Journal of Applied Analysis and Computation
Volume 13, Number 2, April 2023, 913–927
Website:http://www.jaac-online.com
DOI:10.11948/20220230

THE GREEDY RANDOMIZED EXTENDED KACZMARZ ALGORITHM FOR NOISY LINEAR SYSTEMS

Na Chen[1,†] and Deliang Zhu[1]

Abstract For solving the noisy linear systems, we propose a new greedy randomized extended Kaczmarz algorithm by introducing an effective greedy criterion for selecting the working rows and a randomized orthogonal projection for reducing the influence of the noisy term. We prove that the solution of the proposed greedy randomized extended Kaczmarz algorithm converges in expectation to the least squares solution of the given linear system. Theoretical analysis indicate that the convergence rate of the greedy randomized extended Kaczmarz algorithm is much faster than the randomized extended Kaczmarz method, and numerical results also show that the proposed greedy randomized extended Kaczmarz algorithm is superior to the randomized extended Kaczmarz method. Moreover, for noisy linear systems, the proposed algorithm is more effcient than the greedy randomized Kaczmarz algorithm.

Keywords Kaczmarz algorithm, greedy criterion, randomized orthogonal projection, convergence rate.

MSC(2010) 49N45, 68W20.

1. Introduction

The Kaczmarz method is a popular iterative projection algorithm for solving (overdetermined) systems of linear equations [4]. It has been widely used in the field of image reconstruction, signal processing and sparse recovery in recent years [5,7,11,13].

Given a system of linear equations of the form

$$\mathbf{A}\mathbf{x} = \mathbf{b},$$

where $\mathbf{A} \in \mathbf{R}^{m\times n}$ and $\mathbf{b} \in \mathbf{R}^m$. The classic Kaczmarz algorithm goes through the rows of the coefficient matrix $\mathbf{A}$ and approximates the solution of the given linear systems by

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \frac{\mathbf{b}^{(i)} - \left\langle \mathbf{A}^{(i)}, \mathbf{x}_k \right\rangle}{\left\|\mathbf{A}^{(i)}\right\|_2^2} (\mathbf{A}^{(i)})^\top,$$

where $i = k \bmod m$, $\mathbf{A}^{(i)}$ is the ith row of the matrix $\mathbf{A}$. Here $\langle \cdot, \cdot \rangle$ is the inner product in $\mathbf{R}^n$ and $\|\cdot\|_2$ is the induced norm. Throughout the paper all vectors are assumed to be column vectors.

†The corresponding author. Email: chenna0407@hubu.edu,cn(N. Chen)
[1]Hubei Key Laboratory of Applied Mathematics, Faculty of Mathematics and Statistics, Hubei University, Wuhan 430062, China

It is proved that the convergence of the classic Kaczmarz algorithm depends on the ordering of the rows of matrix **A**, and the convergence rate can be improved by selecting the row with probability proportional to the square of its Euclidean norm. Strohmer and Vershynin proposed the randomized Kaczmarz method and proved that it converges with expected exponential rate [14]. Due to its simplicity and efficiency, the randomized Kaczmarz algorithm gets much wider applications [2, 6, 9, 12, 15]. However, the advantage of the random criterion disappeared when the coefficient matrix is scaled with a diagonal matrix that normalized the Euclidean norm of all of its rows to be a same constant. In order to solve this problem, the greedy randomized Kaczmarz (GRK) was introduced in [1]. It has been shown that the greedy randomized Kaczmarz method can work more efficient than the randomized Kaczmarz method by introducing an effective greedy probability criterion for selecting the working rows from the coefficient matrix. The greedy randomized Kaczmarz method was improved and applied in [17] and [3]. Above all, we conclude that the criterions for selecting the working row have an important influence on the convergence of the algorithm.

On the other hand, the linear systems of equations are easily corrupted by noise in many applications. In [10], Needell extended the randomized Kaczmarz algorithm to the noisy cases, and proved that the randomized Kaczmarz method can reach an error threshold dependent on the coefficient matrix with the same rate as in the noise-free case. Furthermore, the estimate vector is within a fixed distance from the solution, and the distance is proportional to the norm of the noise vector. For noisy systems, Zouzias and Freris proved that it is possible to efficiently reduce the norm of the "noisy" part by the randomized orthogonal projection algorithm [16]. The convergence rates for the extended Kaczmarz method were given in [8, 11, 16].

