Expected Residual Minimization Formulation for a Class of Stochastic Vector Variational Inequalities

Zhao, Yong; ZHANG, Jin; Yang, Xinmin; Lin, Gui Hua

Published in:
Journal of Optimization Theory and Applications

DOI:
10.1007/s10957-016-0939-5

Published: 01/11/2017

Noname manuscript No.
(will be inserted by the editor)

Expected Residual Minimization Formulation for a Class of Stochastic Vector Variational Inequalities

Yong Zhao · Jin Zhang · Xinmin Yang · Gui-Hua Lin

Received: date / Accepted: date

Abstract This paper considers a class of vector variational inequalities. First, we present an equivalent formulation, which is a scalar variational inequality, for the deterministic vector variational inequality. We then concentrate on the stochastic circumstance. By noting that the stochastic vector variational inequality may not have a solution feasible for all realizations of the random variable in general, for tractability, we employ the expected residual minimization approach, which aims at minimizing the expected residual of the so-called regularized gap function. We investigate the properties of the expected residual minimization problem and, furthermore, we propose a sample average approximation method for solving the expected residual minimization problem. Comprehensive convergence analysis for the approximation approach is established as well.

Keywords Stochastic vector variational inequalities · Expected residual minimization formulation · Sample average approximation

Mathematics Subject Classification (2000) 90C33, 90C15

1 Introduction

The vector variational inequality (VVI), which is a generalization of the classical variational inequality to the vector case by virtue of multiobjective optimization consideration, was introduced first by

Communicated by Guang-ya Chen.

Yong Zhao(✉)
Department of Mathematics, Sichuan University, Chengdu 610065, China
E-mail: zhaoyongty@126.com

Jin Zhang
Department of Mathematics, Hong Kong Baptist University, Hongkong, China
E-mail: zhangjin@hkbu.edu.hk

Xinmin Yang
College of Mathematics Science, Chongqing Normal University, Chongqing 401131, China
E-mail: xmyang@cqnu.edu.cn

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

Giannessi [1] in finite-dimensional Euclidean spaces. It has numerous applications in optimization, operations research, economic equilibrium, traffic networks, and so on. The monograph [2] is a good source, which contains a large number of papers on vector variational inequalities. In the last decades, there has been a considerable research effort in the study of vector variational inequalities. Chen et al. [3] studied the existence of solutions of vector variational inequalities and the relations between vector variational inequalities and vector optimization problems. Yang and Yao [4] established necessary and sufficient conditions for the existence of solutions for vector variational inequalities with set-valued mappings. In [6], Yang et al. established some relations between a Minty vector variational inequality and a vector optimization problem. Lee et al. [11] showed that vector variational inequality can be an efficient tool for studying vector optimization problems. It is well-known that, by utilizing the so-called gap functions or merit functions, the classical variational inequalities can be transformed into optimization problems and, along this way, some effective algorithms have been developed. This approach has also been extended to deal with VVI, see [4, 5, 7–10, 12]. Among these papers, a scalar-valued gap function is developed for vector variational inequality in [4, 9, 10, 12], and another set-valued gap function is also investigated in [5, 7, 8].

Note that most of the existing methods for VVI are based on the well-known scalarization techniques (see, e.g., [3, 9–13] and the references therein). However, as in the multiobjective optimization theory, these scalarization formulations are generally not equivalent to the original VVI. In this paper, by treating the fixed scalarization parameters as variables, we transform the VVI into an equivalent scalar variational inequality with some simple constraints. This may be regarded as the first contribution of the paper.

The main focus of this paper is actually on the stochastic version of VVI. Since some elements may involve uncertain data in many practical problems, it is meaningful to study stochastic VVI (SVVI). Actually, stochastic versions of the classical variational inequality have been receiving much attention in the recent literature. Two deterministic formulations have been studied for the stochastic variational inequality (SVI). Expected value formulation [14, 16, 17, 23] can be viewed as a deterministic VI defined by the expected value of the function. The expected residual minimization formulation [15, 18, 20–22, 24–26] consists in finding a robust solution by minimizing the expected value of a residual function of the SVI at almost every scenario. In this paper, we first reformulate the stochastic vector variational inequality as an equivalent stochastic scalar variational inequality. Then, motivated by the works [15, 24] on SVI, we present a deterministic formulation, called the expected residual minimization (ERM) formulation, based on the qeuivalent stochastic scalar variational inequality. Some properties of the ERM formulation are discussed. Furthermore, we propose a sample average approximation method for solving the ERM problem, and the limiting behaviors of optimal solutions and stationary points are investigated as well.

2 Equivalent Formulation for VVI

Throughout the paper, the transpose of a vector or matrix x is denoted by x^T, and the Euclidean norm of a vector or matrix is denoted by $\|\cdot\|$. Given a set $A \subset \mathbb{R}^n$, let $\mathrm{conv}(A)$ denotes the convex hull of A. In addition, we denote by

$$\Lambda := \{\lambda \in \mathbb{R}^m : \lambda_j \geq 0, \sum_{j=1}^{m} \lambda_j = 1\}.$$

We consider the VVI of finding $x^* \in K$ such that

$$\big((y-x^*)^T F_1(x^*), \cdots, (y-x^*)^T F_m(x^*)\big) \notin -\mathrm{int}\mathbb{R}^m_+, \quad \forall y \in K, \tag{1}$$

where K is a nonempty, closed and convex set in $\mathbb{R}^n$ and $F_j : \mathbb{R}^n \to \mathbb{R}^n$ $(j = 1, 2, \cdots, m)$ are vector-valued functions. For simplicity, we let $F := (F_1, \cdots, F_m)$, and the set of solutions of the above VVI is denoted by $\mathrm{Sol}(F, K)$.

If we set for each $j = 1, 2, \cdots, m$, $F_j = \nabla f_j$ where each f_j is a differentiable convex function, then a solution of the VVI (1) is a weak Pareto minimum of the convex vector optimization problem

$$\text{(VOP)} \qquad \min f(x) = \big(f_1(x), \cdots, f_m(x)\big) \quad \text{s.t. } x \in K$$

and, conversely, any weak Pareto minimum of (VOP) is also a solution of the VVI (1) with $F_j = \nabla f_j$ for each $j = 1, 2, \cdots, m$. See [2, 3, 6, 11] for further relations between VVI and vector optimization problems under some convexity or generalized convexity assumptions.

One popular approach to deal with the VVI (1) is to transform it into the following standard variational inequality by choosing some weight vector $\lambda \geq 0$ with $\sum_{j=1}^{m} \lambda_j = 1$: Find $x^* \in K$ such that

$$(y-x^*)^T \sum_{j=1}^{m} \lambda_j F_j(x^*) \geq 0, \quad \forall y \in K.$$

This problem is obviously not equivalent to (1). Therefore, as in the multiobjective optimization theory, the above weighted approach may lead us to miss some solutions.

We next present an equivalent formulation for the VVI (1). Note that (1) can be rewritten as follows: Find $x^* \in K$ such that

$$\max_{1 \leq j \leq m} (y-x^*)^T F_j(x^*) \geq 0, \quad \forall y \in K.$$

We further observe that $y = x^*$ is right the optimal solution to the problem

$$\begin{aligned} \min \quad & V(y) := \max_{1 \leq j \leq m} (y-x^*)^T F_j(x^*) \\ \text{s.t.} \quad & y \in K. \end{aligned}$$

By the Danskin's theorem, there holds the optimality conditions

$$0 \in \partial V(x^*) + N_K(x^*) \subset \mathrm{conv}\{F_j(x^*) : j = 1, \cdots, m\} + N_K(x^*),$$

which means that there exists $\lambda^* \geq 0$ with $\sum_{j=1}^{m} \lambda_j^* = 1$ such that

$$-\sum_{j=1}^{m} \lambda_j^* F_j(x^*) \in N_K(x^*),$$

or equivalently,

$$(y-x^*)^T \sum_{j=1}^{m} \lambda_j^* F_j(x^*) \geq 0, \quad \forall y \in K.$$

In consequence, we have the following result immediately.

