Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL
Pacific Journal of Optimization
Vol. 14, No. 4, pp. 703–722, 2018

EXPECTED RESIDUAL MINIMIZATION METHOD FOR STOCHASTIC MIXED VARIATIONAL INEQUALITY PROBLEMS

Lin Chen, Guihua Lin* and Xinmin Yang†

Abstract: In this paper, we study a class of stochastic mixed variational inequality problems (SMVIPs) in finite dimensional spaces. It can be observed that the SMVIP may have no common solution for almost every realization in general. In order to get a reasonable resolution, we first present a deterministic formulation, the expected residual minimization (ERM) formulation, for the SMVIP by means of some merit function. Then we establish some basic properties of the ERM problem and propose a quasi-Monte Carlo approximation approach to solve it. We also obtain some convergence results of optimal solutions and stationary solutions of the approximation problem to their true counterparts.

Key words: *stochastic mixed variational inequality problem, expected residual minimization formulation, gap function, quasi-Monte Carlo method, convergence analysis.*

Mathematics Subject Classification: *90C15, 90C33*

---

1 Introduction

Let $S \subseteq R^n$ be a nonempty closed convex subset, $F : R^n \to R^n$ be a continuous vector-valued mapping, and $g : R^n \to R \cup \{+\infty\}$ be a lower semicontinuous, proper and convex function with closed effective domain. The normal cone operator for S at $x \in S$ is defined as

$$N_S(x) := \{z \in R^n | \langle z, y - x\rangle \le 0, \forall y \in S\},$$

where $\langle \cdot, \cdot \rangle$ denotes the standard inner product in R^n, and $N_S(x) := \emptyset$ for any $x \notin S$. The subdifferential operator for g at x is defined as

$$\partial g(x) := \{\tau \in R^n | g(y) \ge g(x) + \langle \tau, y - x\rangle, \forall y \in R^n\}.$$

See [29, Definition 8.3] for details. Consider the generalized variational inequality problem (GVIP(F, g, S) for short, see [10, 24]) of finding a vector $x \in S$ such that

$$0 \in F(x) + \partial g(x) + N_S(x),$$

which is also known as a generalized equation [26]. This problem and its various special cases have a large variety of applications in partial differential equations, equilibrium problems in

*Research of the second author was partially supported by the Humanity and Social Science Foundation of Ministry of Education of China (No. 15YJA630034) and the Innovation Program of Shanghai Municipal Education Commission (No. 14ZS086).

†Research of the third author was partially supported by the Natural Science Foundation of China (No. 11431004)

games, economics and transportation analysis, nonlinear programming, etc., see[4, 5, 9, 13, 16, 22, 27, 32].

If $\text{int}\{\text{dom}(g)\} \cap S \neq \emptyset$, the problem GVIP$(F, g, S)$ can be stated in terms of the function g rather than its subdifferential mapping (see [24]), i.e. GVIP(F, g, S) can be written as the problem of finding $x \in S$ such that

$$\langle F(x), y - x \rangle + g(y) - g(x) \geq 0, \quad \forall y \in S.$$

In this paper, we consider such an equivalent variational inequality formulation, called mixed variational inequality problem (denoted by MVIP(F, g, S)), which was originally studied by Duvaut and Lions [8]. Since the function F may involve some random factors or uncertainties in many practical problems, in this paper, we focus on the following stochastic MVIP (SMVIP): Find $x \in S$ such that

$$\langle f(x, \xi(\omega)), y - x \rangle + g(y) - g(x) \geq 0, \quad \forall y \in S, \ a.e. \ \xi(\omega) \in \Xi. \tag{1.1}$$

or equivalently

$$P\{\xi(\omega) \in \Xi : \langle f(x, \xi(\omega)), y - x \rangle + g(y) - g(x) \geq 0, \ \ \forall y \in S\} = 1,$$

where $f : R^n \times \Xi \to R^n$ is a real vector-valued mapping, $\xi : \Omega \to \Xi \subseteq R^n$ is a random vector defined on the probability space $(\Omega, \mathcal{F}, P)$ supported on closed set Ξ, and "*a.e.*" is the abbreviation for "almost every". To simplify the notations, we will use ξ to denote either the random vector $\xi(\omega)$ or an element of R^n depending on the context. Due to the existence of the randomness, problem (1.1) may not have a solution in general, which means that (1.1) is not well-defined if we want to solve (1.1) before knowing the realization of ξ. Therefore, our first step is to find a reasonable deterministic formulation for the above SMVIP.

The SMVIP (1.1) is obviously a generalization of the stochastic variational inequality problem (SVIP) of finding $x \in S$ such that

$$\langle f(x, \xi), y - x \rangle \geq 0, \quad \forall y \in S, \ a.e. \ \xi \in \Xi,$$

which has received much attention in the recently [3, 12, 17, 18, 19, 36]. One popular deterministic formulation for SVIP is the so-called expected residual minimization (ERM) formulation, which was firstly presented by Chen and Fukushima for the stochastic complementarity problems in [3] and then was extended by Luo and Lin to the general SVIP in [18], by minimizing the expectation of some residual function. Motivated by these works on SVIP, in this paper, we study the ERM formulation for the SMVIP (1.1).

The rest of the paper is organized as follows. The ERM formulation of SMVIP based on a regularized gap function is presented in the next Section. In Section 3, we give the approximation problem generated by quasi-Monte Carlo method for ERM formulation, and we show the convergence results for the global optimal solutions and the stationary solutions of the approximation problem under some moderate assumptions. In Section 4, we present a uniform exponential convergence theorem for stationary solutions of sample average approximation problem when the sample size is sufficiently large.

In what follows, $\|\cdot\|$ denotes the Euclidean norm of a vector or the spectral norm of a matrix, whereas $\|A\|_F$ denotes the Frobenius norm. Moreover, we always assume that $g(x)$ is a lower semicontinuous, proper and convex function on $S \subseteq \text{dom } g(x)$. In addition, we suppose that the following assumptions hold throughout:

Assumption 1.1 (A1)**.** The function $f(x, \xi)$ is (Borel) measurable in ξ for every $x \in R^n$ and continuously differentiable in x for a.e. $\xi \in \Xi$;

Assumption 1.2 (A2)**.** The function $f(x, \xi)$ is integrably bounded with respect to ξ, i.e. the expected value function $F(x) := E[f(x, \xi)]$ is well defined and finite valued, where E denotes the expectation with respect to the random variable $\xi \in \Xi$.

2 ERM Reformulation for SMVIP

Consider the classical variational inequality problem (VIP): find $z \in S$ such that

$$\langle F(z), y - z\rangle \geq 0, \quad \forall z \in S.$$

Recall that $x \in S$ is a solution of the VIP if and only if

$$0 = x - P_S(x - \alpha^{-1}F(x)),$$

where P_S is the orthogonal projector onto S and $\alpha > 0$ is a given scalar. The norm of the right-hand side in the above equation can serve as a residual function for the VIP, which is usually called the natural residual. In addition, Fukushima introduced the following regularized gap function for the VIP in [11]:

$$\Gamma_\alpha(x) := \max_{y \in S}\{\langle F(x), x - y\rangle - \frac{\alpha}{2}\|x - y\|^2\}.$$

This function has a number of interesting properties, in particularly, it has better smoothness properties comparing to the natural residual.

Now we introduce a mapping $p_g^\alpha : R^n \to S$, called the restricted proximal map, that is given by

$$p_g^\alpha(z) := \arg\min_{y \in S}\{g(y) + \frac{\alpha}{2}\|z - y\|^2\},$$

where α is a positive prox-parameter and the point z is called the prox-center. If $S = R^n$, $p_g^\alpha(\cdot)$ becomes the well-known proximal map, which was first introduced by Moreau in [21] and became a fundamental tool in convex and nonconvex optimizations, see [28, 29] for details. In [33], the author investigated the merit functions and error bounds for a class of generalized variational inequalities by using the proximal map. In this paper, we use the restricted proximal map to investigate the properties of SMVIP. Note that, since $g(x)$ is proper and convex, the objective function above is proper and strongly convex, hence p_g^α is single-valued. It is not hard to show that the restricted proximal map is nonexpansive (i.e. Lipschitz constant is 1, see [28, Section 31]).

In order to get a reasonable formulation for the SMVIP, we now consider the related scenario-based mixed variational inequality problem. For a given $\xi \in \Xi$, the related scenario-based MVIP, denoted by MVIP$(f(\cdot, \xi), g, S)$, is to find $x \in S$ such that

$$\langle f(x, \xi), y - x\rangle + g(y) - g(x) \geq 0, \quad \forall y \in S.$$

Using the restricted proximal map p_g^α, we can define a function $h : R^n \times \Xi \to R^n$ by

$$h_\alpha(x, \xi) := x - p_g^\alpha(x - \alpha^{-1}f(x, \xi)).$$

The following lemma shows that $\|h_\alpha(\cdot, \xi)\|$ plays the similar role for the MVIP$(f(\cdot, \xi), g, S)$ as the natural residual for the classic VIP.

