RANDOMIZED ITERATIVE METHODS FOR GENERALIZED ABSOLUTE VALUE EQUATIONS: SOLVABILITY AND ERROR BOUNDS

JIAXIN XIE, HOUDUO QI, AND DEREN HAN

ABSTRACT. Randomized iterative methods, such as the Kaczmarz method and its variants, have gained growing attention due to their simplicity and efficiency in solving large-scale linear systems. Meanwhile, absolute value equations (AVE) have attracted increasing interest due to their connection with the linear complementarity problem. In this paper, we investigate the application of randomized iterative methods to generalized AVE (GAVE). Our approach differs from most existing works in that we tackle GAVE with non-square coefficient matrices. We establish more comprehensive sufficient and necessary conditions for characterizing the solvability of GAVE and propose precise error bound conditions. Furthermore, we introduce a flexible and efficient randomized iterative algorithmic framework for solving GAVE, which employs sampling matrices drawn from user-specified distributions. This framework is capable of encompassing many well-known methods, including the Picard iteration method and the randomized Kaczmarz method. Leveraging our findings on solvability and error bounds, we establish both almost sure convergence and linear convergence rates for this versatile algorithmic framework. Finally, we present numerical examples to illustrate the advantages of the new algorithms.

1. Introduction

Absolute value equations (AVE) have recently become an active research area due to their relevance to many different fields, including the linear complementarity problem (LCP) [58], biometrics [15], game theory [77], etc. We consider the following generalized AVE (GAVE)

$$(1) Ax - B|x| = b,$$

where $A, B \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$, $|x| = (|x_1|, \dots, |x_n|)^{\top}$, and \top denotes the transpose. Specifically, when m = n and B is the identity matrix, the GAVE (1) reduces to the standard AVE [58]. When B = 0, the GAVE (1) becomes the system of linear equations. Over the past two decades, numerous works have examined the AVE problem from various perspectives, such as solvability [30, 44, 70, 71], error bounds [72], and the development of iterative

Key words: absolute value equations, solvability, error bound, stochastic methods, Kaczmarz method, convergence rate

Mathematics subject classification (2020): 15A06, 68W20, 90C30, 90C33

methods [2, 9–11, 42, 43, 60]. While most of these works primarily focus on cases where m = n, this paper, however, addresses the problem of non-square matrices.

In particular, we present a generic randomized iterative algorithmic framework for solving the GAVE (1). Starting from a proper x^0 , it iterates with the format

(2)
$$x^{k+1} = x^k - \alpha \frac{A^{\top} S_k S_k^{\top} (Ax^k - B|x^k| - b)}{\|S_k^{\top} A\|_2^2},$$

where $\alpha > 0$ is the stepsize, $S_k \in \mathbb{R}^{q \times m}$ is a random matrix drawn from a user-defined probability space $(\Omega, \mathcal{F}, \mathbf{P})$, and $\|\cdot\|_2$ denotes the Euclidean norm. A typical case of the iteration scheme (2) is the randomized Kaczmarz (RK) method [52,63]. For any $i \in \{1, 2, \cdots, m\}$, let e_i denote the *i*-th unit vector, $A_{i,:}$ denote the *i*-th row of A, and b_i denote the *i*-th entry of b. Suppose that the coefficient matrix B = 0, the stepsize $\alpha = 1$, and the sampling space $\Omega = \{e_i\}_{i=1}^m$ with e_i being sampled with probability $\frac{\|A_{i,:}\|_2^2}{\|A\|_F^2}$. Then the iteration scheme (2) reduces to

$$x^{k+1} = x^k - \frac{A_{i_k,:} x^k - b_{i_k}}{\|A_{i_k,:}\|_2^2} A_{i_k,:}^\top,$$

which is exactly the RK method. See Section 5 for further discussion on the choice of the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ for the recovery of existing methods and the development of new ones.

1.1. Our contribution. The main contributions of this work are as follows.

- 1. We present a necessary and sufficient for the unique solvability of the GAVE (1); see Theorem 3.2. To the best of our knowledge, this is the first result to characterize the unique solvability of GAVE with non-square coefficient matrices. To improve the computational tractability of the proposed necessary and sufficient conditions, we further explore sufficient conditions that ensure the existence and uniqueness of the solution of GAVE. Different from the results in the literature, our conditions incorporate a nonsingular matrix, which makes our criteria more comprehensive; see Theorem 3.6, Corollary 3.9, and Theorem 3.11. In addition, we formulate a convex optimization problem to determine the existence of such nonsingular matrix.
- 2. We establish error bounds for the GAVE (1), inspired by the recent work of Zamani and Hladík [72] on the standard AVE. Our results extend beyond the scope of their work. Not only do we propose error bounds for GAVE with non-square coefficient

- matrices, but our error bounds also incorporate a nonsingular matrix; see Theorem 3.16. It greatly generalizes and enhances the results in [72].
- 3. We introduce a simple and versatile randomized iterative algorithmic framework for solving GAVE. Our general framework encompasses several known methods as special cases, including the Picard iteration method, the RK method, and its variants. This enables us to establish new connections between these methods. Additionally, the flexibility of our approach allows us to develop entirely new methods by adjusting the probability space $(\Omega, \mathcal{F}, \mathbf{P})$. By leveraging the solvability and error bounds, we demonstrate both almost sure convergence and linear convergence for this general algorithmic framework. Finally, numerical examples illustrate the benefits of the new algorithms.

1.2. Related work.

- 1.2.1. Solvability. Rohn [58] first introduced the alternative theorem for the unique solvability of the GAVE. This seminal work sparked a large amount of subsequent research exploring the unique solvability conditions of the GAVE [31, 42, 44, 59, 60, 69, 71]. One may refer to [36, 48] for recent surveys on the solvability of the GAVE. The unique solvability of GAVE is characterized by certain conditions derived from the analysis of the interval matrix, singular values, spectral radius, and norms of the matrices A and B involved in the equation. However, these conditions apply only when A and B are square matrices. In this paper, we delve into the unique solvability of the GAVE with non-square coefficient matrices. We establish more comprehensive conditions for characterizing the solvability of the GAVE, expanding the scope of previous research and providing a more inclusive understanding of the problem.
- 1.2.2. Error bound. Error bounds play a crucial role in theoretical and numerical analysis of linear algebraic and optimization [27,51,55,76]. They provide a measure of the accuracy of an approximation, helping elucidate the stability and reliability of numerical methods and algorithms. In [65,66], the authors established the numerical validation for solutions of standard AVE (m = n and B = I) based on the interval methods. Hladík [29] devised outer approximation techniques and derived an array of bounds for the solution set of the GAVE with square coefficient matrices. More recently, Zamani and Hladík [72] studied

error bounds for standard AVE under the assumption that uniqueness of the solution of AVE is guaranteed. They then computed upper bounds for $||x - x^*||$, i.e., the distance to the solution x^* of the AVE, using a computable residual function. In this paper, we extend this line of research by proposing error bounds for GAVE with non-square coefficient matrices. Furthermore, our error bounds incorporate a nonsingular matrix, which significantly broadens and enriches the results presented in [72].

1.2.3. Existing methods for GAVE. In recent years, numerous algorithms have been developed to solve the GAVE with square coefficient matrices. Generally, these algorithms fall into four categories: Newton methods [4,8,12,23,32,43], Picard iteration methods [11,60,61], splitting iteration method [11,18,35], and concave minimization approach [1,45]. For some further comments, please refer to [2,48]. Nevertheless, to our knowledge, there are merely two methods that can handle the GAVE with non-square coefficient matrices. The first is the successive linearization algorithm via concave minimization proposed in [42]. The second is the method of alternating projections (MAP) proposed in [2]. See sections 5 and 6 for more details and discussion of these methods. Our numerical results illustrate that, in comparison to these existing methods, our approach significantly outperforms in resolving non-square GAVE problems.

1.2.4. Stochastic algorithms. Stochastic algorithms, such as the RK method or the stochastic gradient descent (SGD) method [56], have gained popularity recently due to their small memory footprint and good theoretical guarantees [52,56,63]. The Kaczmarz method [34], also known as the algebraic reconstruction technique (ART) [20,26], is a classical iterative algorithm used to solve the large-scale linear system of equations. The method alternates between choosing a row of the system and updating the solution based on the projection onto the hyperplane defined by that row. In the seminal paper [63], Strohmer and Vershynin studied the RK method and proved that if the linear system is consistent, then RK converges linearly in expectation. Since then, a large amount of work has been developed on Kaczmarz-type methods, including accelerated RK methods [25,37,39], randomized block Kaczmarz methods [22,49,50,53,54], greedy RK methods [3,21], randomized sparse Kaczmarz methods [13,62], etc.

In fact, the RK method can be seen as a variant of the SGD method [24, 33, 41, 52, 74]. The SGD method aims to minimize a separable objective function $f(x) = \frac{1}{m} \sum_{i=1}^{m} f_i(x)$ by

stochastically accessing selected components of the objective and taking a gradient step for that component. That is, SGD employs the update rule

$$x^{k+1} = x^k - \alpha_k \nabla f_{i_k}(x^k),$$

where α_k is the step-size and i_k is selected randomly. This approach allows SGD to make progress towards the minimum of the function using only a subset of the gradient information at each step, which can be computationally advantageous, especially for large-scale problems. When the objective function $f(x) = \frac{1}{2m} ||Ax - b||_2^2 = \frac{1}{2m} \sum_{i=1}^m (A_{i,i}x - b_i)^2$, then SGD reduces to the RK method. Furthermore, the randomized iterative method proposed in this paper can also be seen as an application of SGD. By incorporating information from the previous iteration, we formulate a stochastic optimization problem and then solve it by using a single-step of SGD; see section 4 for further discussion.

1.3. **Organization.** The remainder of the paper is organized as follows. After introducing some notations and preliminaries in Section 2, we study the solvability and error bounds in Section 3. In Section 4, we present the randomized iterative algorithmic framework for solving GAVE and establish its convergence. In Section 5, we mention that by selecting the probability space $(\Omega, \mathcal{F}, \mathbf{P})$, the algorithmic framework can recover several existing methods as well as obtain new methods. In Section 6, we perform some numerical experiments to show the effectiveness of the proposed algorithms. Finally, we conclude the paper in Section 7.

2. Basic definitions and Preliminaries

2.1. **Basic definitions.** For any random variables ξ_1 and ξ_2 , we use $\mathbb{E}[\xi_1]$ and $\mathbb{E}[\xi_1|\xi_2]$ to denote the expectation of ξ_1 and the conditional expectation of ξ_1 given ξ_2 . For vector $x \in \mathbb{R}^n$, we use x_i, x^{\top} , and $||x||_2$ to denote the *i*-th entry, the transpose, and the Euclidean norm of x, respectively. For $d \in \mathbb{R}^n$, diag(d) stands for the diagonal matrix whose entries on the diagonal are the components of d.

Let A and B be $m \times n$ matrices. We use $A_{i,j}$, $A_{i,:}$, A^{\top} , A^{\dagger} , $||A||_F$, and Range(A) to denote the (i,j)-th component, the i-th row, the transpose, the Moore-Penrose pseudoinverse, the Frobenius norm, and the column space of A, respectively. If A is nonsingular, then $A^{\dagger} = A^{-1}$. The singular values of A are $\sigma_1(A) \geq \sigma_2(A) \geq \ldots \geq \sigma_{\min\{m,n\}}(A) \geq 0$. Given

 $\mathcal{J} \subseteq [m] := \{1, \dots, m\}$, the complementary set of \mathcal{J} is denoted by \mathcal{J}^c , i.e. $\mathcal{J}^c = [m] \setminus \mathcal{J}$. We use $A_{\mathcal{J},:}$ and $A_{:,\mathcal{J}}$ to denote the row and column submatrix indexed by \mathcal{J} , respectively. The matrix inequality $A \leq B$ is understood entrywise and the interval matrix [A, B] is defined as $[A, B] := \{C \mid A \leq C \leq B\}$. We use $I_n \in \mathbb{R}^{n \times n}$ to denote the identity matrix.