Theorem 2.1 *The VVI (1) is equivalent to the following variational inequalities with simple constraints:*

$$\text{Find } (x^*, \lambda^*) \in K \times \Lambda \text{ such that } (y - x^*)^T \sum_{j=1}^{m} \lambda_j^* F_j(x^*) \geq 0, \quad \forall y \in K. \tag{2}$$

Proof. It is easy to obtain the necessary part from the above analysis. We next show the sufficiency. Suppose that there exists $(x^*, \lambda^*) \in K \times \Lambda$ such that

$$(y - x^*)^T \sum_{j=1}^{m} \lambda_j^* F_j(x^*) \geq 0, \quad \forall y \in K.$$

Since $\lambda^* \in \Lambda$, for any fixed $y \in K$, there exists $j \in \{1, \cdots, m\}$ such that

$$(y - x^*)^T F_j(x^*) \geq 0.$$

Thus, we have

$$\big((y - x^*)^T F_1(x^*), \cdots, (y - x^*)^T F_m(x^*)\big) \notin -\mathrm{int}\mathbb{R}_+^m, \quad \forall y \in K,$$

which implies that x^* is a solution of VVI (1). The proof is complete. □

Following [27], we introduce a regularized gap function as follows:

$$g(x, \lambda) := \max_{y \in K} \{(x - y)^T \sum_{j=1}^{m} \lambda_j F_j(x) - \frac{\alpha}{2} \|x - y\|_G^2\}, \tag{3}$$

where $\alpha > 0$ is a given parameter, G is an $n \times n$ symmetric positive-definite matrix, and $\|\cdot\|_G$ stands for the G-norm, defined by $\|x\|_G = \sqrt{x^T G x}$ for each $x \in \mathbb{R}^n$. It follows from [27] that, for any $(x, \lambda) \in K \times \Lambda$,

$$g(x, \lambda) = (x - H(x, \lambda))^T \sum_{j=1}^{m} \lambda_j F_j(x) - \frac{\alpha}{2} \|x - H(x, \lambda)\|_G^2,$$

where

$$H(x, \lambda) := \mathrm{Proj}_{K,G}\big(x - \alpha^{-1} G^{-1} \sum_{j=1}^{m} \lambda_j F_j(x)\big).$$

It is easy to show that

- $g(x, \lambda) \geq 0$ for each $(x, \lambda) \in K \times \Lambda$;
- $g(x^*, \lambda^*) = 0$ with $(x^*, \lambda^*) \in K \times \Lambda$ if and only if (x^*, λ^*) solves (2).

Hence, (2) can be transformed into the minimization problem

$$\min \ g(x, \lambda) \ \text{ s.t. } x \in K, \lambda \in \Lambda. \tag{4}$$

Theorem 2.2 *Suppose that each function F_j $(j = 1, \cdots, m)$ is continuously differentiable and $\nabla F_j(x)$ is positive-definite for each $x \in K$. If $z^* = (x^*, \lambda^*) \in K \times \Lambda$ is a stationary point of problem (4), that is,*

$$(z - z^*)^T \nabla_{(x,\lambda)} g(x^*, \lambda^*) \geq 0, \quad \forall z = (x, \lambda) \in K \times \Lambda.$$

Then, (x^*, λ^*) solves (2), and hence it solves the VVI (1).

Proof. Note that

$$\nabla_{(x,\lambda)} g(x,\lambda) = \begin{pmatrix} \sum_{j=1}^m \lambda_j F_j(x) - \big(\sum_{j=1}^m \lambda_j \nabla_x F_j(x) - \alpha G\big)(H(x,\lambda) - x) \\ F(x)^T (x - H(x,\lambda)) \end{pmatrix}.$$

Since (x^*, λ^*) is a stationary point, for any $z = (x,\lambda) \in K \times \Lambda$, we have

$$\Big(\sum_{j=1}^m \lambda_j^* F_j(x^*) - \big(\sum_{j=1}^m \lambda_j^* \nabla_x F_j(x^*) - \alpha G\big)(H(x^*,\lambda^*) - x^*)\Big)^T (x - x^*) \\ + \big(F(x^*)^T (x^* - H(x^*,\lambda^*))\big)^T (\lambda - \lambda^*) \geq 0.$$

In a similar way to Theorem 3.3 of [27], we can show the conclusion by letting $z^* = (H(x^*,\lambda^*), \lambda^*)$ in the above inequality. □

3 SVVI and its Deterministic Reformulation

Consider the SVVI of finding $x^* \in K$ such that

$$\big((y - x^*)^T F_1(x^*, \xi(\omega)), \cdots, (y - x^*)^T F_m(x^*, \xi(\omega))\big) \notin -\mathrm{int}\mathbb{R}^m_+, \quad \forall y \in K, \text{ a.e. } \xi \in \varXi, \tag{5}$$

where K is a nonempty, closed and convex set in $\mathbb{R}^n$, $F_j : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^n$ $(j = 1, \cdots, m)$ are vector-valued functions, $\xi : \varOmega \to \varXi$ is a vector of random variables defined on the probability space $(\varOmega, \mathcal{F}, \mathbb{P})$ with support set $\varXi \subset \mathbb{R}^r$, and 'a.e.' is the abbreviation for "almost every" under the given probability measure. To ease the notation, we will use ξ to denote either the random vector $\xi(\omega)$ or an element of $\mathbb{R}^r$ depending on the context.

Similar to the previous section, we can rewrite the above SVVI as the following equivalent formulation: Find $(x^*, \lambda^*) \in K \times \Lambda$ such that

$$(y - x^*)^T \sum_{j=1}^m \lambda_j^* F_j(x^*, \xi) \geq 0, \quad \forall y \in K, \text{ a.e. } \xi \in \varXi. \tag{6}$$

In general, we cannot expect that problem (6) has a common solution for almost every $\xi \in \varXi$. Therefore, in order to get a reasonable resolution, an appropriate deterministic reformulation for problem (6) becomes an important issue in our study. For the remainder of the paper, we choose a positive parameter α and an $n \times n$ symmetric positive-definite matrix G. First of all, we define the regularized gap function $g : \mathbb{R}^n \times \Lambda \times \varXi \to [0, \infty[$ for the SVVI as follows:

$$g(x,\lambda,\xi) := \max_{y \in K} \{(x - y)^T \sum_{j=1}^m \lambda_j F_j(x,\xi) - \frac{\alpha}{2}\|x - y\|_G^2\}. \tag{7}$$

Then, for any $(x,\lambda) \in K \times \Lambda$ and any $\xi \in \varXi$, we have

$$g(x,\lambda,\xi) = (x - H(x,\lambda,\xi))^T \sum_{j=1}^m \lambda_j F_j(x,\xi) - \frac{\alpha}{2}\|x - H(x,\lambda,\xi)\|_G^2, \tag{8}$$

where

$$H(x,\lambda,\xi) := \mathrm{Proj}_{K,G}\big(x - \alpha^{-1} G^{-1} \sum_{j=1}^m \lambda_j F_j(x,\xi)\big).$$

Motivated by the works [15, 24] on SVI, we propose the expected residual minimization (ERM) formulation for the above SVVI as follows:

$$\min\ \theta(x,\lambda) := \mathbb{E}[g(x,\lambda,\xi)] \ \text{ s.t. } \ x\in K, \lambda\in\Lambda, \tag{9}$$

where $\mathbb{E}$ stands for the mathematical expectation with respect to the probability distribution of $\xi \in \Xi$.

Next, we make some assumptions that will be used later on:

(A1) For every $\xi \in \Xi$, $F_j(x,\xi)$ $(j=1,\cdots,m)$ are continuously differentiable with respect to $x\in K$.

(A2) There exists an integrable function $\kappa(\xi)$ such that

$$\mathbb{E}[\kappa^2(\xi)] < +\infty \ \text{ and } \ \sum_{j=1}^m \|F_j(x,\xi)\| + \sum_{j=1}^m \|\nabla_x F_j(x,\xi)\|_F \le \kappa(\xi)$$

hold for any $x\in K$ and almost every $\xi\in\Xi$. Here, $\|\cdot\|_F$ denotes the Frobenius norm of a matrix.

(A3) Each function $F_j(x,\xi)$ $(j=1,\cdots,m)$ is Lipschitz continuous in x over K with Lipschitz constant $C_j(\xi)$ satisfying $\mathbb{E}[C_j^2(\xi)]<+\infty$, that is,

$$\|F_j(y,\xi)-F_j(x,\xi)\| \le C_j(\xi)\|y-x\|, \quad \forall x,y\in K.$$

Moreover, note that, for any $x\in\mathbb{R}^n$, we have

$$\sqrt{\lambda_{\min}}\|x\| \le \|x\|_G \le \sqrt{\lambda_{\max}}\|x\|, \tag{10}$$

where $\lambda_{\min}$ and $\lambda_{\max}$ indicate the smallest and largest eigenvalues of G, respectively.