Lemma 2.1. *For a given $\xi \in \Xi$ and any $\alpha > 0$, the vector $x \in S$ solves MVIP$(f(\cdot, \xi), g, S)$ if and only if $h_\alpha(x, \xi) = 0$.*

Proof. It is obvious that $h_\alpha(x,\xi)=0$ is equivalent to

$$x=\arg\min_{y\in S}\{g(y)+\frac{\alpha}{2}\|y-(x-\alpha^{-1}f(x,\xi))\|^2\},$$

which is equivalent to

$$0\in\partial g(x)+\alpha(x-(x-\alpha^{-1}f(x,\xi))+N_S(x)=\partial g(x)+f(x,\xi)+N_S(x),$$

or

$$-f(x,\xi)-\tau\in\partial g(x)\ \text{ for some } \tau\in N_S(x).$$

By the definition of the subgradient of g, we have that, for any $y\in S$,

$$g(y)\geq g(x)-\langle f(x,\xi)+\tau,y-x\rangle\geq g(x)-\langle f(x,\xi),y-x\rangle,$$

where the last inequality follows from the normal cone $N_S(x)$ at $x\in S$. This means that x solves MVIP$(f(\cdot,\xi),g,S)$.

On the other hand, if $x\in S$ solves MVIP$(f(\cdot,\xi),g,S)$, it is equivalent to that x is a solution of the convex programming problem

$$\min_{y\in S}\ \langle f(x,\xi),y\rangle+g(y)$$

for the fixed $\xi\in\Xi$. We can get from its optimality conditions that

$$0\in\partial g(x)+f(x,\xi)+N_S(x),$$

which means that $h_\alpha(x,\xi)=0$ holds. □

On account of the restricted proximal map p_g^α, we can also define a regularized gap function $r:R^n\times\Xi\to[0,\infty]$ by

$$r_\alpha(x,\xi):=\max_{y\in S}\{\langle f(x,\xi),x-y\rangle+g(x)-g(y)-\frac{\alpha}{2}\|x-y\|^2\},\tag{2.1}$$

where α is a positive parameter. Then for any $x\in R^n$ and $\xi\in\Xi$, we have

$$\begin{aligned}r_\alpha(x,\xi)\ &=\langle f(x,\xi),x-p_g^\alpha(x-\alpha^{-1}f(x,\xi))\rangle+g(x)-g\big(p_g^\alpha(x-\alpha^{-1}f(x,\xi))\big)\\&\quad-\frac{\alpha}{2}\|x-p_g^\alpha(x-\alpha^{-1}f(x,\xi))\|^2.\end{aligned}\tag{2.2}$$

In fact, since $p_g^\alpha(y)\in\text{dom } g$ for any $y\in S$, (2.2) is obvious if $x\notin\text{dom } g$. If $x\in\text{dom } g$, we may suppose that y is the unique solution of the right side of (2.1) because of the strong concavity. It follows that

$$0\in f(x,\xi)+\partial g(y)+\alpha(y-x)+N_S(y),$$

which implies that y uniquely characterizes the solution of the problem

$$\min_{y\in S}\{g(y)+\frac{\alpha}{2}\|y-(x-\alpha^{-1}f(x,\xi))\|^2\}.$$

So we have $y=p_g^\alpha(x-\alpha^{-1}f(x,\xi))$ by the definition of p_g^α, from which we obtain (2.2).

From (2.2), we see that, for each $x\in S\subseteq\text{dom } g$ and $\xi\in\Xi$, $r_\alpha(x,\xi)$ is finite valued and continuous in x under the assumptions (A1) and (A2). The following theorem shows that $r_\alpha(x,\xi)$ can also serve as a merit function for MVIP$(f(\cdot,\xi),g,S)$.

Theorem 2.2. *For a given* $\xi \in \Xi$ *and any* $\alpha > 0$*, the vector* $x \in S$ *solves MVIP*$(f(\cdot,\xi), g, S)$ *if and only if* $r_\alpha(x,\xi) = 0$.

Proof. First, we show that, for any $x \in S$ and fixed $\xi \in \Xi$,

$$r_\alpha(x,\xi) \geq \frac{\alpha}{2}\|h_\alpha(x,\xi)\|^2. \tag{2.3}$$

In fact, from the definition of p_g^α, $p_g^\alpha(x - \alpha^{-1}f(x,\xi))$ satisfies the optimality condition

$$0 \in \partial g\big(p_g^\alpha(x-\alpha^{-1}f(x,\xi))\big)+\alpha\big(p_g^\alpha(x-\alpha^{-1}f(x,\xi)\big)-(x-\alpha^{-1}f(x,\xi))+N_S\big(p_g^\alpha(x-\alpha^{-1}f(x,\xi))\big).$$

This implies

$$-f(x,\xi) - \alpha\big(p_g^\alpha(x - \alpha^{-1}f(x,\xi)) - x\big) - \tau \in \partial g\big(p_g^\alpha(x - \alpha^{-1}f(x,\xi))\big), \tag{2.4}$$

where $\tau \in N_S(p_g^\alpha(x - \alpha^{-1}f(x,\xi)))$. It follows from (2.4) that, for any $y \in S$,

$$g(y)-g\big(p_g^\alpha(x-\alpha^{-1}f(x,\xi))\big)+\langle f(x,\xi)+\alpha\big(p_g^\alpha(x-\alpha^{-1}f(x,\xi))-x\big), y-p_g^\alpha\big(x-\alpha^{-1}f(x,\xi)\big)\rangle \geq 0,$$

Picking $y = x$, we have

$$g(x) - g(p_g^\alpha(x - \alpha^{-1}f(x,\xi))) + \langle f(x,\xi), h_\alpha(x,\xi)\rangle \geq \alpha\|h_\alpha(x,\xi)\|^2,$$

which implies (2.3) together with (2.2).

As a result, for fixed ξ, if $x \in S$ satisfies $r_\alpha(x,\xi) = 0$, we have $h_\alpha(x,\xi) = 0$ and hence, from Lemma 2.1, x solves MVIP$(f(\cdot,\xi), g, S)$. On the other hand, if $x \in S$ solves MVIP$(f(\cdot,\xi), g, S)$, then $h_\alpha(x,\xi) = 0$ holds, which implies $x = p_g^\alpha(x - \alpha^{-1}f(x,\xi))$. Due to (2.2), $r_\alpha(x,\xi) = 0$ holds. ☐

Based on the previous theorem and motivated by the work of Chen and Fukushima [3], we suggest the following ERM formulation for problem (1.1):

$$\min_{x\in S}\ \theta(x) := E[r_\alpha(x,\xi)] = \int_\Xi r_\alpha(x,\xi)\rho(\xi)d\xi \tag{2.5}$$

where ρ is the probability density function of the random variable ξ and is supposed to be continuous on Ξ throughout. It is obvious that, if ρ is known and the objective function θ can be integrated out explicitly, we do not require any discretization procedure in dealing with the above ERM problem. Unfortunately, θ generally can not be calculated in a closed form or is difficult to evaluate exactly, so we have to approximate it by means of some discretization techniques.

We discuss some properties of the ERM problem (2.5) below and study approximation methods for solving (2.5) in the next section. First, we let S_0 and S^* be the solution set of problem (1.1) and the optimal solution set of problem (2.5), respectively.

Definition 2.3 ([15])**.** A bivariate function $f(x,\xi)$ is monotone on S at $\bar{x}$ uniformly in Ξ, if

$$\langle f(\bar{x},\xi) - f(x,\xi), \bar{x} - x\rangle \geq 0, \quad \forall x \in S,\ a.e.\ \xi \in \Xi;$$

$f(x,\xi)$ is strictly monotone on S at $\bar{x}$ uniformly in Ξ, if

$$\langle f(\bar{x},\xi) - f(x,\xi), \bar{x} - x\rangle > 0, \quad \forall x \in S, x \neq \bar{x},\ a.e.\ \xi \in \Xi;$$

$f(x,\xi)$ is strongly monotone on S at $\bar{x}$ uniformly in Ξ with modulus $\sigma > 0$, if

$$\langle f(\bar{x},\xi) - f(x,\xi), \bar{x} - x\rangle \geq \sigma\|\bar{x} - x\|^2, \quad \forall x \in S,\ a.e.\ \xi \in \Xi.$$

To simplify the notation, we say $f(x,\xi)$ is uniformly (strictly, strongly) monotone on S if it is (strictly, strongly) monotone at every point of S uniformly in Ξ. Obviously, the uniform strong monotonicity implies the uniform strict monotonicity, which implies the uniform monotonicity.

