsilon)] = \nabla_x \mathbb{E}[\hat{f}(x,\xi,\epsilon)]$ and hence (3.12) coincides with the usual first order necessary condition

$$0 \in \nabla\mathbb{E}\left[\hat{f}(x,\xi,\epsilon)\right] + \mathcal{N}_{\mathcal{X}}(x). \tag{3.13}$$

A point satisfying the above equation is called a *stationary point* of the smoothed problem. In this paper, we will need the integrable boundedness condition for $\nabla_x \hat{f}(x,\xi,\epsilon)$, hence the set of weak stationary points coincides with that of stationary points of the smoothed problem. Let S denote the set of weak stationary points of (1.1) and $S(\epsilon)$ denote the set of weak stationary points of (1.2).

Theorem 3.1 *Let $\hat{f}(x,\xi,\epsilon)$ be a smoothing of $f(x,\xi)$. Suppose that there exists an integrable function $\kappa(\xi)$ such that the Lipschitz module of $\hat{f}(x,\xi,\epsilon)$ with respect to x is bounded by $\kappa(\xi)$ and*

$$\overline{\lim_{x'\to x,\epsilon\to 0}} \left\{\nabla_x \hat{f}(x',\xi,\epsilon)\right\} \subset \partial_x f(x,\xi) \tag{3.14}$$

for almost every ξ. Suppose also that $S(\epsilon)$ is nonempty. Then

$$\overline{\lim_{\epsilon\to 0}}\, S(\epsilon) \subset S. \tag{3.15}$$

In particular if f is convex with respect to x on $\mathcal{X}$ and $\mathcal{X}$ is a convex set, then any point in set $\overline{\lim}_{\epsilon\to 0} S(\epsilon)$ is an optimal solution of the true problem (1.1).

Proof Let $x(\epsilon) \in S(\epsilon)$. Observe first that since the Lipschitz module of $\hat{f}(x,\xi,\epsilon)$ with respect to x is bounded by an integrable function $\kappa(\xi)$, $S(\epsilon)$ coincides with the set of stationary points of the smoothed problem. By taking a subsequence if necessary, we assume for the simplicity of notation $\lim_{\epsilon\to 0} x(\epsilon) = x$. Since $\nabla_x \hat{f}(x(\epsilon),\xi,\epsilon)$ is bounded by the Lipschitz modulus of $\hat{f}$ in x and hence $\kappa(\xi)$, by the Lebesgue dominated convergence theorem and (3.14),

$$\overline{\lim_{\epsilon\to 0}} \left\{\mathbb{E}[\nabla_x \hat{f}(x(\epsilon),\xi,\epsilon)]\right\} = \mathbb{E}\left[\overline{\lim_{\epsilon\to 0}} \left\{\nabla_x \hat{f}(x(\epsilon),\xi,\epsilon)\right\}\right] \subset \mathbb{E}\left[\partial_x f(x,\xi)\right]. \tag{3.16}$$

Here we need to explain how the Lebesgue dominated convergence theorem is applied, that is, equality (3.16). Let $\hbar \in \mathbb{E}\left[\overline{\lim}_{\epsilon\to 0} \left\{\nabla_x \hat{f}(x(\epsilon),\xi,\epsilon)\right\}\right]$. Then there exists $h \in \overline{\lim}_{\epsilon\to 0} \left\{\nabla_x \hat{f}(x(\epsilon),\xi,\epsilon)\right\}$ such that $\hbar = \mathbb{E}[h]$. By the definition of the outer limit, there exists a subsequence $\left\{\nabla_x \hat{f}(x(\epsilon_k),\xi,\epsilon_k)\right\}$ such that $\left\{\nabla_x \hat{f}(x(\epsilon_k),\xi,\epsilon_k)\right\} \to h$ w.p.1 as $k\to\infty$. Therefore we have

$$\hbar = \mathbb{E}\left[\lim_{k\to\infty} \nabla_x \hat{f}(x(\epsilon_k),\xi,\epsilon_k)\right].$$

Since $\left\{\nabla_x \hat{f}(x(\epsilon_k),\xi,\epsilon_k)\right\}$ is bounded by an integrable function, by the Lebesgue dominated convergence theorem, we can exchange the limit with the expectation operator in the equation above, that is,

$$\hbar = \lim_{k\to\infty} \mathbb{E}\left[\nabla_x \hat{f}(x(\epsilon_k),\xi,\epsilon_k)\right].$$

This shows $\hbar \in \overline{\lim}_{\epsilon\to 0}\left\{\mathbb{E}\left[\nabla_x \hat{f}(x(\epsilon),\xi,\epsilon)\right]\right\}$ and hence

$$\mathbb{E}\left[\overline{\lim_{\epsilon\to 0}}\left\{\nabla_x \hat{f}(x(\epsilon),\xi,\epsilon)\right\}\right] \subset \overline{\lim_{\epsilon\to 0}}\left\{\mathbb{E}\left[\nabla_x \hat{f}(x(\epsilon),\xi,\epsilon)\right]\right\}.$$

Conversely, let $\wp \in \overline{\lim}_{\epsilon\to 0}\left\{\mathbb{E}[\nabla_x \hat{f}(x(\epsilon),\xi,\epsilon)]\right\}$. Then, there exists a convergent subsequence $\left\{\mathbb{E}[\nabla_x \hat{f}(x(\epsilon_k),\xi,\epsilon_k)]\right\} \to \wp$ as $k \to \infty$. By taking a further subsequence if necessary, we may assume for the simplicity of notation that $\nabla_x \hat{f}(x(\epsilon_k),\xi,\epsilon_k) \to p$ w.p.1, as $k \to \infty$ (p depends on x and ξ). The integrable boundedness of the sequence allows us to apply the Lebesgue dominated convergence theorem, that is,

$$\begin{aligned}\wp &= \lim_{k\to\infty} \mathbb{E}\left[\nabla_x \hat{f}(x(\epsilon_k),\xi,\epsilon_k)\right] = \mathbb{E}\left[\lim_{k\to\infty} \nabla_x \hat{f}(x(\epsilon_k),\xi,\epsilon_k)\right]\\ &= \mathbb{E}[p] \in \mathbb{E}\left[\overline{\lim_{\epsilon\to 0}}\left\{\nabla_x \hat{f}(x(\epsilon),\xi,\epsilon)\right\}\right].\end{aligned}$$

This shows

$$\overline{\lim_{\epsilon\to 0}}\left\{\mathbb{E}\left[\nabla_x \hat{f}(x(\epsilon),\xi,\epsilon)\right]\right\} \subset \mathbb{E}\left[\overline{\lim_{\epsilon\to 0}}\left\{\nabla_x \hat{f}(x(\epsilon),\xi,\epsilon)\right\}\right],$$

hence the equality in (3.16) holds. The conclusion follows immediately from this and the fact that the normal cone is upper semi-continuous. □

4 Convergence of SAA stationary points

In this section, we study the convergence of the stationary points of the smoothed SAA problem (1.3). For this purpose, we consider the first order necessary condition of (1.3) which can be written as follows

$$0 \in \frac{1}{N}\sum_{i=1}^{N} \nabla_x \hat{f}\left(x,\xi^i,\epsilon\right) + \mathcal{N}_{\mathcal{X}}(x). \tag{4.17}$$

We assume that for almost every (a.e. for short) $\omega \in \Omega$ there exists $N(\omega) > 0$ such that for all $N > N(\omega)$ (4.17) has a solution. This is guaranteed when $\mathcal{X}$ is compact. In the case when $\mathcal{X}$ is unbounded, other conditions may be needed to ensure the existence of a solution of (4.17). See [25] and references therein. We will discuss this issue in Sect. 5 in the context of a CVaR problem.

In numerical implementation, there are two ways to set ϵ in (1.3): one is to fix ϵ, the other is to let ϵ vary as N increases, that is, let $\epsilon = \epsilon_N$ where $\epsilon_N \to 0$ as $N \to \infty$. In the following theorem, we establish the convergence results of stationary points in both cases. We need the following assumption.

Assumption 4.1 *There exists a small positive constant $\epsilon_0 > 0$ and a measurable function $\kappa(\xi)$ such that*

$$\sup_{x\in\mathcal{C},\epsilon\in[0,\epsilon_0]} \left\|\partial_x \hat{f}(x,\xi,\epsilon)\right\| \le \kappa(\xi),$$

for all $\xi \in \Xi$, where $\mathbb{E}[\kappa(\xi)] < \infty$ and $\mathcal{C}$ is a compact subset of $\mathcal{X}$.

We first consider the case when ϵ is fixed.