Theorem 3.1 *Suppose that conditions (A1) and (A2) are satisfied. Then, both g and θ are continuously differentiable with respect to (x,λ) and, for any $(x,\lambda)\in K\times\Lambda$, we have*

$$\nabla\theta(x,\lambda) = \mathbb{E}[\nabla_{(x,\lambda)} g(x,\lambda,\xi)].$$

Proof. Since $F_j(x,\xi)$ $(j=1,\cdots,m)$ are continuously differentiable with respect to x, it is easy to know that $g(\cdot,\cdot,\xi)$ is continuously differentiable on $K\times\Lambda$ for any $\xi\in\Xi$ and

$$\nabla_{(x,\lambda)} g(x,\lambda,\xi) = \begin{pmatrix} \sum_{j=1}^m \lambda_j F_j(x,\xi) - \big(\sum_{j=1}^m \lambda_j \nabla_x F_j(x,\xi) - \alpha G\big)(H(x,\lambda,\xi)-x) \\ (x-H(x,\lambda,\xi))^T F_1(x,\xi) \\ \vdots \\ (x-H(x,\lambda,\xi))^T F_m(x,\xi) \end{pmatrix}. \tag{11}$$

On the other hand, since $g(x,\lambda,\xi)\ge 0$ for any $(x,\lambda)\in K\times\Lambda$ and $\xi\in\Xi$, we have from (8) and (10) that

$$\begin{aligned} \frac{\alpha}{2}\|x-H(x,\lambda,\xi)\|_G^2 &\le (x-H(x,\lambda,\xi))^T \sum_{j=1}^m \lambda_j F_j(x,\xi) \\ &\le \frac{1}{\sqrt{\lambda_{\min}}}\|x-H(x,\lambda,\xi)\|_G \sum_{j=1}^m \lambda_j \|F_j(x,\xi)\| \\ &\le \frac{1}{\sqrt{\lambda_{\min}}}\|x-H(x,\lambda,\xi)\|_G \sum_{j=1}^m \|F_j(x,\xi)\|, \end{aligned}$$

which implies

$$\|x-H(x,\lambda,\xi)\|_G \le \frac{2}{\alpha\sqrt{\lambda_{\min}}}\sum_{j=1}^m \|F_j(x,\xi)\|$$

and hence

$$\|x - H(x,\lambda,\xi)\| \le \frac{2}{\alpha\lambda_{\min}} \sum_{j=1}^{m} \|F_j(x,\xi)\|.$$

Therefore, we have

$$\begin{aligned}
\|\nabla_{(x,\lambda)} g(x,\lambda,\xi)\| \le & \|\sum_{j=1}^{m} \lambda_j F_j(x,\xi) - (\sum_{j=1}^{m} \lambda_j \nabla_x F_j(x,\xi) - \alpha G)(H(x,\lambda,\xi) - x)\| \\
& + \sum_{j=1}^{m} \|F_j(x,\xi)\| \|x - H(x,\lambda,\xi)\| \\
\le & \sum_{j=1}^{m} \|F_j(x,\xi)\| + \big(\sum_{j=1}^{m} \|\nabla_x F_j(x,\xi)\| + \alpha\|G\|\big) \frac{2}{\alpha\lambda_{\min}} \sum_{j=1}^{m} \|F_j(x,\xi)\| \\
& + \sum_{j=1}^{m} \|F_j(x,\xi)\| \frac{2}{\alpha\lambda_{\min}} \sum_{j=1}^{m} \|F_j(x,\xi)\| \\
\le & \Big(1 + \frac{2}{\alpha\lambda_{\min}} \big(\sum_{j=1}^{m} \|\nabla_x F_j(x,\xi)\| + \alpha\|G\| + \sum_{j=1}^{m} \|F_j(x,\xi)\|\big)\Big) \sum_{j=1}^{m} \|F_j(x,\xi)\|.
\end{aligned}$$

By assumption (A2) and Theorem 16.8 of [31], the function θ is continuously differentiable and

$$\nabla\theta(x,\lambda) = \mathbb{E}[\nabla_{(x,\lambda)} g(x,\lambda,\xi)].$$

This completes the proof. □

Next, we consider the boundedness of the level set of problem (9) defined by

$$L_\theta(c) := \{(x,\lambda) \in K \times \Lambda : \theta(x,\lambda) \le c\},$$

where c is a non-negative number.

Definition 3.1 [16] Let $F : K \to \mathbb{R}^n$.
(i) F is said to be monotone on K iff

$$(F(y) - F(x))^T (y - x) \ge 0, \quad \forall x, y \in K;$$

(ii) F is said to be strongly monotone on K with modulus $\sigma > 0$ iff for any $x, y \in K$,

$$(F(y) - F(x))^T (y - x) \ge \sigma \|y - x\|_G^2.$$

Note that, in the rest of this section, $\mathbb{P}(V)$ denotes the probability of an event V.

Theorem 3.2 *Suppose conditions (A1) and (A2) hold and each function $F_j : K \times \Xi \to \mathbb{R}^n$ $(j = 1, \cdots, m)$ is monotone for almost every $\xi \in \Xi$. Further assume that F_j $(j = 1, \cdots, m)$ are uniformly strongly monotone on K with modulus $\mu_j > 0$ over $V_j \subset \Xi$ with $\mathbb{P}(V_j) > 0$. Let $\mu = \min_{1\le j\le m} \mu_j$, $\nu = \min_{1\le j\le m} \mathbb{P}(V_j)$ and choose $\alpha \in]0, 2\nu\mu[$. Then, the level set $L_\theta(c)$ is bounded for any $c \ge 0$.*

Proof. Suppose that there exists a non-negative number $\bar{c}$ such that $L_\theta(\bar{c})$ is unbounded, which implies that there exists a sequence $\{(x^k, \lambda^k)\} \subset L_\theta(\bar{c})$ such that

$$\lim_{k\to\infty} \|(x^k, \lambda^k)\| = +\infty.$$

Since Λ is a compact set and $\{\lambda^k\} \subset \Lambda$, we have

$$\lim_{k\to\infty} \|x^k\| = +\infty.$$

Using the definitions of monotonicity and strong monotonicity, for all $x, y \in K$, we have

$$\begin{aligned}
&\Big(\sum_{j=1}^m \lambda_j^k \mathbb{E}[F_j(y,\xi)] - \sum_{j=1}^m \lambda_j^k \mathbb{E}[F_j(x,\xi)]\Big)^T (y-x) \\
&= \int_{V_1} \lambda_1^k (F_1(y,\xi) - F_1(x,\xi))^T (y-x)\mathbb{P}(d\xi) + \int_{\Xi\setminus V_1} \lambda_1^k (F_1(y,\xi) - F_1(x,\xi))^T (y-x)\mathbb{P}(d\xi) \\
&+ \cdots \\
&+ \int_{V_m} \lambda_m^k (F_m(y,\xi) - F_m(x,\xi))^T (y-x)\mathbb{P}(d\xi) + \int_{\Xi\setminus V_m} \lambda_m^k (F_m(y,\xi) - F_m(x,\xi))^T (y-x)\mathbb{P}(d\xi) \\
&\geq \nu\mu \|y-x\|_G^2.
\end{aligned}$$

Therefore, $\sum_{j=1}^m \lambda_j^k \mathbb{E}[F_j]$ is strongly monotone with modulus $\nu\mu > 0$, and hence there is a unique $\bar{x} \in K$ such that

$$(y-\bar{x})^T \sum_{j=1}^m \lambda_j^k \mathbb{E}[F_j(\bar{x},\xi)] \geq 0, \quad \forall y \in K.$$

As a result, we have

$$\begin{aligned}
\bar{c} \geq \theta(x^k,\lambda^k) &= \mathbb{E}\big[\max_{y\in K}\{(x^k-y)^T \sum_{j=1}^m \lambda_j^k F_j(x^k,\xi) - \frac{\alpha}{2}\|x^k-y\|_G^2\}\big] \\
&\geq \max_{y\in K}\{(x^k-y)^T \sum_{j=1}^m \lambda_j^k \mathbb{E}[F_j(x^k,\xi)] - \frac{\alpha}{2}\|x^k-y\|_G^2\} \\
&\geq (x^k-\bar{x})^T \sum_{j=1}^m \lambda_j^k \mathbb{E}[F_j(x^k,\xi)] - \frac{\alpha}{2}\|x^k-\bar{x}\|_G^2 \\
&\geq (x^k-\bar{x})^T \sum_{j=1}^m \lambda_j^k \mathbb{E}[F_j(\bar{x},\xi)] + (\nu\mu - \frac{\alpha}{2})\|x^k-\bar{x}\|_G^2 \\
&\xrightarrow{k\to\infty} +\infty,
\end{aligned}$$

where the second inequality follows from Jensen's inequality and the fourth inequality follows from the strong monotonicity of $\sum_{j=1}^m \lambda_j^k \mathbb{E}[F_j]$. This is a contradiction and hence $L_\theta(c)$ is bounded for any $c \geq 0$. □

Note that problem (9) has at least one solution if it has a nonempty and bounded level set (In fact, for any $c > 0$ large enough, the level set $L_\theta(c)$ is nonempty).