Theorem 2.4. *Assume that for a.e. $\xi \in \Xi$, $\bar{x}$ is a solution of the related scenario-based MVIPs, or $\bar{x}$ is an optimal solution of problem* (2.5) *with zero optimal value, then $\bar{x}$ is a solution of problem* (1.1)*. Furthermore, if $f(x,\xi)$ is strictly monotone on S at $\bar{x}$ uniformly in Π with $\Pi \subseteq \Xi$ and $P(\Pi)$ be any positive scalar, then $\bar{x}$ is the unique solution of problem* (1.1)*.*

Proof. By assumptions, we have $0 = \theta(\bar{x}) = E[r_\alpha(\bar{x},\xi)]$. Noticing that $r_\alpha(x,\xi) \geq 0$ always holds, we have $r_\alpha(\bar{x},\xi) = 0$ for any $\bar{x} \in S$ and a.e. $\xi \in \Xi$. By Theorem 2.2, $\bar{x}$ is a solution of problem (1.1).

If problem (1.1) has another solution $\hat{x}$, then we have

$$\langle f(\bar{x},\xi), \hat{x} - \bar{x}\rangle + g(\hat{x}) - g(\bar{x}) \geq 0, \ \ a.e.\ \xi \in \Xi,$$

$$\langle f(\hat{x},\xi), \bar{x} - \hat{x}\rangle + g(\bar{x}) - g(\hat{x}) \geq 0, \ \ a.e.\ \xi \in \Xi.$$

Adding the above two inequalities, we have

$$\langle f(\bar{x},\xi) - f(\hat{x},\xi), \bar{x} - \hat{x}\rangle \leq 0, \ \ a.e.\ \xi \in \Xi,$$

which is a contradiction to the strictly monotonicity on S at $\bar{x}$ uniformly in Π. As a result, $\bar{x}$ is the unique solution of problem (1.1). □

Next we discuss the error bound conditions and the boundedness of the level set defined by

$$L_\theta^S(c) := \{x \in S \,|\, \theta(x) \leq c\},$$

where c is any nonnegative number.

Proposition 2.5. *Assume that S_0 is nonempty and $f(x,\xi)$ is monotone on S uniformly in Ξ and strongly monotone on S uniformly in Θ with modulus $\sigma > 0$, where Θ is a subset of Ξ with $P(\Theta) = m > 0$. If $\alpha \in (0, 2m\sigma)$, then*

$$d(x, S_0) \leq \sqrt{(m \cdot \sigma - \frac{\alpha}{2})^{-1}\ \theta(x)}, \quad \forall x \in S. \tag{2.6}$$

Moreover, the level set $L_\theta^S(c)$ is bounded for any $c \geq 0$.

Proof. Let $\bar{x}$ be a solution of problems (1.1), we have from the definition of $\theta(x)$ that

$$\begin{aligned}
\theta(x) &= E[r_\alpha(x,\xi)] \\
&\geq \max_{y\in S}\{\langle E[f(x,\xi)], x - y\rangle + g(x) - g(y) - \frac{\alpha}{2}\|x - y\|^2\} \\
&\geq \langle E[f(x,\xi)], x - \bar{x}\rangle + g(x) - g(\bar{x}) - \frac{\alpha}{2}\|x - \bar{x}\|^2 \\
&= \int_\Theta [\langle f(x,\xi), x - \bar{x}\rangle + g(x) - g(\bar{x})] d\xi \\
&\quad + \int_{\Xi\backslash\Theta} [\langle f(x,\xi), x - \bar{x}\rangle + g(x) - g(\bar{x})] d\xi - \frac{\alpha}{2}\|x - \bar{x}\|^2
\end{aligned}$$

$$\geq \int_{\Theta} [\langle f(\bar{x},\xi), x-\bar{x}\rangle + g(x) - g(\bar{x})]d\xi + m\cdot\sigma\|x-\bar{x}\|^2$$
$$+ \int_{\Xi\setminus\Theta} [\langle f(\bar{x},\xi), x-\bar{x}\rangle + g(x) - g(\bar{x})]d\xi - \frac{\alpha}{2}\|x-\bar{x}\|^2$$
$$\geq \big(m\cdot\sigma - \frac{\alpha}{2}\big)\|x-\bar{x}\|^2$$
$$\geq \big(m\cdot\sigma - \frac{\alpha}{2}\big)d^2(x,S_0),$$

where the inequalities above follow from the Jensen's inequality, the strong monotonicity of f uniformly in Θ, i.e.

$$\langle f(x,\xi) - f(\bar{x},\xi), x-\bar{x}\rangle > 0, \quad \forall \xi \in \Xi,$$

and the monotonicity uniformly in Ξ, i.e.

$$\langle f(x,\xi) - f(\bar{x},\xi), x-\bar{x}\rangle \geq \sigma\|x-\bar{x}\|^2, \quad \forall \xi \in \Theta.$$

It concludes that (2.6) is true. The second part follows from (2.6) immediately. ☐

As mentioned in the introduction, problem (1.1) may not have a solution in general. So we can't apply Proposition 2.5 in very many situations to obtain the desired robust error bound results, which is important in our theoretical feasibility of our proposed method. Fortunately, we have the following error bound results by means of the related scenario-based MVIP.

Theorem 2.6. *Assume that Ξ is a finite set, and for each $\xi \in \Xi$, $f(\cdot,\xi)$ is strongly monotone on S with modulus $\sigma > 0$. Let $S_0(\xi)$ be the solution set of MVIP$(f(\cdot,\xi), g, S)$, which is supposed to be nonempty. Then*

$$E[d(x, S_0(\xi))] \leq \sqrt{(\sigma - \frac{\alpha}{2})^{-1}\ \theta(x)},$$

for any $\alpha \in (0, 2\sigma)$.

Proof. By the proof of Proposition 2.5, for each $\xi \in \Xi$, we have

$$r_\alpha(x,\xi) \geq \big(\sigma - \frac{\alpha}{2}\big)d^2(x, S_0(\xi)).$$

Integrating on both sides of the above inequality at the same time, the conclusion is obtained by Jensen's inequality immediately. ☐

Remark 2.7. Suppose x^* is a solution of problem (2.5)(not necessarily zero valued). Theorem 2.6 shows that

$$E[d(x^*, S_0(\xi))] \leq \sqrt{(\sigma - \frac{\alpha}{2})^{-1} \cdot \min_{x\in S}\theta(x)}, \tag{2.7}$$

Unlike an error bound for the deterministic counterparts, the left-hand side of (2.7) is in a high probability to be positive. If $\theta(x^*)$ equals to zero, then x^* is a solution of problem (1.1). Otherwise, the inequality (2.7) suggests that the expected distance to the solution set $S_0(\xi)$ for the related scenario-based MVIP is also likely to be small at solutions of ERM formulation (2.5). In other words, we may expect that a solution of ERM formulation (2.5) has a minimum sensitivity with respect to random parameter variations in SMVIP. In this sense, solutions of ERM formulation (2.5) can be regarded as robust solutions for SMVIP. Thanks to the discretization of the expectation operator in next section, it's rational to require the finiteness of the support set Ξ.

3 Quasi-Monte Carlo Method

In the rest of the paper, we suppose that Ξ is a nonempty closed set (not necessarily compact), $g(\cdot)$ is twice continuously differentiable and Lipschitz continuous on S with Lipschitz modulus L. And, besides the assumptions (A1) and (A2), we further suppose the following assumption :

Assumption 3.1 (A3). $\|f(x,\xi) - f(y,\xi)\| \leq \kappa(\xi)\|x-y\|$ for any $x, y \in S$ with $E[\kappa(\xi)] < +\infty$.

Moreover, the probability density function ρ of the random variable ξ is supposed to be known and subsequently, we apply the well-known quasi-Monte Carlo simulation techniques to approximate the expectations involved in (2.5). That is, we first employ a quasi-Monte Carlo method to generate a set of observations $\Xi_k = \{\xi^i | i = 1, 2, ..., N_k\} \subseteq \Xi$ with $N_k \to \infty$ when $k \to \infty$, then treat the following problem as an approximation of (2.5):

$$\min_{x\in S} \ \theta^k(x) := \frac{1}{N_k} \sum_{\xi^i \in \Xi_k} r_\alpha(x, \xi^i)\rho(\xi^i) \tag{3.1}$$

See [23] for more details about the quasi-Monte Carlo approximation techniques. The following results will be used later on.

