Expected Residual Minimization Formulation for Stochastic Absolute Value Equations

Jingyong Tang[1] · Jinchuan Zhou[2]

Abstract

In this paper we investigate a class of stochastic absolute value equations (SAVE). After establishing the relationship between the stochastic linear complementarity problem and SAVE, we study the expected residual minimization (ERM) formulation for SAVE and its Monte Carlo sample average approximation. In particular, we show that the ERM problem and its sample average approximation have optimal solutions under the condition of R_0 pair, and the optimal value of the sample average approximation has uniform exponential convergence. Furthermore, we prove that the solutions to the ERM problem are robust for SAVE. For a class of SAVE problems, we use its special structure to construct a smooth residual and further study the convergence of the stationary points. Finally, a smoothing gradient method is proposed by simultaneously considering sample sampling and smooth techniques for solving SAVE. Numerical experiments exhibit the effectiveness of the method.

Keywords Stochastic absolute value equations · Expected residual minimization formulation · Monte Carlo approximation · Smoothing gradient method

✉ Jingyong Tang
tangjy@xynu.edu.cn

Jinchuan Zhou
jinchuanzhou@163.com

1 School of Mathematics and Statistics, Xinyang Normal University, Xinyang 464000, China

2 School of Mathematics and Statistics, Shandong University of Technology, Zibo 255049, China

Springer

1 Introduction

The absolute value equation (AVE) is to find a vector $x \in \Re^n$ such that

$$Ax + B|x| = b, \tag{1}$$

where $A \in \Re^{n \times n}$, $B \in \Re^{n \times n}$, $b \in \Re^n$ and $|x|$ denotes the component-wise absolute value of x. If $B = -I$, where I denotes an identity matrix, then AVE (1) reduces to the special form:

$$Ax - |x| = b. \tag{2}$$

The main significance of AVE arises from its wide applications in the different fields of mathematics and engineering applications. For example, the general linear complementarity problem [3] (including linear programs), quadratic programs, bimatrix games and some other mathematical programming problems can be formulated as AVE; see [13] for more information. Particularly, if $B = 0$, then AVE (1) reduces to the system of linear equations $Ax = b$, the most foundamental problem in scientific computing. Over the last ten years, AVE (1) and its special form (2) have been extensively studied from both theoretical analysis and numerical algorithms, see [2, 9–12, 20, 26, 29, 30, 32].

In this paper, we mainly study a class of stochastic absolute value equation (SAVE), which is to find a vector $x \in \Re^n$ such that

$$A(\omega)x + B(\omega)|x| = b(\omega), \tag{3}$$

where $\omega \in \Omega$ and $(\Omega, \mathcal{F}, P)$ is a probability space, $A(\omega) \in \Re^{n \times n}$, $B(\omega) \in \Re^{n \times n}$ and $b(\omega) \in \Re^n$ for each $\omega \in \Omega$. If Ω only contains one realization, then SAVE (3) would reduce to AVE (1). One of our motivations for studying SAVE comes from the fact that the well-known stochastic linear complementarity problem (SLCP)

$$u \geq 0, \ v \geq 0, \ v = M(\omega)u + q(\omega), \ u^T v = 0 \tag{4}$$

with $M(\omega) \in \Re^{n \times n}$ and $q(\omega) \in \Re^n$ can be formulated as a SAVE (see Sect. 2 below). In the past few decades, due to its widespread application, SLCP has been extensively studied. Here we only list a few literatures that are closely related to this paper. For example, Chen and Fukushima [4] introduced the Expected Residual Minimization (ERM) formula for SLCP and studied some properties of this new reformulation. Chen et al. [6] further proved that the solutions of the ERM formulation are robust for the monotone SLCP. Zhang and Chen [33] investigated a smooth projected gradient method and applied this algorithm to solve SLCP. Zhou and Caccetta [34] proposed a feasible semismooth Newton method for solving a class of SLCPs with finite many realizations. More research and discussions on SLCP can be found in [7, 16, 17] and references therein. For stochastic absolute value equation, Du [8] proved that the SAVE $A(\omega)x - |x| = b(\omega)$, a special case of SAVE (3), is equivalent to the stochastic bilinear programming and the stochastic generalized linear complementarity problem.

Due to the existence of the randomness of ω, we do not expect the existence of a vector x satisfying SAVE (3) for all $\omega \in \Omega$. Therefore, in order to obtain a reasonable solution of (3), we need to search for suitable reformulations for SAVE. It is well-known that the deterministic Expected Residual Minimization (ERM) formulation is one of the most effective techniques to deal with stochastic optimization problems. The ERM formulation was firstly proposed by Chen and Fukushima [4] for solving SLCP, where the objective is to minimize an expected residual defined by NCP functions. Subsequently, the ERM formulation has been further studied to deal with stochastic variational inequalities [5], stochastic complementarity problems [14], stochastic generalized complementarity problems [18] and stochastic second-order cone complementarity problems [15, 19, 31].

Our goal is to study the ERM deterministic formulation for SAVE (3) and then propose a smoothing gradient method to find its approximate solution. Our main contributions of the paper can be summarized as follows.

(i) We first show that the SLCP can be reformulated as a SAVE, while the converse direction holds as well provided that $A(\omega) + B(\omega)$ is nonsingular for each $\omega \in \Omega$.

(ii) We introduce the ERM formulation for SAVE and its Monte Carlo sample average approximation. The existence of optimal solutions to ERM problem and its discrete approximation is established when $\{A(\bar{\omega}), B(\bar{\omega})\}$ is an R_0 pair for an observation $\bar{\omega}$. Our results show that the solution of the approximate ERM problem converges to that of ERM problem with probability one, and the optimal value of the sample average approximation achieves a uniform exponential convergence.

(iii) Under some regularity conditions, we prove that the solutions of ERM problem are robust in the sense that they may have a minimum sensitivity with respect to random parameter variations in SAVE. This also indicates from a new perspective that ERM is an effective reformulation of SAVE.

(iv) In general, SAVE is nonsmooth due to the existence of absolute values. However, for a special class of SAVE problems, by exploiting its structure we provide a smooth (even twice continuously differentiable) residual and further study the convergence of the stationary points.

(v) For a general SAVE, we propose a new smooth approximation formulation by simultaneously considering sampling and smoothing techniques (in which the number of samples can also be treated as a smoothing factor). This new smooth approximation allows us to use a smooth gradient method to solve SAVE. Our numerical experiments indicate the algorithm is very effective.

The rest of our paper is organized as follows. In the next section, we discuss the relationship between SAVE and SLCP. In Sect. 3, we study the ERM formulation for SAVE and investigate the boundedness of its level sets, the existence of optimal solutions as well as the exponential convergence rate. The robustness of the solutions to the ERM problem is given in Sect. sec4. Section 5 is devoted to a smoothing reformation for a class of special SAVE. Section 6 proposes a smoothing gradient method for solving general SAVE. Conclusions are drawn in Sect. 7.

2 Relations of SAVE and SLCP

Theorem 1 *The* SLCP *(4) can be formulated as a* SAVE *(3) with $A(\omega) = I + M(\omega)$, $B(\omega) = I - M(\omega)$ and $b(\omega) = 2q(\omega)$.*

Proof Note that SLCP (4) can be written as

$$u \geq 0, \ v \geq 0, \ (I + M(\omega))(v - u) + (I - M(\omega))(v + u) = 2q(\omega), \ u^T v = 0.$$

Let $x := v - u$. Then, for any $i = 1, ..., n$, since $u_i \geq 0$, $v_i \geq 0$ and $u_i v_i = 0$, we have $x_i^2 = (v_i - u_i)^2 = (v_i + u_i)^2$ and so $|x_i| = v_i + u_i$. This implies that $|x| = v + u$. So, SLCP (4) reduces to SAVE (3) with $A(\omega) = I + M(\omega)$, $B(\omega) = I - M(\omega)$ and $b(\omega) = 2q(\omega)$. $\qquad \square$

Theorem 2 *If the matrix $A(\omega) + B(\omega)$ is nonsingular for each $\omega \in \Omega$, then* SAVE *(3) is equivalent to* SLCP *(4).*

Proof According to Theorem 1, it only needs to show that SAVE (3) can be formulated as a SLCP (4). Note that SAVE (3) can be written as

$$(A(\omega) + B(\omega))\frac{x + |x|}{2} + (A(\omega) - B(\omega))\frac{x - |x|}{2} = b(\omega).$$

Since $A(\omega) + B(\omega)$ is nonsingular for each $\omega \in \Omega$, the above equation is equivalent to

$$\frac{|x| + x}{2} = (A(\omega) + B(\omega))^{-1}(A(\omega) - B(\omega))\frac{|x| - x}{2} + (A(\omega) + B(\omega))^{-1}b(\omega). \tag{5}$$

By setting $v := \frac{|x|+x}{2}$, $u := \frac{|x|-x}{2}$, $M(\omega) := (A(\omega) + B(\omega))^{-1}(A(\omega) - B(\omega))$ and $q(\omega) := (A(\omega) + B(\omega))^{-1}b(\omega)$, SAVE (5) is reformulated as SLCP (4). $\qquad \square$

Remark 1 Theorem 1 provides an equivalent way to solve SLCP (4). Specifically, we may first find a solution x by solving the following SAVE

$$(I + M(\omega))x + (I - M(\omega))|x| = 2q(\omega).$$

Then setting $v := \frac{|x|+x}{2}$ and $u := \frac{|x|-x}{2}$ yields a solution (u, v) for SLCP (4).