Except the ERM formulation considered in this paper, another deterministic formulation, called expected valued (EV) formulation, is also known in the literature on SVI [17]. The following result, which provides a global error bound for EV formulation, generalizes Theorem 2.3 of [24].

Theorem 3.3 *Suppose conditions (A1) and (A2) hold and each function $F_j : K \times \Xi \to \mathbb{R}^n (j = 1, \cdots, m)$ is monotone for almost every $\xi \in \Xi$. Further assume that F_j $(j = 1, \cdots, m)$ are uniformly strongly monotone on K with modulus $\mu_j > 0$ over $V_j \subset \Xi$ with $\mathbb{P}(V_j) > 0$. Let $\mu = \min_{1\leq j\leq m} \mu_j, \nu = \min_{1\leq j\leq m} \mathbb{P}(V_j)$ and choose $\alpha \in]0, 2\nu\mu[$. Then, for any $\lambda \in \Lambda$, we have*

$$dist(x, Sol(\mathbb{E}[F(x,\xi)], K)) \leq \sqrt{\frac{\theta(x,\lambda)}{\lambda_{\min}(\nu\mu - \frac{\alpha}{2})}}.$$

Proof. Note that from Theorem 3.2, for any given $\lambda \in \Lambda$, the function $\sum_{j=1}^{m} \lambda_j \mathbb{E}[F_j(\cdot,\xi)]$ is strongly monotone with modulus $\nu\mu$ by the assumptions. Therefore, there exists $x^* \in K$ such that

$$\big(x - x^*\big)^T \sum_{j=1}^{m} \lambda_j \mathbb{E}[F_j(x^*,\xi)] \geq 0, \ \forall x \in K. \tag{12}$$

It is obvious that $x^* \in \mathrm{Sol}(\mathbb{E}[F(x,\xi)], K)$. By the strong monotonicity of $\sum_{j=1}^{m} \lambda_j \mathbb{E}[F_j(\cdot,\xi)]$, we have

$$\Big(\sum_{j=1}^{m} \lambda_j \mathbb{E}[F_j(x,\xi)] - \sum_{j=1}^{m} \lambda_j \mathbb{E}[F_j(x^*,\xi)]\Big)^T (x - x^*) \geq \nu\mu \|x - x^*\|_G^2, \ \forall x \in K. \tag{13}$$

It follows from (12) and (13) that

$$\big(x - x^*\big)^T \sum_{j=1}^{m} \lambda_j \mathbb{E}[F_j(x,\xi)] \geq \nu\mu \|x - x^*\|_G^2, \ \forall x \in K.$$

On the other hand, we have

$$\begin{aligned}
\theta(x,\lambda) &= \mathbb{E}\big[\max_{y \in K}\{(x-y)^T \sum_{j=1}^{m} \lambda_j F_j(x,\xi) - \frac{\alpha}{2}\|x-y\|_G^2\}\big] \\
&\geq \max_{y \in K}\{(x-y)^T \sum_{j=1}^{m} \lambda_j \mathbb{E}[F_j(x,\xi)] - \frac{\alpha}{2}\|x-y\|_G^2\} \\
&\geq (x-x^*)^T \sum_{j=1}^{m} \lambda_j \mathbb{E}[F_j(x,\xi)] - \frac{\alpha}{2}\|x-x^*\|_G^2 \\
&\geq (\nu\mu - \frac{\alpha}{2})\lambda_{\min}\|x-x^*\|^2 \\
&\geq (\nu\mu - \frac{\alpha}{2})\lambda_{\min}\mathrm{dist}^2(x, \mathrm{Sol}(\mathbb{E}[F(x,\xi)], K)).
\end{aligned}$$

This completes the proof. □

It is not difficult to notice that the solution given by the ERM model may not be the best, or may be even infeasible for each individual event. In practice, we should take risk into account to make a priori decision in many cases. Naturally, it is necessary to know how good or how bad our decisions can be. We next give a result about the robustness of solutions of the ERM model for the SVVI (5).

Theorem 3.4 *Suppose conditions (A1) and (A2) hold, and $\Xi = \{\xi_1, \xi_2, \cdots, \xi_N\}$ is a finite set with the probability p_i of ξ_i to be positive for each i. Assume that each function $F_j(x,\xi)$ $(j = 1, \cdots, m)$ is strongly monotone with modulus $\mu_j > 0$ for all $\xi \in \Xi$. Let $\mu = \min_{1 \leq j \leq m} \mu_j$ and choose $\alpha \in]0, 2\mu[$. Then, for any $\lambda \in \Lambda$, we have*

$$\mathbb{E}[dist(x, Sol(F(x,\xi), K))] \leq \sqrt{\frac{\theta(x,\lambda)}{\lambda_{\min}(\mu - \frac{\alpha}{2})}}.$$

Proof. Let $\lambda \in \Lambda$ be given. Then, for any $\xi_i \in \Xi$, since the function $\sum_{j=1}^{m} \lambda_j F_j(\cdot,\xi_i)$ is strongly monotone, there exists a unique $x^*(\xi_i) \in K$ such that

$$\big(x - x^*(\xi_i)\big)^T \sum_{j=1}^{m} \lambda_j F_j(x^*(\xi_i), \xi_i) \geq 0, \ \forall x \in K. \tag{14}$$

It is easy to see from the proof of Theorem 3.3 that

$$\|x - x^*(\xi_i)\| \leq \sqrt{\frac{g(x,\lambda,\xi_i)}{\lambda_{\min}(\mu - \frac{\alpha}{2})}}. \tag{15}$$

Note that, by the assumptions, $\mathrm{Sol}(F(x,\xi),K)$ is nonempty for each $\xi \in \Xi$. Hence, we have

$$\begin{aligned}\mathbb{E}\big[\mathrm{dist}(x, \mathrm{Sol}(F(x,\xi),K))\big] &\le \sum_{i=1}^{N} p_i \cdot \|x - x^*(\xi_i)\| \\ &\le \sum_{i=1}^{N} p_i \sqrt{\frac{g(x,\lambda,\xi_i)}{\lambda_{\min}(\mu - \frac{\alpha}{2})}} \\ &\le \sqrt{\frac{\theta(x,\lambda)}{\lambda_{\min}(\mu - \frac{\alpha}{2})}},\end{aligned}$$

where the last inequality follows from the Jensen's inequality. This completes the proof. □

Theorem 3.4 reveals that the expected distance to the solution set $\mathrm{Sol}(F(x,\xi),K)$ is likely to be small at the solution of (9). Therefore, we may expect that a solution of the ERM formulation (9) has a minimum sensitivity with respect to random parameter variations in SVVI (5). In this case, solutions of (9) can be regarded as robust solutions for SVVI (5).

4 Convergence Analysis

Since problem (9) involves the mathematical expectation in the objective function and the distribution of the random variables may be unknown in practice, we apply the well-known sample average approximation techniques to deal with the expected value.