Motivated by these observations and the lack of research on noisy systems, we construct a greedy randomized extended Kaczmarz algorithm (GREK) for solving the noisy system of linear equations in this paper. In the proposed algorithm, we introduce an effective greedy criterion for selecting the working rows and a randomized orthogonal projection for reducing the influence of the noisy term. The convergence of the proposed algorithm indicate that the estimator converges in expectation to the minimum Euclidean norm least squares solution of the given linear system of equations. In addition, numerical experiments also show that the greedy randomized extended Kaczmarz algorithm significantly outperforms the randomized extended Kaczmarz method in terms of both iterations and relative standard error.

Compared to the greedy randomized Kaczmarz method and the randomized extended Kaczmarz method for solving the large-scale systems of linear equations, the main contributions below highlight several features of this paper.

First, the convergence rate of the Kaczmarz algorithm intensively depends on the selecting of the working row, the original Kaczmarz algorithm and the randomized Kaczmarz algorithm adopt sequential criterion and random criterion respectively. In this paper, we select the iterative row by a greedy criterion, which consider not only the ratio between the norms of the row of the coefficient matrix and the coefficient matrix itself as in randomized Kaczmarz method, but also the residual of the current iterate. In each iteration, the row farthest from the current point is preferentially selected, which can greatly improve the convergence rate.

Second, the greedy randomized Kaczmarz method provides obvious advantages over the standard randomized Kaczmarz method in many cases. It is proved that the solution of the greedy randomized Kaczmarz method converges to the least-norm solution of the linear system, and the convergence rate is much fast than the randomized Kaczmarz method [1]. Nevertheless, there is no such research for the more realistic cases where the systems are corrupted by noise. In this paper, we first provide convergence result of the greedy randomized Kaczmarz method for noisy linear systems, and prove that the estimate is within a fixed distance from the solution, and the distance is proportional to the norm of the noise vector.

Third, in order to decrease the effect of the noisy, we apply the randomized orthogonal projection technique to reduce the norm of the noise, and prove that the greedy randomized extended Kaczmarz algorithm converges with expected exponential rate. Furthermore, the theoretic analysis is deeply depend on the property of the index set and the probability criterion, which is different from the randomized Kaczmarz method.

The rest of this paper is organized as follows. In Section 2, we analyze the greedy randomized Kaczmarz method for the case where the system is corrupted by noise. In Section 3, we introduce the greedy randomized extended Kaczmarz method for noisy linear systems, and then establish the convergence theory. The numerical results are reported in Section 4. Finally, we provide a brief conclusion in Section 5.

2. Greedy randomized Kaczmarz solver for noisy linear systems

In this paper, we consider the noisy linear systems

$$\mathbf{Ax} = \mathbf{b}, \quad \mathbf{b} := \mathbf{y} + \mathbf{w}, \tag{2.1}$$

for any fixed vector $\mathbf{w} \in \mathbf{R}^m$. We first specify all the required notations. Let $\mathbf{A} \in \mathbf{R}^{m\times n}$ be the coefficient matrix, we denote the rows and columns of matrix $\mathbf{A}$ by $\mathbf{A}^{(1)}, \mathbf{A}^{(2)}, \cdots, \mathbf{A}^{(m)}$ and $\mathbf{A}_{(1)}, \mathbf{A}_{(2)}, \cdots, \mathbf{A}_{(n)}$, respectively. $\|\mathbf{A}\|_{\mathbf{F}}$ and $\|\mathbf{A}\|_2$ denote the Frobenius norm and the spectral norm. Let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\mathbf{rank(A)}}$ be the non-zero singular values of $\mathbf{A}$, we denote $\sigma_{\max} = \sigma_1$ and $\sigma_{\min} = \sigma_{\mathbf{rank(A)}}$. The Moore-Pensore pseudo-inverse of $\mathbf{A}$ is denoted by $\mathbf{A}^\dagger$, then $\|\mathbf{A}^\dagger\|_2 = 1/\sigma_{\min}$. For any non-zero matrix $\mathbf{A}$, the condition number of $\mathbf{A}$ is $\kappa(\mathbf{A}) := \|\mathbf{A}\|_2\|\mathbf{A}^\dagger\|_2 = \sigma_{\max}/\sigma_{\min}$, the scaled condition number is $\kappa_{\mathbf{F}}(\mathbf{A}) := \|\mathbf{A}\|_{\mathbf{F}}\|\mathbf{A}^\dagger\|_2$. We can easily check that $\kappa^2(\mathbf{A}) \leq \kappa_{\mathbf{F}}^2(\mathbf{A}) \leq \mathrm{rank}(\mathbf{A}) \cdot \kappa^2(\mathbf{A})$.