A symmetric matrix $M \in \mathbb{R}^{n \times n}$ is said to be positive semidefinite if $x^{\top}Mx \geq 0$ holds for any $x \in \mathbb{R}^n$, and M is positive definite if $x^{\top}Mx > 0$ holds for any nonzero $x \in \mathbb{R}^n$. We use $\lambda_1(M) \geq \lambda_2(M) \geq \ldots \geq \lambda_n(M)$ to denote the eigenvalues of M. We can see that $\|A\|_2 = \sigma_1(A), \|A\|_F = \sqrt{\sum_{i=1}^n \sigma_i^2(A)}$, and $\sigma_i(A) = \sqrt{\lambda_i(A^{\top}A)}$ for any $i \in [n]$. Letting $\Lambda = \operatorname{diag}(\lambda_1(M), \ldots, \lambda_n(M))$ and $M = U\Lambda U^{\top}$ denote the eigenvalue decomposition of M, we denote $M^{\frac{1}{2}} = U\Lambda^{\frac{1}{2}}U^{\top}$ with $\Lambda^{\frac{1}{2}} = \operatorname{diag}(\sqrt{\lambda_1(M)}, \ldots, \sqrt{\lambda_n(M)})$. For any two matrices M and N, we write $M \succeq N$ ($M \succ N$) to represent M - N is positive semidefinite (definite). For any $M \succeq 0$, we define $\|x\|_M := \sqrt{\langle x, Mx \rangle} = \|M^{\frac{1}{2}}x\|_2$.

We use \mathcal{X}^* to denote the solution set of the GAVE (1). The generalized Jacobian matrices [14] are used in the presence of nonsmooth functions. Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a locally Lipschitz function. The generalized Jacobian of F at \hat{x} , denoted by $\partial F(\hat{x})$, is defined as

$$\partial F(\hat{x}) := \operatorname{co}\left(\left\{\lim_{n \to \infty} \nabla F(x_n) : x_n \to \hat{x}, x_n \notin \mathcal{X}_f\right\}\right),$$

where \mathcal{X}_f is the set of points at which f is not differentiable and $co(\mathcal{S})$ denotes the convex hull of a set \mathcal{S} . We use $dist_{\mathcal{S}}(x)$ to denote the distance from x to the set \mathcal{S} .

2.2. Some useful lemmas. In this subsection, we recall some known results that we will need later on. By the definition of singular value, we know that for any $P, Q \in \mathbb{R}^{m \times n}$, $\sigma_i^2(P^\top Q) = \lambda_i(Q^\top P P^\top Q) = \lambda_i(Q^\top (P P^\top)^{\frac{1}{2}}(P P^\top)^{\frac{1}{2}}Q) = \sigma_i^2((P P^\top)^{\frac{1}{2}}Q)$. Therefore, we can conclude the following lemma.

Lemma 2.1. Let $P,Q \in \mathbb{R}^{m \times n}$ and $\ell = \min\{m,n\}$. Then for any $i \in [\ell]$, $\sigma_i(P^\top Q) = \sigma_i((PP^\top)^{\frac{1}{2}}Q)$.

Lemma 2.2 ([75]). Let $P, Q \in \mathbb{R}^{m \times n}$. Then $\sigma_i(P+Q) \geq \sigma_i(P) - \|Q\|_2$.

The following mean value theorem is useful for proving the error bound.

Lemma 2.3 ([28], Theorem 8). Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be a locally Lipschitz function on open subset Ω of \mathbb{R}^n and let $x, y \in \Omega$. Then there exist $\xi_i \geq 0$ with $\sum_{i=1}^m \xi_i = 1$, vectors

 $z_i \in \operatorname{co}(\{x,y\})$, and matrices $M_i \in \partial F(z_i), i = 1, \dots, m$, such that

$$F(x) - F(y) = \sum_{i=1}^{m} \xi_i M_i(x - y).$$

In order to proceed, we shall propose a basic assumption on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ used in this paper.

Assumption 2.1. Let (Ω, \mathcal{F}, P) be the probability space from which the sampling matrices are drawn. We assume that $\mathbb{E}_{S \in \Omega} [SS^{\top}]$ is a positive definite matrix.

The following two lemmas are crucial for our convergence analysis.

Lemma 2.4 (Lemma 2.3, [40]). Let $S \in \mathbb{R}^{m \times q}$ be a real-valued random variable defined on a probability space (Ω, \mathcal{F}, P) . Suppose that $\mathbb{E}\left[SS^{\top}\right]$ is a positive definite matrix and $A \in \mathbb{R}^{m \times n}$ with $A \neq 0$. Then

$$H := \mathbb{E}\left[\frac{SS^{\top}}{\|S^{\top}A\|_2^2}\right]$$

is well-defined and positive definite, here we define $\frac{0}{0} = 0$.

Lemma 2.5 ([67], Supermartingale convergence lemma). Let v^k and u^k be sequences of nonnegative random variables such that $\mathbb{E}[v^{k+1} \mid \mathcal{F}_k] \leq v^k - u^k$ a.s. for all $k \geq 0$, where \mathcal{F}_k denotes the collection $\{v^0, \ldots, v^k, u^0, \ldots, u^k\}$. Then, v^k converges to a random variable v a.s. and $\sum_{k=0}^{\infty} u^k < \infty$.

3. Solvability and error bounds

3.1. Solvability. For the case where m = n, Theorem 3.1 proposed by Wu and Shen [71] offers a characterization of the unique solvability of the GAVE (1).

Theorem 3.1 ([71], Theorem 3.2). Suppose that m = n. The GAVE (1) has a unique solution for any $b \in \mathbb{R}^n$ if and only if for any $D \in [-I_n, I_n]$, the matrix A + BD is nonsingular.

For arbitrary values of m and n, the unique solvability of the GAVE (1) can be characterized by the following result, which can be viewed as a generalization of Theorem 3.1. To the best of our knowledge, this is the first result to characterize the unique solvability of GAVE with non-square coefficient matrices.

Theorem 3.2. The GAVE (1) has a unique solution for any $b \in \mathbb{R}^m$ if and only if m = n and for any $D \in [-I_n, I_n]$, the matrix A + BD is nonsingular.

To prove Theorem 3.2, let us first prove a useful lemma. For any $A, B \in \mathbb{R}^{m \times n}$, we define the map $F_{A,B} : \mathbb{R}^n \to \mathbb{R}^m$ as

$$(3) F_{A,B}(x) := Ax - B|x|.$$

We have the following conclusion for the map $F_{A,B}$.

Lemma 3.3. Suppose that $m \neq n$. Then for any A and B, the map $F_{A,B}$ defined in (3) is not a bijection map.

Proof. For the case where m < n, we will prove that for any A and B, the map $F_{A,B}$ is not injective. Define $c_1 := A\mathbf{e} - B\mathbf{e}$, where $\mathbf{e} = (1, \dots, 1)^{\top} \in \mathbb{R}^n$. So in the first orthant, there exists a vector $x^* = \mathbf{e}$ which satisfies $F_{A,B}(x^*) = c_1$. Since

$$\dim\left(\operatorname{Null}(A-B)\right) \ge n - m > 0,$$

we know that the null space of A - B is non-empty. Note that $x^* = \mathbf{e}$ lies in the interior of the first orthant, hence there exist infinitely many vector \tilde{x} such that $F_{A,B}(\tilde{x}) = c_1$. This implies that the map $F_{A,B}$ is not injective.

For the case where m > n, we will prove that for any A and B, the map $F_{A,B}$ is not surjective. For any $s \in \{-1,1\}^n$, we define

$$E_s := \{Ax - B|x| \mid x \in \mathbb{R}^n, D_s x \ge 0\},\$$

where $D_s = \operatorname{diag}(s)$. Since

$$\dim(E_s) \leq n < m$$
,

we have that the Lebesgue measure of E_s is equal to zero. Thus,

$$\operatorname{Image}(F_{A,B}) = \bigcup_{s \in \{-1,1\}^n} E_s \neq \mathbb{R}^m,$$

which implies that the map $F_{A,B}$ is not surjective.

Now, we are ready to prove Theorem 3.2.

Proof of Theorem 3.2. "If". Obvious in view of Theorem 3.1. "Only if". If $m \neq n$, Lemma 3.3 indicates that there cannot exist matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{m \times n}$ such that the

GAVE (1) has a unique solution for any $c \in \mathbb{R}^m$. This implies that m must be equal to n. Combining this result with Theorem 3.1 leads to the conclusion of this theorem.

For the case m > n, based on the proof of Lemma 3.3, we can conclude the following result.

Corollary 3.4. Suppose that m > n. For any $A, B \in \mathbb{R}^{m \times n}$, there exist $b \in \mathbb{R}^m$ such that the GAVE (1) is unsolvable.

Corollary 3.4 implies that the GAVE (1) is generally unsolvable. The following theorem presents a necessary and sufficient condition for ensuring the solvability of the GAVE (1).

Theorem 3.5. The GAVE (1) is solvable if and only if there is $s \in \{-1,1\}^n$ such that

$$\{x \mid (A - BD_s) x = b\} \cap \{x \mid D_s x \ge 0\} \ne \emptyset,$$

where $D_s = diag(s)$.

Proof. The "If" part is obvious, so we focus on the "Only if" part. Since GAVE (1) is solvable, we know that there exist x^* and s^* such that $Ax^* - BD_{s^*}x^* = b$ and $D_{s^*}x^* \ge 0$. Hence, $\emptyset \ne \{x \mid (A - BD_{s^*}) x = b\} \cap \{x \mid D_{s^*}x \ge 0\} \supseteq \{x^*\}$.

If B=0, Theorem 3.5 reduces to the fundamental result in linear algebra: the linear system Ax=b is solvable (consistent) if and only if $b \in \text{Range}(A)$. However, checking the condition from Theorem 3.5 might not be easy. Therefore, a more efficiently computable condition is of interest. In the next subsection, we will investigate such conditions.

3.1.1. Sufficient conditions. Let us first consider the case where $m \leq n$. We have the following result.

Theorem 3.6. Suppose that $m \leq n$ and there exists a nonsingular matrix $M \in \mathbb{R}^{m \times m}$ such that $\sigma_m(MA) > ||MB||_2$. Then for any $b \in \mathbb{R}^m$, the GAVE (1) is solvable.

If $\sigma_m(A) > \|B\|_2$, it is evident that a nonsingular matrix M, for example M = I, exists such that $\sigma_m(MA) > \|MB\|_2$. Moreover, the following example indicates that even if $\sigma_m(A) < \|B\|_2$, there might still exist a nonsingular matrix M such that $\sigma_m(MA) > \|MB\|_2$.

Example 3.7. Let

$$A = \begin{bmatrix} 2 & 2 & 6 \\ -3 & -6 & 8 \end{bmatrix}, B = \begin{bmatrix} 4 & 1 & 0 \\ 3 & -1 & -4 \end{bmatrix}, and M = \begin{bmatrix} 8 & -1 \\ -1 & 8 \end{bmatrix}.$$

We have $\sigma_2(A) \approx 5.6807 < ||B||_2 \approx 5.7780$ and $\sigma_2(MA) \approx 47.7762 > ||MB||_2 \approx 33.6650$.

We note that there also exist matrices A and B, such as A = I and B = 2I, for which there is no matrix M satisfying $\sigma_m(MA) > \|MB\|_2$. Actually, for any given A and B, the existence of the matrix M can be reformulated as a semidefinite programming (SDP) [64,68], which will be further discussed in Section 3.1.2. Moreover, if a nonsingular matrix M exists such that $\sigma_m(MA) > \|MB\|_2$, Lemma 2.1 implies that $M_1 := (M^\top M)^{\frac{1}{2}}$ satisfies $\sigma_m(M_1A) > \|M_1B\|_2$, and for any $\zeta \neq 0$, $M_2 := \zeta M$ also satisfies $\sigma_m(M_2A) > \|M_2B\|_2$. In other words, if there exists a matrix M that satisfies $\sigma_m(MA) > \|MB\|_2$, then there are multiple matrices that satisfy the same condition.