In general, for an integrable function $\phi : \Xi \to \mathbb{R}$, we estimate the expected value $\mathbb{E}[\phi(\xi)]$ with the sample average $\frac{1}{N}\sum_{i=1}^{N} \phi(\xi_i)$, where $\xi_1, \cdots, \xi_N$ be an independent and identically distributed sampling of ξ. The strong law of large numbers guarantees that this procedure converges with probability one (abbreviated by "w.p.1" below), i.e.,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \phi(\xi_i) = \mathbb{E}[\phi(\xi)], \quad \text{w.p.1.} \tag{16}$$

Thus, let $\xi_1, \cdots, \xi_N$ be an independent and identically distributed sampling of ξ, we can get the following sample average approximation (SAA) problem of (9):

$$\min\ \theta_N(x,\lambda) := \frac{1}{N}\sum_{i=1}^{N} g(x,\lambda,\xi_i), \text{ s.t. } x \in K,\ \lambda \in \Lambda. \tag{17}$$

In what follows, we investigate the limiting behavior of this approximation problem.

4.1 Convergence and Exponential Convergence Rate of Optimal Solutions

We first study the limiting behavior of optimal solutions. From now on, we denote by S^* and S_N^* the sets of optimal solutions of problems (9) and (17), respectively.

Theorem 4.1 *Suppose conditions (A1) and (A2) hold, and each function $F_j(x,\xi)(j = 1, \cdots, m)$ is strongly monotone on K with modulus $\mu_j > 0$ for all ξ_i, $i = 1, \cdots, N$. Let $\mu = \min_{1\le j\le m} \mu_j$ and choose $\alpha \in]0, 2\mu[$. Let $(x^N, \lambda^N) \in S_N^*$ for each sufficiently large N. Then, every accumulation point of $\{(x^N, \lambda^N)\}$ is contained in S^*.*

Proof. For all large N, we can show that S_N^* is nonempty and bounded in a similar manner to Theorem 3.2. Let (x^*, λ^*) be an accumulation point of $\{(x^N, \lambda^N)\}$. Without any loss of generality, we assume that $\{(x^N, \lambda^N)\}$ itself converges to (x^*, λ^*). It is obvious that $(x^*, \lambda^*) \in K \times \Lambda$. It follows from the mean-value theorem, the convexity of K, Λ and (11) that, for each (x^N, λ^N) and each ξ_i, there exist $y^{Ni} = \alpha_{Ni} x^N + (1-\alpha_{Ni})x^* \in K$, $\lambda^{Ni} = \alpha_{Ni}\lambda^N + (1-\alpha_{Ni})\lambda^* \in \Lambda$ with $\alpha_{Ni} \in [0,1]$ such that

$$\begin{aligned}
|g(x^N, \lambda^N, \xi_i) - g(x^*, \lambda^*, \xi_i)| &= |\nabla_x g(y^{Ni}, \lambda^{Ni}, \xi_i)^T (x^N - x^*) + \nabla_\lambda g(y^{Ni}, \lambda^{Ni}, \xi_i)^T (\lambda^N - \lambda^*)| \\
&\le \big(\sum_{j=1}^m \|F_j(y^{Ni}, \xi_i)\|\big)\Big(1 + \frac{2}{\alpha\lambda_{\min}}\big(\sum_{j=1}^m \|\nabla_x F_j(y^{Ni}, \xi_i)\| + \alpha\|G\|\big)\Big) \\
&\quad \times \|x^N - x^*\| + \frac{2}{\alpha\lambda_{\min}}\big(\sum_{j=1}^m \|F_j(y^{Ni}, \xi_i)\|\big)^2 \|\lambda^N - \lambda^*\|.
\end{aligned}$$

It then follows that

$$\begin{aligned}
|\theta_N(x^N, \lambda^N) - \theta_N(x^*, \lambda^*)| &\le \frac{1}{N}\sum_{i=1}^N |g(x^N, \lambda^N, \xi_i) - g(x^*, \lambda^*, \xi_i)| \\
&\le \frac{1}{N}\sum_{i=1}^N \big(\sum_{j=1}^m \|F_j(y^{Ni}, \xi_i)\|\big)\Big(1 + \frac{2}{\alpha\lambda_{\min}}\big(\sum_{j=1}^m \|\nabla_x F_j(y^{Ni}, \xi_i)\| + \alpha\|G\|\big)\Big) \\
&\quad \times \|x^N - x^*\| + \frac{2}{\alpha\lambda_{\min}}\frac{1}{N}\sum_{i=1}^N \big(\sum_{j=1}^m \|F_j(y^{Ni}, \xi_i)\|\big)^2 \|\lambda^N - \lambda^*\|.
\end{aligned}$$

Since $\{(x^N, \lambda^N)\}$ converges to (x^*, λ^*), it follows from the assumption (A2) that each term in the last inequality converges to zero. Therefore, we have

$$\lim_{N\to\infty} (\theta_N(x^N, \lambda^N) - \theta_N(x^*, \lambda^*)) = 0. \tag{18}$$

Since

$$|\theta_N(x^N, \lambda^N) - \theta(x^*, \lambda^*)| \le |\theta_N(x^N, \lambda^N) - \theta_N(x^*, \lambda^*)| + |\theta_N(x^*, \lambda^*) - \theta(x^*, \lambda^*)|,$$

by (16) and (18), we have

$$\lim_{N\to\infty} \theta_N(x^N, \lambda^N) = \theta(x^*, \lambda^*) \quad \text{w.p.1.}$$

Note that $(x^N, \lambda^N) \in S_N^*$, which implies that

$$\theta_N(x^N, \lambda^N) \le \theta_N(x, \lambda), \quad \forall (x, \lambda) \in K \times \Lambda.$$

Letting $N \to \infty$ in the above inequality, we have

$$\theta(x^*, \lambda^*) \le \theta(x, \lambda),\ \forall (x, \lambda) \in K \times \Lambda \quad \text{w.p.1,}$$

which indicates that $(x^*, \lambda^*) \in S^*$ with probability one. □

The following result shows that under some mild conditions, the sequence of optimal solutions of the SAA problem (17) converges to an optimal solution of (9) at an exponential rate.

Theorem 4.2 *Suppose conditions (A1)-(A3) hold. Let (x^N, λ^N) denotes an optimal solution of (17) and (x^*, λ^*) denotes an optimal solution of (9). Suppose that the following conditions hold:*

(i) *The sequence $\{(x^N, \lambda^N)\}$ is located in a compact subset $\mathcal{C} \times \Lambda$ of $K \times \Lambda$ with probability one and $(x^*, \lambda^*) \in \mathcal{C} \times \Lambda$;*

(ii) *For every $(x, \lambda) \in \mathcal{C} \times \Lambda$, the moment generating function $\mathbb{E}\big[e^{t\left(g(x,\lambda,\xi)-\theta(x,\lambda)\right)}\big]$ is finite valued for all t in a neighbourhood of zero;*

(iii) *The moment generating function $\mathbb{E}\big[e^{t(\tilde{\kappa}(\xi)-\mathbb{E}[\tilde{\kappa}(\xi)])}\big]$ is finite valued for all t in a neighbourhood of zero, where $\tilde{\kappa}(\xi) := \frac{(\alpha\lambda_{\min}+2\alpha\|G\|)\kappa(\xi)+2\kappa^2(\xi)}{\alpha\lambda_{\min}}$.*

Then, with probability approaching one exponentially fast with the increase of sample size N, (x^N, λ^N) becomes an approximate optimal solution of (9).

Proof. We have from conditions (A1)-(A3) and the proof of Theorem 4.1 that

$$|g(x', \lambda', \xi) - g(x, \lambda, \xi)| \le \tilde{\kappa}(\xi)(\|x' - x\| + \|\lambda' - \lambda\|), \forall x, x' \in \mathcal{C}, \lambda, \lambda' \in \Lambda \tag{19}$$

and $\mathbb{E}[\tilde{\kappa}(\xi)] < \infty$.