In [1], Bai and Wu proposed a greedy randomized Kaczmarz (GRK) algorithm, which can work more efficient than the randomized Kacmarz method by introducing an effective greedy probability criterion for selecting the working rows instead of the random criterion in randomized Kaczmarz method. The greedy randomized Kaczmarz algorithm can be described by Algorithm 1.

The greedy randomized Kaczmarz method provides obvious advantages over the standard randomized Kaczmarz method in many cases [1]. It converges significantly faster than the randomized Kaczmarz algorithm while ensuring shorter running time. Nevertheless, there is no such research for the more realistic cases where the systems are corrupted by noise. In this section, we provide convergence result of the greedy randomized Kaczmarz method for noisy linear systems, and prove that the estimate is within a fixed distance from the solution, and the distance is proportional to the norm of the noise vector.

Algorithm 1 Greedy Randomized Kaczmarz(GRK) Algorithm

Input: $\mathbf{A} \in \mathbf{R}^{m \times n}, \mathbf{b} \in \mathbf{R}^m, \ell$
Output: $\mathbf{x}^{(\ell)}$

Initialize $\mathbf{x}^{(0)} = \mathbf{0}$
for $k = 0, 1, 2, ..., \ell - 1$ **do**
Computer

$$\epsilon_k = \frac{1}{2}\left(\frac{1}{\left\|\mathbf{b} - \mathbf{A}\mathbf{x}^{(k)}\right\|_2^2} \max_{1 \le i_k \le m}\left\{\frac{\left|\mathbf{b}^{(i_k)} - \mathbf{A}^{(i_k)}\mathbf{x}^{(k)}\right|^2}{\left\|\mathbf{A}^{(i_k)}\right\|_2^2}\right\} + \frac{1}{\|\mathbf{A}\|_{\mathbf{F}}^2}\right)$$

Determine the index set of positive integers $\mathbf{U}_k$

$$\mathbf{U}_k = \left\{ i_k \middle| |\mathbf{b}^{(i_k)} - \mathbf{A}^{(i_k)}\mathbf{x}^{(k)}|^2 \ge \epsilon_k \left\|\mathbf{b} - \mathbf{A}\mathbf{x}^{(k)}\right\|_2^2 \left\|\mathbf{A}^{(i_k)}\right\|_2^2 \right\} \tag{2.2}$$

Computer the ith entry $\tilde{r}_k^{(i)}$ of the vector $\tilde{r}_k$ according to

$$\tilde{r}_k^i = \begin{cases} \mathbf{b}^{(i)} - \mathbf{A}^{(i)}\mathbf{x}^{(k)}, & i \in \mathbf{U}_k; \\ 0, & \text{otherwise.} \end{cases}$$

Select $i_k \in \mathbf{U}_k$ with probability $p_{i_k} = |\tilde{r}_k^{(i_k)}|^2 / \|\tilde{r}_k\|_2^2$
Set

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \frac{\mathbf{b}^{(i_k)} - \left\langle \mathbf{x}^{(k)}, \mathbf{A}^{(i_k)} \right\rangle}{\left\|\mathbf{A}^{(i_k)}\right\|_2^2} (\mathbf{A}^{(i_k)})^\top$$

return

Theorem 2.1 (Noisy greedy randomized Kaczmar)**.** *Assume that the system* $\mathbf{A}\mathbf{x} = \mathbf{y}$ *has a solution for some* $\mathbf{y} \in \mathbf{R}^m$*, denoted by* $\mathbf{x}^* := \mathbf{A}^\dagger \mathbf{y}$*. Let* $\hat{\mathbf{x}}^{(k)}$ *denote the kth iterate of the greedy randomized Kaczmarz algorithm applied to the linear system* $\mathbf{A}\mathbf{x} = \mathbf{b}$ *with* $\mathbf{b} := \mathbf{y} + \mathbf{w}$ *for any fixed* $\mathbf{w} \in \mathbf{R}^m$*. In exact arithmetic, it follows that*