3 ERM Formulation for SAVE

Let

$$F(x, \omega) := A(\omega)x + B(\omega)|x| - b(\omega). \tag{6}$$

We introduce the following ERM model for the SAVE (3):

$$\min_{x \in \Re^n} \Theta(x) := \mathbb{E}[\|F(x, \omega)\|^2], \tag{7}$$

where $\mathbb{E}$ denotes the expectation with respect to the random variable ω. To proceed, we first give the assumption required in this section.

Assumption A Let ω be a continuous random variable. Let $A(\omega)$, $B(\omega)$ and $b(\omega)$ be continuous functions of ω satisfying

$$\mathbb{E}[(\|A(\omega)\| + \|B(\omega)\| + \|b(\omega)\|)^2]$$
$$= \int_{\Omega} (\|A(\omega)\| + \|B(\omega)\| + \|b(\omega)\|)^2 \rho(\omega) d\omega < +\infty,$$

where $\rho : \Omega \to \Re_+$ denotes the continuous probability density function of ω.

To analyze the existence of optimal solutions to the ERM problem (7), we introduce the following definition of R_0 pair.

Definition 1 For the given matrices $A, B \in \Re^{n \times n}$, we define $\{A, B\}$ to be an R_0 pair if

$$Ax + B|x| = 0 \implies x = 0.$$

The definition of R_0 pair is motivated by the original definition of R_0 matrix for the LCP [4]. Obviously, if AVE (1) is uniquely solvable for any $b \in \Re^n$, then $\{A, B\}$ is an R_0 pair. According to existing results on the unique solvability of AVE (1), $\{A, B\}$ is an R_0 pair if one of the following conditions holds:

(a) $\sigma_{\max}(|B|) < \sigma_{\min}(A)$, where $\sigma_{\max}$ and $\sigma_{\min}$ denote the maximal and minimal singular value respectively [25];

(b) The inequality $|Ax| \le |B||x|$ only has trivial solutions [24];

(c) The interval matrix $[A - |B|, A + |B|]$ is regular [30];

(d) $\{A + B, A - B\}$ has the column $\mathcal{W}$-property [20].

More sufficient conditions for the unique solvability of AVE (1) can be found in [26].

Lemma 1 If $\{A(\bar{\omega}), B(\bar{\omega})\}$ is an R_0 pair for some $\bar{\omega} \in \Omega$, then there exists a closed sphere $\mathcal{B}(\bar{\omega}, \delta) := \{\omega| \|\omega - \bar{\omega}\| \le \delta\}$ with $\delta > 0$ such that $\{A(\omega), B(\omega)\}$ is an R_0 pair for every $\omega \in \mathcal{B}(\bar{\omega}, \delta) \cap \Omega$.

Proof Assume that the lemma is not true. Then there is a sequence $\{\omega^k\} \subset \Omega$ satisfying $\lim_{k \to \infty} \omega^k = \bar{\omega}$ and, for each k, there exists $x^k \in \Re^n$ such that

$$A(\omega^k)x^k + B(\omega^k)|x^k| = 0, \quad x^k \ne 0.$$

Let $t^k := \frac{x^k}{\|x^k\|}$. Then $\|t^k\| = 1$ and

$$A(\omega^k)t^k + B(\omega^k)|t^k| = 0.$$

This implies the existence of $\bar{t} \in \Re^n$ such that

$$\|\bar{t}\| = 1, \quad A(\bar{\omega})\bar{t} + B(\bar{\omega})|\bar{t}| = 0.$$

This leads to a contradiction with the assumption that $\{A(\bar{\omega}), B(\bar{\omega})\}$ is an R_0 pair. $\square$

Theorem 3 *Suppose that there exists some* $\bar{\omega} \in \Omega$ *such that* $\rho(\bar{\omega}) > 0$ *and* $\{A(\bar{\omega}), B(\bar{\omega})\}$ *is an* R_0 *pair. Then for any* $\gamma > 0$, *the level set defined by*

$$\mathcal{L}(\gamma) := \{x \in \Re^n | \Theta(x) \leq \gamma\}$$

is bounded.

Proof According to the continuity of ρ and Lemma 1, there exist a constant $\bar{\rho} > 0$ and a closed sphere $\mathcal{B}(\bar{\omega}, \delta)$ such that $\rho(\omega) \geq \bar{\rho}$ and $\{A(\omega), B(\omega)\}$ is an R_0 pair for any $\omega \in \bar{\mathcal{B}} := \mathcal{B}(\bar{\omega}, \delta) \cap \Omega$. Suppose on the contrary that there exists a constant $\gamma > 0$ and a sequence $\{x^k\} \subset \mathcal{L}(\gamma)$ such that $\lim_{k \to \infty} \|x^k\| = +\infty$. The continuity of $A(\omega), B(\omega)$ and $b(\omega)$ ensures that for every k there exists an $\omega^k \in \bar{\mathcal{B}}$ such that

$$\|F(x^k, \omega^k)\| = \min_{\omega \in \bar{\mathcal{B}}} \|F(x^k, \omega)\|.$$

Hence

$$\gamma \geq \Theta(x^k) \geq \int_{\bar{\mathcal{B}}} \|F(x^k, \omega)\|^2 \rho(\omega) d\omega \geq \|F(x^k, \omega^k)\|^2 \bar{\rho} \int_{\bar{\mathcal{B}}} d\omega,$$

which together with $\int_{\bar{\mathcal{B}}} d\omega > 0$ implies that $\{\|F(x^k, \omega^k)\|\}$ is bounded. So, there exists a constant $C > 0$ such that

$$\|F(x^k, \omega^k)\| = \|A(\omega^k)x^k + B(\omega^k)|x^k| - b(\omega^k)\| \leq C. \tag{8}$$

Since the sequences $\left\{\frac{x^k}{\|x^k\|}\right\}$ and $\{\omega_k\}$ are bounded, we can assume by taking a subsequence if necessary that they are convergent and denote the limits as $x^* := \lim_{k \to \infty} \frac{x^k}{\|x^k\|}$ and $\omega^* := \lim_{k \to \infty} \omega^k$. Clearly, $\|x^*\| = 1$ and $\omega^* \in \bar{\mathcal{B}}$. By (8), we have

$$\left\| A(\omega^k) \frac{x^k}{\|x^k\|} + B(\omega^k) \left| \frac{x^k}{\|x^k\|} \right| - \frac{b(\omega^k)}{\|x^k\|} \right\| \leq \frac{C}{\|x^k\|}. \tag{9}$$

Since $\{b(\omega^k)\}$ is bounded due to the continuity of $b(\omega)$, by letting k tend towards infinity in (9), it holds that

$$A(\omega^*)x^* + B(\omega^*)|x^*| = 0.$$

Hence $x^* = 0$ because $\{A(\omega^*), B(\omega^*)\}$ is an R_0 pair. This yields a contradiction with $\|x^*\| = 1$. The proof is complete. $\square$

As is well known, for an integrable function $\varphi : \Omega \to \Re$, the Monte Carlo sampling estimate for $\mathbb{E}[\varphi(\omega)]$ is obtained by taking independently and identically distributed random samples $\Omega_k := \{\omega^1, ..., \omega^{N_k}\}$ from Ω and letting

$$\mathbb{E}[\varphi(\omega)] \approx \frac{1}{N_k} \sum_{\omega^i \in \Omega_k} \varphi(\omega^i).$$

The strong law of large numbers guarantees that this procedure converges with probability one (abbreviated by "w.p.1" below), that is,

$$\lim_{k \to \infty} \frac{1}{N_k} \sum_{\omega^i \in \Omega_k} \varphi(\omega^i) = \mathbb{E}[\varphi(\omega)], \ \text{w.p.1}, \tag{10}$$

where $N_k \to +\infty$ as $k \to \infty$.

By generating independently and identically distributed random samples $\Omega_k = \{\omega^1, ..., \omega^{N_k}\}$ from Ω, we can obtain the following approximation of the ERM problem (7):

$$\min_{x \in \Re^n} \Theta_k(x) := \frac{1}{N_k} \sum_{\omega^i \in \Omega_k} \|F(x, \omega^i)\|^2. \tag{11}$$

According to (6), the function $\Theta_k(x)$ can be further written as

$$\Theta_k(x) = \frac{1}{N_k} \sum_{\omega^i \in \Omega_k} \|A(\omega^i)x + B(\omega^i)|x| - b(\omega^i)\|^2. \tag{12}$$

We now give the existence of global optimal solutions to (7) and (11). To this end, we denote the set of optimal solutions to the ERM problem (7) by $\mathcal{S}^*$ and those of the approximate ERM problem (11) by $\mathcal{S}_k^*$ respectively

Theorem 4 *Let Assumption* A *hold. Suppose that there exists an $\bar{\omega} \in \Omega$ such that $\rho(\bar{\omega}) > 0$ and $\{A(\bar{\omega}), B(\bar{\omega})\}$ is an R_0 pair. Then $\mathcal{S}_k^*$ is nonempty and bounded for all large k. Let $x^k \in \mathcal{S}_k^*$ for each k and x^* be an accumulation point of the sequence $\{x^k\}$. Then $x^* \in \mathcal{S}^*$ with probability one.*

Proof By Lemma 1, there exist an index $\bar{k} > 0$ and a closed sphere $\mathcal{B}(\bar{\omega}, \delta)$ such that for all $k \geq \bar{k}$, $\Omega_k \cap \mathcal{B}(\bar{\omega}, \delta)$ are nonempty and $\{A(\omega), B(\omega)\}$ is an R_0 pair for each $\omega \in \Omega_k \cap \mathcal{B}(\bar{\omega}, \delta)$. Hence, by following a similar argument as in Theorem 3, we can show that S_k^* is nonempty and bounded as $k \geq \bar{k}$.

