Mathematics

https://www.aimspress.com/journal/Math

AIMS Mathematics, 9(7): 17453–17463.
DOI: 10.3934/math.2024848
Received: 30 March 2024
Revised: 08 May 2024
Accepted: 14 May 2024
Published: 23 May 2024

Research article

Randomized symmetric Gauss-Seidel method for solving linear least squares problems

Fan Sha[1,*] **and Jianbing Zhang**[2]

[1] School of Mathematics, East China Normal University, Shanghai 200241, China
[2] School of Artificial Intelligence, Nanjing University, Nanjing 210023, China

* **Correspondence:** Email: shafan826@gmail.com.

Abstract: We introduced a random symmetric Gauss-Seidel (RSGS) method, which was designed to handle large scale linear least squares problems involving tall coefficient matrices. This RSGS method projected the approximate residual onto the subspace spanned by two symmetric columns at each iteration. These columns were sampled from the coefficient matrix based on an effective probability criterion. Our theoretical analysis indicated that RSGS converged when the coefficient matrix had full column rank. Furthermore, numerical experiments demonstrated that RSGS outperformed the baseline algorithms in terms of iteration steps and CPU time.

Keywords: randomized symmetric Gauss-Seidel; linear least squares; randomized sampling; project; converge in expectation; probability criterion
Mathematics Subject Classification: 65F10

1. Introduction

We consider solving the linear least squares problems given by

$$\min_{x\in\mathbb{R}^n} \|Ax - b\|_2, \tag{1.1}$$

where $A \in \mathbb{R}^{m\times n}$ has full column rank and $m \geq n$. Problem (1.1) is prevalent in several fields, including ridge regression, machine learning, optimal control, and others. To address this problem in a resourceful and efficient manner, a substantial body of research has been conducted on iterative methods [1–3]. Notably, the randomized Kaczmarz method [4] and the randomized Gauss-Seidel method [5–9] have amassed considerable interest due to their capacity to handle large volumes of data.

Strohmer and Vershynin [4] put forth a randomized version of the Kaczmarz method. This method uniquely selects a row in proportion to the squared Euclidean norm, resulting in a fast convergence.

Inspired by this, Leventhal and Lewis [10] developed a randomized Gauss-Seidel (RGS) method that samples a column of A based on an appropriately chosen probability. Subsequently, Ma et al. [11] established a comprehensive convergence theory.

To expedite convergence, researchers have extensively investigated the two-step Gauss-Seidel method. Liu et al. [12] proposed the 2SGS method, a deterministic iteration scheme based on the maximum residual rule. Liao et al. [13] introduced RGS2, a method that merges two single-step iterations into one, sampling two distinct indices simultaneously. In the same literature, the TRGS method, an enhanced two-step approach, projects the approximate solution onto the solution space using two random columns. Mustafa and Saha [14] devised D2RGS, a two-dimensional coordinate descent method employing uniform sampling to randomly select two distinct columns of the coefficient matrix in each iteration.

In this paper, we concentrate on the result found by Niu and Zheng [8], which presents a novel randomized Gauss-Seidel (NRGS) method as follows:

$$x_{k+1} = x_k + \frac{A_{i_k}^\top (b - Ax_k)}{\|A_{i_k}\|_2^2} e_{i_k}, \ k = 0, 1, 2, \ldots,$$

where A_{i_k} is the i_k-th column of A, e_{i_k} is the i_k-th column of the $n \times n$ identity matrix, $r_k = b - Ax_k$, and the column index is chosen with probability $p_{i_k} = \frac{|A_{i_k}^\top r_k|^2}{\|A^\top r_k\|_2^2}$. The convergence result was given by

$$\mathbb{E}\|x_{k+1} - A^\dagger b\|_{A^\top A}^2 \le \left(1 - \frac{\lambda_{\min}(A^\top A)}{\|A\|_F^2}\right)\left(1 - \frac{\lambda_{\min}(A^\top A)}{\beta}\right)^{k-1} \|x_0 - A^\dagger b\|_{A^\top A}^2, k = 0, 1, 2, \ldots, \tag{1.2}$$

where $\beta = \|A\|_F^2 - \min_{1 \le i \le n} \|A_i\|_2^2$, $A^\dagger$ is the Moore-Penrose pseudoinverse of A, and $A^\dagger b$ is the unique solution of (1.1).

For solving (1.1) more efficiently, we consider projecting the update residual to two random symmetric columns. Borrowing the idea of the probability criterion given by NRGS in [8], we introduce Random Symmetric Gauss-Seidel (RSGS) method.

The organization of this paper is as follows. In Section 2, we present the related work. The RSGS method and its convergence analysis are presented in Section 3. Several numerical examples are displayed in Section 4 to show that our proposed RSGS method performs better than NRGS and other methods compared. Finally, the conclusion is drawn in Section 5.

2. Related work

A large number of methods have been designed to solve linear least squares problems. Here, we review only the most relevant and recent research works.

Niu and Zheng [8] proposed a single-step Gauss-Seidel method, namely NRGS. This algorithm adopts an effective probabilistic criterion, allowing it to efficiently capture the larger elements within the residual vector. Inspired by its probabilistic criterion, we extend this algorithm to the two-dimensional scenario. While ensuring the simplification of the sampling method, we achieve faster convergence rates.