$$\mathbf{E}|\hat{\mathbf{x}}^{(k)} - \mathbf{x}^*\|_2^2 \le \left[1 - \frac{1}{2}\left(\frac{1}{\gamma}\|A\|_{\mathbf{F}}^2 + 1\right)\frac{\sigma_{\min}^2(\mathbf{A})}{\|A\|_{\mathbf{F}}^2}\right] \mathbf{E}\|\hat{\mathbf{x}}^{(k-1)} - \mathbf{x}^*\|_2^2 + \frac{\|\mathbf{w}\|_2^2}{\left(\|\mathbf{A}\|_F - \frac{\gamma}{\|\mathbf{A}\|_F}\right)^2}, k = 2, 3, \cdots$$

where $\gamma = \max\limits_{1 \le i \le m} \sum\limits_{\substack{j=1 \\ j \ne i}}^{m} \|\mathbf{A}^{(j)}\|_2^2.$

Proof. Let $\mathcal{H}_i = \left\{\mathbf{x} : \left\langle \mathbf{A}^{(i)}, \mathbf{x} \right\rangle = \mathbf{y}^{(i)}\right\}$ be the affine hyper-plane consisting of the solutions to the $i-$th equation, and $\mathcal{H}_i^* = \left\{\mathbf{x} : \left\langle \mathbf{A}^{(i)}, \mathbf{x} \right\rangle = \mathbf{y}^{(i)} + \mathbf{w}^{(i)}\right\}$ be the solution spaces of the corresponding noisy equations.

Assume that at the $k-$th iterate of the greedy randomized Kaczmarz algorithm applied to the linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$, the $i-$ row is selected, and the solution is $\hat{\mathbf{x}}^{(k)}$. Denote $\mathbf{x}^{(k)}$ be the projection of $\hat{\mathbf{x}}^{(k-1)}$ on $\mathcal{H}_i$, we know that $\hat{\mathbf{x}}^{(k)}$ is the projection of $\hat{\mathbf{x}}^{(k-1)}$ on $\mathcal{H}_i^*$.

Since $\mathbf{x}^* \in \mathcal{H}_i$, we have

$$\|\hat{\mathbf{x}}^{(k)} - \mathbf{x}^*\|_2^2 = \|\mathbf{x}^{(k)} - \mathbf{x}^*\|_2^* + \|\hat{\mathbf{x}}^{(k)} - \mathbf{x}^{(k)}\|_2^2 \tag{2.3}$$

by orthogonality. Since the two hyper-planes $\mathcal{H}_i$ and $\mathcal{H}_i^*$ are parallel and the distance between $\mathcal{H}_i$ and $\mathcal{H}_i^*$ equals $\mathbf{w}^{(i)}/\|\mathbf{A}^{(i)}\|_2$. Note that $\mathbf{x}^{(k)}$ is the projection of $\hat{\mathbf{x}}^{(k-1)}$ on $\mathcal{H}_i$, the convergence property of the greedy randomized Kaczmarz method given by Theorem 3.1 in [1] tells us that

$$\mathbf{E}\|\mathbf{x}^{(k)} - \mathbf{x}^*\|_2^2 \leq \left[1 - \frac{1}{2}\left(\frac{1}{\gamma}\|A\|_{\mathbf{F}}^2 + 1\right)\frac{\sigma_{\min}^2(\mathbf{A})}{\|A\|_{\mathbf{F}}^2}\right]\mathbf{E}\|\hat{\mathbf{x}}^{(k-1)} - \mathbf{x}^*\|_2^2. \tag{2.4}$$

For the second term in (2.3), we have $\|\hat{\mathbf{x}}^{(k)} - \mathbf{x}^{(k)}\|_2^2 = \frac{\mathbf{w}^{(i)2}}{\|\mathbf{A}^{(i)}\|_2^2}$, then

$$\mathbf{E}\|\hat{\mathbf{x}}^{(k)} - \mathbf{x}^{(k)}\|_2^2 = \sum_{i=1}^{m} p_i \frac{{\mathbf{w}^{(i)}}^2}{\|\mathbf{A}^{(i)}\|_2^2},$$

where $p_i = |\tilde{r}_k^{(i)}|^2 / \|\tilde{r}_k\|_2^2$ and

$$\tilde{r}_k^i = \begin{cases} \mathbf{b}^{(i)} - \mathbf{A}^{(i)}\mathbf{x}^{(k)}, & i \in \mathbf{U}_k; \\ 0, & \text{otherwise.} \end{cases}$$