Next, we prove $x^* \in \mathcal{S}^*$. In fact, for any fixed $x \in \Re^n$,

$$\begin{aligned} \|F(x, \omega)\| &= \|A(\omega)x + B(\omega)|x| - b(\omega)\| \\ &\leq (\|A(\omega)\| + \|B(\omega)\|)\|x\| + \|b(\omega)\| \\ &\leq (\|A(\omega)\| + \|B(\omega)\| + \|b(\omega)\|)(\|x\| + 1). \end{aligned} \tag{13}$$

This together with Assumption A implies that the function $\|F(x, \omega)\|^2$ is integrable over Ω. Thus, according to (10), for any fixed $x \in \Re^n$,

$$\lim_{k \to +\infty} \Theta_k(x) = \Theta(x), \quad \text{w.p.1.} \tag{14}$$

We now assume by taking a subsequence if necessary that $\lim_{k \to \infty} x^k = x^*$. Pick up a constant $\zeta > 0$ satisfying $\Theta(x^*) < \zeta$. Since

$$\|F(x, \omega)\|^2 \leq (\|A(\omega)\| + \|B(\omega)\| + \|b(\omega)\|)^2 (\|x\| + 1)^2,$$

and $\mathbb{E}[(\|A(\omega)\| + \|B(\omega)\| + \|b(\omega)\|)^2] < +\infty$ by Assumption A, by [22, Theorem 16.8], we know that $\Theta(x) = \mathbb{E}[\|F(x, \omega)\|^2]$ is continuous at any $x \in \mathbb{R}^n$. Hence $\Theta(x^k) \leq \zeta$ for all sufficiently large k. Since the level set $\mathcal{L}(\zeta) = \{x \in \Re^n | \Theta(x) \leq \zeta\}$ is closed and bounded by Theorem 3, the constant

$$\tilde{C} := \max\{\|x\| \mid x \in \mathcal{L}(\zeta)\} \tag{15}$$

is well-defined. For all sufficiently large k and any $\omega \in \Omega$, we have

$$\begin{aligned}
&\left| \|A(\omega)x^k + B(\omega)|x^k| - b(\omega)\| - \|A(\omega)x^* + B(\omega)|x^*| - b(\omega)\| \right| \\
&\quad \leq \|A(\omega)\| \|x^k - x^*\| + \|B(\omega)\| \| |x^k| - |x^*| \| \\
&\quad \leq (\|A(\omega)\| + \|B(\omega)\|) \|x^k - x^*\| \\
&\quad \leq (\|A(\omega)\| + \|B(\omega)\| + \|b(\omega)\|) \|x^k - x^*\|.
\end{aligned} \tag{16}$$

Thus, by (13), (15) and (16), it follows from (12) that

$$\begin{aligned}
&|\Theta_k(x^k) - \Theta_k(x^*)| \\
&\leq \frac{1}{N_k} \sum_{\omega^i \in \Omega_k} \left| \|A(\omega^i)x^k + B(\omega^i)|x^k| - b(\omega^i)\|^2 \right. \\
&\qquad \left. - \|A(\omega^i)x^* + B(\omega^i)|x^*| - b(\omega^i)\|^2 \right| \\
&\leq \frac{1}{N_k} \sum_{\omega^i \in \Omega_k} 2(\tilde{C} + 1)(\|A(\omega^i)\| + \|B(\omega^i)\| + \|b(\omega^i)\|)^2 \|x^k - x^*\| \\
&\to 0, \quad \text{w.p.1,}
\end{aligned} \tag{17}$$

where the limit holds because $\{x^k\}$ converges to x^* and

$$\begin{aligned}
&\lim_{k \to \infty} \frac{1}{N_k} \sum_{\omega^i \in \Omega_k} (\|A(\omega^i)\| + \|B(\omega^i)\| + \|b(\omega^i)\|)^2 \\
&\quad = \mathbb{E}[(\|A(\omega)\| + \|B(\omega)\| + \|b(\omega)\|)^2] < +\infty, \quad \text{w.p.1}
\end{aligned}$$

by Assumption A and (10). Further, putting (14) and (17) together yields

$$|\Theta_k(x^k) - \Theta(x^*)| \leq |\Theta_k(x^k) - \Theta_k(x^*)| + |\Theta_k(x^*) - \Theta(x^*)| \to 0, \text{ w.p.1. } \quad (18)$$

On the other hand, since x^k is the optimal solution of the problem (11), we have

$$\Theta_k(x^k) \leq \Theta_k(x), \quad \forall x \in \Re^n.$$

This together with (14) and (18) gives

$$\Theta(x^*) = \lim_{k \to \infty} \Theta_k(x^k) \leq \lim_{k \to \infty} \Theta_k(x) = \Theta(x), \ \forall x \in \Re^n$$

with probability one. Hence $x^* \in \mathcal{S}^*$ with probability one. $\qquad \square$

The following result shows that the optimal values of the sample average approximations have a uniform exponential convergence.

Theorem 5 *Suppose that the support set Ω is a compact set. Let x^k be an optimal solution of (11) for each k and x^* be an accumulation point of the sequence $\{x^k\}$. Then, for every $\varepsilon > 0$, there exist positive constants $D(\varepsilon)$ and $\beta(\varepsilon)$, independent of N_k, such that*

$$\text{Prob}\left\{|\Theta(x^k) - \Theta(x^*)| \geq \varepsilon\right\} \leq D(\varepsilon)e^{-N_k\beta(\varepsilon)}.$$

Proof We assume again without loss of generality that $\{x^k\}$ itself converges to x^*. Let $\mathcal{Z}$ be a compact set that contains the whole sequence $\{x^k\}$. Then $x^* \in \mathcal{Z}$. First, since $\mathcal{Z}$ and Ω are compact sets, by the continuity of F, there exists a constant $c > 0$ such that

$$\|F(x, \omega)\| \leq c, \ \forall (x, \omega) \in \mathcal{Z} \times \Omega. \quad (19)$$

So for every $x \in \mathcal{Z}$, the moment generating function $\mathbb{E}\left[e^{t(\|F(x,\omega)\|^2 - \mathbb{E}[\|F(x,\omega)\|^2])}\right]$ of the random variable $\|F(x, \omega)\|^2 - \mathbb{E}[\|F(x, \omega)\|^2]$ is finite-valued whenever t near zero. Second, for any $(x', \omega), (x, \omega) \in \mathcal{Z} \times \Omega$, it follows from (19) that

$$\begin{aligned}
|\|F(x', \omega)\|^2 - \|F(x, \omega)\|^2| &\leq 2c|\|F(x', \omega)\| - \|F(x, \omega)\|| \\
&\leq 2c\|F(x', \omega) - F(x, \omega)\| \\
&= 2c\|A(\omega)(x' - x) + B(\omega)(|x'| - |x|)\| \\
&\leq 2c(\|A(\omega)\| + \|B(\omega)\|)\|x' - x\|. \quad (20)
\end{aligned}$$

Let $\kappa(\omega) := 2c(\|A(\omega)\| + \|B(\omega)\|)$. Then $\kappa(\omega)$ is bounded on Ω due to the continuity of $A(\omega)$ and $B(\omega)$. This ensures that $\mathbb{E}[\kappa(\omega)] < +\infty$ and the moment generating function $\mathbb{E}[e^{t\kappa(\omega)}]$ of the random variable $\kappa(\omega)$ is finite-valued for all t near of zero.