The following lemma is essential for proving Theorem 3.6.

Lemma 3.8. Suppose that $m \leq n$. If there exists a subset $S \subset [n]$ with |S| = m such that for any $D \in [-I_m, I_m]$, the matrix $A_{:,S} + B_{:,S}D$ is nonsingular, then for any $b \in \mathbb{R}^m$, the GAVE (1) is solvable. In particular, the GAVE (1) has a unique solution for any $b \in \mathbb{R}^m$ if and only if m = n.

Proof. We can rewrite (1) as

$$[A_{:,\mathcal{S}} \ A_{:,\mathcal{S}^c}] \begin{bmatrix} x_{\mathcal{S}} \\ x_{\mathcal{S}^c} \end{bmatrix} - [B_{:,\mathcal{S}} \ B_{:,\mathcal{S}^c}] \begin{bmatrix} |x_{\mathcal{S}}| \\ |x_{\mathcal{S}^c}| \end{bmatrix} = b$$

It is follows from Theorem 3.1 that

$$A_{:.S}x_S - B_{:.S}|x_S| = b$$

has a unique solution, denoted as $x_{\mathcal{S}}^*$. Therefore, $(x_{\mathcal{S}}^*, 0)^{\top}$ is a solution to (4), i.e., a solution to the GAVE (1). Hence, the GAVE (1) is solvable. If m = n, Theorem 3.1 indicates that the GAVE (1) now has unique solution. On the other hand, if the GAVE (1) has unique solution for any $b \in \mathbb{R}^m$, Theorem 3.2 ensures that m = n. This completes the proof of this theorem.

Now we are ready to prove Theorem 3.6.

Proof Theorem 3.6. Suppose that there exists a $\bar{b} \in \mathbb{R}^m$ such the GAVE (1) is unsolvable. By Lemma 3.8, we know that for every subset $S \subset [n]$ with |S| = m, there exists a $D_S \in [-I_m, I_m]$ such that $A_{:,S} + B_{:,S}D_S$ is singular. Then we have

$$0 = \sigma_m \left(MA_{:,\mathcal{S}} + MB_{:,\mathcal{S}}D_{\mathcal{S}} \right) \ge \sigma_m \left(MA_{:,\mathcal{S}} \right) - \|MB_{:,\mathcal{S}}D_{\mathcal{S}}\|_2 \ge \sigma_m \left(MA \right) - \|MB\|_2,$$

where the last inequality follows from the Cauchy interlacing theorem. This contradicts to the assumption that $\sigma_m(MA) > ||MB||_2$. Hence, we know that the GAVE (1) is solvable for each $b \in \mathbb{R}^m$.

For the case m = n, Theorem 3.6 can derive the following corollary.

Corollary 3.9. Suppose that m = n and there exists a nonsingular matrix $M \in \mathbb{R}^{n \times n}$ such that $\sigma_n(MA) > ||MB||_2$. Then for any $b \in \mathbb{R}^n$, the GAVE (1) has a unique solution.

Clearly, the condition in Corollary 3.9 is weaker than the condition $\sigma_n(A) > \|B\|_2$ provided in [70, Theorem 2.1]. If $M = A^{-1}$, Corollary 3.9 indicates that if $\|A^{-1}B\|_2 < 1$, then for any $b \in \mathbb{R}^n$, the GAVE (1) has a unique solution. Furthermore, the following example shows that even if $\sigma_n(A) < \|B\|_2$, there exists a nonsingular matrix M such that $\sigma_n(MA) > \|MB\|_2$.

Example 3.10. Let

$$A = \begin{bmatrix} -7 & 11 \\ 10 & -2 \end{bmatrix}, B = \begin{bmatrix} -2 & 2 \\ 6 & 0 \end{bmatrix}, and M = \begin{bmatrix} 13 & 2 \\ 2 & 11 \end{bmatrix}.$$

We have $\sigma_2(A) \approx 6.2658 < ||B||_2 \approx 6.3592$ and $\sigma_2(MA) \approx 81.2427 > ||MB||_2 \approx 63.5926$.

If m > n, Corollary 3.4 implies that the GAVE (1) now is typically unsolvable. However, once it is solvable, the following result provides a sufficient condition to guarantee the uniqueness of the solution set \mathcal{X}^* .

Theorem 3.11. Suppose that m > n and \mathcal{X}^* is non-empty. If there exists a nonsingular matrix $M \in \mathbb{R}^{m \times m}$ such that $\sigma_n(MA) > ||MB||_2$, then \mathcal{X}^* is singleton.

Proof. Let $x_1, x_2 \in \mathcal{X}^*$, we have

$$0 = ||MAx_1 - MB|x_1| - MAx_2 + MB|x_2||_2$$

$$= ||MA(x_1 - x_2) - MB(|x_1| - |x_2|)||_2$$

$$\geq ||MA(x_1 - x_2)||_2 - ||MB(|x_1| - |x_2|)||_2$$

$$\geq (\sigma_n(MA) - ||MB||_2) ||x_1 - x_2||_2$$

$$\geq 0.$$

Hence, $x_1 = x_2$. This completes the proof of this theorem.

The following example considers that case where m > n and $\sigma_n(A) < ||B||_2$, yet there exists a nonsingular matrix M such that $\sigma_n(MA) > ||MB||_2$.

Example 3.12. Let

$$A = \begin{bmatrix} -6 & -9 \\ 6 & -4 \\ 5 & -2 \end{bmatrix}, B = \begin{bmatrix} -2 & 6 \\ 2 & -4 \\ -6 & -5 \end{bmatrix}, and M = \begin{bmatrix} 45 & 13 & 0 \\ 13 & 33 & 24 \\ 0 & 24 & 24 \end{bmatrix}.$$

We have $\sigma_2(A) \approx 8.8826 < ||B||_2 \approx 8.9327$ and $\sigma_2(MA) \approx 401.4896 > ||MB||_2 \approx 360.9529$.

3.1.2. The existence of the matrix M. In this subsection, we demonstrate that the problem of finding a nonsingular matrix M such that $\sigma_{\ell}(MA) > \|MB\|_2$ can be reformulated as a convex optimization problem, where $\ell = \min\{m, n\}$. For convenience, we assume that $m \geq n$ in the following discussion.

As discussed in Section 3.1.1, if a nonsingular matrix M exists such that $\sigma_n(MA) > \|MB\|_2$, then it holds that $M_1 := (M^\top M)^{\frac{1}{2}}$ satisfies $\sigma_n(M_1A) > \|M_1B\|_2$, and for any $\zeta \neq 0$, $M_2 := \zeta M$ also satisfies $\sigma_n(M_2A) > \|M_2B\|_2$. Therefore, we only need to consider the constraint

$$\mathcal{M} := \{M: M \succeq I\}$$

to find a nonsingular matrix M that satisfies $\sigma_n(M^{\frac{1}{2}}A) > \|M^{\frac{1}{2}}B\|_2$. Besides, it can be verified that \mathcal{M} is a convex set.

For any given $A, B \in \mathbb{R}^{m \times n}$, we define $\Phi_{A,B} : \mathbb{R}^{m \times m} \to \mathbb{R}$ as

$$\Phi_{A,B}(M) := \lambda_1(B^\top M B) - \lambda_n(A^\top M A).$$

Since for any fixed B, $\lambda_1(B^{\top}MB)$ is a convex function, and for any fixed A, $\lambda_n(A^{\top}MA)$ is a concave function; see [5, Section 3.2.3]. Hence, we have that $\Phi_{A,B}$ is a convex function.

Therefore, we can solve the convex optimization problem

(5)
$$\min \Phi_{A,B}(M)$$
 subject to $M \in \mathcal{M}$

to determine the existence of the nonsingular matrix M. If M^* is the optimal solution of (5) and $\Phi_{A,B}(M^*) < 0$, then M^* is the desired nonsingular matrix such that

$$\sqrt{\lambda_n(A^{\top}M^*A)} = \sigma_n((M^*)^{\frac{1}{2}}A) > \|(M^*)^{\frac{1}{2}}B\|_2 = \sqrt{\lambda_1(B^{\top}M^*B)}.$$

Otherwise, we know that there does not exist a nonsingular matrix that satisfies $\sigma_n(MA) > \|MB\|_2$. We note that (5) is indeed a SDP, which can be solved efficiently.

3.2. Error bounds. Inspired by the recent work of Zamani and Hladík [72], this subsection investigates the error bounds for the solvable GAVE (1). In fact, based on the locally upper Lipschitzian property of polyhedral set-valued mappings as described in Proposition 1 in [57], the GAVE (1) exhibits the local error bounds property. Specifically, there exist $\varepsilon > 0$ and $\kappa > 0$ such that when $||Ax - B|x| - b||_2 < \varepsilon$, it holds

(6)
$$\kappa \operatorname{dist}_{\mathcal{X}^*}(x) \le ||Ax - B|x| - b||_2.$$

However, in general, the global error bounds property does not hold necessarily, as demonstrated in Example 3 in [72]. Next, we will provide several sufficient conditions under which the global error bounds property holds.

Theorem 3.13. Let \mathcal{X}^* be non-empty. If zero is the unique solution of Ax - B|x| = 0, then there exists $\kappa > 0$ such that

(7)
$$\kappa \operatorname{dist}_{\mathcal{X}^*}(x) \le ||Ax - B|x| - b||_2, \ \forall x \in \mathbb{R}^n.$$

Proof. The idea of the proof is similar to that of Theorem 12 in [72] or Theorem 2.1 in [46]. Suppose to the contrary that (7) does not hold. Hence, for each $k \in \mathbb{N}$, there exists x^k such that

(8)
$$||x^k - \bar{x}||_2 \ge \operatorname{dist}_{\mathcal{X}^*}(x^k) > k||Ax^k - B|x^k| - b||_2,$$

where $\bar{x} \in \mathcal{X}^*$. Due to the local error bounds property (6), there exists ε such that $||Ax^k - B|x^k| - c||_2 > \varepsilon$ for each $k \geq k_0$, where k_0 is sufficiently large. Consequently, $||x^k - \bar{x}||_2$ tends to infinity as $k \to \infty$. Choosing subsequences if necessary, we may assume that $\frac{x^k}{||x^k||_2}$

goes to a non-zero vector d. By dividing both sides of (8) by $k||x^k||_2$ and taking the limit as k goes to infinity, we get

$$Ad - B|d| = 0,$$

which contradicts the assumptions.

Corollary 3.9 and Theorem 3.11 have already provided sufficient conditions to guarantee that zero is the unique solution of Ax - B|x| = 0. In addition to those results, the following proposition provides a simpler sufficient condition to ensure the uniqueness of the solution set of Ax - B|x| = 0.

Proposition 3.14. Suppose there exists an index $i \in [m]$ such that $|A_{i,:}| < B_{i,:}$. Then zero is the unique solution of Ax - B|x| = 0.

Proof. Based on the assumption, for any $j \in [n]$, it holds that $A_{i,j}x_j < B_{i,j}|x_j|$ if $x_j \neq 0$. Consequently, if $x \neq 0$, we have $A_{i,:}x - B_{i,:}|x| < 0$. Thus, if $x \neq 0$, we have $Ax - B|x| \neq 0$. Therefore, zero is the unique solution of Ax - B|x| = 0.

To derive other types of error bounds, the following lemma is essential.