Then, we can obtain

$$\begin{aligned} &\mathbf{E}\|\hat{\mathbf{x}}^{(k)} - \mathbf{x}^{(k)}\|_2^2 \\ &= \sum_{i \in \mathbf{U}_k} \frac{|\mathbf{b}^{(i)} - \mathbf{A}^{(i)}\mathbf{x}^{(k)}|^2}{\sum_{i \in \mathbf{U}_k} |\mathbf{b}^{(i)} - \mathbf{A}^{(i)}\mathbf{x}^{(k)}|^2} \frac{{\mathbf{w}^{(i)}}^2}{\|\mathbf{A}^{(i)}\|_2^2} \\ &\leq \frac{\|\mathbf{A}\|_F^2}{\|\mathbf{b} - \mathbf{A}\mathbf{x}^{(k)}\|_2^2} \cdot \frac{1}{\sum_{i \in \mathbf{U}_k} \|\mathbf{A}^{(i)}\|_2^2} \cdot \sum_{i \in \mathbf{U}_k} \frac{|\mathbf{b}^{(i)} - \mathbf{A}^{(i)}\mathbf{x}^{(k)}|^2}{\|\mathbf{A}^{(i)}\|_2^2} {\mathbf{w}^{(i)}}^2 \\ &\leq \frac{\|\mathbf{A}\|_F^2}{\|\mathbf{b} - \mathbf{A}\mathbf{x}^{(k)}\|_2^2} \cdot \frac{1}{\min\limits_{1 \leq i \leq m} \|\mathbf{A}^{(i)}\|_2^2} \cdot \max_{1 \leq i \leq m} \frac{|\mathbf{b}^{(i)} - \mathbf{A}^{(i)}\mathbf{x}^{(k)}|^2}{\|\mathbf{A}^{(i)}\|_2^2} \sum_{i \in \mathbf{U}_k} {\mathbf{w}^{(i)}}^2 \\ &\leq \frac{\|\mathbf{A}\|_F^2}{(\|\mathbf{A}\|_F^2 - \gamma)^2} \cdot \|\mathbf{w}\|_2^2 = \frac{1}{\left(\|\mathbf{A}\|_F - \frac{\gamma}{\|\mathbf{A}\|_F}\right)^2} \cdot \|\mathbf{w}\|_2^2. \end{aligned} \tag{2.5}$$

The first inequality follows by the fact that

$$|\mathbf{b}^{(i)} - \mathbf{A}^{(i)}\mathbf{x}^{(k)}|^2 \geq \epsilon_k \|\mathbf{b} - \mathbf{A}\mathbf{x}^{(k)}\|_2^2 \|\mathbf{A}^{(i)}\|_2^2, \quad \forall i \in \mathbf{U}_k$$

and $\epsilon_k \geq \frac{1}{\|\mathbf{A}\|_F^2}$. The third inequality follows by the fact that $\min\limits_{1 \leq i \leq m} \|\mathbf{A}^{(i)}\|_2^2 = \|\mathbf{A}\|_F^2 - \gamma$.

Combining the equality (2.3) and the inequalities (2.4) and (2.5), the result in Theorem 2.1 can be obtained. □

Remark 2.1. For the noisy systems, the convergence factor of the greedy randomized Kaczmarz algorithm is smaller than that of randomized Kaczmarz algorithm in [10] since $\frac{1}{2}\left(\frac{1}{\gamma}\|A\|_{\mathbf{F}}^2 + 1\right) > 1$ and $1 - \frac{1}{2}\left(\frac{1}{\gamma}\|A\|_{\mathbf{F}}^2 + 1\right)\frac{\sigma_{\min}^2(\mathbf{A})}{\|A\|_{\mathbf{F}}^2} < 1 - \frac{\sigma_{\min}^2(\mathbf{A})}{\|A\|_{\mathbf{F}}^2}$. At the same time, the additive noise term $\|\mathbf{w}\|_2^2/\|\mathbf{A}\|_F^2$ of Theorem 7 in [16] is improved to $\|\mathbf{w}\|_2^2/\left(\|\mathbf{A}\|_F - \frac{\gamma}{\|\mathbf{A}\|_F}\right)^2$.