Under the conditions of (i)-(iii) and (19), it follows from [30, Theorem 5.1] that, for any $\epsilon > 0$, there exist positive constants $C(\epsilon)$ and $\beta(\epsilon)$, independent of N, such that

$$\mathrm{Prob}\Big\{\sup_{(x,\lambda)\in\mathcal{C}\times\Lambda} |\frac{1}{N}\sum_{i=1}^N g(x, \lambda, \xi_i) - \theta(x, \lambda)| > \epsilon\Big\} \le C(\epsilon)e^{-N\beta(\epsilon)}.$$

Hence, it is easy to obtain

$$\begin{aligned} &\mathrm{Prob}\Big\{|\frac{1}{N}\sum_{i=1}^N g(x^N, \lambda^N, \xi_i) - \theta(x^*, \lambda^*)| > \epsilon\Big\} \\ &\le \mathrm{Prob}\Big\{\sup_{(x,\lambda)\in\mathcal{C}\times\Lambda} |\frac{1}{N}\sum_{i=1}^N g(x, \lambda, \xi_i) - \theta(x, \lambda)| > \epsilon\Big\} \le C(\epsilon)e^{-N\beta(\epsilon)}. \end{aligned}$$

Note that

$$|\theta(x^N, \lambda^N) - \theta(x^*, \lambda^*)| \le |\theta(x^N, \lambda^N) - \frac{1}{N}\sum_{i=1}^N g(x^N, \lambda^N, \xi_i)| + |\frac{1}{N}\sum_{i=1}^N g(x^N, \lambda^N, \xi_i) - \theta(x^*, \lambda^*)|.$$

Combining the above inequalities, we have

$$\begin{aligned} &\mathrm{Prob}\Big\{|\theta(x^N, \lambda^N) - \theta(x^*, \lambda^*)| > 2\epsilon\Big\} \\ &\le \mathrm{Prob}\Big\{|\theta(x^N, \lambda^N) - \frac{1}{N}\sum_{i=1}^N g(x^N, \lambda^N, \xi_i)| > \epsilon\Big\} + \mathrm{Prob}\Big\{|\frac{1}{N}\sum_{i=1}^N g(x^N, \lambda^N, \xi_i) - \theta(x^*, \lambda^*)| > \epsilon\Big\} \\ &\le 2C(\epsilon)e^{-N\beta(\epsilon)}, \end{aligned}$$

which means that, with probability at least $1 - 2C(\epsilon)e^{-N\beta(\epsilon)}$, an optimal solution of (17) becomes a 2ϵ-approximation optimal solution of (9). This completes the proof. □

4.2 Convergence of Stationary Points

Since problem (17) is generically non-convex, it is necessary to study the limiting behavior of the stationary points of (17). For convenience, we suppose

$$K = \{x \in \mathbb{R}^n : \ c_i(x) \le 0, i = 1, \cdots, p; h_j(x) = 0, j = 1, \cdots, q\},$$

where $c_i : \mathbb{R}^n \to \mathbb{R}$ $(i = 1, \cdots, p)$ are continuously differentiable convex functions and $h_j : \mathbb{R}^n \to \mathbb{R}$ $(j = 1, \cdots, q)$ are affine functions. We first recall some definitions.

- (x^*, λ^*) is said to be stationary to (9) iff there exist Lagrange multiplier vectors $\mu^* \in \mathbb{R}^p, \zeta^* \in \mathbb{R}^q$, $\varsigma^* \in \mathbb{R}$, and $v^* \in \mathbb{R}^m$ such that

$$\nabla_x \theta(x^*, \lambda^*) + \sum_{i=1}^{p} \mu_i^* \nabla c_i(x^*) + \sum_{j=1}^{q} \zeta_j^* \nabla h_j(x^*) = 0, \tag{20}$$

$$\nabla_\lambda \theta(x^*, \lambda^*) + \varsigma^* e - \sum_{j=1}^{m} v_j^* e^j = 0, \tag{21}$$

$$0 \le \mu^* \perp c(x^*) \le 0, \tag{22}$$

$$0 \le v^* \perp \lambda^* \ge 0, \tag{23}$$

$$h(x^*) = 0, \quad \lambda^* e - 1 = 0. \tag{24}$$

- (x^N, λ^N) is said to be stationary to (17) iff there exist Lagrange multiplier vectors $\mu^N \in \mathbb{R}^p, \zeta^N \in \mathbb{R}^q$, $\varsigma^N \in \mathbb{R}$, and $v^N \in \mathbb{R}^m$ such that

$$\nabla_x \theta_N(x^N, \lambda^N) + \sum_{i=1}^{p} \mu_i^N \nabla c_i(x^N) + \sum_{j=1}^{q} \zeta_j^N \nabla h_j(x^N) = 0, \tag{25}$$

$$\nabla_\lambda \theta_N(x^N, \lambda^N) + \varsigma^N e - \sum_{j=1}^{m} v_j^N e^j = 0, \tag{26}$$

$$0 \le \mu^N \perp c(x^N) \le 0, \tag{27}$$

$$0 \le v^N \perp \lambda^N \ge 0, \tag{28}$$

$$h(x^N) = 0, \quad \lambda^N e - 1 = 0, \tag{29}$$

where $e = (1, \cdots, 1) \in \mathbb{R}^m$, and $u \perp v$ means $u^T v = 0$.

Let $I(x^*)$ be the set of indices of active inequality constraints with respect to x^* and $J(\lambda^*)$ be the set of indices of active inequality constraints with respect to λ^*. Before we start our convergence analysis, we introduce the concept of approximate KKT (see [33] for more details) for problem (9).

Definition 4.1 We say that a feasible point (x, λ) of problem (9) satisfies the approximate KKT (AKKT) iff there exist sequences $\{(x^N, \lambda^N)\} \subset \mathbb{R}^n \times \mathbb{R}^m$, $\mu^N \in \mathbb{R}^p, \zeta^N \in \mathbb{R}^q, v^N \in \mathbb{R}^m$, and $\varsigma^N \in \mathbb{R}$ such that

$\lim_{N\to\infty}(x^N,\lambda^N)=(x,\lambda)$,

$$\lim_{N\to\infty}\nabla_x\theta(x^N,\lambda^N)+\sum_{i=1}^{p}\mu_i^N\nabla c_i(x^N)+\sum_{j=1}^{q}\zeta_j^N\nabla h_j(x^N)=0, \tag{30}$$

$$\lim_{N\to\infty}\nabla_\lambda\theta(x^N,\lambda^N)+\varsigma^N e-\sum_{i=1}^{m}v_i^N e^i=0, \tag{31}$$

$$\lim_{N\to\infty}\min\{\mu_i^N,-c_i(x^N)\}=0,\quad i=1,\cdots,p, \tag{32}$$

$$\lim_{N\to\infty}\min\{v_i^N,\lambda_i^N\}=0,\quad i=1,\cdots,m. \tag{33}$$

The AKKT condition can be written in a more compact form as in [32], which says that, the AKKT condition holds at $(x,\lambda)\in K\times\varLambda$ iff there exist $\{(x^N,\lambda^N)\}\subset\mathbb{R}^n\times\mathbb{R}^m$, $\{\mu^N\}\subset\mathbb{R}^p_+$ with $\mu_i^N=0$ for $i\notin I(x)$, $\{\zeta^N\}\subset\mathbb{R}^q$, $\{v^N\}\subset\mathbb{R}^m_+$ with $v_i^N=0$ for $i\notin J(\lambda)$, and $\{\varsigma^N\}\in\mathbb{R}$ such that $\lim_{N\to\infty}(x^N,\lambda^N)=(x,\lambda)$ and

$$\lim_{N\to\infty}\nabla_x\theta(x^N,\lambda^N)+\sum_{i\in I(x)}\mu_i^N\nabla c_i(x^N)+\sum_{j=1}^{q}\zeta_j^N\nabla h_j(x^N)=0,$$
$$\lim_{N\to\infty}\nabla_\lambda\theta(x^N,\lambda^N)+\varsigma^N e-\sum_{i\in J(\lambda)}v_i^N e^i=0.$$

Given $(x^*,\lambda^*)\in K\times\varLambda$, we define

$$K(x,\lambda):=\left\{\begin{array}{ll}\sum_{i\in I(x^*)}\mu_i^*\nabla c_i(x)+\sum_{j=1}^{q}\zeta_j^*\nabla h_j(x): & \mu_i^*\in\mathbb{R}_+,\zeta_j^*\in\mathbb{R}\\ \varsigma^*e-\sum_{i\in J(\lambda^*)}v_i^*e^i: & v_i^*\in\mathbb{R}_+,\varsigma^*\in\mathbb{R}\end{array}\right\}. \tag{34}$$