Theorem 4.1 *Let $\epsilon \neq 0$ be fixed and $\{x_N(\epsilon)\}$ be a sequence of stationary points which satisfies (4.17). Let $x^*(\epsilon)$ be an accumulation point of the sequence as N tends to infinity. If there exists a compact set $\mathcal{C}$ such that w.p.1 it contains a neighborhood of $x^*(\epsilon)$ and Assumption 4.1 holds, then w.p.1 $x^*(\epsilon) \in S(\epsilon)$.*

Proof For $\epsilon > 0$, $\nabla_x \hat{f}(\cdot,\xi,\epsilon)$ is continuous on $\mathcal{X}$ for every $\xi \in \Xi$. By Assumption 4.1, $\nabla_x \hat{f}(x,\xi,\epsilon)$ is bounded by $\kappa(\xi)$ for $x \in \mathcal{C}$. Since $\mathcal{C}$ is compact and ϵ is a constant, by applying [31, Lemma A1] to $\nabla_x \hat{f}(x,\xi,\epsilon)$ componentwise, we have

$$\lim_{N\to\infty} \max_{x\in\mathcal{C}} \left\| \frac{1}{N}\sum_{i=1}^{N} \nabla_x \hat{f}(x,\xi^i,\epsilon) - \mathbb{E}[\nabla_x \hat{f}(x,\xi,\epsilon)] \right\| = 0, \quad \text{w.p.1.} \tag{4.18}$$

By taking a subsequence if necessary, we assume for the simplicity of notation that $x_N(\epsilon) \to x^*(\epsilon)$ as $N \to \infty$. Since $x_N(\epsilon)$ is a stationary point, then

$$0 \in \frac{1}{N}\sum_{i=1}^{N} \nabla_x \hat{f}\left(x_N(\epsilon),\xi^i,\epsilon\right) + \mathcal{N}_{\mathcal{X}}\left(x_N(\epsilon)\right),$$

which can be rewritten as

$$\begin{aligned} 0 \in\ & \frac{1}{N}\sum_{i=1}^{N} \nabla_x \hat{f}\left(x^*(\epsilon),\xi^i,\epsilon\right) \\ & + \left[\frac{1}{N}\sum_{i=1}^{N} \nabla_x \hat{f}\left(x_N(\epsilon),\xi^i,\epsilon\right) - \mathbb{E}\left[\nabla_x \hat{f}\left(x_N(\epsilon),\xi,\epsilon\right)\right]\right] \\ & + \left[\mathbb{E}\left[\nabla_x \hat{f}\left(x_N(\epsilon),\xi,\epsilon\right)\right] - \frac{1}{N}\sum_{i=1}^{N} \nabla_x \hat{f}\left(x^*(\epsilon),\xi^i,\epsilon\right)\right] + \mathcal{N}_{\mathcal{X}}(x_N(\epsilon)). \end{aligned}$$

Let $N \to \infty$. By the strong law of large numbers, the first term on the right hand side of the equation above tends to $\mathbb{E}\left[\nabla_x \hat{f}\left(x^*(\epsilon),\xi,\epsilon\right)\right]$ w.p.1 and the third term tends to zero w.p.1. Now let us look at the second term. By the uniform convergence (4.18),

$$\lim_{N\to\infty}\left\|\frac{1}{N}\sum_{i=1}^{N}\nabla_x\hat{f}\left(x_N(\epsilon),\xi^i,\epsilon\right)-\mathbb{E}\left[\nabla_x\hat{f}\left(x_N(\epsilon),\xi,\epsilon\right)\right]\right\|$$
$$=\lim_{N\to\infty}\max_{x\in\mathcal{C}}\left\|\frac{1}{N}\sum_{i=1}^{N}\nabla_x\hat{f}\left(x,\xi^i,\epsilon\right)-\mathbb{E}\left[\nabla_x\hat{f}\left(x,\xi,\epsilon\right)\right]\right\|=0$$

w.p.1.The proof is complete. □

Under some moderate conditions, it is possible to obtain an error bound $d(x_N(\epsilon), S(\epsilon))$. To this end, we need a notion of metric regularity of set-valued mapping.

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

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

Here the inverse mapping Γ^{-1} is defined as $\Gamma^{-1}(y)=\{x\in\mathcal{X}:y\in\Gamma(x)\}$ and the minimal σ which makes the above inequality hold is called regularity modulus [9]. The metric regularity is equivalent to the surjectivity of coderivative of Γ at $\bar{x}$ for $\bar{y}$ or Aubin's property of F^{-1} at $\bar{y}$ and sufficed by the graphic convexity of F [9]. For a comprehensive discussion of the history and recent development of the notion, see [9], [29, Chap. 9] and references therein.

Theorem 4.2 *Let*

$$\Gamma(x):=\mathbb{E}\left[\nabla_x\hat{f}(x,\xi,\epsilon)\right]+\mathcal{N}_{\mathcal{X}}(x).$$

Let the conditions of Theorem 4.1 hold. If $\{x_N(\epsilon)\}$ converges to $x^*(\epsilon)$ w.p.1 and Γ is metrically regular at $x^*(\epsilon)$ for 0, then for a.e. $\omega\in\Omega$, there exists $N(\omega)>0$ such that for $N>N(\omega)$

$$d(x_N(\epsilon),S(\epsilon))\le\sigma\min\left(\left\|\mathbb{E}\left[\nabla_x\hat{f}(x_N(\epsilon),\xi,\epsilon)\right]\right\|,\left\|\mathbb{E}\left[\nabla_x\hat{f}(x_N(\epsilon),\xi,\epsilon)\right]-\nabla_x\hat{f}_N(x_N(\epsilon),\epsilon)\right\|\right),\tag{4.19}$$

where

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

and σ is the regularity modulus of Γ at $x^*(\epsilon)$ for 0.

Proof Let N be sufficiently large such that $x_N(\epsilon)$ is close to $x^*(\epsilon)$. Since Γ is metrically regular at $x^*(\epsilon)$ for 0, there exists a constant $\sigma > 0$ such that

$$d\,(x_N(\epsilon), S(\epsilon)) \leq \sigma d(0, \Gamma(x_N(\epsilon))). \tag{4.20}$$

Observe that

$$\begin{aligned} d\,(0, \Gamma(x_N(\epsilon))) &\leq d\left(0, \mathbb{E}\left[\nabla_x \hat{f}(x_N(\epsilon), \xi, \epsilon)\right] + 0\right) \\ &= \left\|\mathbb{E}\left[\nabla_x \hat{f}(x_N(\epsilon), \xi, \epsilon)\right]\right\|. \end{aligned} \tag{4.21}$$

Moreover, since

$$0 \in \nabla_x \hat{f}_N(x_N(\epsilon), \epsilon) + \mathcal{N}_{\mathcal{X}}(x_N(\epsilon)),$$

then

$$\begin{aligned} d(0, \Gamma(x_N(\epsilon))) &\leq \mathbb{D}\left(\nabla_x \hat{f}_N(x_N(\epsilon), \epsilon) + \mathcal{N}_{\mathcal{X}}(x_N(\epsilon)), \mathbb{E}\left[\nabla_x \hat{f}(x_N(\epsilon), \xi, \epsilon)\right]\right. \\ &\qquad \left. + \mathcal{N}_{\mathcal{X}}(x_N(\epsilon))\right) \\ &\leq \left\|\nabla_x \hat{f}_N(x_N(\epsilon), \epsilon) - \mathbb{E}\left[\nabla_x \hat{f}(x_N(\epsilon), \xi, \epsilon)\right]\right\|. \end{aligned} \tag{4.22}$$

Combining (4.20)–(4.22), we obtain (4.19). The proof is complete. □

Remark 4.1 By Mordukhovich criterion, the regularity condition assumed in Theorem 4.2 is equivalent to the nonsingularity of the coderivative of Γ at $x^*(\epsilon)$ for 0 and detailed characterization on this and the regularity modulus in deterministic case can be found in [9, Theorem 5.1]. In the case when $\mathbb{E}[\nabla_x \hat{f}(x, \xi, \epsilon)]$ is continuously differentiable, the regularity condition is satisfied if $\nabla\mathbb{E}[\nabla_x \hat{f}(x, \xi, \epsilon)]$ is nonsingular at $x^*(\epsilon)$, see [29, Exercise 9.44]. A stronger result may be obtained if $\mathbb{E}[\nabla_x \hat{f}(x, \xi, \epsilon)]$ is strongly monotone at $x^*(\epsilon)$. See [12, Corollary 5.1.8].

The error bound obtained in Theorem 4.2 may be used to discuss exponential convergence of $\{x_N(\epsilon)\}$. To see this, note that the right hand side of (4.19) is bounded by

$$\sigma \sup_{x \in B(x^*(\epsilon), \delta)} \min\left(\left\|\mathbb{E}\left[\nabla_x \hat{f}(x, \xi, \epsilon)\right]\right\|, \left\|\mathbb{E}\left[\nabla_x \hat{f}(x, \xi, \epsilon)\right] - \nabla_x \hat{f}_N(x, \epsilon)\right\|\right).$$

Under some appropriate conditions such as Hölder continuity of $\nabla_x \hat{f}(x, \xi, \epsilon)$ with respect to x and finite expected value of the moment function $\mathbb{E}\left[e^{\|\nabla_x \hat{f}(x,\xi,\epsilon) - \mathbb{E}\left[\nabla_x \hat{f}(x,\xi,\epsilon)\right]\| t}\right]$ for t close to zero, we can use [36, Theorem 5.1] to show that with probability approaching 1 exponentially fast with the increase of N, $\nabla_x \hat{f}_N(x, \epsilon)$ uniformly approximates $\mathbb{E}[\nabla_x \hat{f}(x, \xi, \epsilon)]$ in $B(x^*(\epsilon), \delta)$ and consequently we can show that with probability approaching 1 exponentially fast, $\{x_N(\epsilon)\}$ converges to $x^*(\epsilon)$. We omit the details of derivation since they are purely technical.

Note also that

$$d(x_N(\epsilon), S(\epsilon)) \le d(x_N(\epsilon), x^*(\epsilon))$$

and equality holds when F is strongly metric regular at $x^*(\epsilon)$ for 0 in which case $S(\epsilon)$ reduces to $\{x^*(\epsilon)\}$ in a small neighborhood of $x^*(\epsilon)$. See discussions in [9, Sect. 5] for details about strong metric regularity.

Combining Theorem 4.1 with Theorem 3.1, we may expect that $x_N(\epsilon)$ is a reasonable approximation of the weak stationary point of the true problem. We now consider the case when ϵ is driven to zero as N tends to infinity, that is, $\epsilon = \epsilon_N \to 0$ as $N \to \infty$. We need a uniform strong law of large numbers for random set-valued mappings which is established by Shapiro and Xu in [37, Theorem 2].