To accelerate the convergence, numerous researchers have conducted studies on the two-step Gauss-Seidel method. Liu et al. [12] proposed a two-step iteration Gauss-Seidel deterministic method named the 2SGS method, which is based on the maximum residual rule. As a deterministic algorithm, 2SGS is particularly suitable for medium-sized problems. However, for large-scale datasets, it is necessary to investigate stochastic algorithms. 2SGS is a highly effective deterministic approach, serving as an inspiring foundation for our subsequent design of efficient stochastic algorithms in large-scale datasets.

Liao et al. [13] introduced the two-step iteration method RGS2, which effectively combines two single-step iterations into one step, where two different indices are sampled simultaneously. The advantage of this approach is the avoidance of the possibility of repeating the same index in two consecutive single-step iterations, thereby improving convergence efficiency. In the same literature, the TRGS is the improved two-step method that projects the approximate solution onto the solution space by two random columns. Differently, each iteration of our method employs a stochastic two-dimensional coordinate space least squares approach. Intuitively, under the same conditions, the two-dimensional minimization method outperforms two-step minimization approach (i.e., the combination of two separated singe steps).

Mustafa and Saha [14] developed a two-dimensional coordinate descent method D2RGS, which employs uniform sampling to randomly select two different columns of the coefficient matrix in each iteration. While two-dimensional uniform sampling can save time compared to non-uniform sampling, it fails to consider the importance of different columns. The challenge lies in balancing the fast convergence of the algorithm with the efficiency of sampling. We aim to achieve local minimization in the two-dimensional coordinate space through symmetric sampling. Our method requires only non-uniform sampling of one index at a time while obtaining two indices. This approach is both simple to implement and efficient.

3. The randomized symmetric Gauss-Seidel method

In this section, we introduce the randomized symmetric Gauss-Seidel method (RSGS), which can also be explained as the randomized symmetric coordinate projection method. The iteration scheme is given by

$$x_{k+1} = x_k + \alpha_k e_{i_k} + \beta_k e_{n-i_k+1}, \tag{3.1}$$

where α_k and β_k are parameters which are chosen dynamically such that

$$A_{i_k}^\top r_{k+1} = A_{n-i_k+1}^\top r_{k+1} = 0, \tag{3.2}$$

where $r_{k+1} = b - Ax_{k+1}$. Substituting (3.1) into (3.2), we have

$$\begin{cases} \alpha_k \|A_{i_k}\|_2^2 + \beta_k A_{i_k}^\top A_{n-i_k+1} &= A_{i_k}^\top r_k, \\ \alpha_k A_{n-i_k+1}^\top A_{i_k} + \beta_k \|A_{n-i_k+1}\|_2^2 &= A_{n-i_k+1}^\top r_k. \end{cases} \tag{3.3}$$

Thus,

$$\alpha_k = \begin{cases} \dfrac{A_{i_k}^\top r_k}{2\|A_{i_k}\|^2}, & \text{if} \quad i_k = n - i_k + 1, \\ \dfrac{A_{i_k}^\top r_k \|A_{n-i_k+1}\|_2^2 - A_{i_k}^\top A_{n-i_k+1} A_{n-i_k+1}^\top r_k}{\|A_{i_k}\|_2^2 \|A_{n-i_k+1}\|_2^2 - (A_{i_k}^\top A_{n-i_k+1})^2}, & \text{if} \quad i_k \neq n - i_k + 1, \end{cases} \tag{3.4}$$

and

$$\beta_k = \begin{cases} \dfrac{A_{i_k}^\top r_k}{2\|A_{i_k}\|^2}, & \text{if} \quad i_k = n - i_k + 1, \\ \dfrac{\|A_{i_k}\|_2^2 A_{n-i_k+1}^\top r_k - A_{i_k}^\top r_k A_{n-i_k+1}^\top A_{i_k}}{\|A_{i_k}\|_2^2 \|A_{n-i_k+1}\|_2^2 - (A_{i_k}^\top A_{n-i_k+1})^2}, & \text{if} \quad i_k \neq n - i_k + 1. \end{cases} \tag{3.5}$$

Here, A_j and e_j are the j-th column of A, and the identity matrix I_n, respectively. The sampled index $i_k \in [n] \triangleq \{1, 2, \cdots, n\}$ is chosen with the probability

$$\mathbb{P}(i = i_k) = \frac{|A_{i_k}^\top r_k|^2 + |A_{n-i_k+1}^\top r_k|^2}{2\|A^\top r_k\|_2^2}. \tag{3.6}$$

Based on this construction and criterion, we give Algorithm 1.

Algorithm 1 Randomized Symmetric Gauss-Seidel (RSGS)

Input: A, b, x_0 and $r_0 = b - Ax_0$;
Output:x_k.
for $k = 0, 1, 2, \ldots$ **do**
Pick the index $i_k \in [n]$ with probability (3.6).
Choose α_k,β_k as (3.4) and (3.5).
Update the approximate solution and residual,

$$\begin{aligned} x_{k+1} &= x_k + \alpha_k e_{i_k} + \beta_k e_{n-i_k+1}, \\ r_{k+1} &= r_k - \alpha_k A_{i_k} - \beta_k A_{n-i_k+1}, \end{aligned}$$

until termination criterion is satisfied.
end for

To discuss the convergence, we first give two Lemmas.

