QUANTITATIVE STABILITY OF THE ERM FORMULATION FOR A CLASS OF STOCHASTIC LINEAR VARIATIONAL INEQUALITIES

JIANXUN LIU, SHENGJIE LI* AND YINGRANG XU

College of Mathematics and Statistics, Chongqing University
Chongqing 401331, China

(Communicated by Kok Lay Teo)

Abstract. This paper focuses on the quantitative stability analysis of the expected residual minimization (ERM) formulation for a class of stochastic linear variational inequalities. Firstly, the existence of solutions of the ERM formulation and its perturbed problem is discussed. Then, the quantitative stability of the ERM formulation is derived under suitable probability metrics. Finally, the sample average approximation (SAA) problem of the ERM formulation is studied, and the rates of convergence of optimal solution sets are obtained under different assumptions.

1. Introduction. We consider the following stochastic linear variational inequality (SLVI) problem: find a vector $x^* \in S$ such that

$$
(x - x^*)^T (M(\xi)x^* + q(\xi)) \geq 0, \quad \forall x \in S, \xi \in \Xi \text{ a.s.,}
$$

where $S \subseteq \mathbb{R}^n$ is a nonempty closed convex set, $\xi : \Omega \rightarrow \Xi$ is a random variable defined in a probability space $(\Omega, \mathcal{F}, P)$ with support set $\Xi \subseteq \mathbb{R}^l$, $M(\cdot) : \mathbb{R}^l \rightarrow \mathbb{R}^{n \times n}$ is a matrix-valued mapping, $q(\cdot) : \mathbb{R}^l \rightarrow \mathbb{R}^n$ is a vector-valued mapping, and “a.s.” is the abbreviation for “almost surely” under the given probability measure.

In general, there is no $x^* \in S$ satisfying (1) for all $\xi \in \Xi$. Assume that $\xi$ obeys a probability distribution $P$, the following ERM formulation, see [2, 3, 1, 5, 13, 14, 23, 12, 11] and the references therein, has been considered to deal with the SLVI problem:

$$
\min_{x \in S} f_P(x) := \mathbb{E}_P[f(x, \xi)] = \int_{\Xi} f(x, \xi) P(d\xi),
$$

where $f(\cdot, \xi) : S \rightarrow R_+$ is a residual function [4] for problem (1) defined by

$$
f(x, \xi) := \max_{y \in S} \left\{ (x - y)^T (M(\xi)x + q(\xi)) - \frac{\alpha}{2} \|x - y\|^2 \right\},
$$

and $\alpha$ is a positive scalar. We make a blanket assumption that for every $x \in S$, $\mathbb{E}_P[f(x, \xi)]$ is well defined.

As far as we know, almost all the papers studied problem (1) by employing the ERM formulation (2) under the assumption that the underlying probability

2020 Mathematics Subject Classification. 90C15, 90C33.

Key words and phrases. Stochastic variational inequality problem, ERM formulation, quantitative stability analysis, sample average approximation, rate of convergence.

* Corresponding author: Shengjie Li.
distribution is fixed. However, these probability measures only reflect the available knowledge on the randomness at hand. This leads us to consider the stability analysis of the problem with respect to changes in the underlying probability distribution.

Stability analysis of optimization problems is important for not only theoretical study but also numerical approximation, see [16, 18, 17, 6, 7, 8] and the references therein. We usually need consider the discrete or empirical approximation to the high dimensional integrals when we handle a stochastic optimization problem numerically. Then we must investigate the quantitative relationship between the original continuous problem and its discrete approximation. Furthermore, we need study whether the optimal value or optimal solution set of the approximation problem converge to those of the original problem. All these questions can be answered through stability analysis. On the other hand, the rate of convergence of the optimal solutions of the discrete approximation problems is helpful for us to decide the sample size. Hence in this paper, we mainly focus on the quantitative stability analysis of problem (2) with respect to the perturbation of the distribution $P$, and we also investigate the rate of convergence of the optimal solutions of the SAA problems to that of the original problem (2).

The main contributions of this paper can be summarized as follows. We first discuss conditions for the existence of solutions to problem (2) and its perturbed distribution problem, which can be regarded as an extension of the corresponding result in [13]. Then, we give a quantitative stability analysis of problem (2) and obtain the quantitative relationship between the solution sets of problem (2) and its perturbed distribution problem under Fortet-Mourier metric. In the last, we study the SAA problem to problem (2), and establish the exponential and polynomial rates of convergence of the solutions of the SAA problems to that of problem (2) under different assumptions, respectively.

Throughout this paper, let $I$ denote the identity matrix, $B$ denote the closed unit ball centered on zero and $P_S(\cdot)$ denote the Euclidean projection onto $S$. The notation $\|\cdot\|$ stands for the Euclidean norm of a vector or the induced matrix norm of a matrix, $P(\Xi)$ denotes the set of probability distributions on the support set $\Xi$. For a vector (or a matrix) $K$, $K^T$ denotes the transpose of $K$. Define $d(a, B) := \inf_{b \in B} \|a - b\|$ and $d(A, B) := \sup_{a \in A} \inf_{b \in B} \|a - b\|$ for any $a \in \mathbb{R}^n$ and $A, B \subset \mathbb{R}^n$.