Lemma 4.1 *Let $V \subset \mathbb{R}^m$ be a compact set, and $\mathcal{A}(v, \xi) : V \times \Xi \to 2^{\mathbb{R}^m}$ be a measurable, compact set-valued mapping that is upper semicontinuous with respect to v on V for almost every ξ, and $\xi : \Omega \to \Xi \subset \mathbb{R}^k$ is a random vector defined on probability space $(\Omega, \mathcal{F}, P)$. Let $\xi^1, \ldots, \xi^N$ be an i.i.d. sample of ξ and*

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

Suppose that there exists $\sigma(\xi)$ such that

$$\|\mathcal{A}(v, \xi)\| := \sup_{A \in \mathcal{A}(v,\xi)} \|A\| \le \sigma(\xi), \tag{4.23}$$

for all $v \in V$ and $\xi \in \Xi$, where $\mathbb{E}[\sigma(\xi)] < \infty$. Then for any $\delta > 0$

$$\overline{\lim_{N\to\infty}}\, \mathcal{A}_N(v) \subset \mathbb{E}[conv\ \mathcal{A}_\delta(v, \xi)], \text{ w.p.1} \tag{4.24}$$

uniformly for $v \in V$, where $\mathcal{A}_\delta(v, \xi) := \bigcup_{w \in B(v,\delta)} \mathcal{A}(w, \xi)$.

Note that by the property of the Aumann's integral [13, Theorem 5.4 (d)], when the probability measure P is nonatomic, $\mathbb{E}[\text{conv}\ \mathcal{A}_\delta(v, \xi)] = \mathbb{E}[\mathcal{A}_\delta(v, \xi)]$.

Theorem 4.3 *Let $\mathcal{A}(v, \xi)$ and $\mathcal{A}_N(v)$ be defined as in Lemma 4.1. Assume that all conditions in the lemma are satisfied. Consider the following stochastic generalized equation*

$$0 \in \mathbb{E}[conv\ \mathcal{A}(v, \xi)] + \mathcal{N}_V(v) \tag{4.25}$$

and its sample average approximation

$$0 \in \mathcal{A}_N(v) + \mathcal{N}_V(v), \tag{4.26}$$

where $\mathcal{N}_V(v)$ denotes the normal cone of V at v, that is,

$$\mathcal{N}_V(v) := [T_V(v)]^- = \left\{\varsigma \in \mathbb{R}^m \mid \langle \varsigma, d\rangle \le 0 \text{ for all } d \in T_V(v)\right\},$$

where $T_V(v) := \lim_{t\downarrow 0}(V - v)/t$. *Suppose that both* (4.25) *and* (4.26) *have nonempty solution sets. Let* v^N *be a solution of* (4.26). *Then w.p.1, an accumulation point of* $\{v^N\}$ *is a solution of the true problem* (4.25).

Note that when $(\Omega, \mathcal{F}, P)$ is nonatomic, then (4.25) reduces to

$$0 \in \mathbb{E}[\mathcal{A}(v, \xi)] + \mathcal{N}_V(v).$$

Proof of Theorem 4.3 By Lemma 4.1, for any $\delta > 0$

$$\overline{\lim_{N\to\infty}}\, \mathcal{A}_N(v) \subset \mathbb{E}\left[\text{conv}\ \mathcal{A}_\delta(v, \xi)\right] \ \text{w.p.1} \tag{4.27}$$

uniformly with respect to $v \in V$. By taking a subsequence if necessary, we assume for the simplicity of notation that $v^N \to v^*$ w.p.1 as $N \to \infty$. We prove that w.p.1

$$\mathbb{D}\left(\mathcal{A}_N(v^N), \mathbb{E}\left[\text{conv}\ \mathcal{A}_{2\delta}(v^*)\right]\right) \to 0, \quad \text{as } N \to \infty.$$

Note that

$$\mathbb{D}\left(\mathcal{A}_N(v^N), \mathbb{E}\left[\text{conv}\ \mathcal{A}_{2\delta}(v^*)\right]\right) \leq \mathbb{D}\left(\frac{1}{N}\sum_{i=1}^N \mathcal{A}\left(v^N, \xi^i\right), \frac{1}{N}\sum_{i=1}^N \mathcal{A}_\delta(v^*, \xi^i)\right)$$
$$+\mathbb{D}\left(\frac{1}{N}\sum_{i=1}^N \mathcal{A}_\delta(x^*, \xi^i), \mathbb{E}\left[\text{conv}\ \mathcal{A}_{2\delta}(x^*, \xi)\right]\right).$$

It follows from (4.27) that the second term on the right hand side of the equation above tends to zero w.p.1 as $N \to \infty$. On the other hand, since $\mathcal{A}(v^N, \xi^i) \subset \mathcal{A}_\delta(v^*, \xi^i)$, for $v^N \in B(v^*, \delta)$, then there exists $N(\omega) > 0$ such that for $N > N(\omega)$, the first term on the right is zero. This shows that

$$0 \in \mathbb{E}\left[\text{conv}\ \mathcal{A}_{2\delta}(v^*, \xi)\right] + \mathcal{N}_V(v^*).$$

Driving δ to zero, we have by the Lebesgue dominated convergence theorem that

$$\lim_{\delta\to 0} \mathbb{E}\left[\text{conv}\ \mathcal{A}_{2\delta}(v^*, \xi)\right] = \mathbb{E}\left[\lim_{\delta\to 0} \text{conv}\ \mathcal{A}_{2\delta}(v^*, \xi)\right],$$

hence

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

The proof is complete. □

Theorem 4.3 may be regarded as extension of consistency analysis of generalized equations in [34, Sect. 7.1] to set valued mappings. In the case when $\mathcal{A}$ is single valued, the theorem is covered by [34, Proposition 19]. Here we use the result to analyze the convergence of stationary points of the smoothed SAA problem.

Recall that a set-valued mapping $\Gamma : \mathcal{X} \to 2^{\mathbb{R}^m}$ 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 $\bigcup_{x \in B(\bar{x})} \Gamma(x)$ is compact. The following result was established by Hogan [14].

Lemma 4.2 *Let $\Gamma : \mathcal{X} \to 2^{\mathbb{R}^m}$ be uniformly compact near $\bar{x}$. Then Γ is upper semi-continuous at $\bar{x}$ if and only if Γ is closed.*

Theorem 4.4 *Let $\{x(\epsilon_N)\}$ be a sequence of stationary points satisfying (4.17) with $\epsilon = \epsilon_N$ and $\epsilon_N \to 0$ as $N \to \infty$, let x^* be an accumulation point of $\{x(\epsilon_N)\}$. Let $\mathcal{C}$ be a compact subset of $\mathcal{X}$ which contains a neighborhood of x^* w.p.1 and Assumption 4.1 holds on $\mathcal{C}$. Then w.p.1 x^* satisfies*

$$0 \in \mathbb{E}\left[\partial_x \hat{f}\left(x^*, \xi, 0\right)\right] + \mathcal{N}_{\mathcal{X}}(x^*). \tag{4.28}$$

If, in addition, $\hat{f}$ satisfies the gradient consistency, that is,

$$\partial_x \hat{f}(x^*, \xi, 0) \subset \partial_x f(x^*, \xi),$$

then x^* is a weak stationary point of the true problem.

Proof We prove the conclusion by applying Theorem 4.3. Define the set-valued mapping as follows

$$\mathcal{A}(x, \xi, \epsilon) := \begin{cases} \nabla_x \hat{f}(x, \xi, \epsilon), & \epsilon \neq 0, \\ \text{conv}\left\{ \overline{\lim\limits_{(x',\epsilon) \to (x,0)}} \partial_x \hat{f}(x', \xi, \epsilon) \right\}, & \epsilon = 0, \end{cases}$$

where $\partial_x \hat{f}(x, \xi, \epsilon)$ is the Clarke generalized gradient of $\hat{f}$ with respect to x.

Note that $(x(\epsilon_N), \epsilon_N)$ satisfies

$$0 \in \frac{1}{N} \sum_{i=1}^{N} \mathcal{A}\left(x(\epsilon_N), \xi^i, \epsilon_N\right) + \mathcal{N}_{\mathcal{X}}(x(\epsilon_N)). \tag{4.29}$$

To apply Theorem 4.3, we need to verify that the set-valued mapping defined above satisfies the conditions of Theorem 4.3.

Observe first that $\mathcal{A}(\cdot, \xi, \cdot) : \mathcal{X} \times [-\epsilon_0, \epsilon_0] \to 2^{\mathbb{R}^m}$ is a random compact set-valued mapping. Let $\epsilon_0 > 0$ be fixed. In what follows, we investigate the upper semi-continuity of $\mathcal{A}(\cdot, \xi, \cdot)$ on $\mathcal{X} \times [-\epsilon_0, \epsilon_0] \to 2^{\mathbb{R}^m}$ with respect to variable (x, ϵ) for every $\xi \in \Xi$. Let $(x, \epsilon) \in \mathcal{X} \times [-\epsilon_0, \epsilon_0]$. If $\epsilon \neq 0$, then for (x', ϵ') close to (x, ϵ), $\mathcal{A}(x', \xi, \epsilon')$ coincides with $\nabla_x \hat{f}(x', \xi, \epsilon')$ and the latter is continuous at (x, ϵ). It therefore suffices to consider the case when $\epsilon = 0$. By the definition of $\mathcal{A}$, it is easy to observe that $\mathcal{A}(\cdot, \xi, \cdot)$ is closed at $(x, 0)$. Moreover, since $\hat{f}(x, \xi, \epsilon)$ is locally Lipschitz with respect to x by assumption, the mapping $\partial_x \hat{f}(x', \xi, \epsilon')$ is bounded for all (x', ϵ') close to $(x, 0)$. Therefore the closure of $\bigcup_{(x',\epsilon')\in B((x,0))} \mathcal{A}(x', \xi, \epsilon')$ is a compact set. By Lemma 4.2, $\mathcal{A}(\cdot, \xi, \cdot)$ is upper semi-continuous at $(x, 0)$. This shows that $\mathcal{A}(\cdot, \xi, \cdot)$ is upper semi-continuous in set $\mathcal{X} \times [-\epsilon_0, \epsilon_0]$. The conclusion follows from Theorem 4.3 and the fact that $\partial_x \hat{f}(x^*, \xi, 0)$ is convex compact set-valued. □

Our convergence analysis focuses on the stationary points. In some circumstances, however, we may obtain an optimal solution $x(\epsilon_N)$ of the smoothed SAA problem (1.3). It is then natural to ask whether an accumulation points of the sequence $\{x(\epsilon_N)\}$ is an optimal solution of the true problem (1.1). In what follows, we address this question.