Lemma 3.1 ([23]). *If $\psi(\xi)$ is integrably bounded over Ξ, then*

$$\lim_{k\to\infty} \frac{1}{N_k} \sum_{\xi^i\in\Xi_k} \psi(\xi^i)\rho(\xi^i) = E[\psi(\xi)].$$

Lemma 3.2. *For any $\alpha > 0$, the regularized gap function $r_\alpha(x,\xi)$ and its gradient $\nabla_x r_\alpha(x,\xi)$ are measurable in ξ for every $x \in R^n$.*

Proof. It follows from [30, Chapter 1, Theorem 19] under assumption (A1). □

From Lemma 3.1 and Lemma 3.2, together with the integrability of $r_\alpha(x,\cdot)$ on Ξ for each $x \in S$ under assumption (A2), it yields

$$\theta(x) = \lim_{k\to\infty} \theta^k(x). \tag{3.2}$$

Theorem 3.3. *For a.e. $\xi \in \Xi$, $r_\alpha(x,\xi)$ is continuously differentiable with respect to x. Moreover, for any $x \in S$, we have*

$$\nabla\theta(x) = E[\nabla_x r_\alpha(x,\xi)]. \tag{3.3}$$

Consequently, $\theta(x)$ is continuously differentiable, too.

Proof. Since for a.e. $\xi \in \Xi$, $f(x,\xi)$ is continuously differentiable with respect to x and $g(x)$ is twice continuously differentiable, in a similar way to [11, Theorem 3.2], we can show that for a.e. $\xi \in \Xi$, $r_\alpha(x,\xi)$ is continuously differentiable with respect to x. Furthermore, we have

$$\nabla_x r_\alpha(x,\xi) = f(x,\xi) + \nabla g(x) - (\nabla_x f(x,\xi) - \alpha I)\big(p_g^\alpha(x - \alpha^{-1}f(x,\xi)) - x\big) \tag{3.4}$$

for any $x \in S$, a.e. $\xi \in \Xi$. Since $r_\alpha(x,\xi) \geq 0$ for any $x \in S$, it follows from (2.2) that

$$\frac{\alpha}{2}\|x - p_g^\alpha(x - \alpha^{-1}f(x,\xi))\|^2$$

$$\leq \langle f(x,\xi), x - p_g^\alpha(x - \alpha^{-1} f(x,\xi))\rangle + g(x) - g(p_g^\alpha(x - \alpha^{-1} f(x,\xi)))$$
$$\leq \|x - p_g^\alpha(x - \alpha^{-1} f(x,\xi))\|(\|f(x,\xi)\| + L).$$

Then we have

$$\|x - p_g^\alpha(x - \alpha^{-1} f(x))\| \leq \frac{2}{\alpha}(\|f(x,\xi)\| + L). \tag{3.5}$$

From (3.4) and (3.5), we have

$$\|\nabla_x r_\alpha(x,\xi)\| \leq \|f(x,\xi)\| + \|\nabla g(x)\| + (\|\nabla_x f(x,\xi)\| + \alpha)\|x - p_g^\alpha(x - \alpha^{-1} f(x,\xi))\|$$
$$\leq \|f(x,\xi)\| + L + \frac{2}{\alpha}(\|\sqrt{n}\cdot\kappa(\xi)\| + \alpha)(\|f(x,\xi)\| + L). \tag{3.6}$$

Therefore, θ is continuously differentiable and (3.3) holds from the Lebesgue's control convergence theorem under the assumptions (A1)–(A3). □

However, there is no guarantee that such a solution is a global optimal solution of the ERM problem (2.5). So, we establish the convexity of the regularized gap function and show that the resulting ERM problem (2.5) is a convex problem. Now, we established some sufficient conditions for the convexity of the regularized gap function when f is affine. Suppose that f is in a special structure of $f(x,\xi) = M(\xi)x + q(\xi)$, where $M(\xi) \in R^{n\times n}$, $q(\xi) \in R^n$ for any $\xi \in \Xi$.

Definition 3.4. We call $M(\xi)$ uniformly positive definite with modulus ϑ_0 if there exists a positive constant ϑ_0 such that

$$\inf_{\xi\in\Xi, \|x\|=1} x^T M(\xi)\, x \geq \vartheta_0.$$

Theorem 3.5. *Suppose that $M(\xi)$ is uniformly positive definite with modulus ϑ_0, then the regularized gap function $r_\alpha(x,\xi)$ is convex in x for all $\alpha \geq \frac{1}{2\vartheta_0}$, and strongly convex with modulus ϑ for all $\alpha \geq \frac{1}{2\vartheta_0}(1+\vartheta)$ with $\vartheta > 0$.*

Proof. The proof is similar to [2, Theorem 2.1], so we omit the details. □

Since the sum of (strongly) convex functions is also (strongly) convex, as a consequence, the (strongly) convexity of $\theta(x)$ is the same as $r_\alpha(x,\xi)$. In what follows, we investigate the limiting behavior of the approximation method mentioned above. We denote by S_k^* the optimal solution sets of problems (3.1).

Theorem 3.6. *Suppose that $x^k \in S_k^*$ for each k and x^* is an accumulation point of the sequence $\{x^k\}$, then $x^* \in S^*$.*

Proof. Without loss of generality, we can assume that $\{x^k\}$ itself converges to x^*, which belongs to S obviously. Let the sequence $\{x^k\}$ be contained in a compact convex set $C \subseteq S$. By the mean-value theorem, for each x^k and ξ^i, there exists $z^{ki} = \lambda_{ki}x^k + (1-\lambda_{ki})x \in C$ with $\lambda_{ki} \in [0,1]$ such that

$$r_\alpha(x^k,\xi^i) - r_\alpha(x,\xi^i) = \nabla_x r_\alpha(z^{ki},\xi^i)^T(x^k - x).$$

Since the functions $\nabla g(x)$, $f(x,\xi)$ and $\nabla_x f(x,\xi)$ are continuous with respect to x under the assumptions (A1)–(A3), it then follows that

$$|\theta^k(x^k) - \theta^k(x^*)| = |\frac{1}{N_k}\sum_{\xi^i\in\Xi_k}(r_\alpha(x^k,\xi^i) - r_\alpha(x^*,\xi^i))\rho(\xi^i)|$$

$$\begin{aligned} &\leq \frac{1}{N_k} \sum_{\xi^i \in \Xi_k} \rho(\xi^i)|(r_\alpha(x^k, \xi^i) - r_\alpha(x^*, \xi^i))| \\ &= \frac{1}{N_k} \sum_{\xi^i \in \Xi_k} \rho(\xi^i)|\nabla_x r_\alpha(z^{ki}, \xi^i)^T (x^k - x^*)| \\ &\leq \|x^k - x^*\| \cdot \frac{1}{N_k} \sum_{\xi^i \in \Xi_k} \rho(\xi^i)\|\nabla_x r_\alpha(z^{ki}, \xi^i)\|, \end{aligned} \tag{3.7}$$

which tends to zero as $k \to +\infty$ by (3.6) and the assumptions (A1)–(A3). Notice that

$$|\theta^k(x^k) - \theta(x^*)| \leq |\theta^k(x^k) - \theta^k(x^*)| + |\theta^k(x^*) - \theta(x^*)|,$$

we have from (3.2) and (3.7) that

$$\theta(x^*) = \lim_{k \to \infty} \theta^k(x^k).$$

Since $x^k \in S_k^*$ for each k , it follows that

$$\theta^k(x^k) \leq \theta^k(x), \quad \forall x \in S.$$

Letting $k \to +\infty$, we obtain

$$\theta(x^*) \leq \theta(x), \quad \forall x \in S,$$

that is, $x^* \in S^*$. □

Since problems (2.5) and (3.1) are generally nonconvex, next we consider the convergence of stationary points. To this end, we suppose that

$$S := \{x \in R^n \mid b(x) \leq 0, c(x) = 0\},$$

where $b(x) = (b_1(x), b_2(x), \ldots, b_p(x))$, $c(x) = (c_1(x), c_2(x), \ldots, c_q(x))$, $b_i : R^n \to R$ $(i = 1, 2, ..., p)$ are differentiable convex functions and $c_j : R^n \to R$ $(j = 1, 2, ..., q)$ are affine functions. Denote $I(x) := \{i \mid b_i(x) = 0, 1 \leq i \leq p\}$.

Definition 3.7. (1) x^k is said to be stationary to (3.1) if there exist Lagrange multiplier vectors $\zeta^k \in R^p$ and $\eta^k \in R^q$ such that

$$\nabla\theta^k(x^k) + \sum_{i=1}^{p} \zeta_i^k \nabla b_i(x^k) + \sum_{j=1}^{q} \eta_j^k \nabla c_j(x^k) = 0, \tag{3.8}$$

$$0 \leq \zeta^k \perp b(x^k) \leq 0, \ c(x^k) = 0. \tag{3.9}$$

where $u \perp v$ means $u^T v = 0$.