The rest of this paper is organized as follows. In Section 2, we introduce the $\zeta$-structure metrics and present some properties of the functions $f(x, \xi)$ and $f_P(x)$. In Section 3, we first discuss the conditions for the existence of solutions to problem (2) and its perturbed distribution problem, and then we give the quantitative relationship between the solution sets of problem (2) and its perturbed distribution problem. The rates of convergence of the solutions of the SAA problems to those of problem (2) are established in Section 4. Finally, we conclude the paper in Section 5.

2. Preliminaries. To establish the quantitative stability results, we need to introduce the so-called pseudo metrics between two probability measures or distributions. In this paper, we adopt the $\zeta$-structure metrics.

**Definition 2.1.** ([16], $\zeta$-structure metrics) Let $\mathcal{G}$ be a set of real-valued measurable functions on $\Xi$. For any two probability measures $P, Q \in P(\Xi)$, the function

\begin{align*}
&\zeta \left( \mathcal{G}, P, Q \right) := \sup_{f \in \mathcal{G}} \left\{ \int f(x) dP(x) \right\} - \int f(x) dQ(x), \\
&\zeta \left( \mathcal{G}, P, Q \right) := \sup_{f \in \mathcal{G}} \left\{ \int f(x) dP(x) \right\} - \int f(x) \mu(dx), \\
&\zeta \left( \mathcal{G}, P, P \right) := \sup_{f \in \mathcal{G}} \left\{ \int f(x) dP(x) \right\} - \int f(x) dP(x), \\
&\zeta \left( \mathcal{G}, P, Q \right) := \sup_{f \in \mathcal{G}} \left\{ \int f(x) dP(x) \right\} - \int f(x) dQ(x).
\end{align*}
\[ \mathcal{D}_G(P,Q) := \sup_{g \in \mathcal{G}} |\mathbb{E}_P[g(\xi)] - \mathbb{E}_Q[g(\xi)]| \]

is called the \( \zeta \)-structure metric between \( P \) and \( Q \) induced by \( G \). Also, \( G \) is called the generator of \( \mathcal{D}_G(\cdot,\cdot) \).

\( \mathcal{D}_G(\cdot,\cdot) \) is a pseudo metric because \( \mathcal{D}_G(P,Q) = 0 \) usually fails to imply \( P = Q \) unless \( G \) is rich enough. Different \( \zeta \)-structure metrics can be obtained by choosing different \( G \)s. For example, for \( p \geq 1 \), if we take
\[ G_{FM_p} := \{ g : \Xi \to \mathbb{R} : |g(\xi_1) - g(\xi_2)| \leq \max\{1, \|\xi_1\|^p, \|\xi_2\|^p\}\|\xi_1 - \xi_2\| \}, \]

then the corresponding \( \zeta \)-structure metric
\[ \zeta_p(P,Q) := \sup_{g \in G_{FM_p}} |\mathbb{E}_P[g(\xi)] - \mathbb{E}_Q[g(\xi)]| \]

is called the \( p \)-th order Fortet-Mourier metric, which is widely used in the quantitative stability analysis of stochastic programming problems. As we can see, this metric requires some locally Lipschitz continuity conditions for the objective function. Especially, for \( p = 1 \), we arrive at the Kantorovich metric
\[ \zeta_1(P,Q) := \sup_{g \in G_{FM_1}} |\mathbb{E}_P[g(\xi)] - \mathbb{E}_Q[g(\xi)]|. \]

There are some other \( \zeta \)-structure metrics, one can refer to [18] for more details.

The \( \zeta \)-structure metrics are closely related with the weak convergence of measures in \( \mathcal{P}(\Xi) \).

**Definition 2.2.** A sequence \((P_n)\) in \( \mathcal{P}(\Xi) \) is said to converge weakly to \( P \in \mathcal{P}(\Xi) \), shortly \( P_n \wep P \) if
\[ \lim_{n \to \infty} \int_\Xi g(\xi) dP_n(\xi) = \int_\Xi g(\xi) dP(\xi) \]
holds for any continuous and bounded function \( g : \Omega \to \mathbb{R} \).

It is known that a sequence \((P_n)\) converges to \( P \) in \((\mathcal{P}(\Xi),\zeta_p)\) iff it converges weakly and
\[ \lim_{n \to \infty} \int_\Xi \|\xi\|^p dP_n(\xi) = \int_\Xi \|\xi\|^p dP(\xi) \]
holds. One can refer to [16] for more details.

We denote the optimal solution set and optimal value of problem (2) by \( S^*_P \) and \( v(P) \), respectively. The function \( \psi_P(\cdot) : \mathbb{R}^+ \to \mathbb{R} \) defined by
\[ \psi_P(\pi) := \min\{f_P(x) - v(P) : d(x,S^*_P) \geq \pi, x \in S\} \]
is called the growth function of problem (2). It is easy to verify that \( \psi_P(\cdot) \) is nondecreasing and lower semicontinuous. Its inverse function is defined as
\[ \psi_P^{-1}(t) := \sup\{\pi \in \mathbb{R}^+ : \psi_P(\pi) \leq t\}, \]
which is nondecreasing too. The function
\[ \Psi_P(t) := t + \psi_P^{-1}(2t), t \in \mathbb{R}^+ \]
is called the conditioning function of problem (2). Obviously, \( \Psi_P \) is lower semicontinuous and increasing. Moreover, it has \( \Psi_P(t) \to 0 \) as \( t \to 0 \).