By definition, $\hat{f}(x, \xi, \epsilon)$ is locally Lipschitz continuous with respect to (x, ϵ), and hence globally Lipschitz on $\mathcal{C} \times [0, \epsilon_0]$ for any compact subset $\mathcal{C}$ of $\mathcal{X}$ and positive number $\epsilon_0 > 0$. Therefore, there exists $\tilde{\kappa}(\xi) > 0$ such that

$$\left\| \hat{f}(x', \xi, \epsilon') - \hat{f}\left(x'', \xi, \epsilon''\right) \right\| \leq \tilde{\kappa}(\xi) \left(\|x' - x''\| + |\epsilon' - \epsilon''| \right), \tag{4.30}$$

for all $x', x'' \in \mathcal{C}$ and $\epsilon', \epsilon'' \in [0, \epsilon_0]$.

Theorem 4.5 *Let x^* denote an optimal solution of (1.1) and $x(\epsilon_N)$ an optimal solution of (1.3). Suppose that the following conditions hold:*

(a) *w.p.1. the sequence $\{x(\epsilon_N)\}$ is located in a compact subset $\mathcal{C}$ of $\mathcal{X}$ such that (4.30) holds;*
(b) $\mathbb{E}[\tilde{\kappa}(\xi)] < \infty$;
(c) *the moment generating function $\mathbb{E}\left[e^{\tilde{\kappa}(\xi)t}\right]$ of the random variable $\tilde{\kappa}(\xi)$ is finite valued for t close to 0;*
(d) *for every $x \in \mathcal{C}$ and $\epsilon \in [0, \epsilon_0]$, the moment generating function*

$$\mathbb{E}\left[e^{\left(\hat{f}(x,\xi,\epsilon) - \mathbb{E}\left[\hat{f}(x,\xi,\epsilon)\right]\right)t}\right]$$

of the random variable $\hat{f}(x, \xi, \epsilon) - \mathbb{E}\left[\hat{f}(x, \xi, \epsilon)\right]$ is finite valued for t close to 0.

Then with probability approaching one exponentially fast with the increase of sample size N, $x(\epsilon_N)$ becomes an approximate optimal solution of the true problem (1.1).

Proof We use the uniform strong law of large numbers [36, Theorem 5.1] to prove the result. Under the conditions of (b)–(d), it follows from [36, Theorem 5.1] that for any $\delta > 0$, there exist positive constant $C = C(\delta)$ and $\beta(\delta)$, independent of N, such that

$$\text{Prob}\left\{ \sup_{x\in\mathcal{C},\epsilon\in[0,\epsilon_0]} \left| \hat{f}_N(x, \epsilon) - \mathbb{E}\left[\hat{f}(x, \xi, \epsilon)\right] \right| > \delta \right\} \leq C(\delta) e^{-N\beta(\delta)}. \tag{4.31}$$

Since $\hat{f}(x, \xi, 0) = f(x, \xi)$, we have from (4.30) that

$$\left| \mathbb{E}\left[\hat{f}(x, \xi, \epsilon)\right] - \mathbb{E}[f(x, \xi)] \right| \leq \mathbb{E}\left[\tilde{\kappa}(\xi)\right] \epsilon,$$

for all $x \in \mathcal{C}$ and $\epsilon \in [0, \epsilon_0]$. Let $\sigma := \mathbb{E}[\tilde{\kappa}(\xi)]$. Then

$$\begin{aligned}\left|\hat{f}_N(x,\epsilon) - \mathbb{E}[f(x,\xi)]\right| &\le \left|\hat{f}_N(x,\epsilon) - \mathbb{E}[\hat{f}(x,\xi,\epsilon)]\right| \\ &\quad + \left|\mathbb{E}\left[\hat{f}(x,\xi,\epsilon)\right] - \mathbb{E}[f(x,\xi)]\right| \\ &\le \left|\hat{f}_N(x,\epsilon) - \mathbb{E}\left[\hat{f}(x,\xi,\epsilon)\right]\right| + \sigma\epsilon.\end{aligned}$$

Since $\epsilon_N \to 0$ as $N \to \infty$, then there exists $N(\omega) > 0$ such that for $N > N(\omega)$ and $\epsilon_N \le \delta/\sigma$, we have from (4.31) and the inequality above that

$$\begin{aligned}&\mathrm{Prob}\left\{\sup_{x\in\mathcal{C}}\left|\hat{f}_N(x,\epsilon_N) - \mathbb{E}[f(x,\xi)]\right| > 2\delta\right\} \\ &\quad\le \mathrm{Prob}\left\{\sup_{x\in\mathcal{C},\epsilon\in[0,\epsilon_0]}\left|\hat{f}_N(x,\epsilon) - \mathbb{E}\left[\hat{f}(x,\xi,\epsilon)\right]\right| > \delta\right\} \\ &\quad\le C(\delta)e^{-N\beta(\delta)}.\end{aligned} \tag{4.32}$$

Observe that

$$\left|\hat{f}_N(x(\epsilon_N),\epsilon_N) - \mathbb{E}[f(x^*,\xi)]\right| \le \sup_{x\in\mathcal{C},\epsilon\in[0,\epsilon_0]}\left|\hat{f}_N(x,\epsilon) - \mathbb{E}\left[\hat{f}(x,\xi,\epsilon)\right]\right|. \tag{4.33}$$

To see this, we note that $\mathbb{E}[f(x(\epsilon_N),\xi)] \ge \mathbb{E}[f(x^*,\xi)]$ (since x^* is optimal solution) and hence

$$\begin{aligned}\hat{f}_N(x(\epsilon_N),\epsilon_N) - \mathbb{E}[f(x^*,\xi)] &= \hat{f}_N(x(\epsilon_N),\epsilon_N) - \mathbb{E}[f(x(\epsilon_N),\xi)] \\ &\quad + \mathbb{E}\left[f(x(\epsilon_N),\xi)\right] - \mathbb{E}[f(x^*,\xi)] \\ &\ge -\sup_{x\in\mathcal{C},\epsilon\in[0,\epsilon_0]}|\hat{f}_N(x,\epsilon_N) - \mathbb{E}[f(x,\xi)]|.\end{aligned}$$

Likewise, we can show that

$$\mathbb{E}[f(x^*,\xi)] - \hat{f}_N(x(\epsilon_N),\epsilon_N) \ge -\sup_{x\in\mathcal{C},\epsilon\in[0,\epsilon_0]}\left|\hat{f}_N(x,\epsilon_N) - \mathbb{E}[f(x,\xi)]\right|.$$

Combining the two inequalities above, we obtain (4.33). Using (4.33) and (4.32), we have

$$\mathrm{Prob}\left\{\left|\hat{f}_N(x(\epsilon_N),\epsilon_N) - \mathbb{E}[f(x^*,\xi)]\right| > 2\delta\right\} \le C(\delta)e^{-N\beta(\delta)}. \tag{4.34}$$

Note that

$$\left|\mathbb{E}\left[f(x(\epsilon_N),\xi)\right]-\mathbb{E}[f(x^*,\xi)]\right| \leq \left|\mathbb{E}\left[f\left(x(\epsilon_N),\xi\right)\right]-\mathbb{E}\left[\hat{f}\left(x(\epsilon_N),\xi,\epsilon_N\right)\right]\right| + \left|\mathbb{E}\left[\hat{f}\left(x(\epsilon_N),\xi,\epsilon_N\right)\right]-\hat{f}_N\left(x(\epsilon_N),\epsilon_N\right)\right| + \left|\hat{f}_N\left(x(\epsilon_N),\epsilon_N\right)-\mathbb{E}[f(x^*,\xi)]\right|, \qquad (4.35)$$

the first term on the right side of the above inequality is bounded by $\sigma\epsilon_N$. Combining (4.31) and (4.35) and letting $\epsilon_N \leq \delta/\sigma$, we have

$$\begin{aligned}
&\text{Prob}\left\{\left|\mathbb{E}\left[f(x(\epsilon_N),\xi)\right]-\mathbb{E}[f(x^*,\xi)]\right| > 5\delta\right\}\\
&\quad\leq \text{Prob}\left\{\left|\mathbb{E}\left[f(x(\epsilon_N),\xi)\right]-\hat{f}_N(x(\epsilon_N),\epsilon_N)\right| \geq 2\delta\right\}\\
&\qquad+\text{Prob}\left\{\left|\hat{f}_N\left(x(\epsilon_N),\epsilon_N\right)-\mathbb{E}[f(x^*,\xi)]\right| \geq 2\delta\right\}\\
&\quad\leq 2C(\delta)e^{-N\beta(\delta)}.
\end{aligned}$$

This implies that with probability at least $1-2C(\delta)e^{-N\beta(\delta)}$ a global minimizer of (1.3) becomes a 5δ-global minimizer of (1.1) hence the conclusion. □

The theorem above shows that under some mild conditions, the sequence of optimal solutions of the smoothed SAA problem converges to an optimal solution of the true problem at an exponential rate. Note that condition (a) means that the optimal solution sequence of the SAA problem is bounded hence it can be contained in a compact set. Condition (b) requires that the Lipschitz module of the smoothed function be integrable. Conditions (c) and (d) mean that the probability distribution of the random variables $\tilde{\kappa}(\xi)$ and $\hat{f}(x,\xi,\epsilon)$ die exponentially fast in the tails. In particular, they hold if ξ has a distribution supported on a bounded subset of $\mathbb{R}^k$. This kind of conditions have been used in the literature. See for instance [36, Sect. 5.1].