Lemma 3.15. For any $x, y \in \mathbb{R}^n$, there exists $\hat{D} := \hat{D}_{x,y} \in [-I_n, I_n]$ such that

$$F_{A,B}(x) - F_{A,B}(y) = (A + B\hat{D})(x - y).$$

Proof. The idea of the proof is similar to that of Theorem 7 in [72]. By the mean value theorem, see Lemma 2.3, there exist $z_i \in \text{co}(\{x,y\}), M_i \in \partial F_{A,B}(z_i)$, and $\xi_i \geq 0, i = 1, \ldots, m$ with $\sum_{i=1}^m \xi_i = 1$ such that

$$F_{A,B}(x) - F_{A,B}(y) = \sum_{i=1}^{m} \xi_i M_i(x-y).$$

Note that for any $z \in \mathbb{R}^n$, $\partial F_{A,B}(z) \subseteq \{A + B \operatorname{diag}(d) \mid \|d\|_{\infty} \le 1\}$. Combing this result with the convexity of $\{A + B \operatorname{diag}(d) \mid \|d\|_{\infty} \le 1\}$, we know that there exists $\hat{D} \in [-I_n, I_n]$ such that $\sum_{i=1}^m \xi_i M_i = A + B\hat{D}$ and hence

$$F_{A,B}(x) - F_{A,B}(y) = (A + B\hat{D})(x - y).$$

This completes the proof of this lemma.

Theorem 3.16. Suppose that \mathcal{X}^* is non-empty and for any $D \in [-I_n, I_n]$, the matrix A + BD is full column rank. Then for any $x^* \in \mathcal{X}^*$ and nonsingular matrix $M \in \mathbb{R}^{m \times m}$,

$$||x - x^*|| \le \max_{D \in [-I_n, I_n]} || (MA + MBD)^{\dagger} || \cdot || M(Ax - B|x| - b) ||, \ \forall x \in \mathbb{R}^n,$$

where $\|\cdot\|$ represents any vector norm and its induced norm.

Proof. From Lemma 3.15, we know that there exists a \hat{D} such that $M(A+B\hat{D})(x-x^*)=M(Ax-B|x|-b)$. Since A+BD is full column rank, we have

$$x - x^* = (MA + MB\hat{D})^{\dagger} M(Ax - B|x| - b).$$

Thus

$$||x - x^*|| = ||(MA + MB\hat{D})^{\dagger} M (Ax - B|x| - b)||$$

$$\leq \max_{D \in [-I_n, I_n]} || (MA + MBD)^{\dagger} || \cdot || M (Ax - B|x| - b) ||,$$

which completes the proof.

Note that under the assumption of Theorem 3.16, it holds that $m \geq n$. Assuming m = n and if the matrix A+BD is full column rank (nonsingular) for every $D \in [-I_n, I_n]$, Theorem 3.1 ensures a unique solution to the GAVE (1). Therefore, Theorem 3.16 yields the following corollary.

Corollary 3.17. Assume that m = n and A + BD is nonsingular for any $D \in [-I_n, I_n]$. Then for any nonsingular matrix $M \in \mathbb{R}^{n \times n}$,

$$||x - x^*|| \le \max_{D \in [-I_n, I_n]} || (MA + MBD)^{-1} || \cdot || M(Ax - B|x| - b) ||, \ \forall x \in \mathbb{R}^n,$$

where $\|\cdot\|$ represents any vector norm and its induced norm.

If B = I and M = I, Corollary 3.17 recovers Theorem 7 in [72], which provides an error bound for the standard system of AVE. Theorem 3.16 can also yield the following corollary.

Corollary 3.18. Let \mathcal{X}^* be non-empty and $M \in \mathbb{R}^{m \times m}$ is a nonsingular matrix. Suppose that $m \geq n$ and $\sigma_n(MA) > ||MB||_2$. Then

$$||x - x^*||_2 \le \frac{1}{\sigma_n(MA) - ||MB||_2} ||Ax - B|x| - b||_{M^\top M}.$$

Proof. For any $D \in [-I_n, I_n]$, we have

$$\sigma_n(MA + MBD) \ge \sigma_n(MA) - ||MBD||_2 \ge \sigma_n(MA) - ||MB||_2 > 0.$$

Therefore, the matrix MA + MBD has full column rank, and thus, A + BD is also full column rank since M is positive definite. Moreover, we have

$$\max_{D \in [-I_n, I_n]} \| (MA + MBD)^{\dagger} \|_2 = \sigma_n^{-1} (MA + MBD) \le \frac{1}{\sigma_n(MA) - \|MB\|_2}.$$

Hence from Theorem 3.16, we have

$$||x - x^*||_2 \le \frac{1}{\sigma_{\min}(MA) - ||MB||_2} ||M(Ax - B|x| - b)||_2$$
$$= \frac{1}{\sigma_n(MA) - ||MB||_2} ||Ax - B|x| - b||_{M^\top M}$$

as desired.

4. Randomized iterative methods for GAVE

In this section, we present our randomized iterative method for solving the GAVE (1) and analyze its convergence properties. Randomization is incorporated into our method through a user-defined probability space $(\Omega, \mathcal{F}, \mathbf{P})$ that describes an ensemble of random matrices $S \in \mathbb{R}^{m \times q}$. The selection of the probability space should ideally be based on the specific problem at hand, as it can impact the error bounds and convergence rates of the method. Our approach and underlying theory accommodate a wide range of probability distributions; see Section 5 for further discussion.

At the k-th iteration, we consider the following stochastic optimization problem

(9)
$$\min_{x \in \mathbb{R}^n} f^k(x) := \mathop{\mathbb{E}}_{S \in \Omega} \left[f_S^k(x) \right],$$

where $f_S^k(x) := \frac{1}{2} \|S^\top (Ax - B|x^k| - b)\|_2^2$ with S being a random variable in $(\Omega, \mathcal{F}, \mathbf{P})$. Starting from x^k , we employ only one step of the SGD method to solve the stochastic optimization problem (9)

$$x^{k+1} = x^k - \alpha_k \nabla f_{S_k}^k(x^k) = x^k - \alpha_k A^{\top} S_k S_k^{\top} (Ax^k - B|x^k| - b),$$

where S_k is drawn from the sample space Ω and α_k is the step-size. Particularly, we choose $\alpha_k = \alpha/\|S_k^\top A\|_2^2$ with $0 < \alpha \le 0$. Now, we are ready to state the randomized iterative method for solving the GAVE (1) described in Algorithm 1.

4.1. Convergence analysis. Let x^* be any element of \mathcal{X}^* . For Algorithm 1, we establish the following convergence result. The parameters α and ξ are chosen to ensure that $0 < \alpha \le 1$, which is a necessary condition for convergence.

Algorithm 1 Randomized iterative method (RIM) for GAVE

Input: $A, B \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, probability spaces $(\Omega, \mathcal{F}, \mathbf{P})$, $\alpha \in (0, 1]$, k = 0, and the initial point $x^0 \in \mathbb{R}^n$.

- 1. Randomly select a sampling matrix $S_k \in \Omega$.
- 2. Update

(10)
$$x^{k+1} = x^k - \alpha \frac{A^\top S_k S_k^\top (Ax^k - B|x^k| - b)}{\|S_k^\top A\|_2^2}.$$

3. If a stopping rule is satisfied, stop and go to output. Otherwise, set k = k + 1 and return to Step 1.

Output: The approximate solution x^k .

Theorem 4.1. Assume that \mathcal{X}^* is nonempty and the probability spaces $(\Omega, \mathcal{F}, \mathbf{P})$ satisfy Assumption 2.1. Let $H = \mathbb{E}\left[\frac{SS^\top}{\|S^\top A\|_2^2}\right]$ and $\{x^k\}_{k\geq 0}$ be the iteration sequence generated by Algorithm 1.

- (i) If $\sigma_n(H^{\frac{1}{2}}A) = \|H^{\frac{1}{2}}B\|_2$ and $\alpha \in (0,1)$, then at least one subsequence of $\{x^k\}_{k\geq 0}$ converges a.s. to a point in the set \mathcal{X}^* and $Ax^k B|x^k| b$ converges a.s. to zero.
- (ii) If $\sigma_n(H^{\frac{1}{2}}A) > \|H^{\frac{1}{2}}B\|_2$ and $\alpha = (2-\xi)\frac{\sigma_n(H^{\frac{1}{2}}A)}{\sigma_n(H^{\frac{1}{2}}A) \|H^{\frac{1}{2}}B\|_2}$ with $\xi \in \left[\frac{\sigma_n(H^{\frac{1}{2}}A) + \|H^{\frac{1}{2}}B\|_2}{\sigma_n(H^{\frac{1}{2}}A)}, 2\right)$, then x^* is the unique solution and

$$\mathbb{E}\left[\|x^k - x^*\|_2^2\right] \le \left(1 - (2 - \xi)\xi\sigma_n^2(H^{\frac{1}{2}}A)\right)^k \|x^0 - x^*\|_2^2.$$

(iii) If zero is the unique solution of Ax - B|x| = 0, $\sigma_n^2(H^{\frac{1}{2}}A) > \|H^{\frac{1}{2}}B\|_2^2 - \lambda_{\min}(H)\kappa^2$ with κ being given in Theorem 3.13, and $\alpha = (2 - \xi) \frac{\sigma_n^2(H^{\frac{1}{2}}A) - \|H^{\frac{1}{2}}B\|_2^2 + \lambda_{\min}(H)\kappa^2}{2\lambda_{\min}(H)\kappa^2}$ with $\xi \in \left[2 \frac{\sigma_n^2(H^{\frac{1}{2}}A) - \|H^{\frac{1}{2}}B\|_2^2}{\sigma_n^2(H^{\frac{1}{2}}A) - \|H^{\frac{1}{2}}B\|_2^2 + \lambda_{\min}(H)\kappa^2}, 2\right)$, then

$$\mathbb{E}\left[\operatorname{dist}_{\mathcal{X}^*}^{2}(x^{k})\right] \leq \left(1 - (2 - \xi)\xi \frac{\sigma_{n}^{2}(H^{\frac{1}{2}}A) - \|H^{\frac{1}{2}}B\|_{2}^{2} + \lambda_{\min}(H)\kappa^{2}}{4\lambda_{\min}(H)\kappa^{2}}\right)^{k} \operatorname{dist}_{\mathcal{X}^{*}}^{2}(x^{0}).$$

Proof of Theorem 4.1. (i) For any $x^* \in \mathcal{X}^*$, we have

$$\begin{split} \|x^{k+1} - x^*\|_2^2 &= \left\| x^k - x^* - \alpha \frac{A^\top S_k S_k^\top (Ax^k - B|x^k| - b)}{\|S_k^\top A\|_2^2} \right\|_2^2 \\ &= \|x^k - x^*\|_2^2 + \alpha^2 \frac{\|A^\top S_k S_k^\top (Ax^k - B|x^k| - b)\|_2^2}{\|S_k^\top A\|_2^4} \\ &- \frac{2\alpha}{\|S_k^\top A\|_2^2} \left\langle S_k^\top (Ax^k - B|x^k| - b), S_k^\top A(x^k - x^*) \right\rangle \\ &\leq \|x^k - x^*\|_2^2 + \alpha^2 \frac{\|S_k^\top (Ax^k - B|x^k| - b)\|_2^2}{\|S_k^\top A\|_2^2} \\ &- \frac{2\alpha}{\|S_k^\top A\|_2^2} \left\langle S_k^\top (Ax^k - B|x^k| - b), S_k^\top A(x^k - x^*) \right\rangle \\ &= \|x^k - x^*\|_2^2 - (\alpha - \alpha^2) \frac{\|S_k^\top (Ax^k - B|x^k| - b)\|_2^2}{\|S_k^\top A\|_2^2} \\ &+ \alpha \frac{\|S_k^\top B(|x^k| - |x^*|)\|_2^2}{\|S_k^\top A\|_2^2} - \alpha \frac{\|S_k^\top A(x^k - x^*)\|_2^2}{\|S_k^\top A\|_2^2}, \end{split}$$

where the inequality follows from $||A^{\top}S_kS_k^{\top}(Ax^k - B|x^k| - b)||_2 \leq ||A^{\top}S_k||_2||S_k^{\top}(Ax^k - B|x^k| - b)||_2$ and the last equality follows from the facts that $-2\langle a,b\rangle = ||a-b||_2^2 - ||a||_2^2 - ||b||_2^2$ and $Ax^* - B|x^*| = b$. Taking expectations, we have