We next present some properties of the functions \( f(x,\xi) \) and \( f_P(x) \). It follows from [4] that \( f(x,\xi) \geq 0 \) for any fixed \( \xi \in \Xi \) and \( x \in S \), and \( f(x,\xi) \) can be rewritten as
\[ f(x,\xi) = (x - H(x,\xi))^T(M(\xi)x + q(\xi)) - \frac{\alpha}{2}\|x - H(x,\xi)\|^2, \]
where
\[ H(x, \xi) := P_S(x - \alpha^{-1}(M(\xi)x + q(\xi))). \]
Moreover, \( f(x, \xi) \) is differentiable with respect to \( x \) and
\[ \nabla_x f(x, \xi) = M(\xi)x + q(\xi) - [M(\xi) - \alpha I](H(x, \xi) - x). \]

In this paper, we suppose that \( \mathbb{E}_P[\|M(\xi)\| + \|q(\xi)\|^2] < +\infty \). Then, by [13], the following proposition holds.

**Proposition 1.** ([13]) The function \( f_P(x) \) is differentiable with respect to \( x \). In particular, for any \( x \in S \), it has
\[ \nabla f_P(x) = \mathbb{E}_P[\nabla_x f(x, \xi)]. \]

### 3. Existence of solutions and quantitative stability analysis.
Consider the perturbed problem of problem (2) under probability distribution \( Q \in \mathbb{P}(\Xi) \), i.e.,
\[ \min_{x \in S} f_Q(x) := \mathbb{E}_Q[f(x, \xi)] = \int_\Xi f(x, \xi)Q(d\xi). \quad (3) \]

In this section, we first discuss the existence of solution of problem (2) and then establish the quantitative relationship between the solution sets of problem (2) and (3).

For \( O = P, Q \) and \( c > 0 \), define the level set of \( f_O(x) \) by \( L_O(c) := \{x \in S | f_O(x) \leq c\} \). Let \( \bar{M}_O := \mathbb{E}_O[M(\xi)] \) and \( \bar{Q}_O := \mathbb{E}_O[q(\xi)] \), where the mathematical expectation \( \mathbb{E}_O \) is taken in componentwise. We introduce the following assumption in [13].

**Assumption 3.1.** Suppose that \( \bar{M}_P \) is positive definite.

Denote \( \mu_P \) the smallest eigenvalue of \( (\bar{M}_P^T + \bar{M}_P)/2 \). Then, it has \( \mu_P > 0 \) under Assumption 3.1. In [13], the authors obtained the boundedness of the level set \( L_P(c) \) under Assumption 3.1.

**Theorem 3.1.** ([13]) Let Assumption 3.1 hold and \( \alpha \in (0, 2\mu_P) \). Then the level set \( L_P(c) \) is bounded for any \( c \geq 0 \).

Obviously, it follows from the continuity of \( f_P(x) \) on \( S \) and Theorem 3.1 that \( S^*_P \) is nonempty and bounded.

For further discussion, we need the following assumption.

**Assumption 3.2.** Assume that the following statements hold.

(i). The random coefficients \( M(\xi) \) and \( q(\xi) \) depend affine linearly on \( \xi = (\xi_1, \xi_2, \ldots, \xi_l) \in \Xi \), i.e., \( M(\xi) = M_0 + \xi_1 M_1 + \ldots + \xi_l M_l, \) \( q(\xi) = q_0 + \xi_1 q_1 + \ldots + \xi_l q_l, \) where \( M_i, q_i \in \mathbb{R}^{n \times n}, i = 0, \ldots, l; \)

(ii). For a given \( r > 0 \), there exists \( \tilde{r} > 0 \) such that \( \|H(x, \xi)\| < \tilde{r} \) for any \( x \in S \cap rB \) and \( \xi \in \Xi \).

We now give some lemmas.

**Lemma 3.2.** Under Assumption 3.2 (i), there holds that \( \bar{M}_Q \to \bar{M}_P \) and \( \bar{Q}_Q \to \bar{Q}_P \) as \( \zeta_1(P, Q) \to 0 \). Furthermore, let \( \mu_Q \) be the smallest eigenvalue of \( (\bar{M}_Q^T + \bar{M}_Q)/2 \). Then, we have \( \mu_Q \to \mu_P \) as \( \zeta_1(P, Q) \to 0 \).

**Proof.** For the first conclusion, we only prove that \( \bar{M}_Q \to \bar{M}_P \) as \( \zeta_1(P, Q) \to 0 \), and the proof of convergence of \( \bar{Q}_Q \to \bar{Q}_P \) is similar. Without loss of generality, we prove it under the matrix norm \( \| \cdot \|_1 \). Since \( \|\mathbb{E}_P[M(\xi)] - \mathbb{E}_Q[M(\xi)]\|_1 = \max_{1 \leq j \leq n} \{\sum_{i=1}^n |\mathbb{E}_P[M_{ij}(\xi)] - \mathbb{E}_Q[M_{ij}(\xi)]|\} \) and \( M_{ij}(\xi) \) depends affine linearly
on $\xi$, we know that there exists $C > 0$ such that $|E_P[M_{ij}(\xi)] - E_Q[M_{ij}(\xi)]| \leq C_\xi(P,Q)$. Therefore, we have $\|E_P[M(\xi)] - E_Q[M(\xi)]\|_1 \leq nC_\xi(P,Q)$. This completes the proof of the first conclusion.