Lemma 1. *Let α_k,β_k satisfy (3.4), (3.5) and (3.3). Then it holds that*

$$\alpha_k^2 \|A_{i_k}\|_2^2 + 2\alpha_k \beta_k A_{i_k}^\top A_{n-i_k+1} + \beta_k^2 \|A_{n-i_k+1}\|_2^2 \geq \frac{(A_{i_k}^\top r_k)^2}{\|A_{i_k}\|_2^2}, \tag{3.7}$$

and

$$\alpha_k^2 \|A_{i_k}\|_2^2 + 2\alpha_k \beta_k A_{i_k}^\top A_{n-i_k+1} + \beta_k^2 \|A_{n-i_k+1}\|_2^2 \geq \frac{(A_{n-i_k+1}^\top r_k)^2}{\|A_{n-i_k+1}\|_2^2}, \tag{3.8}$$

where if $i_k = n - i_k + 1$, equalities hold in (3.7) and (3.8).

Proof. Substituting the first equality of (3.3) into the right-hand side of (3.7), we obtain

$$\frac{(A_{i_k}^\top r_k)^2}{\|A_{i_k}\|_2^2} = \alpha_k^2 \|A_{i_k}\|_2^2 + 2\alpha_k \beta_k A_{i_k}^\top A_{n-i_k+1} + \beta_k^2 \frac{(A_{i_k}^\top A_{n-i_k+1})^2}{\|A_{i_k}\|_2^2}.$$

According to Cauchy-Schwarz inequality $A_{i_k}^\top A_{n-i_k+1} \leq \|A_{i_k}\|_2 \|A_{n-i_k+1}\|_2$, (3.7) is proved.

Similarly, substituting the second inequality of (3.3) into the right-hand side of (3.8) and by Cauchy-Schwarz inequality, (3.8) is proved.

Note that if $i_k = n - i_k + 1$, then $A_{i_k} = A_{n-i_k+1}$ and $A_{i_k}^\top A_{n-i_k+1} = \|A_{i_k}\|_2 \|A_{n-i_k+1}\|_2$, then equalities hold in (3.7) and (3.8).

The proof is finished. □

Lemma 2. *[15] The following inequality holds*

$$\sum_{i=1}^{n} \frac{x_i^{l+1}}{y_i^l} \geq \frac{\left(\sum\limits_{i=1}^{n} x_i\right)^{l+1}}{\left(\sum\limits_{i=1}^{n} y_i\right)^{l}},$$

for $x_i \geq 0, y_i > 0, l > 0, i = 1, 2, \ldots, n$. The equality holds if and only if $\frac{x_1}{y_1} = \cdots = \frac{x_n}{y_n}$.

The convergence of Algorithm 1 is given in the following theorem.

Theorem 1. *Let $A \in \mathbb{R}^{m\times n}$ be of full column rank, and $m \geq n$. For the linear least squares problems (1.1), the iteration series $\{x_k\}_{k=0}^{\infty}$ generated by Algorithm 1 converges in expectation to the unique solution $x_* = A^\dagger b$. Furthermore, we have the following results:*

- *If n is even, for $k \geq 0$, we have*

$$\mathbb{E}\|x_{k+1} - A^\dagger b\|_{A^\top A}^2 \leq \left(1 - \frac{\lambda_{\min}(A^\top A)}{\gamma_2}\right)^k \left(1 - \frac{\lambda_{\min}(A^\top A)}{\|A\|_F^2}\right) \|x_0 - A^\dagger b\|_{A^\top A}^2.$$

- *If n is odd, we have*

$$\begin{aligned} &\mathbb{E}\|x_{k+1} - A^\dagger b\|_{A^\top A}^2 \\ \leq &\begin{cases} \left(1 - \frac{\lambda_{\min}(A^\top A)}{\gamma_2}\right)^{\frac{k}{2}} \left(1 - \frac{\lambda_{\min}(A^\top A)}{\gamma_1}\right)^{\frac{k}{2}} \left(1 - \frac{\lambda_{\min}(A^\top A)}{\|A\|_F^2}\right) \|x_0 - A^\dagger b\|_{A^\top A}^2, & \text{if } k \text{ is even and } k \geq 0, \\ \left(1 - \frac{\lambda_{\min}(A^\top A)}{\gamma_2}\right)^{\frac{k-1}{2}} \left(1 - \frac{\lambda_{\min}(A^\top A)}{\gamma_1}\right)^{\frac{k+1}{2}} \left(1 - \frac{\lambda_{\min}(A^\top A)}{\|A\|_F^2}\right) \|x_0 - A^\dagger b\|_{A^\top A}^2, & \text{if } k \text{ is odd and } k \geq 1, \end{cases} \end{aligned}$$

where $\gamma_1 = \|A\|_F^2 - \min_{1\leq i\leq n} \|A_i\|_2^2$, and $\gamma_2 = \|A\|_F^2 - 2\min_{1\leq i\leq n} \|A_i\|_2^2$.

Proof. According to Algorithm 1, the $k+1$-th ($k \geq 0$) iteration can be represented as the optimization problem:

$$x_{k+1} = \operatorname{argmin}_{x - x_k \in \operatorname{span}\{e_{i_k}, e_{n-i_k+1}\}} \|b - Ax\|_2^2,$$

which is equivalent to the following orthogonal projection,

$$b - Ax_{k+1} \perp \operatorname{span}\{Ae_{i_k}, Ae_{n-i_k+1}\}, \quad x_{k+1} \in x_k + \operatorname{span}\{e_{i_k}, e_{n-i_k+1}\},\ k \geq 0. \tag{3.9}$$