$$\mathbb{E}[\|x^{k+1} - x^*\|_2^2 \mid x^k] \leq \|x^k - x^*\|_2^2 - (\alpha - \alpha^2) \mathbb{E}\left[\frac{\|S^\top (Ax^k - B|x^k| - b)\|_2^2}{\|S^\top A\|_2^2}\right] + \alpha \mathbb{E}\left[\frac{\|S^\top B(|x^k| - |x^*|)\|_2^2}{\|S^\top A\|_2^2}\right] - \alpha \mathbb{E}\left[\frac{\|S^\top A(x^k - x^*)\|_2^2}{\|S^\top A\|_2^2}\right]$$

$$= \|x^k - x^*\|_2^2 - (\alpha - \alpha^2) \|Ax^k - B|x^k| - b\|_H^2$$

$$+ \alpha \|B(|x^k| - |x^*|)\|_H^2 - \alpha \|A(x^k - x^*)\|_H^2,$$

$$\leq \left(1 - \alpha \left(\sigma_n^2 (H^{\frac{1}{2}}A) - \|H^{\frac{1}{2}}B\|_2^2\right)\right) \|x^k - x^*\|_2^2$$

$$- (\alpha - \alpha^2) \|Ax^k - B|x^k| - b\|_H^2,$$

where the last inequality follows from the fact $||B(|x^k| - |x^*|)||_H^2 \le ||H^{\frac{1}{2}}B||_2^2 ||x^k - x^*||_2^2$ and $||A(x^k - x^*)||_H^2 \ge \sigma_n^2 (H^{\frac{1}{2}}A) ||x^k - x^*||_2^2$. According to the supermartingale convergence lemma, $||x^k - x^*||_2$ converges a.s. for every $x^* \in \mathcal{X}^*$. Thus, the sequence $\{x^k\}_{k \ge 0}$ is bounded a.s., leading to the existence of accumulation points for $\{x^k\}_{k \ge 0}$. Furthermore, as $\sum_{k=1}^{\infty} ||Ax^k - B|x^k| - b||_H^2 < \infty$ a.s., it implies that $||Ax^k - B|x^k| - b||_H \to 0$ a.s.. From Lemma 2.4, we know that H is positive definite and hence $||Ax^k - B|x^k| - b||_2 \to 0$ a.s., i.e. $Ax^k - B|x^k| - c \to 0$ a.s.. By the continuity of the function $||Ax - B|x| - b||_2$, for any

accumulation point \tilde{x}^* of $\{x^k\}_{k\geq 0}$, we have $A\tilde{x}^* - B|\tilde{x}^*| - b = 0$ a.s.. This implies $\tilde{x}^* \in \mathcal{X}^*$ a.s..

(ii) From Corollary 3.9 and Theorem 3.11, we know that x^* is the unique solution. By Corollary 3.18, we know that $||Ax^k - B|x^k| - b||_H^2 \ge \left(\sigma_n(H^{\frac{1}{2}}A) - ||H^{\frac{1}{2}}B||_2\right)^2 ||x - x^*||_2^2$, which together with (11) and $\alpha = (2 - \xi) \frac{\sigma_n(H^{\frac{1}{2}}A)}{\sigma_n(H^{\frac{1}{2}}A) - ||H^{\frac{1}{2}}B||_2} \in (0, 1]$ implies

$$\mathbb{E}[\|x^{k+1} - x^*\|_2^2 \mid x^k] \le \left(1 - (2 - \xi)\xi\sigma_n^2(H^{\frac{1}{2}}A)\right)\|x^k - x^*\|_2^2.$$

Taking the expectation over the entire history we have

$$\mathbb{E}[\|x^{k+1} - x^*\|_2^2] \le \left(1 - (2 - \xi)\xi\sigma_n^2(H^{\frac{1}{2}}A)\right)\mathbb{E}[\|x^k - x^*\|_2^2]$$

By induction on the iteration index k, we can obtain the desired result.

(iii) Let
$$x_k^* \in \operatorname{Proj}_{\mathcal{X}^*}(x^k)$$
. Since

$$\operatorname{dist}_{\mathcal{X}^*}^2(x^{k+1}) = \|x^{k+1} - x_{k+1}^*\|_2^2 \le \|x^{k+1} - x_k^*\|_2^2,$$

by using the similar arguments as (11), we have

(12)
$$\mathbb{E}[\operatorname{dist}_{\mathcal{X}^*}^2(x^{k+1}) \mid x^k] \le \left(1 - \alpha \left(\sigma_n^2(H^{\frac{1}{2}}A) - \|H^{\frac{1}{2}}B\|_2^2\right)\right) \operatorname{dist}_{\mathcal{X}^*}^2(x^k) - (\alpha - \alpha^2)\|Ax^k - B|x^k| - b\|_H^2.$$

By the assumptions in this theorem and Theorem 3.13, we know that

$$||Ax^k - B|x^k| - b||_H^2 \ge \lambda_{\min}(H)\kappa^2 \operatorname{dist}_{\mathcal{X}^*}^2(x^k)$$

which together with (12) and $\alpha = (2 - \xi) \frac{\sigma_n^2(H^{\frac{1}{2}}A) - \|H^{\frac{1}{2}}B\|_2^2 + \lambda_{\min}(H)\kappa^2}{2\lambda_{\min}(H)\kappa^2} \in (0, 1]$ leads to

$$\mathbb{E}[\operatorname{dist}_{\mathcal{X}^*}^2(x^{k+1}) \mid x^k] \le \left(1 - (2 - \xi)\xi \frac{\sigma_n^2(H^{\frac{1}{2}}A) - \|H^{\frac{1}{2}}B\|_2^2 + \lambda_{\min}(H)\kappa^2}{4\lambda_{\min}(H)\kappa^2}\right) \operatorname{dist}_{\mathcal{X}^*}^2(x^k).$$

By taking expectation over the entire history and the induction on the iteration index k, we can obtain the desired result.

Remark 4.2. If B=0, Theorem 4.1 (i) aligns with the almost sure convergence result obtained through the properties of stochastic Quasi-Fejér sequences (see Remark 3.2 in [6]). Theorem 4.1 (ii) and (iii) can readily derive the almost sure convergence result. In fact, the proof of Theorem 4.1 (ii) reveals that $\mathbb{E}[\|x^{k+1}-x^*\|_2^2 \mid x^k] \leq \|x^k-x^*\|_2^2 - (2-\xi)\xi\sigma_n^2(H^{\frac{1}{2}}A) \cdot \|x^k-x^*\|_2^2$. Consequently, by applying the supermartingale convergence lemma (Lemma 2.5), we establish that $x^k \to x^*$ a.s. Similarly, we can obtain that $\mathrm{dist}_{\mathcal{X}^*}(x^k) \to 0$ a.s.

Furthermore, Theorem 4.1 (ii) and (iii) indicate Algorithm 1 exhibits the variance reduction property. In fact, supposing $\mathbb{E}[x]$ is bounded for all $x \in \mathbb{R}^n$, by definition, we have

$$\mathbb{E}[\|x - x^*\|_2^2] = \|\mathbb{E}[x - x^*]\|_2^2 + \mathbb{E}[\|x - \mathbb{E}[x]\|_2^2],$$

which implies that the convergence of $\mathbb{E}[\|x-x^*\|_2^2]$ leads to that of $\mathbb{E}\left[\|x-\mathbb{E}[x]\|_2^2\right]$, namely, the reduction of variance.

5. Special cases: Examples

This section provides a brief discussion on the choice of the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ in our method for recovering existing methods and developing new ones. While this list is not exhaustive, it serves to illustrate the flexibility of our algorithm.

5.1. The generalized Picard iteration method. When the sampling space $\Omega = \{(A^{\dagger})^{\top}\}$, i.e. $S = (A^{\dagger})^{\top}$ with probability one, and using the fact that $A^{\top}(A^{\dagger})^{\top}A^{\dagger} = A^{\dagger}$, the iteration scheme (10) of our method becomes

(13)
$$x^{k+1} = x^k - \alpha A^{\dagger} (Ax^k - B|x^k| - b).$$

In particular, when $\alpha = 1$, m = n and A is nonsingular, the iteration scheme (13) recovers the Picard iteration method [60]. Hence, we refer to (13) as the generalized Picard iteration method.

In this case, the parameters in Theorem 4.1 can be simplified as $H = (A^{\dagger})^{\top} A^{\dagger}$, $\sigma_n(H^{\frac{1}{2}}A) = 1$, and $\|H^{\frac{1}{2}}B\|_2 = \|A^{\dagger}B\|_2$. If $\|A^{\dagger}B\|_2 < 1$, then according to Theorem 4.1 (ii), the iteration scheme (13) with $\alpha = 1$ has the following convergence property

$$||x^k - x^*||_2 \le (||A^{\dagger}B||_2)^k ||x^0 - x^*||_2.$$

The above convergence result recovers the result in [72, Proposition 22] when A is nonsingular. We should note that the other convergence results in Theorem 4.1 still hold, although we have only illustrated the result mentioned above.

5.2. The gradient descent method. When the sampling space $\Omega = \{I_m\}$, i.e. $S = I_m$ with probability one, then the iteration scheme (10) of our method becomes

(14)
$$x^{k+1} = x^k - \alpha \frac{A^{\top} (Ax^k - B|x^k| - b)}{\|A\|_2^2}.$$

In particular, when B = 0, the iteration scheme (14) corresponds to the gradient descent method used to solve the least-squares problem.

In this case, the parameters in Theorem 4.1 can be simplified as $H = I/\|A\|_2^2$, $\sigma_n(H^{\frac{1}{2}}A) = \sigma_n(A)/\|A\|_2$, and $\|H^{\frac{1}{2}}B\|_2 = \|B\|_2/\|A\|_2$. If $\sigma_n(A) > \|B\|_2$, then according to Theorem 4.1 (ii), the iteration scheme (13) with $\alpha = 1$ exhibits the following convergence property

$$||x^k - x^*||_2^2 \le \left(1 - \frac{\sigma_n^2(A) - ||B||_2^2}{||A||_2^2}\right)^k ||x^0 - x^*||_2^2.$$

When B=0, the above convergence property aligns with the convergence property of gradient descent for solving the least-squares problem, see e.g. [7, Theorem 3.10].

5.3. Randomized Kaczmarz method for GAVE. When the sampling space $\Omega = \{e_i\}_{i=1}^m$ with e_{i_k} being sampled with probability $\operatorname{Prob}(i_k = i) = \frac{\|A_{i,:}\|_2^2}{\|A\|_F^2}$ at the k-th iteration, then the iteration scheme (10) of our method becomes

(15)
$$x^{k+1} = x^k - \alpha \frac{A_{i_k,:} x^k - B_{i_k,:} |x^k| - b_{i_k}}{\|A_{i_k,:}\|_2^2} A_{i_k,:}^\top.$$

Particularly, when B = 0, the iteration scheme (15) corresponds to the randomized Kaczmarz method for solving linear systems.

In this case, the parameters in Theorem 4.1 can be simplified as $H = I/\|A\|_F^2$, $\sigma_n(H^{\frac{1}{2}}A) = \sigma_n(A)/\|A\|_F$, and $\|H^{\frac{1}{2}}B\|_2 = \|B\|_2/\|A\|_F$. If $\sigma_n(A) > \|B\|_2$, then according to Theorem 4.1 (ii), the iteration scheme (13) with $\alpha = 1$ exhibits the following convergence property

$$\mathbb{E}[\|x^k - x^*\|_2^2] \le \left(1 - \frac{\sigma_n^2(A) - \|B\|_2^2}{\|A\|_F^2}\right)^k \|x^0 - x^*\|_2^2.$$

When B = 0, the above convergence property coincides with the convergence property of the RK method.