(2) x^* is said to be stationary to (2.5) if there exist Lagrange multiplier vectors $\zeta \in R^p$ and $\eta \in R^q$ such that

$$\nabla\theta(x^*) + \sum_{i=1}^{p} \zeta_i \nabla b_i(x^*) + \sum_{j=1}^{q} \eta_j \nabla c_j(x^*) = 0, \tag{3.10}$$

$$0 \leq \zeta \perp b(x^*) \leq 0, \ c(x^*) = 0. \tag{3.11}$$

Definition 3.8 ([35])**.** Let $\phi : R^n \times \Xi \to R^m$ be a real vector-valued mapping, $X \subseteq R^n$ be a closed subset and $x \in X$ be fixed. ϕ is said to be calm at x with modulus $\gamma(\xi)$ if $\phi(x,\xi)$ is finite and there exist a measurable function $\gamma : \Xi \to R_+$ and a positive number δ such that

$$\|\phi(x',\xi) - \phi(x,\xi)\| \le \gamma(\xi)\|x' - x\|$$

for all $x' \in X$ with $\|x' - x\| \le \delta$ and $\xi \in \Xi$. ϕ is said to be calm on X if it is calm at every point of X (the constants $\gamma(\xi), \delta$ may depend on the point x).

Lemma 3.9. *For any* $\{x^k\} \subseteq S$ *with* $x^k \to x^*$ *as* $k \to \infty$, *suppose that each* $\nabla_x f_j(\cdot,\xi)$ $(j = 1,\dots,n)$ *is calm at* x^* *with modulus* $\gamma_j(\xi)$ *integrably bounded over* Ξ. *Then we have*

$$\nabla\theta(x^*) = \lim_{k\to\infty} \nabla\theta^k(x^k).$$

Proof. It follows from (2.2) and (3.4) that

$$\begin{aligned}
&|\nabla\theta^k(x^k) - \nabla\theta^k(x^*)| \\
=&|\frac{1}{N_k}\sum_{\xi^i\in\Xi_k}(\nabla_x r_\alpha(x^k,\xi^i) - \nabla_x r_\alpha(x^*,\xi^i))\rho(\xi^i)| \\
\le&\frac{1}{N_k}\sum_{\xi^i\in\Xi_k}\rho(\xi^i)\Big(\|f(x^k,\xi^i) - f(x^*,\xi^i))\| + \|\nabla g(x^k) - \nabla g(x^*)\| \\
&+ \|\nabla_x f(x^k,\xi^i)p_g^\alpha(x^k - \alpha^{-1}f(x^k,\xi^i)) - \nabla_x f(x^*,\xi^i))p_g^\alpha(x^* - \alpha^{-1}f(x^*,\xi^i))\| \\
&+ \|\nabla_x f(x^k,\xi^i)x^k - \nabla_x f(x^*,\xi^i)x^*\| + \alpha\|x^k - x^*\| \\
&+ \alpha\|p_g^\alpha(x^k - \alpha^{-1}f(x^k,\xi^i)) - p_g^\alpha(x^* - \alpha^{-1}f(x^*,\xi^i))\|\Big).
\end{aligned}$$

It is easy to verify that

$$\lim_{k\to\infty}\frac{1}{N_k}\sum_{\xi^i\in\Xi_k}\rho(\xi^i)\|f(x^k,\xi^i) - f(x^*,\xi^i))\| = 0 \tag{3.12}$$

by the assumption (A3) and Lemma 3.1. Moreover, by the mean-value theorem, for each x^k, there exists $\lambda \in [0,1]$ such that

$$\nabla g(x^k) - \nabla g(x^*) = \int_0^1 \nabla^2 g(z^k(\lambda))(x^k - x^*)\, d\lambda,$$

where $z^k(\lambda) = \lambda x^k + (1-\lambda)x^*$. Notice that there must be a compact convex set $U \subseteq S$ containing the whole sequence $\{x^k\}$. Due to the twice continuous differentiability of g, there exists a positive constant M such that $\max\{\|z\|, \|\nabla^2 g(z)\|\} \le M$ for any $z \in U$. Then we have

$$\begin{aligned}
\frac{1}{N_k}\sum_{\xi^i\in\Xi_k}\rho(\xi^i)\|\nabla g(x^k) - \nabla g(x^*)\| =&\frac{1}{N_k}\sum_{\xi^i\in\Xi_k}\rho(\xi^i)\|\int_0^1 \nabla^2 g(z^k(\lambda))(x^k - x^*)\, d\lambda\| \\
\le&\frac{1}{N_k}\sum_{\xi^i\in\Xi_k}\rho(\xi^i)\int_0^1 \|\nabla^2 g(z^k(\lambda))\|\,\|x^k - x^*\|\, d\lambda \\
\le&\frac{M}{N_k}\sum_{\xi^i\in\Xi_k}\rho(\xi^i)\int_0^1 \|x^k - x^*\|\, d\lambda
\end{aligned}$$

$$=\|x^k - x^*\| \cdot \frac{M}{N_k} \sum_{\xi^i \in \Xi_k} \rho(\xi^i).$$

It then follows that

$$\lim_{k\to\infty} \frac{1}{N_k} \sum_{\xi^i \in \Xi_k} \rho(\xi^i) \|\nabla g(x^k) - \nabla g(x^*)\| = 0. \tag{3.13}$$

Now we prove

$$\lim_{k\to\infty} \frac{1}{N_k} \sum_{\xi^i \in \Xi_k} \rho(\xi^i) \|\nabla_x f(x^k, \xi^i) - \nabla_x f(x^*, \xi^i))\| = 0. \tag{3.14}$$

In fact, from the relation between the spectral norm and the corresponding Frobenius norm of a matrix, we have

$$\begin{aligned}
&\frac{1}{N_k} \sum_{\xi^i \in \Xi_k} \rho(\xi^i) \|\nabla_x f(x^k, \xi^i) - \nabla_x f(x^*, \xi^i))\| \\
\leq &\frac{1}{N_k} \sum_{\xi^i \in \Xi_k} \rho(\xi^i) \|\nabla_x f(x^k, \xi^i) - \nabla_x f(x^*, \xi^i))\|_F \\
\leq &\sum_{j=1}^{n} \frac{1}{N_k} \sum_{\xi^i \in \Xi_k} \rho(\xi^i) \|\nabla_x f_j(x^k, \xi^i) - \nabla_x f_j(x^*, \xi^i))\|.
\end{aligned}$$

By the calmness of $\nabla_x f_j(\cdot, \xi)$ at x^*, we can get

$$\frac{1}{N_k} \sum_{\xi^i \in \Xi_k} \rho(\xi^i) \|\nabla_x f_j(x^k, \xi^i) - \nabla_x f_j(x^*, \xi^i))\| \leq \frac{1}{N_k} \sum_{\xi^i \in \Xi_k} \rho(\xi^i) \cdot \gamma_j(\xi) \|x^k - x^*\|. \tag{3.15}$$

letting $k \to \infty$, we have (3.14) immediately from Lemma 3.1.

It follows from the nonexpansivity of p_g^α that

$$\begin{aligned}
&\frac{1}{N_k} \sum_{\xi^i \in \Xi_k} \rho(\xi^i) \|p_g^\alpha(x^k - \alpha^{-1} f(x^k, \xi^i)) - p_g^\alpha(x^* - \alpha^{-1} f(x^*, \xi^i))\| \\
\leq &\frac{1}{N_k} \sum_{\xi^i \in \Xi_k} \rho(\xi^i) \|(x^k - \alpha^{-1} f(x^k, \xi^i)) - (x^* - \alpha^{-1} f(x^*, \xi^i))\| \\
\leq &\frac{1}{N_k} \sum_{\xi^i \in \Xi_k} \rho(\xi^i) \big(\|x^k - x^*\| + \alpha^{-1} \|f(x^k, \xi^i)) - f(x^*, \xi^i)\|\big) \\
\leq &\frac{1}{N_k} \sum_{\xi^i \in \Xi_k} \rho(\xi^i) \cdot \frac{1}{N_k} \sum_{\xi^i \in \Xi_k} \big(\|x^k - x^*\| + \alpha^{-1} \|f(x^k, \xi^i)) - f(x^*, \xi^i)\|\big),
\end{aligned} \tag{3.16}$$

which tends to zero as $k \to \infty$ taking into account of (3.12). Notice that, from (3.5), we have

$$\begin{aligned}
\|p_g^\alpha(x^* - \alpha^{-1} f(x^*, \xi^i))\| &\leq \|x^*\| + \|x^* - p_g^\alpha(x^* - \alpha^{-1} f(x^*, \xi^i))\| \\
&\leq M + \frac{2}{\alpha}(\|f(x^*, \xi^i)\| + L).
\end{aligned}$$