Then, according to (3.3)–(3.5), for $k \geq 0$, we have

$$\|x_{k+1} - A^\dagger b\|_{A^\top A}^2 - \|x_k - A^\dagger b\|_{A^\top A}^2$$

$$
\begin{aligned}
&= \|b - Ax_{k+1}\|_2^2 - \|b - Ax_k\|_2^2 \\
&= \|r_k - \alpha_k A_{i_k} - \beta_k A_{n-i_k+1}\|_2^2 - r_k^\top r_k \\
&= -2\alpha_k A_{i_k}^\top r_k - 2\beta_k r_k^\top A_{n-i_k+1} + \alpha_k^2 \|A_{i_k}\|_2^2 + \beta_k^2 \|A_{n-i_k+1}\|_2^2 + 2\alpha_k\beta_k A_{i_k}^\top A_{n-i_k+1} \\
&= -\alpha_k^2 \|A_{i_k}\|_2^2 - \beta_k^2 \|A_{n-i_k+1}\|_2^2 - 2\alpha_k\beta_k A_{i_k}^\top A_{n-i_k+1} \\
&\le -\alpha_k^2 \|A_{i_k}\|_2^2 - \beta_k^2 \|A_{n-i_k+1}\|_2^2 + 2|\alpha_k| \cdot |\beta_k| \cdot |A_{i_k}^\top A_{n-i_k+1}| \\
&\le -\alpha_k^2 \|A_{i_k}\|_2^2 - \beta_k^2 \|A_{n-i_k+1}\|_2^2 + 2|\alpha_k| \cdot |\beta_k| \cdot \|A_{i_k}\|_2 \|A_{n-i_k+1}\|_2 \\
&= -(|\alpha_k| \|A_{i_k}\|_2 - |\beta_k| \|A_{n-i_k+1}\|_2)^2 \\
&\le 0,
\end{aligned}
$$

where the second inequality is due to the fact that, for $k \ge 0$, $|A_{i_k}^\top A_{n-i_k+1}| \le \|A_{i_k}\|_2 \|A_{n-i_k+1}\|_2$. Taking the conditional expectation and according to (3.9), for $k \ge 0$, it holds that

$$
\begin{aligned}
&\mathbb{E}_k \|x_{k+1} - A^\dagger b\|_{A^\top A}^2 \\
&= \|x_k - A^\dagger b\|_{A^\top A}^2 - \mathbb{E}_k \|A(x_{k+1} - x_k)\|_2^2 \\
&= \|x_k - A^\dagger b\|_{A^\top A}^2 - \mathbb{E}_k (\alpha_k^2 \|A_{i_k}\|_2^2 + 2\alpha_k\beta_k A_{i_k}^\top A_{n-i_k+1} + \beta_k^2 \|A_{n-i_k+1}\|_2^2).
\end{aligned}
$$

Now, we estimate the lower bound of the second term in the above last equality. By (3.3)–(3.5), for $k \ge 1$, we have

$$
\begin{aligned}
&\mathbb{E}_k (\alpha_k^2 \|A_{i_k}\|_2^2 + 2\alpha_k\beta_k A_{i_k}^\top A_{n-i_k+1} + \beta_k^2 \|A_{n-i_k+1}\|_2^2) \\
&= \sum_{i_k=1}^n \left(\frac{1}{2} \frac{(A_{i_k}^\top r_k)^2}{\|A^\top r_k\|_2^2} + \frac{1}{2} \frac{(A_{n-i_k+1}^\top r_k)^2}{\|A^\top r_k\|_2^2} \right) (\alpha_k^2 \|A_{i_k}\|_2^2 + 2\alpha_k\beta_k A_{i_k}^\top A_{n-i_k+1} + \beta_k^2 \|A_{n-i_k+1}\|_2^2) \\
&= \frac{1}{2} \sum_{i_k=1}^n \frac{(A_{i_k}^\top r_k)^2}{\|A^\top r_k\|_2^2} \left(\frac{(A_{i_k}^\top r_k)^2}{\|A_{i_k}\|_2^2} - \frac{(A_{i_k}^\top r_k)^2}{\|A_{i_k}\|_2^2} + \alpha_k^2 \|A_{i_k}\|_2^2 + 2\alpha_k\beta_k A_{i_k}^\top A_{n-i_k+1} + \beta_k^2 \|A_{n-i_k+1}\|_2^2 \right) \\
&\quad + \frac{1}{2} \sum_{i_k=1}^n \frac{(A_{n-i_k+1}^\top r_k)^2}{\|A^\top r_k\|_2^2} \Big(\frac{(A_{n-i_k+1}^\top r_k)^2}{\|A_{n-i_k+1}\|_2^2} - \frac{(A_{n-i_k+1}^\top r_k)^2}{\|A_{n-i_k+1}\|_2^2} \\
&\quad + \alpha_k^2 \|A_{i_k}\|_2^2 + 2\alpha_k\beta_k A_{i_k}^\top A_{n-i_k+1} + \beta_k^2 \|A_{n-i_k+1}\|_2^2 \Big) \\
&\ge \frac{1}{2} \sum_{i_k=1}^n \frac{(A_{i_k}^\top r_k)^2}{\|A^\top r_k\|_2^2} \frac{(A_{i_k}^\top r_k)^2}{\|A_{i_k}\|_2^2} + \frac{1}{2} \sum_{i_k=1}^n \frac{(A_{n-i_k+1}^\top r_k)^2}{\|A^\top r_k\|_2^2} \frac{(A_{n-i_k+1}^\top r_k)^2}{\|A_{n-i_k+1}\|_2^2} \\
&= \frac{1}{\|A^\top r_k\|_2^2} \left(\frac{1}{2} \sum_{i_k=1}^n \frac{(A_{i_k}^\top r_k)^4}{\|A_{i_k}\|_2^2} + \frac{1}{2} \sum_{i_k=1}^n \frac{(A_{n-i_k+1}^\top r_k)^4}{\|A_{n-i_k+1}\|_2^2} \right) \\
&= \frac{1}{\|A^\top r_k\|_2^2} \sum_{i_k \in [n] \setminus \{i_{k-1}, n-i_{k-1}+1\}} \left(\frac{1}{2} \frac{(A_{i_k}^\top r_k)^4}{\|A_{i_k}\|_2^2} + \frac{1}{2} \frac{(A_{n-i_k+1}^\top r_k)^4}{\|A_{n-i_k+1}\|_2^2} \right) \\
&\ge \frac{1}{\|A^\top r_k\|_2^2} \left(\frac{1}{2} \frac{\left(\sum_{i_k \in [n] \setminus \{i_{k-1}, n-i_{k-1}+1\}} (A_{i_k}^\top r_k)^2 \right)^2}{\sum_{i_k \in [n] \setminus \{i_{k-1}, n-i_{k-1}+1\}} \|A_{i_k}\|_2^2} + \frac{1}{2} \frac{\left(\sum_{i_k \in [n] \setminus \{i_{k-1}, n-i_{k-1}+1\}} (A_{n-i_k+1}^\top r_k)^2 \right)^2}{\sum_{i_k \in [n] \setminus \{i_{k-1}, n-i_{k-1}+1\}} \|A_{n-i_k+1}\|_2^2} \right) \\
&\ge \begin{cases} \frac{\|A^\top r_k\|_2^2}{\|A\|_F^2 - 2\min\{\|A_i\|_2^2, i \in [n]\}} & \text{if} \quad i_{k-1} \ne n - i_{k-1} + 1, \\ \frac{\|A^\top r_k\|_2^2}{\|A\|_F^2 - \min\{\|A_i\|_2^2, i \in [n]\}} & \text{if} \quad i_{k-1} = n - i_{k-1} + 1, \end{cases}
\end{aligned}
$$