For the second conclusion, by Theorem 5 in [15], we know that $\mu_Q \to \mu_P$ as $\bar{M}_Q \to \bar{M}_P$. Together with the first conclusion, we have that $\mu_q \to \mu_P$ as $\zeta_1(P,Q) \to 0$. This completes the proof.

\[ \text{Lemma 3.3.} \quad \text{Let Assumption 3.1 and Assumption 3.2 (i) hold. Then, there exists } \tau > 0 \text{ such that } \bar{M}_Q \text{ is positive definite when } \zeta_1(P,Q) < \tau. \]

\[ \text{Proof.} \quad \text{Suppose that for any } \tau > 0, \text{ there exists } Q \text{ satisfies } \zeta_1(P,Q) < \tau, \text{ such that } \bar{M}_Q \text{ is not positive definite. This implies that} \]

\[ x(\tau)^T \bar{M}_Q x(\tau) \leq 0 \quad (4) \]

for some $x(\tau) \in \mathbb{R}^n$ with $\|x(\tau)\| = 1$. Since $x(\tau)$ is bounded, without loss of generality, we let $\lim_{\tau \to 0} x(\tau) = \bar{x}$. By Lemma 3.2, we know that $\bar{M}_Q \to \bar{M}_P$ as $\tau \to 0$. Therefore,

\[ x(\tau)^T \bar{M}_Q x(\tau) - \bar{x}^T \bar{M}_P \bar{x} = x(\tau)^T (\bar{M}_Q - \bar{M}_P)x(\tau) + x(\tau)^T \bar{M}_P(x(\tau) - \bar{x}) + (x(\tau) - \bar{x})^T \bar{M}_P \bar{x} \to 0. \]

Then we have that

\[ \lim_{\tau \to 0} x(\tau)^T \bar{M}_Q x(\tau) = \bar{x}^T \bar{M}_P \bar{x}. \]

Moreover, it follows from (4) that

\[ \bar{x}^T \bar{M}_P \bar{x} \leq 0, \quad \|\bar{x}\| = 1, \]

which contradicts the assumption that $\bar{M}_P$ is positive definite. This completes the proof. \[ \square \]

In what follows, we prove the boundedness of the level set $L_Q(c)$.

\[ \text{Theorem 3.4.} \quad \text{Suppose that Assumption 3.1 and Assumption 3.2 (i) hold and } \alpha \in (0, 2\mu_P). \text{ Then, there exists } \tau > 0 \text{ such that the level set } L_Q(c) \text{ is bounded for any } c \geq 0 \text{ when } \zeta_1(P,Q) < \tau. \]

\[ \text{Proof.} \quad \text{This proof is similar to that of [13, Theorem 4.1]. For completeness, we give a simple proof. Since } \bar{M}_P \text{ is positive, we know that there exists unique } x^* \text{ such that} \]

\[ (x - x^*)^T (\bar{M}_P x^* + \bar{q}_P) \geq 0, \quad \forall x \in S. \quad (5) \]

It follows from Lemma 3.2 that $\bar{M}_Q x^* + \bar{q}_Q \to \bar{M}_P x^* + \bar{q}_P$ and $\mu_Q \to \mu_P$ as $\zeta_1(P,Q) \to 0$. Noting that $\alpha \in (0, 2\mu_P)$, we can take a scalar $\epsilon \in (0, \frac{2\mu_P - \alpha}{4})$. Then, there exists $\tau > 0$ such that

\[ \|\bar{M}_Q x^* + \bar{q}_Q - (\bar{M}_P x^* + \bar{q}_P)\| < \epsilon, \quad (6) \]

\[ \mu_Q > \mu_P - \epsilon \quad (7) \]

when $\zeta_1(P,Q) < \tau$.

Let $\zeta_1(P,Q) < \tau$. Suppose that there is a $\bar{c} \geq 0$ such that the level set $L_Q(\bar{c})$ is unbounded. This implies that there exists a sequence $\{x^j\} \subseteq L_Q(\bar{c})$ such that
Lemma 3.5. Let \( \xi \in \Xi \) and Assumption 3.2 be satisfied. Then for any \( \xi_1, \xi_2 \in \Xi \) and \( x \in S \) with \( \|x\| \leq r \), there exists \( L > 0 \) such that

\[
|f(x, \xi_1) - f(x, \xi_2)| \leq L \max\{1, \|\xi_1\|, \|\xi_2\|\} \|\xi_1 - \xi_2\|.
\]

Proof. Note that

\[
|f(x, \xi_1) - f(x, \xi_2)| = |(x - H(x, \xi_1))^T(M(\xi_1)x + q(\xi_1)) - (x - H(x, \xi_2))^T(M(\xi_2)x + q(\xi_2))| \leq \\
\frac{\alpha}{2} \|x - H(x, \xi_1)\| \|\xi_1 - \xi_2\| \leq L \max\{1, \|\xi_1\|, \|\xi_2\|\} \|\xi_1 - \xi_2\|.
\]

where the third inequality follows from Lemma 3.3 and the definition of \( \mu_Q \); the fourth inequality follows from (7); and the fifth inequality follows from (6). This is a contradiction and hence we complete the proof. \( \square \)