Then we have

$$\begin{aligned}
&\frac{1}{N_k}\sum_{\xi^i\in\Xi_k}\rho(\xi^i)\|\nabla_x f(x^k,\xi^i)p_g^\alpha(x^k-\alpha^{-1}f(x^k,\xi^i))-\nabla_x f(x^*,\xi^i))p_g^\alpha(x^*-\alpha^{-1}f(x^*,\xi^i))\|\\
=&\frac{1}{N_k}\sum_{\xi^i\in\Xi_k}\rho(\xi^i)\|\nabla_x f(x^k,\xi^i)\big(p_g^\alpha(x^k-\alpha^{-1}f(x^k,\xi^i))-p_g^\alpha(x^*-\alpha^{-1}f(x^*,\xi^i))\big)\\
&+\big(\nabla_x f(x^k,\xi^i)-\nabla_x f(x^*,\xi^i))\big)p_g^\alpha(x^*-\alpha^{-1}f(x^*,\xi^i))\|\\
\leq&\frac{1}{N_k}\sum_{\xi^i\in\Xi_k}\rho(\xi^i)\Big(\|\nabla_x f(x^k,\xi^i)\|\cdot\|(p_g^\alpha(x^k-\alpha^{-1}f(x^k,\xi^i))-p_g^\alpha(x^*-\alpha^{-1}f(x^*,\xi^i))\|\\
&+\|\nabla_x f(x^k,\xi^i)-\nabla_x f(x^*,\xi^i))\|\cdot\|p_g^\alpha(x^*-\alpha^{-1}f(x^*,\xi^i))\|\Big)\\
\leq&\frac{1}{N_k}\sum_{\xi^i\in\Xi_k}\rho(\xi^i)\Big(\|\kappa(\xi^i)\|\cdot\|(p_g^\alpha(x^k-\alpha^{-1}f(x^k,\xi^i))-p_g^\alpha(x^*-\alpha^{-1}f(x^*,\xi^i))\|\\
&+\|\nabla_x f(x^k,\xi^i)-\nabla_x f(x^*,\xi^i))\|\cdot\big(M+\frac{2}{\alpha}(\|f(x^*,\xi^i)\|+L)\big)\Big)\\
\leq&\frac{1}{N_k}\sum_{\xi^i\in\Xi_k}\rho(\xi^i)\|\kappa(\xi^i)\|\cdot\frac{1}{N_k}\sum_{\xi^i\in\Xi_k}\rho(\xi^i)\|(p_g^\alpha(x^k-\alpha^{-1}f(x^k,\xi^i))-p_g^\alpha(x^*-\alpha^{-1}f(x^*,\xi^i))\|\\
&+\frac{1}{N_k}\sum_{\xi^i\in\Xi_k}\rho(\xi^i)\|\nabla_x f(x^k,\xi^i)-\nabla_x f(x^*,\xi^i))\|\cdot\frac{1}{N_k}\sum_{\xi^i\in\Xi_k}\rho(\xi^i)\big(M+\frac{2}{\alpha}(\|f(x^*,\xi^i)\|+L)\big),
\end{aligned}\tag{3.17}$$

which tends to zero as $k\to\infty$ by (3.14)–(3.16) and (A2)–(A3). In addition, we have from (A3) and (3.14) that

$$\begin{aligned}
&\frac{1}{N_k}\sum_{\xi^i\in\Xi_k}\rho(\xi^i)\|\nabla_x f(x^k,\xi^i)x^k-\nabla_x f(x^*,\xi^i)x^*\|\\
=&\frac{1}{N_k}\sum_{\xi^i\in\Xi_k}\rho(\xi^i)\|\nabla_x f(x^k,\xi^i)(x^k-x^*)+(\nabla_x f(x^k,\xi^i)-\nabla_x f(x^*,\xi^i))x^*\|\\
\leq&\frac{1}{N_k}\sum_{\xi^i\in\Xi_k}\rho(\xi^i)\big(\|\nabla_x f(x^k,\xi^i)\|\cdot\|x^k-x^*\|+\|\nabla_x f(x^*,\xi^i)-\nabla_x f(x^k,\xi^i)\|\cdot\|x^*\|\big)\\
\leq&\frac{1}{N_k}\sum_{\xi^i\in\Xi_k}\rho(\xi^i)\big(\|\kappa(\xi^i)\|\cdot\|x^k-x^*\|+M\cdot\|\nabla_x f(x^k,\xi^i)-\nabla_x f(x^*,\xi^i)\|\big)\\
\leq&\frac{1}{N_k}\sum_{\xi^i\in\Xi_k}\rho(\xi^i)\|\kappa(\xi^i)\|\cdot\frac{1}{N_k}\sum_{\xi^i\in\Xi_k}\rho(\xi^i)\cdot\|x^k-x^*\|\\
&+\frac{M}{N_k}\sum_{\xi^i\in\Xi_k}\rho(\xi^i)\|\nabla_x f(x^k,\xi^i)-\nabla_x f(x^*,\xi^i)\|
\end{aligned}\tag{3.18}$$

which tends to zero as $k\to\infty$.

Thus, it follows from (3.12)–(3.18) that

$$\lim_{k\to\infty}\nabla\theta^k(x^k)=\nabla\theta^k(x^*).$$

Since

$$|\nabla\theta^k(x^k)-\nabla\theta(x^*)|\leq|\nabla\theta^k(x^k)-\nabla\theta^k(x^*)|+|\nabla\theta^k(x^*)-\nabla\theta(x^*)|,$$

due to Lemma 3.1 and Theorem 3.6, we obtain the conclusion. □

Definition 3.10 ([25])**.** We say that a feasible point x^* of problem (3.1) conforms to the approximate KKT (AKKT) conditions if there exist a sequence $\{x^k\}$, $\zeta^k \in R^p$, $\eta^k \in R^q$ such that $\lim_{k\to\infty} x^k = x^*$ and

$$\lim_{k\to\infty} \nabla\theta^k(x^k)) + \sum_{i=1}^{p} \zeta_i^k \nabla b_i(x^k) + \sum_{j=1}^{q} \eta_j^k \nabla c_j(x^k) = 0,$$
$$\lim_{k\to\infty} \min\{\zeta_i^k, -b_i(x^k)\} = 0, \quad i = 1, \cdots, p.$$

Given $x \in S$, we define

$$K(x) := \{ \sum_{i\in I(x)} \zeta_i \nabla b_i(x) + \sum_{j=1}^{q} \eta_j \nabla c_j(x) \mid \zeta_i \in \mathbb{R}_+, \eta_j \in \mathbb{R}\}. \tag{3.19}$$

Definition 3.11 ([1])**.** We say that $x \in S$ satisfies the cone-continuity property (CCP) if the set-valued mapping $x \rightrightarrows K(x)$, defined in (3.19), is outer semicontinuous at x, that is,

$$\limsup_{x'\to x} K(x') \subseteq K(x).$$

The CCP has been shown to be the weakest possible strict constraint qualification, under which the AKKT implies the KKT. See [1] for details. Now we investigate the limiting behavior of the stationary points obtained by the quasi-Monte Carlo method.

Theorem 3.12. *Suppose x^k be stationary to* (3.1) *for each k and x^* be an accumulation point of $\{x^k\}$. If the CCP holds as a constraint qualification at x^* and each $\nabla_x f_j(\cdot, \xi)$ $(j = 1, \ldots, n)$ is calm at x^* with modulus $\gamma_j(\xi)$ integrably bounded over Ξ, then x^* is a stationary point of problem* (2.5)*.*

Proof. Without loss of generality, we assume that $\{x^k\}$ itself converges to x^*. Since x^k is a stationary point to (3.1) for each k, then there exist Lagrange multiplier vectors $\zeta^k \in R^p$ and $\eta^k \in R^q$ satisfying (3.8) and (3.9). Let $\varepsilon^k := \nabla\theta(x^k) - \nabla\theta^k(x^k)$. By virtue of Lemma 3.9 and the continuity of the gradient of θ, we have

$$\lim_{k\to\infty} \varepsilon^k = \lim_{k\to\infty} \big(\nabla\theta(x^k) - \nabla\theta(x^*) + \nabla\theta(x^*) - \nabla\theta^k(x^k)\big) = 0.$$

It is easy to see that

$$\varepsilon^k = \nabla\theta(x^k) + \sum_{i=1}^{p} \zeta_i^k \nabla b_i(x^k) + \sum_{j=1}^{q} \eta_j^k \nabla c_j(x^k) \to 0. \tag{3.20}$$