$$\geq \begin{cases} \frac{\lambda_{\min}(A^\top A)\|r_k\|_2^2}{\|A\|_F^2 - 2\min\{\|A_i\|_2^2, i\in[n]\}} & \text{if} \quad i_{k-1} \neq n - i_{k-1} + 1, \\ \frac{\lambda_{\min}(A^\top A)\|r_k\|_2^2}{\|A\|_F^2 - \min\{\|A_i\|_2^2, i\in[n]\}} & \text{if} \quad i_{k-1} = n - i_{k-1} + 1, \end{cases}$$

where the first inequality is obtained by Lemma 1, the second inequality is due to Lemma 2 and $\text{span}\{A_{i_{k-1}}, A_{i_{n-i_{k-1}+1}}\} \perp r_k$. Thus,

$$\mathbb{E}_k\|x_{k+1} - A^\dagger b\|_{A^\top A}^2 \leq \left(1 - \frac{\lambda_{\min}(A^\top A)}{\tilde{\gamma}_k}\right)\|x_k - A^\dagger b\|_{A^\top A}^2, \ k \geq 1, \tag{3.10}$$

where

$$\tilde{\gamma}_k = \begin{cases} \|A\|_F^2 - \min\{\|A_i\|_2^2, i \in [n]\} \triangleq \gamma_1, & \text{if } i_{k-1} = n - i_{k-1} + 1, \\ \|A\|_F^2 - 2\min\{\|A_i\|_2^2, i \in [n]\} \triangleq \gamma_2, & \text{if } i_{k-1} \neq n - i_{k-1} + 1. \end{cases} \tag{3.11}$$

Note that, if n is odd and $k \geq 2$, we have

$$\begin{aligned}
&\mathbb{E}_{k-1}\|x_{k+1} - A^\dagger b\|_{A^\top A}^2 \\
&\leq \mathbb{E}_{k-1}\left(\left(1 - \frac{\lambda_{\min}(A^\top A)}{\tilde{\gamma}_k}\right)\|x_k - A^\dagger b\|_{A^\top A}^2\right) \\
&= \sum_{i_{k-1}\neq\frac{n+1}{2}}\left(\left(1 - \frac{\lambda_{\min}(A^\top A)}{\tilde{\gamma}_k}\right)\|x_k - A^\dagger b\|_{A^\top A}^2 \frac{1}{2}\left(\frac{(A_{i_{k-1}}^\top r_{k-1})^2}{\|A^\top r_{k-1}\|_2^2} + \frac{(A_{n-i_{k-1}+1}^\top r_{k-1})^2}{\|A^\top r_{k-1}\|_2^2}\right)\right) \\
&+ \frac{(A_{\frac{n+1}{2}}^\top r_{k-1})^2}{\|A^\top r_{k-1}\|_2^2}\left(1 - \frac{\lambda_{\min}(A^\top A)}{\gamma_1}\right)\|x_k - A^\dagger b\|_{A^\top A}^2 \\
&\leq \begin{cases} \left(1 - \frac{\lambda_{\min}(A^\top A)}{\gamma_2}\right)\sum\limits_{i_{k-1}=1}^{n}\left(\|x_k - A^\dagger b\|_{A^\top A}^2 \frac{1}{2}\left(\frac{(A_{i_{k-1}}^\top r_{k-1})^2}{\|A^\top r_{k-1}\|_2^2} + \frac{(A_{n-i_{k-1}+1}^\top r_{k-1})^2}{\|A^\top r_{k-1}\|_2^2}\right)\right) & \text{if } i_{k-2} = n - i_{k-2} + 1 = \frac{n+1}{2}, \\ \left(1 - \frac{\lambda_{\min}(A^\top A)}{\gamma_1}\right)\mathbb{E}_{k-1}\|x_k - A^\dagger b\|_{A^\top A}^2 & \text{if } i_{k-2} \neq n - i_{k-2} + 1, \end{cases} \\
&\leq \begin{cases} \left(1 - \frac{\lambda_{\min}(A^\top A)}{\gamma_2}\right)\mathbb{E}_{k-1}\|x_k - A^\dagger b\|_{A^\top A}^2 & \text{if } i_{k-2} = n - i_{k-2} + 1 = \frac{n+1}{2}, \\ \left(1 - \frac{\lambda_{\min}(A^\top A)}{\gamma_1}\right)\mathbb{E}_{k-1}\|x_k - A^\dagger b\|_{A^\top A}^2 & \text{if } i_{k-2} \neq n - i_{k-2} + 1, \end{cases} \\
&\leq \left(1 - \frac{\lambda_{\min}(A^\top A)}{\gamma_1}\right)\left(1 - \frac{\lambda_{\min}(A^\top A)}{\gamma_2}\right)\|x_{k-1} - A^\dagger b\|_{A^\top A}^2,
\end{aligned}$$