Definition 4.2 [33] We say that $(x^*,\lambda^*)\in K\times\varLambda$ satisfies the cone-continuity property (CCP) iff the set-valued mapping $(x,\lambda)\rightrightarrows K(x,\lambda)$, defined in (34), is outer semicontinuous at (x^*,λ^*), that is

$$\limsup_{(x,\lambda)\to(x^*,\lambda^*)}K(x,\lambda)\subset K(x^*,\lambda^*). \tag{35}$$

Lemma 4.1 *Suppose conditions (A1)–(A3) hold and $\lim_{N\to\infty}(x^N,\lambda^N)=(x^*,\lambda^*)$. Then, we have*

$$\lim_{N\to\infty}\nabla\theta_N(x^N,\lambda^N)=\nabla\theta(x^*,\lambda^*)\quad w.p.1.$$

Proof. By the definitions, we have

$$\nabla\theta(x,\lambda)=\begin{pmatrix}\sum_{j=1}^{m}\lambda_j\mathbb{E}[F_j(x,\xi)]-\mathbb{E}\big[(\sum_{j=1}^{m}\lambda_j\nabla_xF_j(x,\xi)-\alpha G)(H(x,\lambda,\xi)-x)\big]\\ \mathbb{E}[(x-H(x,\lambda,\xi))^TF_1(x,\xi)]\\ \vdots\\ \mathbb{E}[(x-H(x,\lambda,\xi))^TF_m(x,\xi)]\end{pmatrix}$$

and

$$\nabla\theta_N(x,\lambda)=\begin{pmatrix}\frac{1}{N}\sum_{i=1}^{N}\Big(\sum_{j=1}^{m}\lambda_jF_j(x,\xi_i)-(\sum_{j=1}^{m}\lambda_j\nabla_xF_j(x,\xi_i)-\alpha G)(H(x,\lambda,\xi_i)-x)\Big)\\ \frac{1}{N}\sum_{i=1}^{N}\big((x-H(x,\lambda,\xi_i))^TF_1(x,\xi_i)\big)\\ \vdots\\ \frac{1}{N}\sum_{i=1}^{N}\big((x-H(x,\lambda,\xi_i))^TF_m(x,\xi_i)\big)\end{pmatrix}.$$

It follows from conditions (A1)-(A3) that we have, for any compact set $D \subset K \times \Lambda$,

$$\lim_{N\to\infty} \max_{(x,\lambda)\in D} \|\nabla\theta_N(x,\lambda) - \nabla\theta(x,\lambda)\| = 0 \quad \text{w.p.1.} \tag{36}$$

It follows from (A1) and (A2) that $\nabla\theta(\cdot,\cdot)$ is a continuous function. Since $\lim_{N\to\infty}(x^N,\lambda^N) = (x^*,\lambda^*)$, the sequence $\{(x^N,\lambda^N)\}$ is contained in a closed neighbourhood, denoted by $B \subset K \times \Lambda$, of (x^*,λ^*) and, for any given $\epsilon > 0$,

$$\|\nabla\theta(x^N,\lambda^N) - \nabla\theta(x^*,\lambda^*)\| < \frac{\epsilon}{2}$$

holds when N is sufficiently large. By (36), there exists $N_0 > 0$ such that, for all $N \geq N_0$, we obtain that $(x^N,\lambda^N) \in B$ and

$$\|\nabla\theta_N(x^N,\lambda^N) - \nabla\theta(x^N,\lambda^N)\| < \frac{\epsilon}{2},$$

which implies that $\|\nabla\theta_N(x^N,\lambda^N) - \nabla\theta(x^*,\lambda^*\| < \epsilon$. Therefore, we have

$$\lim_{N\to\infty} \nabla\theta_N(x^N,\lambda^N) = \nabla\theta(x^*,\lambda^*) \quad \text{w.p.1.}$$

This completes the proof. □

Theorem 4.3 *Suppose conditions (A1)–(A3) hold. Let (x^N,λ^N) be stationary to (17) for each N and (x^*,λ^*) be an accumulation point of $\{(x^N,\lambda^N)\}$. If the CCP holds at (x^*,λ^*), then (x^*,λ^*) is stationary to problem (9) with probability one.*

Proof. Without any loss of generality, we assume that $\{(x^N,\lambda^N)\}$ itself converges to (x^*,λ^*). Since (x^N,λ^N) is stationary to (17) for each N, there exist Lagrange multiplier vectors $\{\mu^N\} \subset \mathbb{R}^p_+, \{\zeta^N\} \subset \mathbb{R}^q, \{\varsigma^N\} \subset \mathbb{R}$, and $\{v^N\} \subset \mathbb{R}^m_+$ such that

$$\nabla_x\theta_N(x^N,\lambda^N) + \sum_{i=1}^p \mu_i^N \nabla c_i(x^N) + \sum_{j=1}^q \zeta_j^N \nabla h_j(x^N) = 0,$$
$$\nabla_\lambda\theta_N(x^N,\lambda^N) + \varsigma^N e - \sum_{j=1}^m v_j^N e^j = 0.$$

By virtue of Lemma 4.1 and combining the continuity of the gradient of θ, we have

$$\lim_{N\to\infty} \nabla_x\theta(x^N,\lambda^N) - \nabla_x\theta(x^*,\lambda^*) + \nabla_x\theta(x^*,\lambda^*) - \nabla_x\theta_N(x^N,\lambda^N) = 0,$$
$$\lim_{N\to\infty} \nabla_\lambda\theta(x^N,\lambda^N) - \nabla_\lambda\theta(x^*,\lambda^*) + \nabla_\lambda\theta(x^*,\lambda^*) - \nabla_\lambda\theta_N(x^N,\lambda^N) = 0$$

with probability one. Thus, we have

$$\lim_{N\to\infty} \nabla_x\theta(x^N,\lambda^N) + \sum_{i=1}^p \mu_i^N \nabla c_i(x^N) + \sum_{j=1}^q \zeta_j^N \nabla h_j(x^N) = 0,$$
$$\lim_{N\to\infty} \nabla_\lambda\theta(x^N,\lambda^N) + \varsigma^N e - \sum_{j=1}^m v_j^N e^j = 0$$

with probability one. Hence, the AKKT conditions are satisfied and, from Definition 4.1, we have $\mu_i^N = 0$ for $i \notin I(x^*)$, and $v_j^N = 0$ for $j \notin J(\lambda^*)$ such that

$$\varepsilon^N := \nabla_x\theta(x^N,\lambda^N) + \sum_{i\in I(x^*)} \mu_i^N \nabla c_i(x^N) + \sum_{j=1}^q \zeta_j^N \nabla h_j(x^N) \to 0,$$
$$\delta^N := \nabla_\lambda\theta(x^N,\lambda^N) + \varsigma^N e - \sum_{j\in J(\lambda^*)} v_j^N e^j \to 0$$

with probability one. It follows that

$$\begin{pmatrix} \sum\limits_{i\in I(x^*)} \mu_i^N \nabla c_i(x^N) + \sum\limits_{j=1}^{q} \zeta_j^N \nabla h_j(x^N) \\ \varsigma^N e - \sum\limits_{j\in J(\lambda^*)} v_j^N e^j \end{pmatrix} \in K(x^N, \lambda^N) \tag{37}$$

and

$$\sum_{i\in I(x^*)} \mu_i^N \nabla c_i(x^N) + \sum_{j=1}^{q} \zeta_j^N \nabla h_j(x^N) = \varepsilon^N - \nabla_x \theta(x^N, \lambda^N), \tag{38}$$

$$\varsigma^N e - \sum_{j\in J(\lambda^*)} v_j^N e^j = \delta^N - \nabla_\lambda \theta(x^N, \lambda^N) \tag{39}$$

with probability one. Taking limits in (37)–(39) and using the continuity of $\nabla\theta$ and $\varepsilon^N \to 0, \delta^N \to 0$, we get

$$\begin{pmatrix} -\nabla_x \theta(x^*, \lambda^*) \\ -\nabla_\lambda \theta(x^*, \lambda^*) \end{pmatrix} = \lim_{N\to\infty} \begin{pmatrix} \sum\limits_{i\in I(x^*)} \mu_i^N \nabla c_i(x^N) + \sum\limits_{j=1}^{q} \zeta_j^N \nabla h_j(x^N) \\ \varsigma^N e - \sum\limits_{j\in J(\lambda^*)} v_j^N e^j \end{pmatrix} \in \limsup_{N\to\infty} K(x^N, \lambda^N)$$

with probability one. Moreover, we have

$$\limsup_{N\to\infty} K(x^N, \lambda^N) \subset \limsup_{x\to x^*, \lambda\to\lambda^*} K(x, \lambda) \subset K(x^*, \lambda^*) \quad \text{w.p.1},$$

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

$$\begin{pmatrix} -\nabla_x \theta(x^*, \lambda^*) \\ -\nabla_\lambda \theta(x^*, \lambda^*) \end{pmatrix} \in K(x^*, \lambda^*) \quad \text{w.p.1},$$

which is equivalent to (20)–(23). Then, taking a limit in (29), we obtain that (24). That is, (x^*, λ^*) is stationary to problem (9) with probability one. □

Remark 4.1 The CCP condition is the weakest constraint qualification under which the approximate KKT points are guaranteed to converge to the KKT points, as investigated in [33]. Any constraint qualification stronger than the CCP (for instance, the widely-used Slater's constraint qualification [24, 25] for convex constraints, etc.) ensures the validity of Theorem 4.3.