5.4. Randomized block Kaczmarz method for GAVE. When the sampling space $\Omega = \left\{ I_{:,\mathcal{J}}(A_{\mathcal{J},:}^{\dagger})^{\top} \right\}_{\mathcal{J} \subset [m]}$, then the iteration scheme (10) of our method becomes

(16)
$$x^{k+1} = x^k - \alpha A_{\mathcal{J}_k,:}^{\dagger} (A_{\mathcal{J}_k,:} x^k - B_{\mathcal{J}_k,:} |x^k| - b_{\mathcal{J}_k}).$$

Particularly, when B=0, the iteration scheme (16) corresponds to the randomized block Kaczmarz (RBK) method for solving linear systems [53]. We briefly review the strategies for selecting the subset \mathcal{J}_k for the RBK method [19,53]. We define the number of blocks denoted by t and divide the rows of the matrix into t subsets, creating a partition $\Gamma = \{\mathcal{J}_1, \ldots, \mathcal{J}_t\}$. Then the block \mathcal{J}_k can be chosen from the partition using one of two strategies: it can be chosen randomly from the partition independently of all previous choices, or it can be sampled without replacement, which Needell and Tropp [53] found to be more effective. If t = m, that is, if the number of blocks is equal to the number of rows, each block consists of a single row and we recover the RK method. The calculation of the pseudoinverse $A_{\mathcal{J}_k,:}^{\dagger}$ in each iteration is computationally expensive. However, if the submatrix $A_{\mathcal{J}_k,:}^{\dagger}$ is well-conditioned, we can use efficient algorithms like conjugate gradient for least-squares (CGLS) to calculate it.

To analyze the convergence of the RBK method, it is necessary to define specific quantities [53]. The row paving (t, β_1, β_2) of a matrix A is a partition $\Gamma = \{\mathcal{J}_1, \dots, \mathcal{J}_t\}$ that satisfies

$$\beta_1 \leq \lambda_{\min}(A_{\mathcal{J},:}A_{\mathcal{J},:}^{\top}) \text{ and } \lambda_{\max}(A_{\mathcal{J},:}A_{\mathcal{J},:}^{\top}) \leq \beta_2, \ \forall \ \mathcal{J} \in \Gamma,$$

where t indicates the size of the paving, and β_1 and β_2 denote the lower and upper paving bounds, respectively. Assume that A is a matrix with full column rank and row paving (t, β_1, β_2) , and the index i is selected with probability 1/t. Let us now consider the parameters in Theorem 4.1. The parameter $H = \frac{1}{t} \sum_{i=1}^{t} I_{:,\mathcal{J}_i} (A_{\mathcal{J}_i,:}^{\dagger})^{\top} A_{\mathcal{J}_i,:}^{\dagger} (I_{:,\mathcal{J}_i})^{\top}$ satisfies $\frac{1}{t\beta_2} I \leq H \leq \frac{1}{t\beta_1} I$, and consequently, $\sigma_n(H^{\frac{1}{2}}A) \geq \sigma_n(A)/\sqrt{t\beta_2}$, and $\|H^{\frac{1}{2}}B\|_2 \leq \|B\|_2/\sqrt{t\beta_1}$. If $\sigma_n(A) > \sqrt{\frac{\beta_2}{\beta_1}} \|B\|_2$, then according to Theorem 4.1 (ii), the iteration scheme (16) with $\alpha = 1$ exhibits the following convergence property

$$\mathbb{E}[\|x^k - x^*\|_2^2] \le \left(1 - \left(\frac{\sigma_n^2(A)}{t\beta_2} - \frac{\|B\|_2^2}{t\beta_1}\right)\right)^k \|x^0 - x^*\|_2^2.$$

When B = 0, this convergence property aligns with that of the RBK method for solving consistent linear systems with full column rank coefficient matrices, as detailed in [53, Theorem 1.2].

In practice, the RBK method may encounter several challenges, including the computational expense at each iteration due to the necessity of applying the pseudoinverse to a vector, which is equivalent to solving a least-squares problem. Additionally, the method presents difficulties in parallelization. To overcome these obstacles, the randomized average block Kaczmarz (RABK) method was introduced [17, 49, 50, 73].

5.5. Randomized average block Kaczmarz method for GAVE. The randomized average block Kaczmarz (RABK) method is a block-parallel approach that computes multiple

updates at each iteration [17, 49, 50, 73]. Specifically, consider the following partition of [m]

$$\mathcal{I}_{i} = \{ \varpi(j) : j = (i-1)p+1, (i-1)p+2, \dots, ip \}, i = 1, 2, \dots, t-1,$$

$$\mathcal{I}_{t} = \{ \varpi(j) : j = (t-1)p+1, (t-1)p+2, \dots, m \}, |\mathcal{I}_{t}| \le p,$$

where ϖ is a uniform random permutation on [m] and p is the block size. We define $\|A\|_{\varpi,p} := \sqrt{\sum_{i=1}^t \|A_{\mathcal{I}_i,:}\|_2^2}$ and select an index $i_k \in [t]$ with the probability $\operatorname{Prob}(i_k = i) = \|A_{\mathcal{I}_i,:}\|_2^2/\|A\|_{\varpi,p}^2$, and then set $S_k = I_{:,\mathcal{I}_{i_k}}$. The iteration scheme (10) of our method becomes

(17)
$$x^{k+1} = x^k - \alpha \frac{A_{\mathcal{I}_{i_k},:}^{\top} \left(A_{\mathcal{I}_{i_k},:} x^k - B_{\mathcal{I}_{i_k},:} |x^k| - b_{\mathcal{I}_{i_k}} \right)}{\|A_{\mathcal{I}_{i_k},:}\|_2^2}.$$

Particularly, when B = 0, the iteration scheme (17) corresponds to the randomized average block Kaczmarz (RABK) method for solving linear systems [17, 49, 50, 73].

In this case, the parameters in Theorem 4.1 can be simplified as $H = 1/\|A\|_{\varpi,p}^2$, $\sigma_n(H^{\frac{1}{2}}A) = \sigma_n(A)/\|A\|_{\varpi,p}$, and $\|H^{\frac{1}{2}}B\|_2 = \|B\|_2/\|A\|_{\varpi,p}$. If $\sigma_n(A) > \|B\|_2$, then according to Theorem 4.1 (ii), the iteration scheme (13) with $\alpha = 1$ exhibits the following convergence property

$$\mathbb{E}[\|x^k - x^*\|_2^2] \le \left(1 - \frac{\sigma_n^2(A) - \|B\|_2^2}{\|A\|_{\varpi, p}^2}\right)^k \|x^0 - x^*\|_2^2.$$

If p = 1, meaning that each block only contains a single row, then the iterative scheme (17) recovers the RK method (15) for GAVE, and the convergence result stated above aligns with the result established in Section 5.3.

We now make a comparison between the cases where p=1 and p=m. For convenience, we assume that $\sigma_n^2(A) - \|B\|_2^2 = 1$. The convergence factors for the cases p=1 and p=m are $1 - \frac{1}{\|A\|_F^2}$ and $1 - \frac{1}{\|A\|_2^2}$, respectively. Since the computational cost for the case p=m at each step is about m-times as expensive as that for the case p=1, we can turn this comparison into a comparison between $\left(1 - \frac{1}{\|A\|_F^2}\right)^m$ and $1 - \frac{1}{\|A\|_2^2}$. Since $\|A\|_F^2 \le m\|A\|_2^2$, it follows that

$$\left(1 - \frac{1}{\|A\|_F^2}\right)^m \le \left(1 - \frac{1}{m\|A\|_2^2}\right)^m < \exp\left(-\frac{1}{\|A\|_2^2}\right) < 1 - \frac{1}{\|A\|_2^2}.$$

This suggests that, theoretically, the RABK method with p=1 is more efficient than the RABK method with p=m for solving the GAVE (1). However, we note that one can use the parallelization technique to speed-up the iteration scheme (17) in terms of the total running time.

Finally, we note that the versatility of our framework and the general convergence theorem (Theorem 4.1) allow for the adaptation of the probability spaces $(\Omega, \mathcal{F}, \mathbf{P})$ to suit specific problems. For instance, random sparse matrices or sparse Rademacher matrices could be particularly appropriate for certain classes of problems.

6. Numerical experiments

In this section, we study the computational behavior of the proposed randomized iterative method. In particular, we focus mainly on the evaluation of the performance of the RABK method for GAVE. We compare RABK with some of the state-of-the-art methods, namely, the generalized Newton method (GNM) [32, 43], the Picard iteration method (PIM) [60], the successive linearization algorithm (SLA) [42], and the method of alternating projections (MAP) [2].

All the methods are implemented in MATLAB R2022a for Windows 11 on a desktop PC with Intel(R) Core(TM) i7-1360P CPU @ 2.20GHz and 32 GB memory. The code to reproduce our results can be found at https://github.com/xiejx-math/GAVE-codes.

We have previously detailed the PIM in Section 5.1. Next, we briefly describe GNM, SLA, and MAP, respectively.

(i) Generalized Newton method (GNM) [32,43]. This algorithm is aimed at solving the GAVE (1) with m = n and the iterations are given by

$$x^{k+1} = (A - BD^k)^{-1}b,$$

where $D^k = \operatorname{diag}(\operatorname{sign}(x_1^k), \dots, \operatorname{sign}(x_n^k))$. The iterations are derived by applying the semismooth Newton method in solving the equation Ax - B|x| = b. As in [43], we use the MATLAB's backslash operator "\" to obtain the iterates.

(ii) Successive linearization algorithm (SLA) [42]. Given an initial point $(x^0, t^0, s^0) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m$ and $\varepsilon > 0$, we solve the linear programming

(18)
$$\min_{\substack{(x,t,s) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \\ \text{s.t.} \quad -s \leq Ax - Bt - b \leq s, \\ -s < t < s,}} \varepsilon \sum_{i=1}^n (-\operatorname{sign}(x_i^k)x_i + t_i) + \sum_{i=1}^m s_i$$

and call its solution $(x^{k+1}, t^{k+1}, s^{k+1})$. To solve (18), we use the MATLAB function linprog.

(iii) Method of alternating projections (MAP) [2]. Let $w := (u, v) \in \mathbb{R}^n \times \mathbb{R}^n$ and consider the following feasibility problem

(19) Find
$$w \in C_1 \cap C_2$$
,

where

$$C_1 := \{w : Tw = \sqrt{2}b\}, C_2 := \{w : u \ge 0, v \ge 0, \text{ and } \langle u, v \rangle\},\$$

and

$$T := [A - B \quad -A - B] \in \mathbb{R}^{m \times 2n}.$$

In [2], the authors showed that the GAVE (1) is equivalent to the feasibility problem (19), i.e., if (u^*, v^*) solves (19), then $x^* = \frac{1}{\sqrt{2}}(u^* - v^*)$ solves the GAVE (1). The method of alternating projections (MAP) can be employed to solve (19). Particularly, starting from a proper w^0 , the MAP iterates with the format

$$w^{k+1} \in (P_{C_1}P_{C_2})(w^k).$$

Here for any $w \in \mathbb{R}^{2n}$, we have

$$P_{C_1}(w) = w - T^{\dagger}(Tw - \sqrt{2}b)$$

and $z \in P_{C_2}(w)$ if and only if for each i = 1, ..., n,

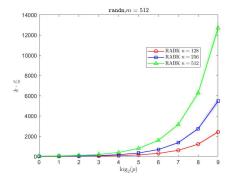
$$(z_i, z_{n+i}) \in \begin{cases} \{(0, (v_i)_+)\}, & u_i < v_i, \\ \{((u_i)_+, 0)\}, & u_i > v_i, \\ \{(0, (v_i)_+), ((u_i)_+, 0)\}, & u_i = v_i. \end{cases}$$

For the case m=n, we compare our algorithm with GNM, PIM, and MAP. We exclude the SLA from this comparison due to its extensive computational time demands when addressing these problems. In the more general case, our algorithm is compared solely with the SLA and the MAP. This is because the other solvers, namely GNM and PIM, are only capable of handling square matrices. All computations are initialized with $x^0 = 0$. The computations are terminated once the relative solution error (RSE), defined as RSE = $||x^k - x^*||_2^2/||x^*||_2^2$, or the relative residual error (RRE), defined as RRE = $||Ax^k - B|x^k| - b||_2^2/||b||_2^2$, is less than a specific error tolerance. For the RABK method, we set $\alpha = 1$ and for the SLA, we set $\epsilon = 1$. All the results below are averaged over 20 trials.