We can easily verify that under Assumption 3.2 (i), there exists \( \tau > 0 \) such that \( \mathbb{E}_{\Xi}([\|M(\xi)\| + \|q(\xi)\|]^2] < +\infty \) when \( \zeta_1(P, Q) < \tau \). Then, we can know that \( f_Q(x) \) is differentiable with respect to \( x \) in a similar way to Proposition 1 under the assumptions in Theorem 3.4. Denote \( S_Q^* \) and \( v(Q) \) the optimal solution set and optimal value of problem (3), respectively. By Theorem 3.4, we obtain that \( S_Q^* \) is nonempty and bounded.

Before we establish the quantitative relationship between \( S_Q^* \) and \( S_P^* \), \( v(Q) \) and \( v(P) \), we present the following property about the function \( f(x, \xi) \) under Assumption 3.2.

Lemma 3.5. Let \( r > 0 \) and Assumption 3.2 be satisfied. Then for any \( \xi_1, \xi_2 \in \Xi \) and \( x \in S \) with \( \|x\| \leq r \), there exists \( L > 0 \) such that

\[
|f(x, \xi_1) - f(x, \xi_2)| \leq L \max\{1, \|\xi_1\|, \|\xi_2\|\} \|\xi_1 - \xi_2\|.
\]
Firstly, by Assumption 3.2 (i), we know that there exists $L_1 > 0$ such that $\|M(\xi_1) - M(\xi_2)\| \leq L_1 \|\xi_1 - \xi_2\|$ and $\|q(\xi_1) - q(\xi_2)\| \leq L_1 \|\xi_1 - \xi_2\|$. Then

$$\|x^T((M(\xi_1) - M(\xi_2))x + q(\xi_1) - q(\xi_2))\| \leq \|x\|\|M(\xi_1) - M(\xi_2)\|\|x\| + \|q(\xi_1) - q(\xi_2)\| \leq L_1 r(r + 1)\|\xi_1 - \xi_2\|. \quad (8)$$

Again by Assumption 3.2 (i), we have that there exists $L_2 > 0$ such that

$$\|M(\xi)x + q(\xi)\| \leq L_2 \max\{1, \|\xi\|\}\|x\| + 1 \leq L_2 \max\{1, \|\xi\|\}(r + 1).$$

Moreover, by the non-expansive property of the projection operator, we get that

$$\|H(x, \xi_1) - H(x, \xi_2)\| = \|P_S(x - \alpha^{-1}(M(\xi_1)x + q(\xi_1))) - P_S(x - \alpha^{-1}(M(\xi_2)x + q(\xi_2)))\| \leq \alpha^{-1}\|\alpha^{-1}(M(\xi_2)x + q(\xi_2)) - (M(\xi_1)x + q(\xi_1))\| \leq \alpha^{-1}(\|M(\xi_1) - M(\xi_2)\|\|x\| + \|q(\xi_1) - q(\xi_2)\|) \leq \alpha^{-1} L_1 (r + 1)\|\xi_1 - \xi_2\|.$$

Therefore, by Assumption 3.2 (ii), we obtain that

$$\|H(x, \xi_1)^T(M(\xi_1)x + q(\xi_1)) - H(x, \xi_2)^T(M(\xi_2)x + q(\xi_2))\| \leq \|H(x, \xi_1)(\|M(\xi_1) - M(\xi_2)\|\|x\| + \|q(\xi_1) - q(\xi_2)\|) + \|H(x, \xi_1) - H(x, \xi_2)\|\|M(\xi_2)x + q(\xi_2)\| \leq L_1 r(r + 1)\|\xi_1 - \xi_2\| + \alpha^{-1} L_1 L_2 (r + 1)^2 \max\{1, \|\xi_2\|\}\|\xi_1 - \xi_2\|. \quad (9)$$

On the other hand, we easily get that

$$\frac{\alpha}{2} \|x - H(x, \xi_2)\|^2 \leq \frac{\alpha}{2} \|H(x, \xi_1) - H(x, \xi_2)\|1(2x - H(x, \xi_2) - H(x, \xi_1)) \leq \frac{\alpha}{2} \|H(x, \xi_1) - H(x, \xi_2)\| 2\|x\| + \|H(x, \xi_2)\| + \|H(x, \xi_1)\| \leq L_1 (r + 1)(r + \tilde{r})\|\xi_1 - \xi_2\|. \quad (10)$$

Finally, combining (8), (9) and (10), we have that

$$|f(x, \xi_1) - f(x, \xi_2)| \leq L_1 r(r + 1)\|\xi_1 - \xi_2\| + L_1 \tilde{r}(r + 1)\|\xi_1 - \xi_2\| + \alpha^{-1} L_1 L_2 (r + 1)^2 \max\{1, \|\xi_2\|\}\|\xi_1 - \xi_2\| + L_1 (r + 1)(r + \tilde{r})\|\xi_1 - \xi_2\| = (2L_1(r + 1)(r + \tilde{r}) + \alpha^{-1} L_1 L_2 (r + 1)^2 \max\{1, \|\xi_2\|\})\|\xi_1 - \xi_2\| \leq (2L_1(r + 1)(r + \tilde{r}) + \alpha^{-1} L_1 L_2 (r + 1)^2) \max\{1, \|\xi_1\|, \|\xi_2\|\}\|\xi_1 - \xi_2\|.$$