Thus, the conditions (C1)–(C3) assumed in [28, Theorem 5.1] are all satisfied. Therefore, according to [28, Theorem 5.1], for every $\varepsilon > 0$, there exist positive constants $D(\varepsilon)$ and $\beta(\varepsilon)$, independent of N_k, such that

$$\text{Prob}\left\{\sup_{x\in\mathcal{Z}}\left|\frac{1}{N_k}\sum_{\omega^i\in\Omega_k}\|F(x,\omega^i)\|^2 - \mathbb{E}[\|F(x,\omega)\|^2]\right| \geq \varepsilon/2\right\} \leq \frac{1}{2}D(\varepsilon)e^{-N_k\beta(\varepsilon)},$$

i.e.,

$$\text{Prob}\left\{\sup_{x\in\mathcal{Z}}|\Theta_k(x) - \Theta(x)| \geq \varepsilon/2\right\} \leq \frac{1}{2}D(\varepsilon)e^{-N_k\beta(\varepsilon)}. \tag{21}$$

For each k, since x^k is an optimal solution of (11), we have $\Theta_k(x^k) \leq \Theta_k(x^*)$ and hence

$$\begin{aligned}
\Theta_k(x^k) - \Theta(x^*) &= \Theta_k(x^k) - \Theta_k(x^*) + \Theta_k(x^*) - \Theta(x^*) \\
&\leq \Theta_k(x^*) - \Theta(x^*) \\
&\leq \sup_{x\in\mathcal{Z}}|\Theta_k(x) - \Theta(x)|.
\end{aligned} \tag{22}$$

On the other hand, by Theorem 4, x^* is an optimal solution of (7) with probability one. So for each k, $\Theta(x^*) \leq \Theta(x^k)$ with probability one and hence

$$\begin{aligned}
\Theta_k(x^k) - \Theta(x^*) &= \Theta_k(x^k) - \Theta(x^k) + \Theta(x^k) - \Theta(x^*) \\
&\geq \Theta_k(x^k) - \Theta(x^k) \\
&\geq -\sup_{x\in\mathcal{Z}}|\Theta_k(x) - \Theta(x)|.
\end{aligned} \tag{23}$$

Combining (22) and (23) together yields

$$|\Theta_k(x^k) - \Theta(x^*)| \leq \sup_{x\in\mathcal{Z}}|\Theta_k(x) - \Theta(x)|.$$

This together with (21) gives

$$\begin{aligned}
\text{Prob}\left\{|\Theta_k(x^k) - \Theta(x^*)| \geq \varepsilon/2\right\} &\leq \text{Prob}\left\{\sup_{x\in\mathcal{Z}}|\Theta_k(x) - \Theta(x)| \geq \varepsilon/2\right\} \\
&\leq \frac{1}{2}D(\varepsilon)e^{-N_k\beta(\varepsilon)}.
\end{aligned} \tag{24}$$

Since

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

it follows from (21) and (24) that

$$\text{Prob}\left\{|\Theta(x^k) - \Theta(x^*)| \geq \varepsilon\right\}$$
$$\leq \text{Prob}\left\{|\Theta(x^k) - \Theta_k(x^k)| \geq \varepsilon/2\right\} + \text{Prob}\left\{|\Theta_k(x^k) - \Theta(x^*)| \geq \varepsilon/2\right\}$$
$$\leq \text{Prob}\left\{\sup_{x \in \mathcal{Z}} |\Theta(x) - \Theta_k(x)| \geq \varepsilon/2\right\} + \text{Prob}\left\{|\Theta_k(x^k) - \Theta(x^*)| \geq \varepsilon/2\right\}$$
$$\leq D(\varepsilon)e^{-N_k\beta(\varepsilon)}.$$

This completes the proof. □

At the end of this section, we give some discussions on the computational complexity of the Monte Carlo approximation method. Denote $\pi(x, \omega) := \|F(x, \omega)\|^2 - \mathbb{E}[\|F(x, \omega)\|^2]$. We further assume that the moment generating function $\mathbb{E}\left[e^{t\pi(x,\omega)}\right]$ of the random variable $\pi(x, \omega)$ satisfies the following condition: there exists $\sigma > 0$ such that for all $x \in \mathbb{R}^n$,

$$\mathbb{E}\left[e^{t(\pi(x,\omega))}\right] \leq \exp(\sigma^2 t^2/2), \ \forall t \in \mathbb{R}. \tag{25}$$

This condition was introduced by Shapiro and Xu [28]. It is noticed that the right side of (25) is just the moment generating function of normal distribution with a variance σ^2. Hence, the inequality (25) holds as an equality as $\pi(x, \omega)$ has normal distribution. In addition, this condition can also be satisfied if $\pi(x, \omega)$ has a subguass distribution.

From the proof process of Theorem 5, we know that the conditions (C1)–(C3) assumed in [28, Theorem 5.1] are all satisfied. Moreover, by (20) we have

$$|\mathbb{E}[\|F(x', \omega)\|^2] - \mathbb{E}[\|F(x, \omega)\|^2]| \leq \mathbb{E}[2c(\|A(\omega)\| + \|B(\omega)\|)]\|x' - x\|, \ \forall x', x \in \mathcal{Z}.$$

Let $L := \mathbb{E}[2c(\|A(\omega)\| + \|B(\omega)\|)]$. Obviously, L is finite on the compact set Ω due to the continuity of $A(\omega)$ and $B(\omega)$. Denote by $D := \sup_{x', x \in \mathcal{Z}} \|x' - x\|$ the diameter of $\mathcal{Z}$. Then, according to (5.14) in [28], we know that for $\alpha \in (0, 1)$, if the sample size N_k satisfies

$$N_k \geq \frac{\mathcal{O}(1)\sigma^2}{\varepsilon^2}\left[n\log\left(\frac{\mathcal{O}(1)DL}{\varepsilon}\right) + \log\left(\frac{2}{\alpha}\right)\right], \tag{26}$$

then

$$\text{Prob}\left\{\sup_{x \in \mathcal{Z}}\left|\frac{1}{N_k}\sum_{\omega^i \in \Omega_k} \|F(x, \omega^i)\|^2 - \mathbb{E}[\|F(x, \omega)\|^2]\right| \geq \varepsilon/2\right\} \leq \frac{\alpha}{2},$$

i.e.,

$$\text{Prob}\left\{\sup_{x \in \mathcal{Z}} |\Theta_k(x) - \Theta(x)| \geq \varepsilon/2\right\} \leq \frac{\alpha}{2}.$$

Then by following the proof of Theorem 5, we can obtain

$$\text{Prob}\left\{|\Theta(x^k) - \Theta(x^*)| \geq \varepsilon\right\} \leq \alpha.$$

This means that x^k is an ε-optimal solution with probability at least $1 - \alpha$. Hence, to obtain an ε-optimal solution, the computational complexity of Monte Carlo approximation method is $\mathcal{O}\left(\frac{n}{\varepsilon^2}\log\frac{1}{\varepsilon}\right)$ by (26). For the computational complexity of Monte Carlo based methods, more research and discussions can be founded in [1, 27]

4 Robustness of ERM Problem

In this section, we discuss the robustness of solutions of the ERM problem (7). This indicates from a new perspective that ERM is an effective reformulation of SAVE. To this end, we first consider the error bound conditions for the original AVE (1). Let

$$F(x) := Ax + B|x| - b.$$

It is easy to see that $F(x)$ is globally Lipschitz continuous and semismooth on $\mathfrak{R}^n$. Hence, $F(x)$ is differentiable almost everywhere by Rademacher's theorem. Let D_F be the set of points at which F is differentiable. The Clarke generalized Jacobian of F at x, denoted by $\partial F(x)$, is defined by

$$\partial F(x) := \text{conv}\left\{\lim_{x^k \to x} F'(x^k)|\ x^k \in D_F\right\},$$

where conv denotes the convex hull and $F'(x)$ is the Jacobian whenever $F(x)$ is differentiable at x. The mapping $F(x)$ is said to be BD-regular at x if all subgradients in $\partial F(x)$ are nonsingular.

Denote the solution set of AVE (1) as X^*, that is,

$$X^* := \{x \in \mathfrak{R}^n | Ax + B|x| = b\}.$$

Assume that X^* is nonempty. A function $r : \mathfrak{R}^n \to \mathfrak{R}_+$ is called a residual function of AVE (1) if $r(x) \geq 0$ for all x and

$$r(x) = 0 \iff x \in X^*.$$

With the help of residual functions, the AVE problem is transformed into finding out the roots of nonlinear equations. The property of error bound plays an important role in theoretical and numerical analysis. Recall that a residual function r has a local error bound for AVE (1) if there exist constants $\kappa > 0$ and $\delta > 0$ such that for each $x \in \{x \in \mathfrak{R}^n | \text{dist}(x, X^*) \leq \delta\}$,

$$\kappa r(x) \geq \text{dist}(x, X^*), \tag{27}$$

where $\text{dist}(x, X^*) := \inf_{y \in X^*} \|x - y\|$. The residual function r has a global error bound for AVE (1) if (27) holds for all $x \in \Re^n$.

Let

$$r(x) := \|Ax + B|x| - b\|. \tag{28}$$

Clearly $r(x)$ is a residual function of the AVE (1).

Lemma 2 *Let $r(x)$ be defined by (28). If $\{A, B\}$ is an R_0 pair, then the level set*

$$\mathcal{L}(\theta) := \{x \in \Re^n | r(x) \leq \theta\}$$

is compact for any $\theta > 0$.