Remark 4.2 An implicit assumption in the above discussion is that the iterated sequences have accumulation points. Actually, how to guarantee this assumption is a very important issue. It obviously holds under the conditions that the feasible set K is nonempty and bounded, or the underlying objective functions is uniformly coercive. As a future topic, we will try to find some weaker sufficient conditions for it. One may see some discussion by Ralph and Xu in [28].

5 Conclusions

We have shown that the VVI (1) is equivalent to the scalar variational inequality (2). Based on this observation, we have presented a deterministic formulation (9) for the SVVI (5). Some properties of the model (9) have been investigated and a sample average approximation method has been proposed for solving (9). Under some moderate conditions, we have established a convergence theory for the proposed approach. Note that the SVVI considered in this paper only characterizes the weak Pareto solution. It is interesting to explore whether we can obtain similar analysis for the SVVI which characterizes the Pareto solution. We will investigate this topic in the near future.

Acknowledgements This work was supported in part by the NSFC Grant (No. 11431004), the Hongkong Baptist University FRG1/15-16/027 and RC-NACAN-ZHANG J, 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). The authors are grateful to Professor F. Giannessi, Professor G.Y. Chen, and two anonymous referees for their valuable comments and constructive suggestions, which help to improve the original manuscript.

References

1. Giannessi, F.: Theorems of alternative, quadratic programs and complementary problems. Cottle, R W et al. (eds.), Variational Inequality and Complementary Problems, Wiley and Sons, New York (1980)
2. Giannessi, F (eds.).: Vector Variational Inequalities and Vector Equilibria. Vol. 38, Kluwer Academic, Dordrecht, The Netherlands (2000)
3. Chen, G.Y., Huang, X.X., Yang, X.Q.: Vector Optimization: Set-Valued and Variational Analysis. Lecture Notes in Economics and Mathematical Systems, vol. 541, Springer, Berlin (2005)
4. Yang, X.Q., Yao, J.C.: Gap functions and existence of solutions to set-valued vector variational inequalities. J. Optim. Theory Appl. **115**, 407-417 (2002)
5. Chen, G.Y., Goh, C.J., Yang, X.Q.: On gap functions for vector variational inequalities. Giannessi, F. (eds.) Vector Variational Inequality and Vector Equilibria: Mathematical Theories, 55-70. Kluwer Academic, Boston (2000)
6. Yang, X.M., Yang, X.Q., Teo, K.L.: Some remarks on the Minty vector variational inequality. J. Optim. Theory Appl. **121**, 193-201 (2004)
7. Li, S.J., Chen, G.Y.: Properties of gap function for vector variational inequality. Giannessi, F. and Maugeri, A. (eds.) Variational Analysis and Applications, 605-631. Springer, Berlin (2005)
8. Li, S.J., Yan, H., Chen, G.Y.: Differential and sensitivity properties of gap functions for vector variational inequalities. Math. Methods Oper. Res. **57**, 377-391 (2003)
9. Huang, N.J., Li, J., Yang, X.Q.: Weak sharpness for gap functions in vector variational inequalities. J. Math. Anal. Appl. **394**, 449-457 (2012)
10. Charitha, C., Dutta, J.: Regularized gap functions and error bounds for vector variational inequality. Pac. J. Optim. **6**, 497-510 (2010)
11. Lee, G.M., Kim, D.S., Lee, B.S., Yen, N.D.: Vector variational inequality as a tool for studying vector optimization problems. Nonlinear Anal. **34**, 745-765 (1998)
12. Charitha, C., Duttab, J., Lalitha, C.S.: Gap functions for vector variational inequalities. Optimization. **64**, 1499-1520 (2015)
13. Ma, H.Q., Huang, N.J., Wu, M., Regan, D.O.: A new gap function for vector variational inequalities with an application. J. Appl. Math. Article ID 423040 (2013)

14. Gürkan, G.G., Özge, A.Y., Robinson, S.M.: Sample-path solution of stochastic variational inequalities. Math. Program. **84**, 313-333 (1999)
15. Chen, X.J., Fukushima, M.: Expected residual minimization method for stochastic linear complementarity problems. Math. Oper. Res. **30**, 1022-1038 (2005)
16. Jiang, H.Y., Xu, H.F.: Stochastic approximation approaches to the stochastic variational inequality problem. IEEE Trans. Autom. Control. **53**, 1462-1475 (2008)
17. Lin, G.H., Fukushima, M.: Stochastic equilibrium problems and stochastic mathematical programs with equilibrium constraints: A survey. Pac. J. Optim. **6**, 455-482 (2010)
18. Lin, G.H., Chen, X.J., Fukushima, M.: New restricted NCP function and their applications to stochastic NCP and stochastic MPEC. Optimization. **56**, 641-753 (2007)
19. Lin, G.H., Fukushima, M.: New reformulations for stochastic nonlinear complementarity problems. Optim. Method Softw. **21**, 551-564 (2006)
20. Chen, X.J., Zhang, C., Fukushima, M.: Robust solution of monotone stochastic linear complementarity problems. Math. Program. **117**, 51-80 (2009)
21. Chen, X.J., Wets, R.J-B., Zhang, Y.F.: Stochastic variational inequalities: residual minimization smoothing sampling average approximations. SIAM J. Optim. **22**, 649-673 (2012)
22. Zhang, Y.F., Chen, X.J.: Regularizations for stochastic linear variational inequalities. J. Optim. Theory Appl. **163**, 460-481 (2014)
23. Xu, H.F.: Sample average approximation methods for a class of stochastic variational inequality problems. Asia Pac J. Oper. Res. **27**, 103-119 (2010)
24. Luo, M.J., Lin, G.H.: Expected residual minimization method for stochastic variational inequality problems. J. Optim. Theory Appl. **140**, 103-116 (2009)
25. Luo, M.J., Lin, G.H: Convergence results of the ERM method for nonlinear stochastic variational inequality problems. J. Optim. Theory Appl. **142**, 569-581 (2009)
26. Luo, M.J., Lin, G.H.: Stochastic variational inequality problems with additional constraints and their applications in supply chain network equilibria. Pac. J. Optim. **7** 263-279 (2011)
27. Fukushima, M.: Equivalent differentiable optimization problems and descent methods for asymmetric variational inequality problems. Math. Program. **53**, 99-110 (1992)
28. Ralph, D., Xu, H.F.: Convergence of stationary points of sample average two stage stochastic programs: a generalized equation approach. Math. Oper. Res. **36**, 568-592 (2011)
29. Xu, H.F, Zhang, D.L: Smooth sample average approximation of stationary points in nonsmooth stochastic optimization and applications. Math. Program. **119**, 371-401 (2009)
30. Shapiro, A., Xu, H.F.: Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation. Optimization. **57**, 395-418, (2008)
31. Patrick, B.: Probability and Measure. Wiley-Interscience, New York (1995)
32. Andreani, R., Haeser, G., Martinez, J.M.: On sequential optimality conditions for smooth constrained optimization. Optimization. **60**, 627-641 (2011)
33. Andreani, R., Martinez, J.M., Ramos, A., Silva, P.J.S: A cone-continuity constraint qualification and algorithmic consequences. SIAM J. Optim. **26**, 96-110 (2016)

34. Shapiro, A., Dentcheva, D., Ruszczynski, A.: Lectures on Stochastic Programming: Modeling and Theory. SIAM, Philadelphia (2009)