We then complete the proof by letting $L := 2L_1(r + 1)(r + \tilde{r}) + \alpha^{-1} L_1 L_2 (r + 1)^2$.  \(\Box\)

By Lemma 3.5, we can see that $f(x, \xi) \in G_{FM_2}$, and that is why we adopt the Fortet-Mourier metric in this paper. We then immediately obtain the following quantitative estimation.
Lemma 3.6. Let Assumption 3.2 be satisfied. Then for any $x \in S$ with $\|x\| \leq r$, it has
\[ |f_P(x) - f_Q(x)| \leq L\zeta_2(P, Q), \]
where $L$ comes from Lemma 3.5.

Now, based on Theorem 3.1, Theorem 3.4 and Lemma 3.6, we have the following quantitative result.

Theorem 3.7. Let Assumptions 3.1 and 3.2 be satisfied and $\alpha \in (0, 2\mu_P)$. Then there exists $L > 0$ and $\tau > 0$ such that
\[ |v(P) - v(Q)| \leq L\zeta_2(P, Q), \]
when $\zeta_2(P, Q) < \tau$. 

Proof. From the previous discussion, we know that $S^*_P$ is nonempty and bounded, and there exists $\tau > 0$ such that $S^*_Q$ is nonempty and bounded when $\zeta_1(P, Q) \leq \zeta_2(P, Q) < \tau$. Without loss of generality, we assume that $S^*_P, S^*_Q \subseteq S \cap rB$ for some $r > 0$. By Lemma 3.6, we get that there exists $L > 0$ such that
\[ |v(P) - v(Q)| \leq \sup_{x \in S \cap rB} |f_P(x) - f_Q(x)| \leq L\zeta_2(P, Q), \]
which verifies the first assertion. For the second assertion, we have that for any $x^*_Q \in S^*_Q$,
\[ 2L\zeta_2(P, Q) \geq L\zeta_2(P, Q) + v(Q) - v(P) \]
\[ = L\zeta_2(P, Q) + f_Q(x^*_Q) - v(P) \]
\[ \geq f_P(x^*_Q) - f_Q(x^*_Q) + f_Q(x^*_Q) - v(P) \]
\[ = f_P(x^*_Q) - v(P) \]
\[ \geq \psi_P(d(x^*_Q, S^*_P)), \]
where the first inequality follows from (11); the second inequality follows from Lemma 3.6; and the last inequality follows from the definition of the function $\psi_P$. Then, we obtain
\[ d(x^*_Q, S^*_P) \leq \psi_P^{-1}(2L\zeta_2(P, Q)) \]
\[ \leq L\zeta_2(P, Q) + \psi_P^{-1}(2L\zeta_2(P, Q)) \]
\[ = \Psi_P(L\zeta_2(P, Q)). \]
Since $x^*_Q \in S^*_Q$ is selected arbitrarily, we actually prove that
\[ S^*_Q \subseteq S^*_P + \Psi_P(L\zeta_2(P, Q))B. \]

4. Rates of convergence. In general, it is difficult to evaluate an expectation function. We consider the discrete approximation to problem (2) in this section. Let $\xi^1, \xi^2, \ldots, \xi^{N_k}$ be the independent and identically distributed samples according
to the probability distribution \( P \), where \( N_k \to \infty \) as \( k \to \infty \). Then we have the following SAA problem to problem (2) with the sample size \( N_k \), i.e.

\[
\min_{x \in S} f_k(x) := \frac{1}{N_k} \sum_{i=1}^{N_k} f(x, \xi^i).
\] (12)

Let \( S^*_k \) and \( v(k) \) be the optimal solution set and optimal value of problem (12), respectively. In [13], under Assumption 3.1 and \( \alpha \in (0, 2\mu_P) \), the authors have given the results that \( S^*_k \) is nonempty and bounded for every sufficiently large \( k \) and \( d(S^*_k, S^*_p) \to 0 \) with probability one as \( k \to \infty \). While they did not discuss the rate of convergence of \( S^*_k \) to \( S^*_p \). We next study the rates of convergence of \( S^*_k \) to \( S^*_p \) under different assumptions.

### 4.1. Exponential rate of convergence.

The exponential rate of convergence is a classical result in investigating stochastic programming or stochastic variational analysis. Theorem 4.1. Let Assumptions 3.1 and 4.1 hold and \( \alpha \in (0, 2\mu_P) \). Then for any \( \epsilon > 0 \), there exists positive constants \( C(\epsilon) \) and \( \beta(\epsilon) \) such that

\[
\Pr \left\{ \sup_{x \in S \cap \mathbb{R}^2} |f_k(x) - f_P(x)| \geq \epsilon \right\} \leq C(\epsilon)e^{-N_k\beta(\epsilon)}.
\]

As we have pointed out that under Assumption 3.1 and \( \alpha \in (0, 2\mu_P) \), \( S^*_k \) is nonempty and bounded for every sufficiently large \( k \). We then have the following exponential rate of convergence result.