Proof By similarly following the arguments given in Theorem 3, we can prove that $\mathcal{L}(\theta)$ is bounded for all $\theta > 0$. This together with the continuity of r yields the desired result. $\qquad\square$

Theorem 6 *Suppose that $\{A, B\}$ is an R_0 pair and $F(x)$ is BD-regular at any $x \in X^*$. Then $r(x)$ provides a global error bound for AVE (1), that is, there exists a constant $\xi > 0$ such that*

$$\xi r(x) \geq \text{dist}(x, X^*), \ \forall x \in \Re^n.$$

Proof We first show that the residual $r(x)$ provides a local error bound for AVE (1). In fact, if this conclusion is not true, there must exist a sequence $\{x^k\}$ such that

$$\lim_{k \to \infty} \frac{r(x^k)}{\text{dist}(x^k, X^*)} = 0,$$

which in turn implies $r(x^k) \to 0$ as $k \to \infty$. Hence there is a constant $\theta > 0$ such that $r(x^k) \leq \theta$ for all $k \geq 0$. By Lemma 2, $\{x^k\}$ has an accumulation point x^*, and we assume without loss of generality that $x^k \to x^*$ as $k \to \infty$. Since $r(x)$ is continuous, $r(x^*) = 0$, i.e., $x^* \in X^*$. Thus

$$\lim_{k \to \infty} \frac{r(x^k)}{\|x^k - x^*\|} \leq \lim_{k \to \infty} \frac{r(x^k)}{\text{dist}(x^k, X^*)} = 0. \tag{29}$$

Because $F(x)$ is semismooth and BD-regular at x^*, according to [23, Proposition 3], there exist $\varrho > 0$ and $\delta > 0$ such that $r(x) \geq \varrho \|x - x^*\|$ for all x satisfying $\|x - x^*\| \leq \delta$, which contradicts (29). Thus, $r(x)$ provides a local error bound for AVE (1).

Next, we show that $r(x)$ further provides a global error bound for AVE (1). Suppose on the contrary that for any $k > 0$, there exists x^k such that

$$\text{dist}(x^k, X^*) > kr(x^k). \tag{30}$$

Choose a fixed solution $\hat{x} \in X^*$. Then

$$\|x^k - \hat{x}\| \geq \mathrm{dist}(x^k, X^*) > kr(x^k). \tag{31}$$

We claim that there exist $K > 0$ and $\delta > 0$ such that $r(x^k) > \delta$ for every $k > K$. If not, then for any integer $K > 0$ and any $\delta > 0$, there exists $k > K$ such that $r(x^k) \leq \delta$. By (30) and the property of local error bound of $r(x)$, we have

$$\frac{\delta}{k}\mathrm{dist}(x^k, X^*) > \delta r(x^k) \geq \delta\mathrm{dist}(x^k, X^*).$$

It implies $\delta/k > 1$, leading to a contradiction as $k \to +\infty$. Thus, by (31) we have $\|x^k - \hat{x}\| > k\delta$ for every $k > K$, which further yields $\|x^k\| \to +\infty$ as $k \to +\infty$. Since the sequence $\left\{\frac{x^k}{\|x^k\|}\right\}$ is bounded, we assume by taking a subsequence if necessary that $\lim_{k \to \infty} \frac{x^k}{\|x^k\|} = x^*$. It then follows from (31) that

$$\begin{aligned}
1 &= \lim_{k \to \infty} \frac{\|x^k - \hat{x}\|}{\|x^k\|} \geq \lim_{k \to \infty} k\frac{r(x^k)}{\|x^k\|} \\
&= \lim_{k \to \infty} k\left\|A\frac{x^k}{\|x^k\|} + B\left|\frac{x^k}{\|x^k\|}\right| - \frac{b}{\|x^k\|}\right\|,
\end{aligned}$$

which gives

$$\lim_{k \to \infty} \left\|A\frac{x^k}{\|x^k\|} + B\left|\frac{x^k}{\|x^k\|}\right| - \frac{b}{\|x^k\|}\right\| = \|Ax^* + B|x^*|\| = 0,$$

i.e., $Ax^* + B|x^*| = 0$. Since $\{A, B\}$ is an R_0 pair, then $x^* = 0$, a contradiction with $\|x^*\| = 1$. This completes the proof. □

Let us discuss the relation between the solution of the ERM problem (7) and that of (3). For a fixed $\omega \in \Omega$, denote by $X^*(\omega)$ the solution set of the SAVE (3), that is,

$$X^*(\omega) := \{x \in \Re^n | A(\omega)x + B(\omega)|x| = b(\omega)\}.$$

We assume that $X^*(\omega)$ is nonempty.

Theorem 7 *Let $\Omega = \{\omega^1, ..., \omega^N\}$. For each $\omega \in \Omega$, suppose that $\mathrm{F}(x, \omega)$ defined by (6) is BD-regular at any $x \in X^*(\omega)$ and $\{A(\omega), B(\omega)\}$ is an R_0 pair. Then there is a constant $\eta > 0$ such that*

$$\mathbb{E}[\mathrm{dist}(x, X^*(\omega))] \leq \eta\sqrt{\Theta(x)}, \quad \forall x \in \Re^n.$$

Proof By Theorem 6, there exists a constant $\eta > 0$ such that

$$\mathrm{dist}(x, X^*(\omega^i)) \leq \eta\|A(\omega^i)x + B(\omega^i)|x| - b(\omega^i)\|, \quad \forall i \in \{1, ..., N\}.$$

Therefore

$$\mathbb{E}[\text{dist}(x, X^*(\omega))] \leq \eta\mathbb{E}[\|A(\omega)x + B(\omega)|x| - b(\omega)\|]$$
$$\leq \eta\sqrt{\mathbb{E}[\|A(\omega)x + B(\omega)|x| - b(\omega)\|^2]}$$
$$= \eta\sqrt{\Theta(x)},$$

where the second inequality comes from the Schwarz inequality. □

Theorem 7 indicates that, for an optimal solution x^* of ERM problem (7), it holds that

$$\mathbb{E}[\text{dist}(x^*, X^*(\omega))] \leq \eta\sqrt{\Theta(x^*)}. \tag{32}$$

Unlike an error bound for AVE (1), the left hand side of the inequality (32) is in general positive at a solution of ERM problem (7). Nevertheless, the inequality (32) suggests that the expectation of the distance from x^* to the solution set $X^*(\omega)$ is bounded by the expectation residual function $\Theta(x^*)$. This nice property tells us that the solutions of ERM problem (7) may have a minimum sensitivity with respect to random variables in SAVE (3). In this sense, the solutions of ERM problem (7) are robust for SAVE.

In our previous paper [30], we proved that the AVE (1) is uniquely solvable for any $b \in \Re^n$ if the interval matrix $[A - |B|, A + |B|]$ is regular. For the definitions of the interval matrix and its regularity, one may refer to [21, 30] for more detailed information. Based on Theorem 7 and regularity conditions, we obtain the following result.

Theorem 8 *Let $\Omega = \{\omega^1, ..., \omega^N\}$. Suppose that the interval matrix $[A(\omega) - |B(\omega)|, A(\omega) + |B(\omega)|]$ is regular for each $\omega \in \Omega$. Then there is a constant $\lambda > 0$ such that*

$$\mathbb{E}[\|x - x^*(\omega)\|] \leq \lambda\sqrt{\Theta(x)}, \ \forall x \in \Re^n,$$

where $x^*(\omega)$ denotes the unique solution of SAVE *(3) for each fixed $\omega \in \Omega$.*

Proof For a fixed $\omega \in \Omega$, by [30, Theorem 2.2], the SAVE (3) is uniquely solvable and so $\{A(\omega), B(\omega)\}$ is an R_0 pair. Thus, according to Theorem 7, it only needs to prove that $F(x, \omega)$ defined by (6) is BD-regular at $x^*(\omega)$. In fact, for $\omega \in \Omega$, the element V in $\partial F(x, \omega)$ takes the format

$$V = A(\omega) + B(\omega)D_x,$$

where $D_x = \text{diag}(d_x)$ with

$$(d_x)_i := \begin{cases} 1, & \text{if } x_i > 0 \\ \tau, & \text{if } x_i = 0 \\ -1, & \text{if } x_i < 0 \end{cases}$$

and τ takes any number in the interval $[-1, 1]$. Since $\text{diag}(d_x) \in [-I, I]$, then $B(\omega)D_x \in [-|B(\omega)|, |B(\omega)|]$ and hence $A(\omega) + B(\omega)D_x \in [A - |B(\omega)|, A + |B(\omega)|]$. This together with the regularity of $[A(\omega) - |B(\omega)|, A(\omega) + |B(\omega)|]$ implies that $A(\omega) + B(\omega)D_x$ is nonsingular, i.e., V is nonsingular. This completes the proof. □

5 A Class of Special SAVE

We consider in this section the following SAVE:

$$A(\omega)x - |x| = b(\omega). \tag{33}$$

This SAVE is a special case of SAVE (3) by letting $B(\omega) = -I$, see [8]. Due to the existence of absolute value, SAVE (33) is nonsmooth. However, it can be equivalently transformed into a smooth SAVE by exploiting its special structure. For $i = 1, ..., n$, denote by $a_i(\omega)^T$ the i-th row of $A(\omega)$. Then

$$\begin{aligned} A(\omega)x - |x| = b(\omega) &\Longleftrightarrow |x| = A(\omega)x - b(\omega) \\ &\Longleftrightarrow |x_i| = a_i(\omega)^T x - b_i(\omega) \\ &\Longleftrightarrow |x_i|^3 = (a_i(\omega)^T x - b_i(\omega))^3 \\ &\Longleftrightarrow |x|^3 = (A(\omega)x - b(\omega))^3, \end{aligned}$$

where $\alpha^3 := (\alpha_1^3, ..., \alpha_n^3)^T$ for $\alpha = (\alpha_1, ..., \alpha_n)^T \in \Re^n$. Thus, solving SAVE (33) is equivalent to solving the following SAVE:

$$(A(\omega)x - b(\omega))^3 - |x|^3 = 0. \tag{34}$$

Since the function $f(t) = |t|^3$ is twice continuously differentiable on $\Re$ with $f'(t) = 3|t|t$ and $f''(t) = 6|t|$, SAVE (34) is smooth. Let

$$\tilde{F}(x, \omega) := (A(\omega)x - b(\omega))^3 - |x|^3.$$

The corresponding ERM model for SAVE (34) takes the form

$$\min_{x \in \Re^n} \tilde{\Theta}(x) := \mathbb{E}[\|\tilde{F}(x, \omega)\|^2], \tag{35}$$

and its sample average approximation is

$$\min_{x \in \Re^n} \tilde{\Theta}_k(x) := \frac{1}{N_k} \sum_{\omega^i \in \Omega_k} \|\tilde{F}(x, \omega^i)\|^2. \tag{36}$$

Since the function $\tilde{F}(\cdot, \omega)$ is twice continuously differentiable on $\Re^n$, the problems (35) and (36) are both smooth and hence some smoothing-type algorithms are applicable. The limiting behavior of the stationary points of (36) is established as below.