where the two inequalities next to the last are due to that $r_{k-1} \perp A_{\frac{n+1}{2}}$, while the last inequality is obtained by (3.11). If n is even, then

$$\mathbb{E}_{k-1}\|x_{k+1} - A^\dagger b\|_{A^\top A}^2 \leq \left(1 - \frac{\lambda_{\min}(A^\top A)}{\gamma_2}\right)^2 \|x_{k-1} - A^\dagger b\|_{A^\top A}^2, \ k \geq 2.$$

By taking the full expectation on both sides of (3.10), for $k \geq 2$, we get that

$$\begin{aligned}
&\mathbb{E}\|x_{k+1} - A^\dagger b\|_{A^\top A}^2 \\
&\leq \begin{cases} \left(1 - \frac{\lambda_{\min}(A^\top A)}{\gamma_2}\right)^2 \mathbb{E}\|x_{k-1} - A^\dagger b\|_{A^\top A}^2, & \text{if } n \text{ is even}, \\ \left(1 - \frac{\lambda_{\min}(A^\top A)}{\gamma_1}\right)\left(1 - \frac{\lambda_{\min}(A^\top A)}{\gamma_2}\right)\mathbb{E}\|x_{k-1} - A^\dagger b\|_{A^\top A}^2, & \text{if } n \text{ is odd}. \end{cases}
\end{aligned}$$

Note that

$$
\begin{aligned}
&\mathbb{E}\|x_1 - A^{\dagger}b\|_{A^\top A}^2 \\
&= \sum_{i_0=1}^{n}\left(\min_{x_1-x_0\in\text{span}\{e_{i_0},e_{n-i_0+1}\}}\|x_1 - A^{\dagger}b\|_{A^\top A}^2 \frac{1}{2}\left(\frac{(A_{i_0}^\top r_0)^2}{\|A^\top r_0\|^2} + \frac{(A_{n-i_0+1}^\top r_0)^2}{\|A^\top r_0\|^2}\right)\right) \\
&\le \frac{1}{2}\sum_{i_0=1}^{n}\left(\min_{x_1-x_0\in\text{span}\{e_{i_0}\}}\|x_1 - A^{\dagger}b\|_{A^\top A}^2 \frac{(A_{i_0}^\top r_0)^2}{\|A^\top r_0\|^2} + \min_{x_1-x_0\in\text{span}\{e_{n-i_0+1}\}}\|x_1 - A^{\dagger}b\|_{A^\top A}^2 \frac{(A_{n-i_0+1}^\top r_0)^2}{\|A^\top r_0\|^2}\right) \\
&\le \left(1 - \frac{\lambda_{\min}(A^\top A)}{\|A\|_F^2}\right)\|x_0 - A^{\dagger}b\|_{A^\top A}^2,
\end{aligned}
$$

where the last inequality is obtained by Theorem 2.2 in [8].

$$
\begin{aligned}
&\mathbb{E}\|x_2 - A^{\dagger}b\|_{A^\top A}^2 \\
&\le \begin{cases} \left(1 - \frac{\lambda_{\min}(A^\top A)}{\gamma_2}\right)\left(1 - \frac{\lambda_{\min}(A^\top A)}{\|A\|_F^2}\right)\|x_0 - A^{\dagger}b\|_{A^\top A}^2, & \text{if } n \text{ is even,} \\ \left(1 - \frac{\lambda_{\min}(A^\top A)}{\gamma_1}\right)\left(1 - \frac{\lambda_{\min}(A^\top A)}{\|A\|_F^2}\right)\mathbb{E}\|x_0 - A^{\dagger}b\|_{A^\top A}^2, & \text{if } n \text{ is odd.} \end{cases}
\end{aligned}
$$