5 Applications

In this section, we apply the smoothing method and convergence result discussed in the previous sections to a CVaR problem and a supply chain problem.

5.1 Conditional value at risk

Value at risk (VaR) and conditional value at risk are two important risk measures in risk management. VaR is defined as a threshold value that the probability of a loss function exceeding the value is limited to a specified level. It is observed [28] that VaR has a number of disadvantages such as being unstable, difficult to work with numerically when losses are not normally distributed and providing no handle on the extent of losses that might be suffered beyond the threshold. CVaR, introduced as an improved risk measure [27], is defined as the expected losses under the condition that the probability of the loss function exceeding VaR value is limited to a specified level. CVaR overcomes the disadvantages of VaR and more remarkably it provides an optimization short-cuts. See [1,27,28].

Minimizing CVaR concerns finding an optimal solution of a nonsmooth stochastic minimization problem. The nonsmoothness in the problem is essentially caused by a max-function in the integrand of the objective function. Alexander et al. [1] propose a smoothing method for getting around the nonsmoothness in the minimization of risk for derivative portfolios, and show with numerical experiments that the smoothing method is more efficient than a standard linear programming approach [27] when either the number of instruments or the number of Monte-Carlo samples dramatically increase. In this section, we present a theoretical convergence analysis of the smoothing SAA method for CVaR using the established results in Sect. 4.

Let $g(x, \xi(\omega))$ denote the loss function associated with the decision vector $x \in \mathcal{X} \subset \mathbb{R}^m$ where $\mathcal{X}$ is a convex set and $\xi : \Omega \to \mathbb{R}^d$ is random vector with a probability density function $\rho(\xi)$. Here x is interpreted as a portfolio and $\mathcal{X}$ as a set of available portfolios. ξ represents market uncertainties that can affect the loss.

For each portfolio x, let $\Gamma(x, \alpha)$ denote the probability of the loss function $g(x, \xi)$ not exceeding a particular value $\alpha \in \mathbb{R}$, that is,

$$\Gamma(x, \alpha) = \text{Prob}\{g(x, \xi) \le \alpha\} = \int_{g(x,\xi)\le\alpha} \rho(\xi)d\xi. \tag{5.36}$$

Obviously $\Gamma(x, \alpha)$ is the cumulative distribution function of random variable $g(x, \xi)$.

Let $\beta \in (0, 1)$ be a confidence level. The value-at-risk is defined as

$$\alpha_\beta(x) = \inf\{\alpha \in \mathbb{R} : \Gamma(x, \alpha) \ge \beta\} \tag{5.37}$$

and the conditional value-at-risk is defined as

$$\phi_\beta(x) = \inf_{\alpha\in\mathbb{R}} \left(\alpha + \frac{1}{1-\beta}\mathbb{E}[[g(x, \xi) - \alpha]_+]\right). \tag{5.38}$$

Note that $[z]_+$ is essentially a max-function. To be consistent with the notation in Example 3.1, we will use $p(z)$ instead of $[z]_+$ later on. Note also that if the random variable $g(x, \xi)$ has continuous distribution, then CVaR is the conditional expected value of the loss under the condition $g(x, \xi) \ge \alpha_\beta(x)$, that is,

$$\phi_\beta(x) = \frac{1}{1-\beta} \int_{g(x,\xi)\ge\alpha_\beta(x)} g(x, \xi)\rho(\xi)d\xi. \tag{5.39}$$

Here it is implicitly assumed that the integral is well defined, that is, finite valued for all $x \in \mathcal{X}$. Taking α as a variable, Rockafellar and Uryasev [27,28] consider an augmented function

$$F_\beta(x,\alpha) = \alpha + \frac{1}{1-\beta}\int\limits_{g(x,\xi)\geq\alpha} g(x,\xi)\rho(\xi)d\xi = \alpha + \frac{1}{1-\beta}\mathbb{E}[p(g(x,\xi)-\alpha)]. \tag{5.40}$$

and show that if $\text{Prob}(g(x,\xi) = \alpha) = 0$, then $\phi_\beta(x) = \min_\alpha F_\beta(x,\alpha)$. See [28, Theorem 10]. Consequently, they demonstrate that minimizing CVaR with respect to x is equivalent to minimizing $F_\beta(x,\alpha)$ with respect to (x,α), that is,

$$\min_{x\in\mathcal{X}} \phi_\beta(x) = \min_{(x,\alpha)\in\mathcal{X}\times\mathbb{R}} F_\beta(x,\alpha). \tag{5.41}$$

As observed in [27,28], if $g(x,\xi)$ is convex in x, then $F_\beta(x,\alpha)$ is convex and continuous in (x,α).

In the following discussion, we do not assume the convexity of $g(x,\xi)$ in x, instead, we assume that $\mathcal{X}$ is a convex compact set. Note that the model is slightly different from Alexander et al. [1] where transaction and management costs are considered since our main focus is on the smoothing method.

Let $\hat{p}_1(z,\epsilon)$ be the smoothing of $p(z)$ as defined in Example 3.1. Alexander, Coleman and Li's smoothing problem can be written as

$$\begin{array}{ll} \min & \hat{F}_\beta(x,\alpha,\epsilon) = \mathbb{E}\left[G_\beta(x,\alpha,\xi,\epsilon)\right] \\ \text{s.t.} & (x,\alpha)\in\mathcal{X}\times\mathbb{R}, \end{array} \tag{5.42}$$

where $\epsilon > 0$ and

$$G_\beta(x,\alpha,\xi,\epsilon) = \alpha + \frac{1}{1-\beta}\hat{p}_1(g(x,\xi)-\alpha,\epsilon).$$

Since $g(x,\xi)-\alpha$ is continuously differentiable on $\mathbb{R}^m\times\mathbb{R}^d\times\mathbb{R}\backslash\{0\}$ and $\hat{p}_1(x,\epsilon)$ is a smoothing, then $\hat{F}_\beta(x,\alpha,\epsilon)$ is a smoothing of $F_\beta(x,\alpha)$ in the sense of Definition 3.1. We first present a result on the compactness of the solution set of (5.42).

Proposition 5.1 *Let $\mathcal{X}$ be a compact convex set. Suppose that there exists $x\in\mathcal{X}$ and $\alpha\in\mathbb{R}$ such that: (a) $\mathbb{E}[|g(x,\xi)|] < \infty$, (b) $\text{Prob}(g(x,\xi)=\alpha) = 0$. Then the optimal solution set of (5.42) is a nonempty compact set.*

Proof Since $\hat{F}_\beta(x,\alpha,\epsilon)$ is continuous in x and $\mathcal{X}$ is compact, then the projection of optimal solution set on x-axis is closed and bounded hence compact. In what follows, we consider the minimizers with respect to α.

When $\epsilon = 0$ and x is fixed, the compactness of the set of minimizers of $F_\beta(x,\alpha)$ in α is established in [27, Theorem 1]. By (3.6)

$$\begin{aligned} &\left|G_\beta(x,\alpha,\xi,\epsilon) - G_\beta(x,\alpha,\xi,0)\right| \\ &\quad\leq \frac{1}{1-\beta}\mathbb{E}\left[\left|\hat{p}_1\left(g(x,\xi)-\alpha,\epsilon\right) - p(g(x,\xi)-\alpha,0)\right|\right] \leq \frac{\epsilon}{4(1-\beta)}. \end{aligned}$$

Since

$$\mathbb{E}[|G_\beta(x,\alpha,\xi,0)|] \le \mathbb{E}\left[|\alpha| + \frac{1}{1-\beta}|g(x,\xi)|\right] < \infty,$$

and

$$\left|G_\beta(x,\alpha,\xi,\epsilon)\right| \le \left|G_\beta(x,\alpha,\xi,0)\right| + \frac{\epsilon}{4(1-\beta)},$$

then $\mathbb{E}\left[\left|G_\beta\,(x,\alpha,\xi,\epsilon)\right|\right] < \infty$ and $\hat{F}_\beta(x,\alpha,\epsilon)$ is well defined. Moreover

$$\begin{aligned}\left|\hat{F}_\beta(x,\alpha,\epsilon) - F_\beta(x,\alpha)\right| &\le \frac{1}{1-\beta}\mathbb{E}[|\hat{p}_1(g(x,\xi)-\alpha,\epsilon) - p(g(x,\xi)-\alpha,0)|]\\ &\le \frac{1}{4(1-\beta)}\epsilon.\end{aligned}$$

Therefore

$$\lim_{\epsilon\to 0}\hat{F}_\beta(x,\alpha,\epsilon) = F_\beta(x,\alpha)$$

uniformly with respect to x and α. By the Berge's stability theorem, the set of global minimizers of $\hat{F}_\beta(x,\alpha,\epsilon)$ with respect to α is contained in the set of δ-global minimizers of $F_\beta(x,\alpha)$, hence it is compact when ϵ is sufficiently small. □

Next we discuss the SAA method for solving (5.42). Let $\xi^1,\ldots,\xi^N$ be an i.i.d sample of ξ. The sample average approximation of (5.42) can be written as

$$\begin{array}{ll}\min & \frac{1}{N}\sum_{i=1}^N G_\beta(x,\alpha,\xi^i,\epsilon)\\ \text{s.t.} & (x,\alpha)\in\mathcal{X}\times\mathbb{R}.\end{array} \tag{5.43}$$

We first investigate the existence of optimal solutions of the SAA problem (5.43).