**Theorem 4.1.** Let Assumptions 3.1 and 4.1 hold and \( \alpha \in (0, 2\mu_P) \). Then for any \( \epsilon > 0 \), there exists positive constants \( C(\epsilon) \) and \( \beta(\epsilon) \) such that

\[
\Pr \{ d(S^*_k, S^*_p) \geq \epsilon \} \leq C(\epsilon)e^{-N_k\beta(\epsilon)}
\]

for sufficiently large \( k \).
Proof. Since $S_k^*$ and $S_p^*$ are nonempty and bounded, we may assume that $S_k^*, S_p^* \subseteq S \cap rB$ for some $r > 0$. We first prove

\[ d(S_k^*, S_p^*) \leq \psi_p^{-1}\left(2 \sup_{x \in S \cap rB} |f_k(x) - f_P(x)|\right). \]  

(13)

Since $\sup_{x \in S \cap rB} |f_P(x) - f_k(x)| \geq |v(P) - v(k)|$, letting $x^k \in S_k^*$, we have

\[ 2 \sup_{x \in S \cap rB} |f_P(x) - f_k(x)| \geq f_P(x^k) - f_k(x^k) + v(k) - v(P) = f_P(x^k) - v(P) \geq \psi_p(d(x^k, S_p^*)). \]

Since $x^k \in S_k^*$ is arbitrary, we obtain $d(S_k^*, S_p^*) \leq \psi_p^{-1}\left(2 \sup_{x \in S \cap rB} |f_P(x) - f_k(x)|\right)$. By (13) and the nondecreasing property of $\psi_p$, we know that

\[ \Pr\{d(S_k^*, S_p^*) \geq \varepsilon\} \leq \Pr\left\{\psi_p^{-1}\left(2 \sup_{x \in S \cap rB} |f_P(x) - f_k(x)|\right) \geq \varepsilon\right\} \leq \Pr\left\{2 \sup_{x \in S \cap rB} |f_P(x) - f_k(x)| \geq \psi_p(\varepsilon)\right\}. \]  

(14)

On the other hand, by Proposition 2, it has

\[ \Pr\left\{2 \sup_{x \in S \cap rB} |f_P(x) - f_k(x)| \geq \psi_p(\varepsilon)\right\} \leq C\left(\frac{\psi_p(\varepsilon)}{2}\right) e^{-N_k\beta\left(\frac{\psi_p(\varepsilon)}{2}\right)}. \]  

(15)

Combining (14) and (15), we obtain

\[ \Pr\{d(S_k^*, S_p^*) \geq \varepsilon\} \leq C\left(\frac{\psi_p(\varepsilon)}{2}\right) e^{-N_k\beta\left(\frac{\psi_p(\varepsilon)}{2}\right)}. \]

We then complete the proof by letting $C(\varepsilon) := C\left(\frac{\psi_p(\varepsilon)}{2}\right)$ and $\beta(\varepsilon) := \beta\left(\frac{\psi_p(\varepsilon)}{2}\right)$. \hfill \Box

The exponential rate convergence is obtained under Assumption 4.1, which stands for light tailed distribution of $f(x, \xi)$ and $\kappa(\xi)$, see [10, Chapter 7]. For the cases that $f(x, \xi)$ and $\kappa(\xi)$ fail to satisfy light tailed distribution, such as heavy tailed distribution, it needs some new assumptions to derive the rate of convergence. Recently, Jiang et al [9] derived the polynomial rate of convergence of SAA under mild conditions. We here employ this result to derive the polynomial rate of convergence of $S_k^*$ to $S_p^*$.

4.2. Polynomial rate of convergence. In this subsection, we do the following assumption.

**Assumption 4.2.** For some positive even number $m$, assume that

(i). $E_P[f(x, \xi)^m] < +\infty$ for every $x \in S$;

(ii). $E_P[\kappa^m(\xi)] < +\infty$.

We immediately have the following proposition by Theorem 3.2 in [9].

**Proposition 3.** Let Assumption 4.2 hold. Then for any $\varepsilon > 0$, there exists positive scalar $C$, independent of $N_k$, such that

\[ \Pr\left\{\sup_{x \in S \cap rB} |f_k(x) - f_P(x)| \geq \varepsilon\right\} \leq \frac{C}{N_k^3 \varepsilon^m} \]  

(16)

for sufficiently large $k$. 

By Proposition 3, we obtain the following polynomial rate of convergence result.

**Theorem 4.2.** Let Assumptions 3.1 and 4.2 hold and $\alpha \in (0, 2\mu_P)$. Then for any $\varepsilon > 0$, there exists a positive constant $C$, independent of $N_k$, such that

$$\text{Pr} \{ d(S_k^*, S_P^*) \geq \Psi_P(\varepsilon) \} \leq \frac{C}{N_k^m \varepsilon^m}$$

for sufficiently large $k$.