6.1. Synthetic data. The coefficient matrices are randomly Gaussian matrices generated by the MATLAB function randn. Letting $r = \min\{m, n\}$, we initially construct a matrix B and subsequently construct A such that $A = UPV^{\top}$, where $U \in \mathbb{R}^{m \times r}$, $V \in \mathbb{R}^{n \times r}$, and $P = \|B\|_2 \cdot I + D \in \mathbb{R}^{r \times r}$. Here, D is a diagonal matrix and $D_{i,i} \in (0,1)$. Using MATLAB notation, these matrices are generated by the following commands: $B=\operatorname{randn}(m,n)$ or $B=\operatorname{eye}(n)$, $[U,\sim]=\operatorname{qr}(\operatorname{randn}(m,r),0)$, $[V,\sim]=\operatorname{qr}(\operatorname{randn}(n,r),0)$, and $P=\operatorname{norm}(B)*\operatorname{eye}(r)+\operatorname{diag}(\operatorname{rand}(r,1))$. We generate the exact solution by $x^*=\operatorname{randn}(n,1)$ and then set $b=Ax^*-B|x^*|$. In this example, $\sigma_r(A)>\|B\|_2$. Therefore, the GAVE (1) has a unique solution if $m \geq n$ (see Corollary 3.9 and Theorem 3.11). The computations are terminated once the RSE is less than 10^{-12} .

First, we explore the influence of the block size p on the convergence of the RABK method in solving the GAVE (1) with randomly generated Gaussian matrices. The performance of the algorithms is measured in both the computing time (CPU) and the number of full iterations $(k \cdot \frac{p}{m})$, which makes the number of operations for one pass through the rows of A are the same for all the algorithms. In this experiment, we fix m = 512 and set n = 512 and set n = 512 and set n = 512. The results are displayed in Figure 1. The bold line illustrates the median value derived from 20 trials. The lightly shaded area signifies the range from the minimum to the maximum values, while the darker shaded one indicates the data lying between the 25-th and 75-th quantiles. From Figure 1, it can be seen that as the value of p increases, both the CPU time used by the algorithm and the number of full iterations increase. This indicates that the smaller the value of p, the better the performance of the RABK method, which is consistent with the discussion in Section 5.5. Therefore, in the subsequent tests, we take p = 1.

In Figure 2, we report the averages CPU times for PIM, GNM, MAP, and RABK when using square coefficient matrices, i.e., m=n. From Figure 2, it can be observed that when n is relatively small, the GNM outperforms the other methods, with RABK and PIM being the least efficient. However, as n increases, the performance of the RABK method gradually improves, eventually surpassing the other algorithms and emerging as the most efficient method. Figure 3 compares the SLA, MAP, and RABK methods when employing non-square coefficient matrices. The results demonstrate that the RABK method outperforms both SLA and MAP.



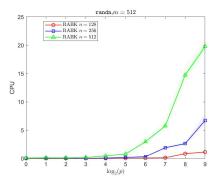
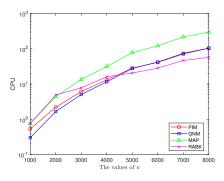


FIGURE 1. Figures depict the evolution of the number of full iterations (left) and the CPU time (right) with respect to the block size p. We fix m = 512 and set n to be 128, 256, and 512 and B is random Gaussian matrix.



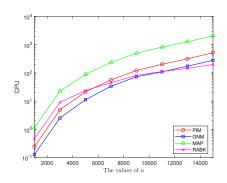
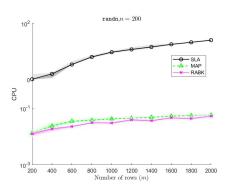


FIGURE 2. Figures depict the evolution of CPU time vs the increasing dimensions of the coefficient matrices. We have m=n, and B is either a random Gaussian matrix (left) or an identity matrix (right).



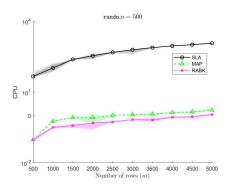


FIGURE 3. Comparison of SLA, MAP, and RABK with non-square coefficient matrices. Figures depict the CPU time (in seconds) vs increasing number of rows. The title of each plot indicates the values of n.

6.2. Ridge Regression. Ridge regression is a popular parameter estimation method used to address the collinearity problem frequently arising in multiple linear regression [16, 47]. We consider an asymmetric ridge regression of the form:

$$\min_{x \in \mathbb{R}^n} h(x) + \sum_{i=1}^n \left(\lambda_i \max\{x_i, 0\}^2 + \mu_i \max\{-x_i, 0\}^2 \right),$$

where the penalty parameters λ_i and μ_i satisfy $\lambda_i \neq \mu_i, i = 1, ..., n$. We note that the case $\lambda_i = \mu_i = \lambda$ for every i corresponds to the classical ridge regression, which will not be considered here. The case $\lambda_i = 0$ for all i and $\mu_i > 0$ corresponds to a penalization of the negativity of the coefficients, promoting solutions with positive coefficients. The necessary condition for optimality is given by:

$$\nabla h(x) + 2\lambda \circ \max\{x, 0\} - 2\mu \circ \max\{-x, 0\} = 0,$$

where \circ denotes the componentwise (or Hadamard) product. Noting that $2 \max\{x, 0\} = |x| + x$ and $2 \max\{-x, 0\} = |x| - x$, we end up with the following problem

(20)
$$\nabla h(x) + (\mu + \lambda) \circ x - (\mu - \lambda) \circ |x| = 0.$$

If we set $\lambda_i = \bar{\lambda}$ and $\mu_i = \bar{\mu}$ for every $i \in [n]$ and consider the loss function $h(x) = \frac{1}{2} ||Lx - c||_2^2$, where $L \in \mathbb{R}^{m \times n}$ and $c \in \mathbb{R}^m$. Then (20) becomes the following GAVE problem

$$\left(L^{\top}L + (\bar{\mu} + \bar{\lambda})I\right)x - (\bar{\mu} - \bar{\lambda})|x| = L^{\top}c.$$

Figure 4 illustrates our experimental results for varying values of $\bar{\lambda}$, and $\bar{\mu}$. The matrix L and vector c are randomly generated with values in [-5,5]. The computations are terminated once the RRE is less than 10^{-12} and set p=1 for the RABK method. The performance of the algorithms is measured in the computing time (CPU) and the number of full iterations. From the iteration schemes of PIM, GNM, and MAP, it can be seen that one iteration of PIM and GNM corresponds to one full iteration, while one iteration of MAP corresponds to two full iterations. It can be observed from Figure 4 that RABK outperforms the other methods in terms of the number of full iterations. However, in terms of CPU time, both PIM and GNM outperform RABK, even though we only show the CPU time taken by RABK to execute (17). This is because the parallel processing capability of MATLAB can help alleviate the computational costs associated with computing the inverse of a matrix.

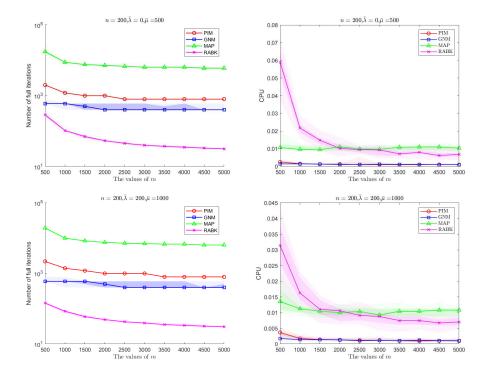


FIGURE 4. Comparison of PIM, GNM, MAP, and RABK for the asymmetric ridge regression. Figures depict the number of full iterations (left) and CPU time (right) vs increasing of m. The title of each plot indicates the values of n, $\bar{\lambda}$, and $\bar{\mu}$.

7. Concluding remarks

We have proposed a simple and versatile randomized iterative algorithmic framework for solving GAVE, applicable to both square and non-square coefficient matrices. By manipulating the probability spaces, our algorithmic framework can recover a wide range of popular algorithms, including the PIM and the RK method. The flexibility of our approach also enables us to modify the parameter matrices to create entirely new methods tailored to specific problems. We have provided conditions that ensure the unique solvability and established error bounds for GAVE. Numerical results confirm the efficiency of our proposed method.

This work not only broadens the applicability of randomized iterative methods, but it also opens up new possibilities for further algorithmic innovations for GAVE. The momentum acceleration technique, a strategy recognized for its effectiveness in enhancing the performance of optimization methods [24,39,73], presents a promising area of exploration. Investigating a momentum variant of randomized iterative methods for the GAVE problem could yield significant advancements in this field. Moreover, real-world applications often present challenges in the form of uncertainty or noise in the input data or computation. Inexact algorithms have been recognized for their ability to provide robustness and flexibility in handling such uncertainties [11,38]. Analyzing inexact variants of randomized iterative methods for solving GAVE would also be a valuable topic.

References

- [1] Lina Abdallah, Mounir Haddou, and Tangi Migot. Solving absolute value equation using complementarity and smoothing functions. *Journal of Computational and Applied Mathematics*, 327:196–207, 2018.
- [2] Jan Harold Alcantara, Jein-Shan Chen, and Matthew K Tam. Method of alternating projections for the general absolute value equation. *Journal of Fixed Point Theory and Applications*, 25(1):39, 2023.
- [3] Zhong-Zhi Bai and Wen-Ting Wu. On greedy randomized Kaczmarz method for solving large sparse linear systems. SIAM J. Sci. Comput., 40(1):A592–A606, 2018.
- [4] JY Bello Cruz, Orizon Pereira Ferreira, and LF Prudente. On the global convergence of the inexact semismooth Newton method for absolute value equation. *Computational Optimization and Applications*, 65(1):93–108, 2016.
- [5] Stephen P Boyd and Lieven Vandenberghe. Convex optimization. Cambridge university press, 2004.
- [6] Luis Briceño-Arias, Julio Deride, and Cristian Vega. Random activations in primal-dual splittings for monotone inclusions with a priori information. J. Optim. Theory Appl., pages 1–26, 2022.
- [7] Sébastien Bubeck et al. Convex optimization: Algorithms and complexity. Foundations and Trends® in Machine Learning, 8(3-4):231–357, 2015.
- [8] Louis Caccetta, Biao Qu, and Guanglu Zhou. A globally and quadratically convergent method for absolute value equations. *Computational optimization and applications*, 48:45–58, 2011.
- [9] Cairong Chen, Bo Huang, Dongmei Yu, and Deren Han. Optimal parameter of the SOR-like iteration method for solving absolute value equations. *Numerical Algorithms*, pages 1–28, 2023.
- [10] Cairong Chen, Yinong Yang, Dongmei Yu, and Deren Han. An inverse-free dynamical system for solving the absolute value equations. *Applied Numerical Mathematics*, 168:170–181, 2021.
- [11] Cairong Chen, Dongmei Yu, and Deren Han. Exact and inexact Douglas-Rachford splitting methods for solving large-scale sparse absolute value equations. IMA Journal of Numerical Analysis, 43(2):1036– 1060, 2023.
- [12] Cairong Chen, Dongmei Yu, Deren Han, and Changfeng Ma. A non-monotone smoothing Newton algorithm for solving the system of generalized absolute value equations. arXiv preprint arXiv:2111.13808, to appear in Journal of Computational Mathematics, 2021.
- [13] Xuemei Chen and Jing Qin. Regularized Kaczmarz algorithms for tensor recovery. SIAM J. Imaging Sci., 14(4):1439–1471, 2021.
- [14] Frank H Clarke. Optimization and nonsmooth analysis. SIAM, 1990.
- [15] Thao Mai Dang, Thuc Dinh Nguyen, Thang Hoang, Hyunseok Kim, Andrew Beng Jin Teoh, and Deokjai Choi. Avet: A novel transform function to improve cancellable biometrics security. *IEEE Transactions* on Information Forensics and Security, 18:758–772, 2022.
- [16] Aris Daniilidis, Mounir Haddou, Tri Minh Le, and Olivier Ley. Solving nonlinear absolute value equations. arXiv preprint arXiv:2402.16439, 2024.
- [17] Kui Du, Wu-Tao Si, and Xiao-Hui Sun. Randomized extended average block Kaczmarz for solving least squares. SIAM J. Sci. Comput., 42(6):A3541-A3559, 2020.