Theorem 9 *Let the support set Ω be a compact set. Assume that x^k is a stationary point of ([36](#)) and x^* is an accumulation point of $\{x^k\}$. Then x^* is a stationary point of ([35](#)) with probability one.*

Proof For the sake of simplicity, we assume without of generality that $\lim_{k \to \infty} x^k = x^*$. Let $\mathcal{D}$ be a compact and convex set containing the whole sequence $\{x^k\}$. For each k, since x^k is a stationary point of ([36](#)), then

$$\frac{2}{N_k} \sum_{\omega^i \in \Omega_k} \nabla_x \tilde{F}(x^k, \omega^i) \tilde{F}(x^k, \omega^i) = 0. \tag{37}$$

By the continuity of $\tilde{F}$, $\nabla_x \tilde{F}$ and $\nabla_x^2 \tilde{F}_j (j = 1, ..., n)$ on the compact set $\mathcal{D} \times \Omega$, there exists a constant $M > 0$ such that

$$\|\tilde{F}(x, \omega)\| \leq M, \quad \|\nabla_x \tilde{F}(x, \omega)\| \leq M, \tag{38}$$
$$\|\nabla_x^2 \tilde{F}_j(x, \omega)\| \leq M (j = 1, ..., n) \tag{39}$$

hold for every $(x, \omega) \in \mathcal{D} \times \Omega$. Therefore

$$\left| \frac{1}{N_k} \sum_{\omega^i \in \Omega_k} \nabla_x \tilde{F}_j(x^k, \omega^i)^T \tilde{F}(x^k, \omega^i) - \frac{1}{N_k} \sum_{\omega^i \in \Omega_k} \nabla_x \tilde{F}_j(x^*, \omega^i)^T \tilde{F}(x^*, \omega^i) \right|$$

$$\leq \frac{1}{N_k} \sum_{\omega^i \in \Omega_k} \|\nabla_x \tilde{F}_j(x^k, \omega^i)\| \|\tilde{F}(x^k, \omega^i) - \tilde{F}(x^*, \omega^i)\|$$

$$+ \frac{1}{N_k} \sum_{\omega^i \in \Omega_k} \|\nabla_x \tilde{F}_j(x^k, \omega^i) - \nabla_x \tilde{F}_j(x^*, \omega^i)\| \|\tilde{F}(x^*, \omega^i)\|$$

$$\leq \frac{M}{N_k} \sum_{\omega^i \in \Omega_k} \int_0^1 \left(\|\nabla_x \tilde{F}(tx^k + (1-t)x^*, \omega^i)\|_F \right.$$

$$\left. + \|\nabla_x^2 \tilde{F}_j(tx^k + (1-t)x^*, \omega^i)\|_F \right) \|x^k - x^*\| dt$$

$$\leq 2M^2 \|x^k - x^*\|$$

$$\to 0 \text{ as } k \to +\infty,$$

where the second inequality comes from the mean-value theorem and ([38](#)), and the third inequality follows from ([38](#)) and ([39](#)). Thus

$$\lim_{k \to \infty} \frac{2}{N_k} \sum_{\omega^i \in \Omega_k} \nabla_x \tilde{F}(x^k, \omega^i) \tilde{F}(x^k, \omega^i)$$

$$= \lim_{k \to \infty} \frac{2}{N_k} \sum_{\omega^i \in \Omega_k} \nabla_x \tilde{F}(x^*, \omega^i) \tilde{F}(x^*, \omega^i)$$

$$= 2\mathbb{E}[\nabla_x \tilde{F}(x^*, \omega) \tilde{F}(x^*, \omega)]$$
$$= \mathbb{E}[\nabla_x \|\tilde{F}(x^*, \omega)\|^2], \text{ w.p.1.} \tag{40}$$

Moreover, it follows from (38) that

$$\|\nabla_x \tilde{F}(x, \omega) \tilde{F}(x, \omega)\| \leq M^2, \quad \forall (x, \omega) \in \mathcal{D} \times \Omega.$$

According to [22, Theorem 16.8], we obtain

$$\lim_{k \to \infty} \frac{2}{N_k} \sum_{\omega^i \in \Omega_k} \nabla_x \tilde{F}(x^k, \omega^i) \tilde{F}(x^k, \omega^i) = \nabla \mathbb{E}[\|\tilde{F}(x^*, \omega)\|^2], \text{ w.p.1.} \tag{41}$$

Then, by letting $k \to \infty$ in (37) and taking (40) and (41) into account, we have

$$\nabla \tilde{\Theta}(x^*) = \nabla \mathbb{E}[\|\tilde{F}(x^*, \omega)\|^2] = 0, \text{ w.p.1,}$$

Hence x^* is a stationary point of (35) with probability one. □

6 A Smoothing Gradient Method

In this section, we present a smoothing gradient method to solve SAVE (3) and give some numerical examples to explain effectiveness of the method. For the ERM problem (7), here we use the well-known sample average approximation (SAA) method to get the approximation of $\Theta(x)$ as

$$\min_{x \in \Re^n} \tilde{\Theta}(x) := \frac{1}{N} \sum_{i=1}^{N} \|\mathrm{F}(x, \omega^i)\|^2$$
$$= \frac{1}{N} \sum_{i=1}^{N} \|A(\omega^i)x + B(\omega^i)|x| - b(\omega^i)\|^2, \tag{42}$$

where $\omega^i (i = 1, ..., N)$ is sample points generated by the given distribution. Since the function $\mathrm{F}(x, \omega) = A(\omega)x + B(\omega)|x| - b(\omega)$ is nonsmooth with respect to x, we can not apply smoothing-type methods to solve the problem (42). To overcome this difficulty, we consider the following smoothing function:

$$\mathrm{F}_N(x, \omega) = A(\omega)x + B(\omega)\sqrt{x^2 + 1/N} - b(\omega),$$

where N is the number of sample points and

$$\sqrt{x^2 + 1/N} := \left(\sqrt{x_1^2 + 1/N}, ..., \sqrt{x_n^2 + 1/N} \right)^T.$$

Obviously, $F_N(x, \omega)$ is smooth with respect to x and $\lim\limits_{N \to \infty} F_N(x, \omega) = F(x, \omega)$ for each $x \in \Re^n$. By using $F_N(x, \omega)$, we define the following smooth approximation of the problem (42):

$$\min_{x \in \Re^n} \Theta_N(x) := \frac{1}{N} \sum_{i=1}^{N} \|F_N(x, \omega^i)\|^2$$

$$= \frac{1}{N} \sum_{i=1}^{N} \|A(\omega^i)x + B(\omega^i)\sqrt{x^2 + 1/N} - b(\omega^i)\|^2. \quad (43)$$

It is easy to see that if x^* is the global optimal solution of the problem (43), then x^* is an approximation to the global optimization solution of the problem (42) when N is sufficiently large. Moreover, $\Theta_N(x)$ is smooth with respect to x and its gradient is

$$\nabla_x \Theta_N(x) = \frac{2}{N} \sum_{i=1}^{N} (A(\omega^i) + B(\omega^i)T_x)^T (A(\omega^i)x + B(\omega^i)\sqrt{x^2 + 1/N} - b(\omega^i))$$

$$(44)$$

with $T_x := \text{diag}\left(\frac{x_i}{\sqrt{x_i^2 + 1/N}}\right)$. Now we present a smoothing gradient method to solve the problem (43).

Algorithm A (A smoothing gradient method)

    Step 0   Given $x^0 \in \Re^n$, $\sigma, \rho \in (0, 1)$ and $\epsilon > 0$. Set $k := 0$.

    Step 1   If $\|\nabla_x \Theta_N(x^k)\| \leq \epsilon$, then stop.

    Step 2   Compute the search direction $d_k = -\nabla_x \Theta_N(x^k)$.

    Step 3   Let the step-size $\alpha_k := \rho^{l_k}$, where l_k is the smallest nonnegative integer l satisfying

$$\Theta_N(x^k + \rho^l d_k) \leq (1 - \sigma\rho^{2l})\Theta_N(x^k).$$

    Step 4   Set $x_{k+1} := x_k + \alpha_k d_k$. Set $k := k + 1$ and go to Step 1.