Thus, we have

$$\sum_{i\in I(x^*)} \zeta_i \nabla b_i(x^k) + \sum_{j=1}^{q} \eta_j \nabla c_j(x^k) \in K(x^k) \tag{3.21}$$

and

$$\varepsilon^k - \nabla\theta(x^k) = \sum_{i\in I(x^*)} \zeta_i^k \nabla b_i(x^k) + \sum_{j=1}^{q} \eta_j^k \nabla c_j(x^k). \tag{3.22}$$

Taking limit in (3.22), we have from the continuity of $\nabla\theta(\cdot)$, (3.20), and (3.21) that

$$-\nabla\theta(x^*) = \lim_{k\to\infty} \sum_{i\in I(x^*)} \zeta_i^k \nabla b_i(x^k) + \sum_{j=1}^{q} \eta_j^k \nabla c_j(x^k) \in \limsup_{k\to\infty} K(x^k).$$

Moreover, we know

$$\limsup_{k\to\infty} K(x^k) \subseteq \limsup_{x'\to x^*} K(x') \subseteq K(x^*),$$

where the last inclusion follows from the CCP assumption. Therefore, we have

$$-\nabla\theta(x^*) \in K(x^*),$$

which is equivalent to (3.10). Then, taking limit in (3.9), we obtain (3.11) immediately. That is, x^* is stationary to problem (2.5). □

4 Exponential Convergence of Stationary Points

We proceed to discuss the rate of convergence of stationary points, that is, how fast x^k converges to x^* in the sense of Definition 3.7. From the computational perspective, it is important because it concerns the efficiency of the SAA method. One of the most important issues concerning the convergence analysis is how to predetermine the sample size in order to estimate an approximate solution within the prescribed precision and confidence. It is a remarkable breakthrough that the classical Cramér¨ s large deviation theorem [7] is found to deliver this. In some practical instances, it is difficult or computationally expensive to obtain an iid sample particularly when the sample size is large while Cramér¨ s large deviation theorem requiring iid sampling. Indeed, the well-known quasi-Monte Carlo method does not require iid sampling and yet it works remarkably well. See an extensive discussion on the benefits of non-iid sampling by Homem-de-Mello [14]. As far as we are concerned, Dai, Chen and Birge [6] seemed to be the first to investigate the convergence of SAA estimators under general sampling (including iid and non-iid). They used the well-known Gärther-Ellis theorem [7] to establish the exponential convergence. Homem-de-Mello [14] presented a comprehensive study of this issue and derived the exponential convergence of statistical estimators of optimal solutions in stochastic programming under non-iid sampling. More recently, Xu [35] studied the uniform exponential convergence of SAA for a class of random functions under general sampling and apply the established convergence results to nonsmooth stochastic optimization, stochastic Nash equilibrium problems and stochastic generalized equations; Sun and Xu[34] discussed the uniform exponential convergence of sample average approximation of random functions in which they extended the similar results to the discontinuous situations. By applying the result in [14] and [35], here we investigate the exponential convergence of stationary points as those presented in Section 3. In this section, the samples are generated by randomized quasi-Monte Carlo method[14].

Lemma 4.1. *Let X be a compact subset of S. Suppose:*

(i) *$\nabla g(x)$ is calm on X with modulus L';*

(ii) *each $\nabla_x f_j(\cdot,\xi)$ $(j=1,\ldots,n)$ is calm on X with modulus $J_i(\xi)$ integrably bounded over Ξ,*

then $\nabla_x r_\alpha(x,\xi)$ is calm on X, whose modulus is bounded by $p(\xi) := \sqrt{n}\cdot\kappa(\xi)(2-\alpha^{-1}\kappa(\xi)) + 2\kappa(\xi) + 2M'\cdot\sum_{i=1}^{n} J_i(\xi) + 2\alpha + L'$, and $M' := \max_{x\in X}\|x\|$.

Proof. For any $x, y \in X$, we have from (3.4) that

$$
\begin{aligned}
&\|\nabla_x r_\alpha(x,\xi) - \nabla_x r_\alpha(y,\xi)\| \\
\leq &\|\big(f(x,\xi) - f(y,\xi)\big) + \big(\nabla g(x) - \nabla g(y)\big) + \big(\nabla_x f(x,\xi)x - \nabla_x f(y,\xi)y\big) \\
&+ \alpha\big(p_g^\alpha(x - \alpha^{-1}f(x,\xi)) - p_g^\alpha(y - \alpha^{-1}f(y,\xi))\big) - \alpha(x-y) \\
&- \big(\nabla_x f(x,\xi)p_g^\alpha(x - \alpha^{-1}f(x,\xi)) - \nabla_x f(y,\xi)p_g^\alpha(y - \alpha^{-1}f(y,\xi))\big)\| \\
\leq &\|f(x,\xi) - f(y,\xi)\| + \|\nabla g(x) - \nabla g(y)\| + \|\nabla_x f(x,\xi)(x-y) + \big(\nabla_x f(x,\xi) - \nabla_x f(y,\xi)\big)y\| \\
&+ \alpha\|(x-y) - \alpha^{-1}(f(x,\xi) - f(y,\xi))\| + \alpha\|x-y\| \\
&+ \|\nabla_x f(x,\xi)\big(p_g^\alpha(x - \alpha^{-1}f(x,\xi)) - p_g^\alpha(y - \alpha^{-1}f(y,\xi))\big) \\
&+ \big(\nabla_x f(x,\xi) - \nabla_x f(y,\xi)\big)p_g^\alpha(y - \alpha^{-1}f(y,\xi))\|
\end{aligned}
$$

Since X is a compact set, the conclusion can be drawn under the assumptions. □

To establish the uniform exponential convergence, we need some assumptions on asymptotic behavior of the sample average of the modulus of the function. Let

$$M_x^N(t) := E\{e^{t[\nabla\theta^N(x) - \nabla\theta(x)]}\}.$$

Lemma 4.2. *For every $x \in X$ and $t \in R$, the limit*

$$M_x(t) := \lim_{N\to+\infty} M_x^N(t)$$

exists as an extended real number and $M_x(t) < \infty$ for t close to 0.

Proof. Consider the random function $f(x,\xi)$, and let

$$M_{x,f}^N(t) := E\{e^{t[f^N(x) - E[f(x,\xi)]]}\},$$

where $f^k(x) := \frac{1}{N_k}\sum_{\xi^i\in\Xi_k} f(x,\xi^i)\rho(\xi^i)$. First, we show that for every $x \in X$ and $t \in R$, the limit

$$M_{x,f}(t) := \lim_{N\to+\infty} M_{x,f}^N(t)$$

exists as an extended real number and $M_{x,f}(t) < \infty$ for t close to 0. Indeed, it is satisfied particularly in the case when the samplings $\{\xi^i | i = 1, 2, ..., N_k\}$ are generated by randomized quasi-Monte Carlo method, see detailed discussions about this issue by Homem-de-Mello [14]. From the definitions of $\theta(x)$ and $\theta^N(x)$, the conclusion is obtained directly.

□

Assumption 4.1. Let $p^k := \frac{1}{N_k}\sum_{\xi^i\in\Xi_k} p(\xi^i)\rho(\xi^i)$. There exists a positive constant λ such that

$$Prob\{p^k \geq \mu\} \leq e^{-k\lambda}$$

for any $\mu \geq E[p(\xi)]$.

A similar assumption is made in [35]. Assumption 4.1 means that probability distribution of the random variable $p(\xi)$ dies exponentially fast in the tails. In particular, it holds if this random variable has a distribution supported on a bounded subset.

Lemma 4.3 (Uniform exponential convergence of sample average $\nabla\theta^k(x)$)**.** *Let X be a compact subset of S and Assumption 4.1 holds. Suppose that the moment generating function $E[e^{p(\xi)t}]$ is finite valued for t close to 0. Then for any small positive number $\epsilon > 0$, there exist positive constants $\hat{c}(\epsilon)$ and $\hat{\beta}(\epsilon)$, independent of k, such that for k sufficiently large,*

$$Prob\Big\{ \sup_{x\in X} \|\nabla\theta^k(x) - \nabla\theta(x)\| \geq \epsilon \Big\} \leq \hat{c}(\epsilon)e^{-k\hat{\beta}(\epsilon)}.$$

Proof. we refer to Shapiro's earlier result [31, Proposition 2.1] which states that if a random function is continuously differentiable w.p.1 and is Lipschitz continuous with an integrably bounded Lipschitz modulus, then the expected value of the function is continuously differentiable. Under these assumptions, we are able to derive the uniform exponential convergence by virtue of [35]. We omit the details of the proof. □

Let T^* and T^k be the sets of stationary points to (2.5) and (3.1), respectively. Assume that both T^* and T^k are nonempty.