3. The greedy randomized extended Kaczmarz algorithm for noisy linear systems

The result in Theorem 2.1 indicates that the greedy randomized Kaczmarz method is effective for the noisy linear systems whose noise is negligible. In this section, we introduce a randomized approximate orthogonal projection to reduce the influence of the noisy term as in [16], and propose the greedy randomized extended Kaczmarz algorithm, see Algorithm 2. We prove that the solution of the proposed algorithm converges in expectation to the minimum ℓ_2- norm solution of the given linear system of equations.

Let $\mathcal{R}(\mathbf{A})$ be the column space of $\mathbf{A}$, and $\mathcal{R}(\mathbf{A})^\perp$ be the orthogonal complement of $\mathcal{R}(\mathbf{A})$. For given $\mathbf{b} \in \mathbf{R}^m$, denote $\mathbf{b}_{\mathcal{R}(\mathbf{A})}$ as the projection of $\mathbf{b}$ onto $\mathcal{R}(\mathbf{A})$, we can uniquely write $\mathbf{b} = \mathbf{b}_{\mathcal{R}(\mathbf{A})} + \mathbf{b}_{\mathcal{R}(\mathbf{A})^\perp}$.

The proposed algorithm consists of two main parts. The first part includes Steps 7 and 8, which maintain an approximation to $\mathbf{b}_{\mathcal{R}(\mathbf{A})}$ by $\mathbf{b} - \mathbf{z}^{(k)}$, it can efficiently reduce the norm of the "noise" part of $\mathbf{b}$. The second part (including Steps 3-6 and 9), applies the greedy randomized Kaczmarz algorithm to the system $\mathbf{A}\mathbf{x} = \mathbf{b} - \mathbf{z}^{(k)}$. Since $\mathbf{b} - \mathbf{z}^{(k)}$ converges to $\mathbf{b}_{\mathcal{R}(\mathbf{A})}$, $\mathbf{x}^{(k)}$ will converge to the minimum Euclidean norm solution of $\mathbf{A}\mathbf{x} = \mathbf{b}_{\mathcal{R}(\mathbf{A})}$, which equals to $\mathbf{x}^* = \mathbf{A}^\dagger \mathbf{y}$.

To illustrate the convergence property of the greedy randomized extended Kaczmarz method, we establish and demonstrate the following theorem.

Theorem 3.1. *Assume that the system* $\mathbf{A}\mathbf{x} = \mathbf{y}$ *has a solution for some* $\mathbf{y} \in \mathbf{R}^m$*, denoted by* $\mathbf{x}^* = \mathbf{A}^\dagger \mathbf{y}$*. Let* $\mathbf{x}^{(T)}$ *denote the* T*th iterate of the greedy randomized extended Kaczmarz algorithm applied to the linear system* $\mathbf{A}\mathbf{x} = \mathbf{y} + \mathbf{w}$ *for any fixed* $\mathbf{w} \in \mathbf{R}^m$*, then the iteration sequence* $\{\mathbf{x}^{(T)}\}_{T=0}^{\infty}$ *converges to* $\mathbf{x}^*$ *in expectation:*

$$
\begin{aligned}
&\mathbf{E}\left\|\mathbf{x}^{(T)} - \mathbf{x}^*\right\|_2^2 \\
&\leq \left(1 - \frac{1}{\kappa_{\mathbf{F}}^2(\mathbf{A})}\right)^{\lfloor T/2 \rfloor} \left(1 + \frac{2}{\left(1 - \frac{\gamma}{\|\mathbf{A}\|_{\mathbf{F}}^2}\right)^2 \left(\frac{1}{\gamma}\|\mathbf{A}\|_{\mathbf{F}}^2 + 1\right)} \kappa^2(\mathbf{A})\right) \|\mathbf{x}^*\|_2^2, \qquad (3.1)
\end{aligned}
$$

where $\gamma = \max\limits_{1 \leq i \leq m} \sum\limits_{\substack{j=1 \\ j \neq i}}^{m} \|\mathbf{A}^{(j)}\|_2^2$.