Proposition 5.2 *Let $\mathcal{X}$ be a compact convex set. Suppose that*: (a) *there exists $x\in\mathcal{X}$ and $\alpha\in\mathbb{R}$ such that* $\mathbb{E}[|g(x,\xi)|] < \infty$; (b) $\text{Prob}(g(x,\xi)=\alpha)=0$; (c) $g(x,\xi)$ *is Lipschitz continuous in x and there exists an integrable function $c(\xi)$ such that*

$$\left|g(x',\xi) - g(x'',\xi)\right| \le c(\xi)\left\|x'-x''\right\|,\quad \forall x',x''\in\mathcal{X}.$$

Then there exists $\alpha_0 > 0$ such that for ϵ sufficiently small and N sufficiently large, w.p.1 (5.43) *has a global minimizer in $\mathcal{X}\times[-\alpha_0,\alpha_0]$.*

Proof By Proposition 5.1, $G_\beta(x,\alpha,\xi,\epsilon)$ is bounded by an integrable function. By the strong law of large numbers,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N G_\beta(x,\alpha,\xi^i,\epsilon) = \mathbb{E}\left[G_\beta(x,\alpha,\xi,\epsilon)\right] = \hat{F}_\beta(x,\alpha,\epsilon).$$

Let $\alpha_0 > 0$ be sufficiently large and $\epsilon > 0$ be sufficiently small such that the optimal solution set of the smoothed problem (5.42) is contained in $\mathcal{X} \times [-\alpha_0, \alpha_0]$. For $\nu > 0$, let $(\bar{x}_1, \bar{\alpha}_1), \ldots, (\bar{x}_M, \bar{\alpha}_M) \in \mathcal{X} \times [-\alpha_0, \alpha_0]$ be such that for any $(x, \alpha) \in \mathcal{X} \times [-\alpha_0, \alpha_0]$, there exists $(\bar{x}_j, \bar{\alpha}_j)$ such that $\|(x, \alpha) - (\bar{x}_1, \bar{\alpha}_1)\| \le \nu$, i.e., $\{(\bar{x}_1, \bar{\alpha}_1), \ldots, (\bar{x}_M, \bar{\alpha}_M)\}$ is a ν-net of $\mathcal{X} \times [-\alpha_0, \alpha_0]$. By assumption and property of $\hat{p}_1$,

$$\left|G_\beta(x, \alpha, \xi, \epsilon) - G_\beta(\bar{x}_j, \bar{\alpha}_j, \xi, \epsilon)\right| \le \left|\alpha - \bar{\alpha}_j\right| + \frac{1}{1-\beta}\left[c(\xi)\left\|x - \bar{x}_j\right\| + \left|\alpha - \bar{\alpha}_j\right|\right]$$

and

$$\left|\hat{F}_\beta(x, \alpha, \epsilon) - \hat{F}_\beta(\bar{x}_j, \bar{\alpha}_j, \epsilon)\right| \le \left|\alpha - \bar{\alpha}_j\right| + \frac{1}{1-\beta}\left[\mathbb{E}[c(\xi)]\left\|x - \bar{x}_j\right\| + \left|\alpha - \bar{\alpha}_j\right|\right].$$

Therefore

$$\begin{aligned}
&\left|\frac{1}{N}\sum_{i=1}^N G_\beta(x, \alpha, \xi^i, \epsilon) - \hat{F}_\beta(x, \alpha, \epsilon)\right| \\
&\quad \le \frac{1}{N}\sum_{i=1}^N \left|G_\beta(x, \alpha, \xi^i, \epsilon) - G_\beta\left(\bar{x}_j, \bar{\alpha}_j, \xi^i, \epsilon\right)\right| \\
&\qquad + \left|\frac{1}{N}\sum_{i=1}^N G_\beta\left(\bar{x}_j, \bar{\alpha}_j, \xi^i, \epsilon_j\right) - \mathbb{E}\left[G_\beta\left(\bar{x}_j, \bar{\alpha}_j, \xi, \epsilon\right)\right]\right| \\
&\qquad + \left|\mathbb{E}\left[G_\beta\left(\bar{x}_j, \bar{\alpha}_j, \xi, \epsilon\right)\right] - \mathbb{E}\left[G_\beta(x, \alpha, \xi, \epsilon)\right]\right|.
\end{aligned}$$

The first term in the right hand side of the above equation is bounded by

$$\left|\alpha - \bar{\alpha}_j\right| + \frac{1}{1-\beta}\left(\frac{1}{N}\sum_{i=1}^N c(\xi^i)\right)\left\|x - \bar{x}_j\right\| + \frac{1}{1-\beta}\left|\alpha - \bar{\alpha}_j\right|;$$

the second term tends to zero as $N \to \infty$ by the strong law of large numbers; the third term is bounded by

$$\left|\alpha - \bar{\alpha}_j\right| + \frac{1}{1-\beta}\left[\mathbb{E}[c(\xi)]\left\|x - \bar{x}_j\right\| + \left|\alpha - \bar{\alpha}_j\right|\right].$$

This shows that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N G_\beta(x, \alpha, \xi^i, \epsilon) = \hat{F}_\beta(x, \alpha, \epsilon)$$

uniformly with respect to $(x, \alpha) \in \mathcal{X} \times [-\alpha_0, \alpha_0]$. The rest is straightforward by the Berge's stability theorem. □

We are now ready to state a convergence result for the stationary points of the smoothed SAA problem (5.43).

Proposition 5.3 *Let $\mathcal{W}(\epsilon_N)$ denote the set of stationary points $(x(\epsilon_N), \alpha(\epsilon_N))$ of (5.43) with $\epsilon = \epsilon_N$ and $\epsilon_N \to 0$ as $N \to \infty$. Let $\mathcal{W}$ denote the set of stationary points of the true problem (5.41). Let conditions in Proposition 5.2 hold. If $\overline{\lim}_{N\to\infty} \mathcal{W}(\epsilon_N)$ is bounded, then w.p.1*

$$\overline{\lim_{N\to\infty}} \mathcal{W}(\epsilon_N) \subset \mathcal{W}. \tag{5.44}$$

Proof Let α_0 be sufficiently large such that $\overline{\lim}_{N\to\infty} \mathcal{W}(\epsilon_N) \subset \mathcal{X} \times [-\alpha_0, \alpha_0]$. We use Theorem 4.4 to prove the result. It therefore suffices to verify the conditions of the theorem. The main condition we need to verify is Assumption 4.1. Let $\mathcal{C} = \mathcal{X} \times [-\alpha_0, \alpha_0]$. We calculate $\nabla_x G_\beta(x, \alpha, \xi, \epsilon)$. It is easy to derive that

$$\nabla_x G_\beta(x, \alpha, \xi, \epsilon) = \frac{1}{1-\beta} \times \frac{d\hat{p}_1(g(x, \xi) - \alpha, \epsilon)}{dz} \times \nabla_x g(x, \xi)$$

and

$$\frac{\partial G_\beta(x, \alpha, \xi, \epsilon)}{\partial \alpha} = 1 - \frac{1}{1-\beta} \times \frac{d\hat{p}_1(g(x, \xi) - \alpha, \epsilon)}{dz}.$$

Following the discussion in Example 3.1, we know that

$$\left| \frac{d\hat{p}_1\left(g(x, \xi) - \alpha, \epsilon\right)}{dz} \right| \leq 1.$$

Consequently

$$\left\| \nabla_{(x,\alpha)} G_\beta(x, \alpha, \xi, \epsilon) \right\| \leq \frac{1}{1-\beta} \|\nabla_x g(x, \xi)\| + 1 + \frac{1}{1-\beta} \leq 1 + \frac{1}{1-\beta}(\kappa(\xi) + 1).$$

Therefore Assumption 4.1 is satisfied on the whole region $\mathcal{X} \times \mathbb{R}$.

Finally, by the gradient consistency of $\hat{p}_1$ (3.8), we obtain the gradient consistency of $\nabla_x G_\beta(x, \alpha, \xi, \epsilon)$ at $\epsilon = 0$, that is,

$$\begin{aligned} &\overline{\lim_{(x',\epsilon)\to(x,0)}} \nabla_{(x,\alpha)} G_\beta(x, \alpha, \xi, \epsilon) \subset \partial_{(x,\alpha)} G(x, \alpha, \xi) \\ &= \left\{ \left(\frac{1}{1-\beta} t, 1 + \frac{1}{1-\beta} t \right)^T : t \in [0, 1] \right\}. \end{aligned}$$

The conclusion follows from Theorem 4.4. □

5.2 An inventory control problem

Objective functions in many supply chain problems are often nonsmooth [7,41,42] and involve random factors. Here we consider an inventory control problem in supply chain where a supplier orders a perishable goods from external suppliers and then sells them to retailers. The sale takes place over a short time period.

Let $D(x, \gamma, \xi(\omega))$ denote retailer's demand function. Here $x \in \mathcal{X} \subset \mathbb{R}^m$ may be interpreted as a vector of decision variables such as the supplier's capital investment on service and advertisement which influence the market demand, and $\mathcal{X}$ as a set of feasible decisions. $\gamma \in [w, \bar{\gamma}]$ denotes the price set by the supplier with lower bound w and upper bound $\bar{\gamma}$ and $\xi(\omega) \in \mathbb{R}^d$ represents a vector of random variables due to various of uncertainties in the market. In practice w is interpreted as the (wholesale) unit purchase cost.