Theorem 10 *For any positive integer N, let $\{x^k\}$ be the iteration sequence generated by Algorithm A. Then any accumulation point x^* of $\{x^k\}$ satisfies $\nabla_x \Theta_N(x^*) = 0$.*

Proof Since x^* is an accumulation point of $\{x^k\}$, there exists an infinite subsequence $\{x^k\}_{k \in K} \subset \{x^k\}$ such that $\lim\limits_{k \in K, k \to \infty} x^k = x^*$. Since $\{\Theta_N(x^k)\}$ is monotonically decreasing, it is convergent and so $\lim\limits_{k \to \infty} \Theta_N(x^k) = \Theta_N(x^*)$. Note that if $\Theta_N(x^*) = 0$, then by (43) we have for $i = 1, ..., N$,

$$A(\omega^i)x^* + B(\omega^i)\sqrt{(x^*)^2 + 1/N} - b(\omega^i) = 0,$$

which together with (44) gives $\nabla_x \Theta_N(x^*) = 0$. Now we assume $\Theta_N(x^*) > 0$. Since

$$\Theta_N(x^{k+1}) \leq (1 - \sigma \alpha_k^2) \Theta_N(x^k),$$

it follows from $\lim_{k \to \infty} \Theta_N(x^k) = \Theta_N(x^*) > 0$ that $\lim_{k \to \infty} \alpha_k = 0$. Moreover, by the line search criterion in Step 3, we have for all $k \in K$,

$$\Theta_N(x^k + \rho^{-1} \alpha_k d_k) > (1 - \sigma(\rho^{-1} \alpha_k)^2) \Theta_N(x^k),$$

i.e.,

$$\frac{\Theta_N(x^k + \rho^{-1} \alpha_k d_k) - \Theta_N(x^k)}{\rho^{-1} \alpha_k} > -\sigma \rho^{-1} \alpha_k \Theta_N(x^k).$$

Since $\Theta_N(x)$ is continuously differentiable everywhere, by letting $k \to \infty$ with $k \in K$ on both sides of the above inequality, we have $\nabla_x \Theta_N(x^*)^T d^* \geq 0$, where $d^* = -\nabla_x \Theta_N(x^*)$. It follows that $-\|\nabla_x \Theta_N(x^*)\|^2 \geq 0$ and so $\nabla_x \Theta_N(x^*) = 0$. We complete the proof. □

In the following, we give some numerical examples to explain effectiveness of Algorithm A. All experiments are carried on a PC with CPU of 12th Gen Intel(R) Core(TM) i9-12,900 2.40 GHz and RAM of 64.0GB. The program codes are written in MATLAB and run in MATLAB R2018a environment. In our experiments, we choose $\rho = 0.2$, $\sigma = 10^{-3}$, $\epsilon = 10^{-5}$. For the purpose of comparison, we also apply the smoothing gradient method proposed by Du [8] to solve these examples. The main difference between two methods is that for the generated sampling $\{\omega^1, ..., \omega^N\}$, our method takes the smooth parameter $\mu = \frac{1}{N}$ which is fixed, while Du's method [8] takes the smooth parameter $\mu = \mu_k$ which is updated in every iteration.

In the following tables, N-SGM denotes Algorithm A, SGM denotes the smoothing gradient method proposed in [8], N is the number of sample points, IT denotes the iteration numbers, CPU denotes the CPU time in seconds and x^k denotes the obtained solution when the algorithms terminate.

Example 1 Consider SAVE (3), in which

$$A(\omega) = \begin{pmatrix} 2 + \omega & 1 \\ 2 & 1 + \omega \end{pmatrix}, \quad B(\omega) = \begin{pmatrix} 3 + \omega & -1 \\ 0 & 2\omega \end{pmatrix}, \quad b(\omega) = \begin{pmatrix} 5 + 2\omega \\ 4 + 6\omega \end{pmatrix}.$$

This SAVE has a solution $x^* = (1, 2)^T$. We generate samples $\omega^i, i = 1, 2, ..., N$, which obey the the uniform distribution of $[0, 1]$. We take $x^0 = \text{rand}(2, 1)$ as the starting point. Numerical results are given in Table 1.

Example 2 Consider SAVE (3), in which $\omega = (\omega_1, \omega_2, \omega_3, \omega_4)^T$ and

$$A(\omega) = \begin{pmatrix} 1 + \omega_1 & 1 & 0 & 0 \\ -1 & 2 + \omega_2 & 1 & 0 \\ 0 & -1 & 3 + \omega_3 & 1 \\ 0 & 0 & -1 & 4 + \omega_4 \end{pmatrix},$$

Table 1 Numerical results of Example 1

	N	IT	CPU	x^k
N-SGM	10^2	53	0.02	(0.9975, 2.0007)
	10^3	56	0.08	(0.9998, 2.0001)
	10^4	40	0.56	(1.0000, 2.0000)
	10^5	40	5.13	(1.0000, 2.0000)
SGM	10^2	50	0.04	(0.9998, 2.0001)
	10^3	50	0.09	(0.9998, 2.0001)
	10^4	48	0.87	(1.0000, 2.0000)
	10^5	27	4.35	(1.0000, 2.0000)

Table 2 Numerical results of Example 2

	N	IT	CPU	x^k
N-SGM	10^2	426	0.12	(0.9987, 0.9994, 0.9986, 0.9978)
	10^3	311	0.58	(0.9999, 0.9999, 0.9999, 0.9998)
	10^4	276	4.36	(1.0000, 1.0000, 1.0000, 1.0000)
	10^5	304	66.13	(1.0000, 1.0000, 1.0000, 1.0000)
SGM	10^2	232	0.11	(0.9987, 0.9994, 0.9986, 0.9978)
	10^3	242	0.23	(0.9999, 0.9999, 0.9999, 0.9998)
	10^4	311	6.62	(0.9999, 0.9999, 0.9999, 0.9998)
	10^5	318	96.34	(1.0000, 1.0000, 1.0000, 1.0000)

$$B(\omega) = \begin{pmatrix} \omega_1 & 0 & 0 & 0 \\ 0 & \omega_2 & 0 & 0 \\ 0 & 0 & \omega_3 & 0 \\ 0 & 0 & 0 & \omega_4 \end{pmatrix}, \quad b(\omega) = \begin{pmatrix} 2\omega_1 + 2 \\ 2\omega_2 + 2 \\ 2\omega_3 + 3 \\ 2\omega_4 + 3 \end{pmatrix},$$

where ω_1 and ω_2 are distributed uniformly, ω_3 is distributed exponentially, ω_4 is distributed normally, respectively, with the following parameters:

$$\omega_1 \sim U(-0.5, 0.5), \quad \omega_2 \sim U(-1, 1), \quad \omega_3 \sim Exp(1.5), \quad \omega_4 \sim N(0, 5).$$

This SAVE has a solution $x^* = (1, 1, 1, 1)^T$. We take $x^0 = \text{rand}(4, 1)$ as the starting point. Numerical results are given in Table 2.

Example 3 Consider SLCP (4), in which $\omega = (\omega_1, \omega_2, \omega_3, \omega_4)^T$ and

$$M(\omega) = \begin{pmatrix} \omega_1 & 2 & 2 & 2 \\ 0 & \omega_2 & 2 & 2 \\ 0 & 0 & \omega_3 & 2 \\ 0 & 0 & 0 & \omega_4 \end{pmatrix}, \quad q(\omega) = \begin{pmatrix} -1 \\ -1 \\ -1 \\ -\omega_4 \end{pmatrix},$$

Table 3 Numerical results of Example 3

	N	IT	CPU	x^k
N-SGM	10^2	91	0.05	$(1.0074, 1.0076, 0.9971, -0.9974)$
	10^3	92	0.16	$(1.0007, 1.0007, 0.9998, -0.9998)$
	10^4	92	1.36	$(1.0001, 1.0001, 1.0000, -1.0000)$
	10^5	93	18.79	$(1.0000, 1.0000, 1.0000, -1.0000)$
SGM	10^2	72	0.05	$(1.0074, 1.0076, 0.9971, -0.9974)$
	10^3	80	0.18	$(1.0007, 1.0007, 0.9998, -0.9998)$
	10^4	101	1.77	$(1.0001, 1.0001, 0.9998, -0.9998)$
	10^5	130	31.13	$(1.0000, 1.0000, 1.0000, -1.0000)$

where ω_1 is distributed uniformly, ω_2 is distributed exponentially, ω_3 and ω_4 are distributed normally, respectively, with the following parameters:

$$\omega_1 \sim U(-1, 1), \ \omega_2 \sim Exp(2), \ \omega_3 \sim N(0, 1), \ \omega_4 \sim N(0, 2).$$

This SLCP has the solution $u^* = (0, 0, 0, 1)^T$ and $v^* = (1, 1, 1, 0)^T$. By Remark 1, to obtain the solution of this SLCP, we may solve the SAVE (3) with

$$A(\omega) = I + M(\omega), \ \ B(\omega) = I - M(\omega), \ \ b(\omega) = 2q(\omega). \tag{45}$$

We take $x^0 = \text{rand}(4, 1)$ as the starting point and the numerical results are given in Table 3.