In the proposed algorithm, we firstly use a randomized orthogonal projection to reduce the norm of the noise as in steps 7 and 8, where $\mathbf{b} - \mathbf{z}^{(k)}$ is an approximation of $\mathbf{b}_{\mathcal{R}(\mathbf{A})}$. As in [16], Zouzias and Freris demonstrated the following lemma.

Lemma 3.1. *Let* $\mathbf{A} \in \mathbf{R}^{m \times n}$ *and* $\mathbf{b} \in \mathbf{R}^m$*. For any integer* $k > 0$*, after* k *iterations of Algorithm 1 it holds that*

$$\mathbf{E}\|\mathbf{z}^{(k)} - \mathbf{b}_{\mathcal{R}(\mathbf{A})^\perp}\|_2^2 \leq \left(1 - \frac{1}{\kappa_{\mathbf{F}}^2(\mathbf{A})}\right)^k \|\mathbf{b}_{\mathcal{R}(\mathbf{A})}\|_2^2.$$

Bai and Wu in [1] proved that the estimate of greedy randomized Kaczmarz method converges to the unique least-norm solution of the noiseless linear system. For noisy linear systems, we can also prove that the estimate of greedy randomized Kaczmarz algorithm is within a fixed distance from the solution, and the distance is proportional to the norm of the noise vector $\mathbf{w}$ as in [10].

Algorithm 2 Greedy Randomized Extended Kaczmarz(GREK) Algorithm

Input: $\mathbf{A} \in \mathbf{R}^{m\times n}, \mathbf{b} \in \mathbf{R}^m, \ell$

Output: $\mathbf{x}^{(\ell)}$

1: Initialize $\mathbf{x}^{(0)} = \mathbf{0}$, $\mathbf{z}^{(0)} = \mathbf{b}$

2: **for** $k = 0, 1, 2, ..., \ell - 1$ **do**

3: Computer

$$\epsilon_k = \frac{1}{2}\left(\frac{1}{\left\|\mathbf{b}-\mathbf{z}^{(k)}-\mathbf{A}\mathbf{x}^{(k)}\right\|_2^2}\max_{1\le i_k\le m}\left\{\frac{\left|\mathbf{b}^{(i_k)}-\mathbf{z}_{i_k}^{(k)}-\mathbf{A}^{(i_k)}\mathbf{x}^{(k)}\right|^2}{\left\|\mathbf{A}^{(i_k)}\right\|_2^2}\right\}\right)+\frac{1}{2\|\mathbf{A}\|_{\mathbf{F}}^2}$$

4: Determine the index set of positive integers $\mathbf{U}_k$:

$$\left\{i_k\left||\mathbf{b}^{(i_k)} - \mathbf{z}_{i_k}^{(k)} - \mathbf{A}^{(i_k)}\mathbf{x}^{(k)}|^2 \ge \epsilon_k \left\|\mathbf{b} - \mathbf{z}^{(k)} - \mathbf{A}\mathbf{x}^{(k)}\right\|_2^2 \left\|\mathbf{A}^{(i_k)}\right\|_2^2\right.\right\} \quad (3.2)$$

5: Computer the ith entry $\tilde{r}_k^{(i)}$ of the vector $\tilde{r}_k$ according to

$$\tilde{r}_k^i = \begin{cases} \mathbf{b}^{(i)} - \mathbf{z}_i^{(k)} - \mathbf{A}^{(i)}\mathbf{x}^{(k)}, & i \in \mathbf{U}_k; \\ 0, & \text{otherwise.} \end{cases}$$

6: Select $i_k \in \mathbf{U}_k$ with probability $p_{i_k} = |\tilde{r}_k^{(i_k)}|^2 / \|\tilde{r}_k\|_2^2$

7: Select $j_k \in [n]$ with probability $q_{j_k} = \left\|\mathbf{A}_{(j_k)}\right\|_2^2 / \|\mathbf{A}\|_{\mathbf{F}}^2$

8: Set

$$\mathbf{z}^{(k+1)} = \mathbf{z}^{(k)} - \frac{\left\langle \mathbf{A}_{(j_k)}, \mathbf{z}^{(k)}\right\rangle}{\left\|\mathbf{A}_{(j_k)}\right\|_2^2}\mathbf{A}_{(j_k)}$$

9: Set

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \frac{\mathbf{b}^{(i_k)} - \mathbf{z}_{i_k}^{(k)} - \left\langle \mathbf{x}^{(k)}, \mathbf{A}^{(i_k)}\right\rangle}{\left\|\mathbf{A}^{(i_k)}\right\|_2^2}(\mathbf{A}^{(i_k)})^\top$$

10: **return**

Based on the above Lemma, we are now prepared to prove Theorem 3.1.