- [18] Vahid Edalatpour, Davod Hezari, and Davod Khojasteh Salkuyeh. A generalization of the Gauss-Seidel iteration method for solving absolute value equations. Applied Mathematics and Computation, 293:156– 167, 2017.
- [19] Inês A Ferreira, Juan A Acebrón, and José Monteiro. Survey of a class of iterative row-action methods: The Kaczmarz method. arXiv preprint arXiv:2401.02842, 2024.
- [20] Richard Gordon, Robert Bender, and Gabor T. Herman. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *Journal of Theoretical Biology*, 29(3):471–481, December 1970.
- [21] Robert M Gower, Denali Molitor, Jacob Moorman, and Deanna Needell. On adaptive sketch-and-project for solving linear systems. SIAM J. Matrix Anal. Appl., 42(2):954–989, 2021.
- [22] Robert M. Gower and Peter Richtárik. Randomized iterative methods for linear systems. SIAM J. Matrix Anal. Appl., 36(4):1660–1690, 2015.
- [23] Farhad Khaksar Haghani. On generalized traub's method for absolute value equations. Journal of Optimization Theory and Applications, 166(2):619–625, 2015.
- [24] Deren Han, Yansheng Su, and Jiaxin Xie. Randomized Douglas-Rachford methods for linear systems: Improved accuracy and efficiency. SIAM J. Optim., 34(1):1045–1070, 2024.
- [25] Deren Han and Jiaxin Xie. On pseudoinverse-free randomized methods for linear systems: Unified framework and acceleration. arXiv preprint arXiv:2208.05437, 2022.
- [26] Gabor T Herman and Lorraine B Meyer. Algebraic reconstruction techniques can be made computationally efficient (positron emission tomography application). *IEEE Trans. Medical Imaging*, 12(3):600–609, 1993.
- [27] Nicholas J Higham. Accuracy and stability of numerical algorithms. SIAM, 2002.
- [28] JB Hiriart-Urruty. Mean value theorems in nonsmooth analysis. Numerical Functional Analysis and Optimization, 2(1):1–30, 1980.
- [29] Milan Hladík. Bounds for the solutions of absolute value equations. Computational Optimization and Applications, 69:243–266, 2018.
- [30] Milan Hladík. Properties of the solution set of absolute value equations and the related matrix classes. SIAM Journal on Matrix Analysis and Applications, 44(1):175–195, 2023.
- [31] Milan Hladík and Hossein Moosaei. Some notes on the solvability conditions for absolute value equations. Optimization Letters, 17(1):211–218, 2023.
- [32] Sheng-Long Hu, Zheng-Hai Huang, and Qiong Zhang. A generalized newton method for absolute value equations associated with second order cones. *Journal of Computational and Applied Mathematics*, 235(5):1490–1501, 2011.
- [33] Zehui Jia, Wenxing Zhang, Xingju Cai, and Deren Han. Stochastic alternating structure-adapted proximal gradient descent method with variance reduction for nonconvex nonsmooth optimization. Mathematics of Computation, 93(348):1677–1714, 2024.
- [34] S Karczmarz. Angenäherte auflösung von systemen linearer glei-chungen. Bull. Int. Acad. Pol. Sic. Let., Cl. Sci. Math. Nat., pages 355–357, 1937.
- [35] Yi-Fen Ke and Chang-Feng Ma. Sor-like iteration method for solving absolute value equations. Applied Mathematics and Computation, 311:195–202, 2017.
- [36] Shubham Kumar, Milan Hladík, Hossein Moosaei, et al. Characterization of unique solvability of absolute value equations: an overview, extensions, and future directions. Optimization Letters, pages 1–19, 2024.
- [37] Ji Liu and Stephen Wright. An accelerated randomized Kaczmarz algorithm. Math. Comp., 85(297):153–178, 2016.
- [38] Nicolas Loizou and Peter Richtárik. Convergence analysis of inexact randomized iterative methods. SIAM Journal on Scientific Computing, 42(6):A3979–A4016, 2020.
- [39] Nicolas Loizou and Peter Richtárik. Momentum and stochastic momentum for stochastic gradient, Newton, proximal point and subspace descent methods. Comput. Optim. Appl., 77(3):653-710, 2020.
- [40] Dirk A Lorenz and Maximilian Winkler. Minimal error momentum Bregman-Kaczmarz. arXiv preprint arXiv:2307.15435, 2023.

- [41] Anna Ma and Deanna Needell. Stochastic gradient descent for linear systems with missing data. Numer. Math. Theory Methods Appl., 12(1):1–20, 2019.
- [42] OL Mangasarian. Absolute value programming. Computational optimization and applications, 36:43–53, 2007.
- [43] OL Mangasarian. A generalized Newton method for absolute value equations. Optimization Letters, 3:101–108, 2009.
- [44] OL Mangasarian and RR Meyer. Absolute value equations. Linear Algebra and Its Applications, 419(2-3):359–367, 2006.
- [45] Olvi L Mangasarian. Absolute value equation solution via concave minimization. Optimization Letters, 1(1):3–8, 2007.
- [46] Olvi L Mangasarian and J Ren. New improved error bounds for the linear complementarity problem. Mathematical Programming, 66(1-3):241–255, 1994.
- [47] Gary C McDonald. Ridge regression. Wiley Interdisciplinary Reviews: Computational Statistics, 1(1):93–100, 2009.
- [48] Hladík Milan, Moosaei Hossein, Hashemi Fakhrodin, Ketabchi Saeed, and Pardalos Panos M. An overview of absolute value equations: From theory to solution methods and challenges. arXiv preprint arXiv:2404.06319, 2024.
- [49] Jacob D Moorman, Thomas K Tu, Denali Molitor, and Deanna Needell. Randomized Kaczmarz with averaging. BIT Numerical Mathematics, 61:337–359, 2021.
- [50] Ion Necoara. Faster randomized block Kaczmarz algorithms. SIAM J. Matrix Anal. Appl., 40(4):1425–1452, 2019.
- [51] Ion Necoara, Yu Nesterov, and Francois Glineur. Linear convergence of first order methods for nonstrongly convex optimization. *Mathematical Programming*, 175:69–107, 2019.
- [52] Deanna Needell, Nathan Srebro, and Rachel Ward. Stochastic gradient descent, weighted sampling, and the randomized Kaczmarz algorithm. *Math. Program.*, 155:549–573, 2016.
- [53] Deanna Needell and Joel A Tropp. Paved with good intentions: analysis of a randomized block Kaczmarz method. *Linear Algebra and its Applications*, 441:199–221, 2014.
- [54] Deanna Needell and Rachel Ward. Two-subspace projection method for coherent overdetermined systems. J. Fourier Anal. Appl., 19(2):256–269, 2013.
- [55] Jong-Shi Pang. Error bounds in mathematical programming. Mathematical Programming, 79(1-3):299–332, 1997.
- [56] Herbert Robbins and Sutton Monro. A stochastic approximation method. Ann. Math. Statistics, pages 400–407, 1951.
- [57] Stephen M Robinson. Some continuity properties of polyhedral multifunctions. pp. 206-214, Springer, 1981.
- [58] Jiri Rohn. A theorem of the alternatives for the equation Ax + B|x| = b. Linear and Multilinear Algebra, 52(6):421–426, 2004.
- [59] Jiri Rohn. On unique solvability of the absolute value equation. Optimization Letters, 3(4):603–606, 2009
- [60] Jiri Rohn, Vahideh Hooshyarbakhsh, and Raena Farhadsefat. An iterative method for solving absolute value equations and sufficient conditions for unique solvability. *Optimization Letters*, 8:35–44, 2014.
- [61] Davod Khojasteh Salkuyeh. The Picard-HSS iteration method for absolute value equations. *Optimization Letters*, 8:2191–2202, 2014.
- [62] Frank Schöpfer and Dirk A Lorenz. Linear convergence of the randomized sparse Kaczmarz method. Math. Program., 173(1):509–536, 2019.
- [63] Thomas Strohmer and Roman Vershynin. A randomized Kaczmarz algorithm with exponential convergence. J. Fourier Anal. Appl., 15(2):262–278, 2009.
- [64] Lieven Vandenberghe and Stephen Boyd. Semidefinite programming. SIAM review, 38(1):49-95, 1996.
- [65] Haijun Wang, Hao Liu, and Suyu Cao. A verification method for enclosing solutions of absolute value equations. *Collectanea mathematica*, 64(1):17–38, 2013.
- [66] HJ Wang, DX Cao, H Liu, and L Qiu. Numerical validation for systems of absolute value equations. Calcolo, 54:669–683, 2017.

- [67] David Williams. Probability with martingales. Cambridge university press, 1991.
- [68] Henry Wolkowicz, Romesh Saigal, and Lieven Vandenberghe. *Handbook of semidefinite programming: theory, algorithms, and applications*, volume 27. Springer Science & Business Media, 2012.
- [69] Shi-Liang Wu and Cui-Xia Li. The unique solution of the absolute value equations. Applied Mathematics Letters, 76:195–200, 2018.
- [70] Shi-Liang Wu and Cui-Xia Li. A note on unique solvability of the absolute value equation. Optimization Letters, 14(7):1957–1960, 2020.
- [71] Shiliang Wu and Shuqian Shen. On the unique solution of the generalized absolute value equation. Optimization Letters, 15:2017–2024, 2021.
- [72] Moslem Zamani and Milan Hladík. Error bounds and a condition number for the absolute value equations. *Mathematical Programming*, 198(1):85–113, 2023.
- [73] Yun Zeng, Deren Han, Yansheng Su, and Jiaxin Xie. On adaptive stochastic heavy ball momentum for solving linear systems. arXiv:2305.05482, to appear in SIAM Journal on Matrix Analysis and Applications. 2023.
- [74] Yun Zeng, Deren Han, Yansheng Su, and Jiaxin Xie. Randomized Kaczmarz method with adaptive stepsizes for inconsistent linear systems. *Numer. Algor.*, 94:1403–1420, 2023.
- [75] Fuzhen Zhang. Matrix theory: basic results and techniques. Springer Science & Business Media, 2011.
- [76] Hui Zhang. New analysis of linear convergence of gradient-type methods via unifying error bound conditions. *Mathematical Programming*, 180(1-2):371–416, 2020.
- [77] Anwa Zhou, Kun Liu, and Jinyan Fan. Semidefinite relaxation methods for tensor absolute value equations. SIAM Journal on Matrix Analysis and Applications, 44(4):1667–1692, 2023.

LMIB of the Ministry of Education, School of Mathematical Sciences, Beihang University, Beijing, 100191, China.

Email address: xiejx@buaa.edu.cn

DEPARTMENT OF APPLIED MATHEMATICS, THE HONG KONG POLYTECHNIC UNIVERSITY, HUNG HOM, KOWLOON, HONG KONG.

Email address: houduo.qi@polyu.edu.hk

LMIB of the Ministry of Education, School of Mathematical Sciences, Beihang University, Beijing, 100191, China.

Email address: handr@buaa.edu.cn