Theorem 4.4 (Exponential convergence of stationary points)**.** *Suppose x^k be a solution of* (3.1) *and the sequence $\{x^k\}$ is contained in a compact subset X of S almost surely. Assume the conditions of Lemma 4.3 to be hold, then for any small positive number $\epsilon > 0$, there exist positive constants $c(\epsilon)$ and $\beta(\epsilon)$, independent of k, such that*

$$Prob\Big\{ d(x^k, T^*) \geq \epsilon \Big\} \leq c(\epsilon)e^{-k\beta(\epsilon)}.$$

Proof. We need to translate the uniform exponential convergence of $\nabla\theta^k(x)$ to $\nabla\theta(x)$ into the exponential convergence of x^k to T^*. To this end, we need some sensitivity analysis of generalized equations discussed in [35]. So, equivalently, we consider stationary points satisfying the following generalized equation rather than those in Definition 3.7, that is,

$$0 \in \nabla\theta(x) + N_S(x), \tag{4.1}$$

and the perturbed equation

$$0 \in \nabla\theta^k(x) + N_S(x). \tag{4.2}$$

Then the conclusion follows from Theorem 3.3, Lemma 4.3 and [35, Lemma 4.2]. □

Remark 4.5. It is important to note that the constants $c(\epsilon)$ and $\beta(\epsilon)$ in Theorem 4.4 may be significantly different from their counterparts in Lemma 4.3. To establish a precise relationship of these constants, we need more information about the sensitivity of the true problem at the stationary points. One possibility is to look into the metric regularity condition for the set-valued mapping $G(x) := \nabla\theta(x) + N_S(x)$. If there exists a constant Q such that

$$d(x, T^*) \leq Q \cdot d(0, G(x))$$

for x close to T^*, then we can establish

$$d(x^N, T^*) \leq Q \cdot \|\nabla\theta^N(x) - \nabla\theta(x)\|.$$

We refer interested readers to [20, 29] for recent discussions on metric regularity. Under this circumstance, the constants $c(\epsilon)$ and $\beta(\epsilon)$ in Theorem 4.4 can be easily expressed in terms of their counterparts in Lemma 4.3. Similar to the discussions in [35], we may estimate the sample size. To this end, we assume that there exists a constant ϱ such that for all $x \in X$

$$M_x(t) \leq e^{\varrho^2 t^2/2}$$

for all $t \in R$. Then, following [35, Remark 3.2], we can obtain an estimation of the sample size, that is, for $\beta \in (0,1)$, $Prob\Big\{d(x^N, T^*) \geq \epsilon\Big\} \leq \beta$ when

$$N \geq \frac{O(1)\varrho^2}{\epsilon^2}\Big[n \ln\Big(\frac{O(1)DE[p(\xi)]}{\epsilon}\Big) + \ln(\frac{1}{\beta})\Big],$$

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

Acknowledgments

We would like to thank two anonymous referees and the chief editor Prof. J. Sun for constructive comments and helpful suggestions which helped to improve the presentation.

References

[1] R. Andreani, J. M. Martinez, A. Ramos and P. J. Silva, A cone-continuity constraint qualification and algorithmic consequences, *SIAM J. Optimiz.* 26 (2016) 96–110.

[2] R. P. Agdeppa, N. Yamashita and M. Fukushima, Convex expected residual models for stochastic affine variational inequality problems and its application to the traffic equilibrium problem, *Pacific J. Optim.* 6 (2010) 3–19.

[3] X. J. Chen and M. Fukushima, Expected residual minimization method for stochastic linear complementarity problems, *Math. Oper. Res.* 30 (2005) 1022–1038

[4] R. W. Cottle, F. Giannessi and J. L. Lions, *Variational Inequalities and Complementarity Problems: Theory and Applications*, John Wiley & Sons, Chichester, 1980.

[5] S. Dafermos, Exchange price equilibria and variational inequalities, *Math. Program.* 46 (1990) 391–402.

[6] L. Dai, C. H. Chen and J. R. Birge, Convergence properties of two stage stochastic programming, *J. Optimiz. Theory App.* 106 (2000) 489–509.

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

[8] G. Duvaut and J. L. Lions, *Les Inéquations en Méchanique et en Physique*, Dunod, Paris, 1972.

[9] F. Facchinei and J. S. Pang, *Finite-Dimensional Variational Inequalities and Complementary Problems*, Springer-Verlag, New York, 2003.

[10] S. C. Fang and E. L. Peterson, Generalized variational inequalities, *J. Optimiz. Theory App.* 38 (1982) 363–383.

[11] M. Fukushima, Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems, *Math. Program.* 53 (1992) 99–110.

[12] G. Gürkan, A. Y. Özge and S. M. Robinson, Sample-path solution of stochastic variational inequalities, *Math. Program.* 64 (1999) 313–333.

[13] P. T. Harker and J. S. Pang, Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications, *Math. Program.* 48 (1990) 161–220.

[14] T. Homem-de-Mello, On rates of convergence for stochastic optimization problems under non-independent and identically distributed sampling, *SIAM J. Optimiz.* 19 (2008) 524–551.

[15] H. Y. Jiang and H. F. Xu, Stochastic Approximation Approaches to the Stochastic Variational Inequality Problem, *IEEE T. Automat. Contr.* 53 (2008) 1462–1475.

[16] S. Karamardian, The nonlinear complementarity problem with applications, part 1, *J. Optimiz. Theory App.* 4 (1969) 87–98.

[17] G. H. Lin, X. J. Chen and M. Fukushima, New restricted NCP function and their applications to stochastic NCP and stochastic MPEC, *Optimization* 56 (2007) 641–753.

[18] M. J. Luo and G. H. Lin, Expected residual minimization method for stochastic variational inequality problems, *J. Optimiz. Theory App.* 140 (2009) 103–116.

[19] M. J. Luo and G. H. Lin, Convergence results of the ERM method for nonlinear stochastic variational inequality problems, *J. Optimiz. Theory App.* 142 (2009) 569–581.

[20] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation, II: Applications*, Grundlehren Math. Wiss. (Fundamental Principles of Mathematical Sciences), vol. 331, Springer, 2006.

[21] J. J. Moreau, Porpriétés des applications "prox", *Comptes Rendus de l'Académie des Sciences de Paris* 256 (1963) 1069–1071.

[22] J. Nash, Non-cooperative games, *Ann. Math.* 54 (1951) 286–295.

[23] H. Niederreiter, *Quasi-Monte Carlo Methods*, Wiley Online Library, 2010.

[24] M. Patriksson, *Nonlinear Programming and Variational Inequalities: A Unified Approach*, Kluwer Academic Publishers, Dordrecht, 1998.

[25] L. Q. Qi and Z. Wei, On the constant positive linear dependence condition and its application to SQP methods, *SIAM J. Optimiz.* 10 (2000) 963–981.

[26] S. M. Robinson, Generalized equations and their solutions, part II: Applications to nonlinear programming, *Math. Program. Study* 19 (1982) 200–221.

[27] R. T. Rockafellar, *Lagrange Multipliers and Variational Inequalities, in Variational Inequalities and Complementarity Problems: Theory and Applications*, John Wiley & Sons, Chichester, 1980,

[28] R. T. Rockafellar, *Convex Analysis*, Princeton Univercity Press, Princeton, 1970.

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

[30] A. Ruszczynski and A. Shapiro, *Stochastic Programming, Handbooks in Operations Research and Management Science*, Elsevier, Amsterdam, 2003.

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

[32] M. J. Smith, The existence, uniqueness and stability of traffic equilibria, *Transport. Sci.* 13B (1979) 295–304.

[33] M. V. Solodov, Merit functions and error bounds for generalized variational inequalities, *J.Math.Anal.Appl.* 287 (2003) 405–414.

[34] H. L. Sun and H. F. Xu, A note on uniform exponential convergence of sample average approximation of random functions, *J.Math.Anal.Appl.* 385 (2012) 698–708.

[35] H. F. Xu, Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming, *J.Math.Anal.Appl.* 368 (2010) 692–710.

[36] C. Zhang and X. J. Chen, Stochastic nonlinear complementarity problem and applications to traffic equilibrium under uncertainty, *J. Optimiz. Theory App.* 137 (2008) 277–295.

Manuscript received 21 June 2016
revised 27 October 2016
accepted for publication 19 May 2017

Lin Chen
College of Mathematics, Sichuan University
Chengdu 610064, China
E-mail address: chinaallenchen@126.com

Guihua Lin
School of Management, Shanghai University
Shanghai 200444, China
E-mail address: guihualin@shu.edu.cn

Xinmin Yang
College of Mathematics Science, Chongqing Normal University
Chongqing 401331, China
E-mail address: xmyang@cqnu.edu.cn