Example 4 Consider SLCP (4), in which $\omega = (\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6)^T$ and

$$M(\omega) = \begin{pmatrix} 1 & 1 & \omega_1 & \omega_2 \\ 1 & 1 & \omega_3 & \omega_4 \\ 2 - \omega_5 & 0 & 2 & 0 \\ 0 & 2 - \omega_6 & 0 & 2 \end{pmatrix}, \ \ q(\omega) = \begin{pmatrix} -2 \\ -2 \\ \omega_5 \\ \omega_6 \end{pmatrix},$$

where ω_1 and ω_2 are distributed uniformly, ω_3 is distributed exponentially, ω_4, ω_5 and ω_6 are distributed normally, respectively, with the following parameters:

$$\omega_1 \sim U(-0.5, 0.5), \ \omega_2 \sim U(-1, 1), \ \omega_3 \sim Exp(2),$$
$$\omega_4 \sim N(0, 2), \ \omega_5 \sim N(0, 4), \ \omega_6 \sim N(0, 6).$$

This SLCP has the solution $u^* = (1, 1, 0, 0)^T$ and $v^* = (0, 0, 2, 2)^T$. Same as Example 3, we solve the SAVE (3) with

$$A(\omega) = I + M(\omega), \ \ B(\omega) = I - M(\omega), \ \ b(\omega) = 2q(\omega).$$

We take $x^0 = (1.1, 1, 1)^T$ as the starting point and the numerical results are given in Table 4.

Table 4 Numerical results of Example 4

	N	IT	CPU	x^k
N-SGM	10^2	493	0.12	$(-0.9977, -0.9976, 2.0017, 2.0014)$
	10^3	412	0.70	$(-0.9998, -0.9998, 2.0002, 2.0001)$
	10^4	410	6.36	$(-1.0000, -1.0000, 2.0000, 2.0000)$
	10^5	409	75.69	$(-1.0000, -1.0000, 2.0000, 2.0000)$
SGM	10^2	219	0.05	$(-0.9977, -0.9976, 2.0017, 2.0014)$
	10^3	295	0.62	$(-0.9998, -0.9998, 2.0002, 2.0001)$
	10^4	276	4.77	$(-0.9998, -0.9998, 2.0001, 2.0001)$
	10^5	301	62.65	$(-1.0000, -1.0000, 2.0000, 2.0000)$

From numerical results listed in Tables 1, 2, 3 and 4, we may observe that, as the sample size N increases, both N-SGM and SGM converge stably, that is, there is clearly convergent trend toward the solutions of the test problems. Moreover, we may see that N-SGM has some advantages over SGM because the former converges faster as the sample size N increases.

7 Conclusions

We have established the relationship between the SAVE (3) and the SLCP (4). We have presented the ERM formulation (7) for SAVE and have given its Monte Carlo approximation. Some results related to existence of solutions, exponential convergence rate and robustness of solutions have been derived. In particular, we have discussed a class of special SAVE (33) and have obtained a result related to convergence of stationary points. Finally, we have proposed a smoothing gradient method to solve SAVE and have given some numerical examples to explain its effectiveness. Although Guass/Subguass distribution satisfies the condition (25), it is worth further studying the computational complexity of Monte Carlo approximation method when the underlying distribution of random vector is not Guass/Subguass or even unknown in advance. Hence how to establish the computational complexity of Monte Carlo approximation method for SAVE under weaker conditions is an important but challenging topic and we prefer to leave this as our future research topic.

Data Availibility The data sets generated during the current study are available from the authors on reasonable request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Anderson, D.F., Higham, D.J., Sun, Y.: Computational complexity analysis for Monte Carlo approximations of classically scaled population processes. Multiscale Model. Sim. **16**(3), 1206–1226 (2018)
2. Ali, R., Pan, K.: The new iteration methods for solving absolute value equations. Appl. Math. **68**(1), 109–122 (2023)
3. Cottle, R.W., Pang, J.-S., Stone, R.E.: The Linear Complementarity Problem. Academic Press, New York (1992)
4. Chen, X., Fukushima, M.: Expected residual minimization method for stochastic linear complementarity problems. Math. Oper. Res. **30**, 1022–1038 (2005)
5. Chen, X., Wets, R.J.-B., Zhang, Y.: Stochastic variational inequalities: residual minimization smoothing sample average approximations. SIAM J. Optim. **22**, 649–673 (2012)
6. Chen, X., Zhang, C., Fukushima, M.: Robust solution of monotone stochastic linear complementarity problems. Math. Progr. **117**(1), 51–80 (2009)
7. Cui, X., Zhang, L.: Stochastic R_0 matrix linear complementarity problems: the Fischer-Burmeister function-based expected residual minimization. Comput. Appl. Math. **40**, 183 (2021)
8. Du, S.Q., Sun, J.J., Niu, S.Q., Zhang, L.P.: Stochastic absolute value equations. J. Appl. Math. Comput. **69**, 921–939 (2023)
9. Haghani, F.K.: On generalized Traub's method for absolute value equations. J. Optim. Theory Appl. **166**, 619–625 (2015)
10. Hladík, M.: Bounds for the solutions of absolute value equations. Comput. Optim. Appl. **69**, 243–266 (2018)
11. Iqbal, J., Iqbal, A., Arif, M.: Levenberg-Marquardt method for solving systems of absolute value equations. J. Comput. Appl. Math. **282**, 134–138 (2015)
12. Mangasarian, O.L.: A hybrid algorithm for solving the absolute value equation. Optim. Lett. **9**(7), 1469–1474 (2015)
13. Mangasarian, O.L., Meyer, R.R.: Absolute value equations. Linear Algebra Appl. **419**, 359–367 (2006)
14. Lin, G.H., Fukushima, M.: New reformulations for stochastic complementarity problems. Optim. Meth. Softw. **21**, 551–564 (2006)
15. Lin, G.H., Luo, M.J., Zhang, D.L., Zhang, J.: Stochastic second-order-cone complementarity problems: expected residual minimization formulation and its applications. Math. Progr. **165**, 197–233 (2017)
16. Liu, H.W., Huang, Y.K., Li, X.L.: Partial projected Newton method for a class of stochastic linear complementarity problems. Numer. Alg. **58**, 593–618 (2011)
17. Liu, Z.M., Du, S.Q., Wang, R.Y.: Nonsmooth Levenberg-Marquardt type method for solving a class of stochastic linear complementarity problems with finitely many elements. Alg. **9**, 83 (2016)
18. Luo, M.J., Wang, L.Z.: The deterministic ERM and CVaR reformulation for the stochastic generalized complementarity problem. Japan J. Indust. Appl. Math. **34**, 321–333 (2017)
19. Luo, M.J., Zhang.Y.: Smoothing sample average approximation method for solving stochastic second-order-cone complementarity problems. J. Inequal. Appl. 2018: 77 (2018)
20. Mezzadri. F.: On the solution of general absolute value equations. Appl. Math. Lett. 107, 106462 (2020)
21. Neumaier, A.: Interval Methods for Systems of Equations. Cambridge University Press, Cambridge (1990)
22. Patrick, B.: Probability and Measure. Wiley, New York (1995)
23. Pang, J.S., Qi, L.: Nonsmooth equations: motivation and algorithms. SIAM J. Optim. **3**, 443–465 (1993)
24. Rohn, J.: A theorem of the alternatives for the equation $Ax + B|x| = b$. Linear Multilinear A. **52**, 421–426 (2004)
25. Rohn, J.: On unique solvability of the absolute value equation. Optim. Lett. **3**, 603–606 (2009)
26. Rohn, J., Hooshyarbakhsh, V., Farhadsefat, R.: An iterative method for solving absolute value equations and sufficient conditions for unique solvability. Optim. Lett. **8**, 35–44 (2014)
27. Shapiro A.: Computational complexity of stochastic programming: Monte Carlo sampling approach. In Proceedings of the International Congress of Mathematicians, pages 2979–2995, Hyderabad, India, 2010
28. Shapiro, A., Xu, H.: Stochastic mathematical programs with equilibrium constraints, modelling and sample average approximation. Optim. **57**, 395–418 (2008)

29. Shen, S., Wu, S.: On the unique solution of the generalized absolute value equation. Optim. Lett. **15**(6), 2017–2024 (2021)
30. Tang, J.Y., Zhou, J.C.: A quadratically convergent descent method for the absolute value equation $Ax + B|x| = b$. Oper. Res. Lett. **47**, 229–234 (2019)
31. Wang, G.X., Zhang, J., Zeng, B., Lin, G.H.: Expected residual minimization formulation for a class of stochastic linear second-order cone complementarity problems. Eur. J. Oper. Res. **265**, 437–447 (2018)
32. Wu, S.L., Li, C.X.: The unique solution of the absolute value equations. Appl. Math. Lett. **76**, 195–200 (2018)
33. Zhang, C., Chen, X.J.: Smoothing projected gradient method and its application to stochastic linear complementarity problems. SIAM J. Optim. **20**, 627–649 (2009)
34. Zhou, G.L., Caccetta, L.: Feasible semismooth Newton method for a class of stochastic linear complementarity problems. J. Optim. Theory Appl. **139**(2), 379–392 (2008)