**Proof.** We can obtain that $d(S_k^*, S_P^*) \leq \Psi_P^{-1}(2 \sup_{x \in S \cap r_B} |f_k(x) - f_P(x)|)$ in a similar manner to the proof of Theorem 4.1. Hence we have

$$d(S_k^*, S_P^*) \leq \Psi_P(\sup_{x \in S \cap r_B} |f_k(x) - f_P(x)|).$$

Then, by (16), it has

$$\text{Pr} \{ d(S_k^*, S_P^*) \geq \Psi_P(\varepsilon) \} \leq \text{Pr} \left\{ \Psi_P \left( \sup_{x \in S \cap r_B} |f_k(x) - f_P(x)| \right) \geq \Psi_P(\varepsilon) \right\}$$

$$= \text{Pr} \left\{ \sup_{x \in S \cap r_B} |f_k(x) - f_P(x)| \geq \varepsilon \right\}$$

$$\leq \frac{C}{N_k^m \varepsilon^m}.$$

This completes the proof. $\square$

Generally, the polynomial rate is slower than the exponential rate. However, the Assumption 4.2 is much weaker than the Assumption 4.1. The Assumption 4.2 only needs some finite moment information, which can be easily satisfied in many applications. It could be a powerful tool in the situation that the underlying distributions are heavy tailed and thus the classical exponential rate of convergence under light tailed distributions fails.

5. **Conclusions.** In this paper, we studied the quantitative stability of the ERM formulation for a class of stochastic linear variational inequalities. We first discussed the existence of solution sets of the ERM formulation and its perturbed distribution problem. Then, we derived the quantitative relationship between this two solution sets. Finally, we investigated the discrete approximation to the ERM formulation, and obtained the rates of convergence of the solution sets under different assumptions.

**Acknowledgments.** The authors thank anonymous referees for their professional and helpful comments. The research was supported by the National Natural Science Foundation of China [grant numbers: 11971078], the China Postdoctoral Science Foundation [grant number: 2019M653332] and the Natural Science Foundation of Chongqing [grant number: cstc2019cjyj-hshX0092].

**REFERENCES**

[1] 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, *Pac. J. Optim.*, 6 (2010), 3–19.

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

[3] X. Chen, R. J.-B. Wets and Y. Zhang, Stochastic variational inequalities: Residual minimization smoothing sample average approximations, *SIAM J. Optim.*, 22 (2012), 649–673.
[4] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer, New York, 2003.

[5] H. Fang, X. Chen and M. Fukushima, Stochastic $R_0$ matrix linear complementarity problems, *SIAM J. Optim.*, 18 (2007), 482–506.

[6] J. Jiang and Z. Chen, Quantitative stability of multistage stochastic programs via calm modifications, *Oper. Res. Lett.*, 46 (2018), 543–547.

[7] J. Jiang and Z. Chen, Quantitative stability analysis of two-stage stochastic linear programs with full random recourse, *Numer. Func. Anal. Optim.*, 40 (2019), 1847–1876.

[8] J. Jiang, X. Chen and Z. Chen, Quantitative analysis for a class of two-stage stochastic linear variational inequality problems, *Comput. Optim. Appl.*, 76 (2020), 431–460.

[9] J. Jiang, Z. Chen and X. M. Yang, Rates of convergence of sample average approximation under heavy tailed distributions, To preprint on Optimization Online.

[10] Z. Lin and Z. Bai, *Probability Inequalities*, Springer Science & Business Media, 2011.

[11] J. Liu and S. Li, Unconstrained optimization reformulation for stochastic nonlinear complementarity problems, *Appl. Anal.*, (2019).

[12] F. Lu and S.-J. Li, Method of weighted expected residual for solving stochastic variational inequality problems, *Appl. Math. Comput.*, 269 (2015), 651–663.

[13] M. J. Luo and G. H. Lin, Expected residual minimization method for stochastic variation inequality problems, *J. Optim. Theory Appl.*, 140 (2009), 103–116.

[14] M. J. Luo and G. H. Lin, Convergence results of the ERM method for nonlinear stochastic variation inequality problems, *J. Optim. Theory Appl.*, 142 (2009), 569–581.

[15] L. Mirsky, Symmetric gauge functions and unitarily invariant norms, *Quart J. Math. Oxford*, 11 (1960), 50–59.

[16] S. T. Rachev, L. B. Klebanov, S. V. Stoyanov and F. J. Fobozzi, *The Methods of Distances in the Theory of Probility and Statistics*, Springer, New York, 2013.

[17] S. T. Rachev and W. Römsch, Quantitative stability in stochastic programming: The method of probability metrics, *Math. Oper. Res.*, 27 (2002), 792–818.

[18] W. Römsch, Stability of stochastic programming problems, *Stochastic programming*, 10 (2003), 483–554.

[19] A. Shapiro, D. Dentcheva and A. Ruszczyński, *Lectures on Stochastic Programming: Modeling and Theory*, SIAM, Philadelphia, 2014.

[20] A. Shapiro and H. Xu, Stochastic mathematical programs with with equilibrium constraints, modelling and sample average approximation, *Optimization*, 57 (2008), 395–418.

[21] H. 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.

[22] H. Xu, Sample average approximation methods for a class of stochastic variational inequality problems, *Asia-Pac. J. Oper. Res.*, 27 (2010), 103–119.

[23] Y. Zhang and X. Chen, Regularizations for stochastic linear variational inequalities, *J. Optim. Theory Appl.*, 163 (2014), 460–481.

[24] Y. Zhao, J. Zhang, X. Yang and G.-H. Lin, Expected residual minimization formulation for a class of stochastic vector variational inequalities, *J. Optim. Theory Appl.*, 175 (2017), 545–566.