The supplier needs to make a decision on x, Q and γ before the realization of the market uncertainties. We assume that the supplier has a knowledge of the distribution of $\xi(\omega)$ either from a prediction or from past experience. Suppose that ξ is a realization of $\xi(\omega)$ and the market demand is $D(x, \gamma, \xi)$. The supplier's total sale to the retailers is $\min(Q, D(x, \gamma, \xi))$ at price γ hence the revenue is $\gamma \min(Q, D(x, \gamma, \xi))$. At the end of selling period, the leftover is $\max(Q - D(x, \gamma, \xi), 0)$. Assuming the leftovers can be sold at a unit price of s. Then the total salvage value is $s \max(Q - D(x, \gamma, \xi), 0)$. The total revenue is therefore

$$\gamma \min(Q, D(x, \gamma, \xi)) + s \max(Q - D(x, \gamma, \xi), 0).$$

The costs we consider in this model involve the usual purchase costs, holding cost and delivery cost as well as capital investment cost and cost for losing a chance to sell. With order size Q, the purchase cost is wQ (with w being unit purchase cost). For the simplicity of discussion, we combine the holding cost and delivery cost which totals πQ with π being the unit cost. The penalty cost for not meeting the demand is $h \max(D(x, \gamma, \xi) - Q, 0)$ where h is unit penalty cost. Finally total investment cost is $K^T x$ where $K \in \mathbb{R}^m$ is a vector with nonnegative components. Consequently the total costs to the supplier is

$$(w + \pi)Q + h \max(D(x, \gamma, \xi) - Q, 0) + K^T x.$$

Let

$$\begin{aligned} v(x, \gamma, Q, \xi) := &\ \gamma \min(Q, D(x, \gamma, \xi)) + s \max(Q - D(x, \gamma, \xi), 0) \\ &- (w + \pi)Q - h \max(D(x, \gamma, \xi) - Q, 0) - K^T x. \end{aligned}$$

$v(x, \gamma, Q, \xi)$ is the net profit by the supplier at scenario ξ. Since the supplier needs to make a decision before the realization of the market uncertainties, what the supplier can best do in selecting an optimal decision is to maximize the expected value of $v(x, \gamma, Q, \xi)$. For the convenience of notation, we let $f(x, \gamma, Q, \xi) := -v(x, \gamma, Q, \xi)$.

Then the supplier's optimal decision problem can be formulated as

$$\begin{aligned}&\min_{(x,\gamma,Q)} \mathbb{E}[f(x,\gamma,Q,\xi)]\\&\quad := \mathbb{E}[h\max(D(x,\gamma,\xi)-Q,0)+(w+\pi)Q\\&\qquad -\gamma\min(Q,D(x,\gamma,\xi))-s\max(Q-D(x,\gamma,\xi),0)+K^Tx]\\&\qquad \text{s.t.}(x,\gamma,Q)\in\mathcal{X}\times[w,\bar{\gamma}]\times[0,\bar{Q}].\end{aligned}$$

Using the max-function, we can rewrite problem as:

$$\begin{aligned}\min_{(x,\gamma,Q)} \mathbb{E}[f(x,\gamma,Q,\xi)] &= \mathbb{E}[(h+\gamma-s)p(D(x,\gamma,\xi)-Q)\\&\quad +(w+\pi-s)Q-(\gamma-s)D(x,\gamma,\xi)+K^Tx]\\ \text{s.t.} \qquad & (x,\gamma,Q)\in\mathcal{X}\times[w,\bar{\gamma}]\times[0,\bar{Q}],\end{aligned} \tag{5.45}$$

where $p(z):=\max(z,0)$. Assume that $D(x,\gamma,\xi)$ is a continuously differentiable on $\mathcal{X}\times[w,\bar{\gamma}]\times\mathbb{R}^k$. We use $\hat{p}_2(z,\epsilon)$ which is introduced in Example 3.1 to smooth $p(z)$. The smooth approximation of problem (5.45) can be written as

$$\begin{aligned}\min_{(x,\gamma,Q)} \mathbb{E}[\hat{f}(x,\gamma,Q,\xi,\epsilon)] &= \mathbb{E}[(h+\gamma-s)\hat{p}_2(D(x,\gamma,\xi)-Q,\epsilon)\\&\quad +(w+\pi-s)Q-(\gamma-s)D(x,\gamma,\xi)+K^Tx]\\ \text{s.t.} \qquad & (x,\gamma,Q)\in\mathcal{X}\times[w,\bar{\gamma}]\times[0,\bar{Q}],\end{aligned} \tag{5.46}$$

where $\epsilon>0$. Let $\xi^1,\ldots,\xi^N$ be an i.i.d sample of ξ. The sample average approximation of (5.46) can be written as

$$\begin{aligned}\min_{(x,\gamma,Q)} \mathbb{E}\left[\hat{f}^N(x,\gamma,Q,\xi,\epsilon)\right] &= \frac{1}{N}\sum_{i=1}^N \hat{f}(x,\gamma,Q,\xi^i,\epsilon)\\ \text{s.t.} \qquad & (x,\gamma,Q)\in\mathcal{X}\times[w,\bar{\gamma}]\times[0,\bar{Q}].\end{aligned} \tag{5.47}$$

Note that since the feasible solution set $\mathcal{X}\times[w,\bar{\gamma}]\times[0,\bar{Q}]$ is a compact set, the existence of stationary point of (5.47) is guaranteed.

Proposition 5.4 *Let $(x(\epsilon_N),\gamma(\epsilon_N),\alpha(\epsilon_N))$ be a stationary point of (5.47) with $\epsilon=\epsilon_N$ and $\epsilon_N\to 0$ as $N\to\infty$. Let $\mathcal{U}(\epsilon_N)$ denote the set of stationary points $\{(x(\epsilon_N),\gamma(\epsilon_N),\alpha(\epsilon_N))\}$ and $\mathcal{U}$ denote the set of stationary points of the true problem (5.45). If there exists an integrable function $\kappa_2(\xi)>0$ such that $\|\nabla_x D(x,\gamma,\xi)\|+|\partial D(x,\gamma,\xi)/\partial\gamma|\le\kappa_2(\xi)$, then w.p.1*

$$\overline{\lim_{N\to\infty}}\ \mathcal{U}(\epsilon_N)\subset\mathcal{U}. \tag{5.48}$$

Proof Similar to Proposition 5.3, we use Theorem 4.4 to prove the result. The main condition to be verified is Assumption 4.1. For this purpose, we calculate $\nabla_x\hat{f}(x,\gamma,$

$Q, \xi, \epsilon)$. It is easy to derive that

$$\nabla_x \hat{f}(x, \gamma, Q, \xi, \epsilon) = (h + \gamma - s) \times \frac{d\hat{p}_2(D(x, \gamma, \xi) - Q, \epsilon)}{dz} \\ \times \nabla_x D(x, \gamma, \xi) - (\gamma - s)\nabla_x D(x, \gamma, \xi) + K,$$

$$\partial \hat{f}(x, \gamma, Q, \xi, \epsilon)/\partial\gamma = \hat{p}_2(D(x, \gamma, \xi) - Q, \epsilon) + (h + \gamma - s) \\ \times d\hat{p}_2(D(x, \gamma, \xi) - Q, \epsilon)/dz \times \partial D(x, \gamma, \xi)/\partial\gamma \\ -(\gamma - s)\partial D(x, \gamma, \xi)/\partial\gamma - D(x, \gamma, \xi),$$

and

$$\partial \hat{f}(x, \gamma, Q, \xi, \epsilon)/\partial Q = (w + \pi - s) - (h + \gamma - s)\frac{d\hat{p}_2(D(x, \gamma, \xi) - Q, \epsilon)}{dz}.$$

Following the discussion in Example 3.1, we know that

$$\left|\frac{d\hat{p}_2(D(x, \gamma, \xi) - Q, \epsilon)}{dz}\right| \le 1 + \ln 2.$$

Consequently

$$\left\|\nabla_{(x,\gamma,Q)} \hat{f}(x, \gamma, Q, \xi, \epsilon)\right\| \\ \le \left[(1 + \ln 2)|h + \gamma - s| + |\gamma - s|\right]\kappa_2(\xi) + \|K\| + |D(x, \gamma, \xi)| \\ +\hat{p}_2(D(x, \gamma, \xi) - Q, \epsilon) + |w + \pi - s| + (1 + \ln 2)|h + \gamma - s|.$$

The right hand side of the equation above is integrable as the demand function is bounded by $\bar{D}$ and $\hat{p}_2$ by $\max(\epsilon \ln 2, \epsilon \ln 2 + D(x, \gamma, \xi) - Q)$.

Finally, by the gradient consistency of $\hat{p}_2$ (3.11), we obtain the gradient consistency $\nabla_x \hat{f}(x, \gamma, Q, \xi, \epsilon)$ at $\epsilon = 0$, that is,

$$\overline{\lim_{(x',\gamma',\epsilon)\to(x,\gamma,0)}} \nabla_{(x,\gamma,Q)} \hat{f}(x, \gamma, Q, \xi, \epsilon) \subset \partial_{(x,\gamma,Q)} f(x, \gamma, Q, \xi).$$

The conclusion follows from Theorem 4.4. □

We have undertaken some numerical tests on the supply chain problem. In what follows we report some preliminary results of the tests.

The tests are carried out in Matlab7.2 installed in a PC with Windows XP operating system. We use the Matlab built-in optimization solver *fmincon* to solve the SAA problem (5.47).