By induction on k, if n is even, for $k \ge 0$, we have

$$
\mathbb{E}\|x_{k+1} - A^{\dagger}b\|_{A^\top A}^2 \le \left(1 - \frac{\lambda_{\min}(A^\top A)}{\gamma_2}\right)^k \left(1 - \frac{\lambda_{\min}(A^\top A)}{\|A\|_F^2}\right)\|x_0 - A^{\dagger}b\|_{A^\top A}^2,
$$

if n is odd, we have

$$
\begin{aligned}
&\mathbb{E}\|x_{k+1} - A^{\dagger}b\|_{A^\top A}^2 \\
&\le \begin{cases} \left(1 - \frac{\lambda_{\min}(A^\top A)}{\gamma_2}\right)^{\frac{k}{2}}\left(1 - \frac{\lambda_{\min}(A^\top A)}{\gamma_1}\right)^{\frac{k}{2}}\left(1 - \frac{\lambda_{\min}(A^\top A)}{\|A\|_F^2}\right)\|x_0 - A^{\dagger}b\|_{A^\top A}^2, & \text{if } k \text{ is even and } k \ge 0, \\ \left(1 - \frac{\lambda_{\min}(A^\top A)}{\gamma_2}\right)^{\frac{k-1}{2}}\left(1 - \frac{\lambda_{\min}(A^\top A)}{\gamma_1}\right)^{\frac{k+1}{2}}\left(1 - \frac{\lambda_{\min}(A^\top A)}{\|A\|_F^2}\right)\|x_0 - A^{\dagger}b\|_{A^\top A}^2, & \text{if } k \text{ is odd and } k \ge 1, \end{cases}
\end{aligned}
$$

the proof is completed. □

Remark 1. *Since* $\|A\|_F^2 - 2\min\{\|A_i\|_2^2, i \in [n]\} < \|A\|_F^2 - \min\{\|A_i\|_2^2, i \in [n]\}$*, our method is superior to NRGS in theory.*

4. Numerical examples

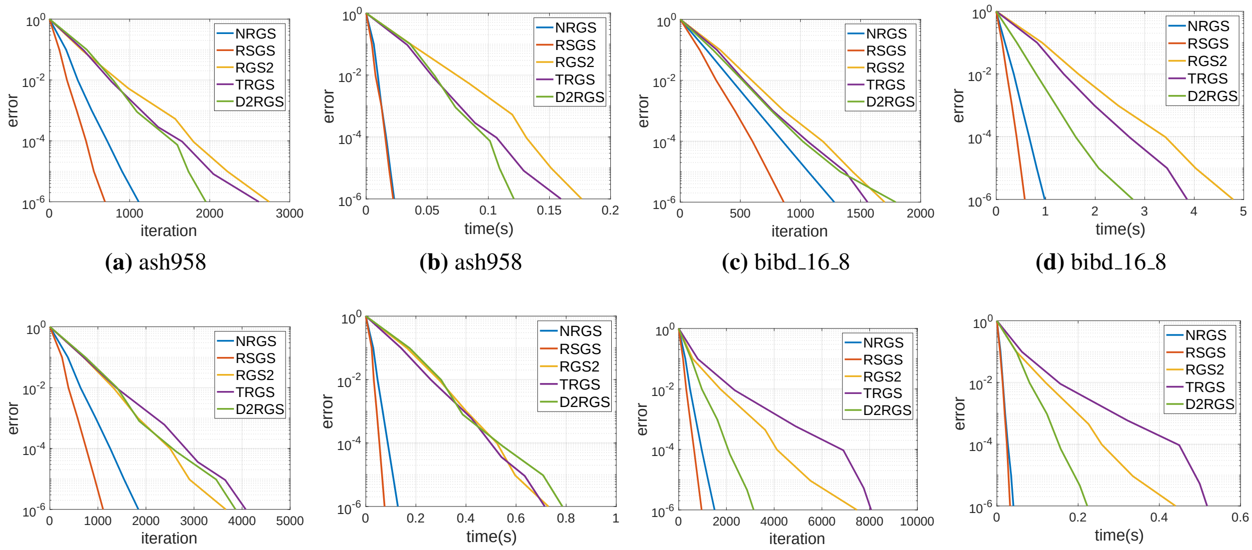
In this section, we present several examples that utilize two groups of real coefficient matrices. These examples are designed to compare the effectiveness of RSGS and NRGS, RGS2 [13], TRGS [13], and D2RGS [14] methods when solving Problem (1.1). The first group of matrices is chosen from the Florida sparse matrix collection [16] and is listed in Table 1, where for the matrices nemsafm, df2177 and bibd_16_8, we set A as their transpose. The second group consists of randomly generated dense matrices, produced using MATLAB's randn function. The inconsistent linear system is constructed by setting $b = Ax + r$, where x is a vector with entries generated from a standard normal distribution, and the residual r belongs to the null space of $A^\top$, which is derived using the null function in MATLAB. For all computations, the initial point is set as $x_0 = 0$, and the stopping criterion is:

$$\text{err} = \frac{\|x_k - x_*\|_2}{\|x_*\|_2} \leq 10^{-6}, \quad \text{where} \quad x_* = A^\dagger b.$$

Table 1. The properties of different sparse matrices.

Name	ash958	nemsafm	df2177	bibd_16_8
$m \times n$	958×292	334×2348	630×10358	120×12870
rank	292	334	630	120
Density	0.68%	0.36%	0.34%	23.33%
Condition number	3.20	4.77	2.01	9.54

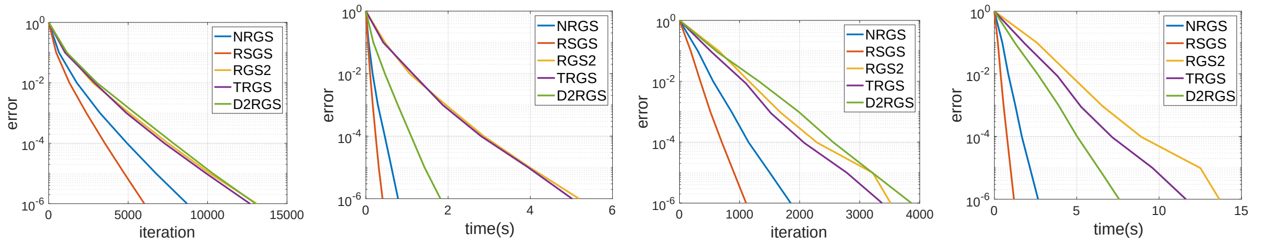
In order to evaluate and compare the performance of RSGS, NRGS and other baseline algorithms, we graph the relative error (err) against the metrics of IT and CPU times. These graphs are depicted in Figures 1 and 2, respectively. As can be observed from the figures, RSGS demonstrates more efficiency than the NRGS, RGS2 [13], TRGS [13], and D2RGS [14] methods.



(a) ash958 **(b)** ash958 **(c)** bibd_16_8 **(d)** bibd_16_8

(e) df2177 **(f)** df2177 **(g)** nemsafm **(h)** nemsafm

Figure 1. Compasions of different baselines in terms of iteration and running time on Florida sparse matrices.



(a) 1000×500 **(b)** 1000×500 **(c)** 5000×500 **(d)** 5000×500

Figure 2. Compasions of different baselines in terms of iteration and running time on dense matrices.

5. Conclusions

We proposed a randomized symmetric Gauss-Seidel method for solving linear least squares problems with the nonuniform sampling on the probability criterion (3.6). Our theoretical analysis indicates that RSGS converges when the coefficient matrix has full column rank. Furthermore, numerical experiments demonstrate that RSGS outperform the baseline algorithms.

Author contributions

Fan Sha: Write original draft, Study conception, Formal analysis; Jianbing Zhang: Supervision, Revision. All authors have read and approved the final version of the manuscript for publication.

Use of AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This research is supported by National Key R&D Program of China (Grant No. 2022YFA1004403) and by funding from oppo: Multi-modal information enhanced Image Caption.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. G. H. Golub, C. F. V. Loan, *Matrix computations*, Johns Hopkins University Press, 2013.
2. A. Hefny, D. Needell, A. Ramdas, Rows versus columns: Randomized Kaczmarz or Gauss-Seidel for ridge regression, *SIAM J. Sci. Comput.*, **39** (2017), S528–S542. https://doi.org/10.1137/16M1077891

3. Y. Saad, *Iterative methods for sparse linear systems*, Society for Industrial and Applied Mathematics, 2003.
4. T. Strohmer, R. Vershynin, A randomized Kaczmarz algorithm with exponential convergence, *J. Fourier Anal. Appl.*, **15** (2009), 262–278. https://doi.org/10.1007/s00041-008-9030-4
5. Z. Z. Bai, W. T. Wu, On greedy randomized coordinate descent methods for solving large linear least-squares problems, *Numer. Linear Algebra Appl.*, **26** (2019), e2237. https://doi.org/10.1002/nla.2237
6. K. Du, Tight upper bounds for the convergence of the randomized extended Kaczmarz and Gauss-Seidel algorithms, *Numer. Linear Algebra Appl.*, **26** (2019), e2233. https://doi.org/10.1002/nla.2233
7. R. M. Gower, P. Richtárik, Randomized iterative methods for linear systems, *SIAM J. Matrix Anal. A.*, **36** (2015), 1660–1690. https://doi.org/10.1137/15M1025487
8. Y. Q. Niu, B. Zheng, A new randomized Gauss-Seidel method for solving linear least-squares problems, *Appl. Math. Lett.*, **116** (2021), 107057. https://doi.org/10.1016/j.aml.2021.107057
9. J. Nutini, M. Schmidt, I. Laradji, M. Friedlander, H. Koepke, Coordinate descent converges faster with the gauss-southwell rule than random selection, *Proc. Int. Conf. Mach. Learn.*, 2015, 1632–1641. http://proceedings.mlr.press/v37/nutini15.pdf
10. D. Leventhal, A. S. Lewis, Randomized methods for linear constraints: Convergence rates and conditioning, *Math. Oper. Res.*, **35** (2010), 641–654. https://doi.org/10.1287/moor.1100.0456
11. A. Ma, D. Needell, A. Ramdas, Convergence properties of the randomized extended Gauss-Seidel and Kaczmarz methods, *SIAM J. Matrix Anal. Appl.*, **36** (2015), 1590–1604. https://doi.org/10.1137/15M1014425
12. Y. Liu, X. L. Jiang, C. Q. Gu, On maximum residual block and two-step Gauss-Seidel algorithms for linear least-squares problems, *Calcolo*, **58** (2015), 1–32. https://doi.org/10.1007/s10092-021-00404-x
13. Y. M. Liao, T. X. Lu, F. Yin, A two-step randomized Gauss-Seidel method for solving large-scale linear least squares problems, *Electron. Res. Arch.*, **30** (2022), 755–779. https://doi.org/10.3934/era.2022040
14. A. Mustafa, M. Saha, A two-dimensional randomized extended Gauss-Seidel algorithm for solving least squares problems, *Numer. Algor.*, (2023), 1–22. https://doi.org/10.1007/s11075-023-01661-3
15. Y. J. Guan, W. G. Li, L. L. Xing, T. T. Qiao, A note on convergence rate of randomized Kaczmarz method, *Calcolo*, **57** (2020), 1–11. https://doi.org/10.1007/s10092-020-00376-4
16. T. A. Davis, Y. F. Hu, The university of Florida sparse matrix collection, *Acm T. Math. Software*, **38** (2011), 1–25. https://doi.org/10.1145/2049662.2049663

AIMS Press