Example 5.1 Consider an inventory model (5.45) with unit penalty cost $h = 6$, unit holding and delivery cost $\pi = 6$, unit salvage value $s = 6$, $w = 40$, $\bar{\gamma} = 90$ and $\bar{Q} = 60$. The demand function is $D(\gamma, \xi) = \bar{\gamma} + \xi - \frac{1}{2}\gamma$, where ξ is assumed to follow a uniform distribution with support set [0, 8]. Note that for the simplicity of tests, we have set $x = 0$. The optimal decision problem becomes

$$\begin{aligned}
&\min \mathbb{E}\left[f(\gamma, Q, \xi)\right] \\
&\quad = \mathbb{E}\left[\gamma p \left(90 + \xi - \frac{1}{2}\gamma - Q\right) + 40Q - (\gamma - 6)\left(90 + \xi - \frac{1}{2}\gamma\right)\right] \\
&\text{s.t. } (\gamma, Q) \in [40, 90] \times [0, 60].
\end{aligned}$$

The objective function can be integrated explicitly, that is,

$$\begin{aligned}
\mathbb{E}\left[f(\gamma, Q, \xi)\right] = \frac{\gamma}{16}\left[p\left(98 - \frac{1}{2}\gamma - Q\right)\right]^2 - \frac{\gamma}{16}\left[p\left(90 - \frac{1}{2}\gamma - Q\right)\right]^2 \\
+ 40Q - (\gamma - 6)\left(94 - \frac{1}{2}\gamma\right).
\end{aligned}$$

The exact solution of the true problem is (90, 445/9) and the optimal value is −2067.1. Note that the negative value means that there is a profit for the supplier.

We use Peng's smoothing function $\hat{p}_2(z, \epsilon)$ as discussed in Example 3.1. The smoothed problem is

$$\begin{aligned}
\hat{p}_2\left(90 + \xi - \frac{1}{2}\gamma - Q, \epsilon\right) &= \epsilon \ln\left(1 + e^{\frac{90+\xi-\frac{1}{2}\gamma-Q}{\epsilon}}\right) \\
&= 90 + \xi - \frac{1}{2}\gamma - Q + \epsilon \ln\left(1 + e^{-\frac{90+\xi-\frac{1}{2}\gamma-Q}{\epsilon}}\right).
\end{aligned}$$

The test results are displayed in Table 1. Note that the objective function of the test problem is nonconvex. However, the Matlab optimization solver fmincon returns an optimal solution when an initial feasible solution is reasonably close to the optimal solution.

Table 1 Numerical test results of Example 5.1

ϵ	N	Optimal value	Q^*	γ^*
2	100	−2003.9	49.9390	88.9404
	500	−1991.0	49.9444	88.9738
	1, 000	−1988.2	49.9878	88.9880
0.2	100	−2088.3	49.7782	90.0000
	500	−2050.7	49.5973	89.0079
	1, 000	−2073.8	49.9252	89.8938
0.02	100	−2077.5	49.9339	88.9757
	500	−2051.8	49.6094	89.0078
	1, 000	−2073.4	49.8038	89.8674

The initial results show that the convergence is not very sensitive to change of the value of ϵ so long as ϵ is sufficiently small. Likewise there is no significant improvement when the sample size is changed from 100 to 500 or 1000, which means that the convergence is very fast against the increase of sample size and there is not much improvement when the size is sufficiently large. This is consistent with the observations obtained in the literature. See [19] and references.

Acknowledgments We would like to thank two anonymous referees and the associate editor for valuable comments.

References

1. Alexander, S., Coleman, T.F., Li, Y.: Minimizing CVaR and VaR for a portfolio of derivatives. J. Bank. Finance **30**, 583–605 (2006)
2. Artstein, Z., Vitale, R.A.: A strong law of large numbers for random compact sets. Ann. Probab. **3**, 879–882 (1975)
3. Artstein, Z., J-B Wets, R.: Consistency of minimizers and the SLLN for stochastic programs. J. Convex Anal **2**, 1–17 (1995)
4. Aumann, R.J.: Integrals of set-valued function. J. Math. Anal. Appl. **12**, 1–12 (1965)
5. Aubin, J.-P., Frankowska, H.: Set-Valued Analysis. Birkhäuser, Boston (1990)
6. Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems, Springer Series in Operations Research, Springer, Heidelberg (2000)
7. Chen, H., Chen, J., Chen, Y.: A coordination mechanism for a supply chain with demand information updating. Int. J. Prod. Econ. **103**, 347–361 (2006)
8. Clarke, F.H.: Optimization and Nonsmooth Analysis. Wiley, New York (1983)
9. Dontchev, A.L., Lewis, A.S., Rockafellar, R.T.: The radius of metric regularity. Trans. Am. Math. Soc. **355**, 493–517 (2004)
10. Ermoliev, Y.M.: Stochastic Programming Methods. Nauka, Moscow (1976)
11. Ermoliev, Y.M., Norkin, V.I.: Solution of nonconvex nonsmooth stochastic optimization problems. Cybern. Syst. Anal. **39**, 701–715 (2003)
12. Facchinei, F., Pang, J.-S.: Finite-dimensional Variational Inequalities and Complementarity Problems, Springer, Heidelberg (2003)
13. Hess, C.: Set-Valued integration and set-valued probability theory: an overview. In: Handbook of Measure Theory, vol. I, II, pp 617–673. North-Holland, Amsterdam (2002)
14. Hogan, W.W.: Point-to-set maps in mathematical programming. SIAM Rev. **15**, 591–603 (1973)
15. Homen-De-Mello, T.: Estimation of derivatives of nonsmooth performance measures in regenerative systems. Math. Oper. Res. **26**, 741–768 (2001)
16. King, A.J., Wets, R. J.-B.: Epi-consistency of convex stochastic programs. Stochast. Stochast. Rep. **34**, 83–92 (1991)
17. Kiwiel, K.C.: An aggregate subgradient method for nonsmooth convex minimization. Math. Program. **27**, 320–341 (1983)
18. Lemaréchal, C.: Bundle methods in nonsmooth optimization. In: Lemaréchal, C., Mifflin, R. (eds.) Nonsmooth Optimization, pp 79–102. Pergamon Press, Oxford (1978)
19. Linderoth, J., Shapiro, A., Wright, S.: The empirical behavior of sampling methods for stochastic programming. Anna. Oper. Res. **142**, 215–241 (2006)
20. Luo, Z.-Q., Pang, J.-S., Ralph, D.: Mathematical Programs with Equilibrium Constraints. Cambridge University Press, Cambridge (1996)
21. Meng, F., Xu, H.: A regularized sample average approximation method for stochastic mathematical programs with nonsmooth equality constraints. SIAM J. Optim. **17**, 891–919 (2006)
22. Peng, J.: A smoothing function and its applications. In: Fukushima, M., Qi, L. (eds.) Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, pp 293–316. Kluwer, Dordrecht (1998)
23. Plambeck, E.L., Fu, B.R., Robinson, S.M., Suri, R.: Sample-path optimization of convex stochastic performances functions. Math. Program. **75**, 137–176 (1996)

24. Ralph, D., Xu, H.: Implicit smoothing and its application to optimization with piecewise smooth equality constraints. J. Optim. Theory Appl. **124**, 673–699 (2005)
25. Robinson, S.M.: Generalized equations, Mathematical programming: the state of the art. Bonn, 1982, pp. 346–367. Springer, Berlin (1983)
26. Robinson, S.M.: Analysis of sample-path optimization. Math. Oper. Res. **21**, 513–528 (1996)
27. Rockafellar, R.T., Uryasev, S.: Optimization of conditional value-of-risk. J. Risk **2**, 21–41 (2000)
28. Rockafellar, R.T., Uryasev, S.: Conditional value-at-risk for general loss distributions. J. Bank. Finance **26**, 1443–1471 (2002)
29. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Springer, Berlin (1998)
30. Römisch, W., Schultz, R.: Lipschitz stability for stochastic programs with complete recourse. SIAM J. Optim. **6**, 531–547 (1996)
31. Rubinstein, R.Y., Shapiro, A.: Discrete Events Systems: Sensitivity Analysis and Stochastic Optimization by the Score Function Methods. Wiley, New York (1993)
32. Rusczyński, A.: A linearization method for nonsmooth stochastic programming problems. Math. Oper. Res. **12**, 32–49 (1987)
33. Rusczyński, A., Shapiro, A.: Stochastic Programming, Handbooks in OR & MS, vol. 10. North-Holland, Amsterdam (2003)
34. Shapiro, A.: Monte Carlo sampling methods. In: Rusczyński, A., Shapiro, A. (eds.) Stochastic Programming, Handbooks in OR & MS, vol. 10. North-Holland, Amsterdam (2003)
35. Shapiro, A., Homem-de-Mello, T.: On rate of convergence of Monte Carlo approximations of stochastic programs. SIAM J. Optim. **11**, 70–86 (2000)
36. Shapiro, A., Xu, H.: Stochastic mathematical programs with equilibrium constraints, modeling and sample average approximation. Optimization (to appear) (2005). http://www.optimization-online.org/DB_HTML/2005/01/1046.html
37. Shapiro, A., Xu, H.: Uniform laws of large numbers for set-valued mappings and subdifferentials of random functions. J. Math. Anal. Appl. **325**, 1390–1399 (2006)
38. Wets, R.: Stochastic Programming. In: Nemhauser, G.L., et al. (eds.) Handbooks in OR & MS, vol. 1, pp. 573–629 (1989)
39. Xu, H.: An implicit programming approach for a class of stochastic mathematical programs with complementarity constraints. SIAM J. Optim. **16**, 670–696 (2006)
40. Xu, H., Meng, F.: Convergence analysis of sample average approximation methods for a class of stochastic mathematical programs with equality constraints. Math. Oper. Res. **32**, 648–669 (2007)
41. Zhang, J.: Transshipment and its impact on supply chain members' performance. Manage. Sci. **51**, 1534–1539 (2005)
42. Zimmer, K.: Supply chain coordination with uncertain just-in-time delivery. Int. J. Prod. Econ. **77**, 1–15 (2002)