Proof of Theorem 3.1. Set $\alpha = 1 - \frac{1}{\kappa_{\mathbf{F}}^2(\mathbf{A})}$, denote $\mathbf{E}[\cdot]$ be the full expectation and $\mathbf{E}_k[\cdot]$ be the conditional expectation with respect to the first k iterations of Algorithm 2, that's $\mathbf{E}_k[\cdot] = \mathbf{E}[\cdot|i_0, j_0, i_1, j_1, \cdots, i_{k-1}, j_{k-1}]$, and from the law of the iterated expectation, we have $\mathbf{E}[\mathbf{E}_k[\cdot]] = \mathbf{E}[\cdot]$. Observe that steps 7 and 8 are independent from steps $3-6$ and 9, so Lemma 3.1 implies for every $\ell \ge 0$, we have

$$\mathbf{E}\|\mathbf{z}^{(\ell)} - \mathbf{b}_{\mathcal{R}(\mathbf{A})^\perp}\|_2^2 \le \alpha^\ell \|\mathbf{b}_{\mathcal{R}(\mathbf{A})}\|_2^2 \le \|\mathbf{b}_{\mathcal{R}(\mathbf{A})}\|_2^2, \quad (3.3)$$

where $\alpha = 1 - \frac{1}{\kappa_{\mathbf{F}}^2(\mathbf{A})}$.

Fix a parameter $k^* = \lfloor T/2 \rfloor$, after the k^*-th iteration of Algorithm 2, it follows from Theorem 2.1 that

$$\mathbf{E}_{(k^*-1)}\|\mathbf{x}^{(k^*)}-\mathbf{x}^*\|_2^2 \leq \beta\|\mathbf{x}^{(k^*-1)}-\mathbf{x}^*\|_2^2 + \frac{\|\mathbf{b}_{\mathcal{R}(\mathbf{A})^\perp}-\mathbf{z}^{(k^*-1)}\|_2^2}{\left(\|\mathbf{A}\|_F - \frac{\gamma}{\|\mathbf{A}\|_F}\right)^2}, \tag{3.4}$$

where $\beta = \left[1-\frac{1}{2}\left(\frac{1}{\gamma}\|A\|_{\mathbf{F}}^2+1\right)\frac{\sigma_{\min}^2(\mathbf{A})}{\|A\|_{\mathbf{F}}^2}\right]$. Indeed the greedy randomized Kaczmarz algorithm (Step 3, 4, 5, 6 and 9) is execute with input $(\mathbf{A}, \mathbf{b}-\mathbf{z}^{(k^*-1)})$ and current estimate vector $\mathbf{x}^{(k^*-1)}$. Set $\mathbf{y}=\mathbf{b}_{\mathcal{R}(\mathbf{A})}$ and $\mathbf{w}=\mathbf{b}_{\mathcal{R}(\mathbf{A})^\perp}-\mathbf{z}^{(k^*-1)}$ in Theorem 2.1.

Then, taking the full expectation on both sides of the inequality (3.4), and using the recursive relation iteratively, it holds that

$$\begin{aligned}
&\mathbf{E}\|\mathbf{x}^{(k^*)}-\mathbf{x}^*\|_2^2\\
&\leq \beta\mathbf{E}\|\mathbf{x}^{(k^*-1)}-\mathbf{x}^*\|_2^2 + \frac{\mathbf{E}\|\mathbf{b}_{\mathcal{R}(\mathbf{A})^\perp}-\mathbf{z}^{(k^*-1)}\|_2^2}{\left(\|\mathbf{A}\|_F - \frac{\gamma}{\|\mathbf{A}\|_F}\right)^2}\\
&\leq \beta\mathbf{E}\|\mathbf{x}^{(k^*-1)}-\mathbf{x}^*\|_2^2 + \frac{\|\mathbf{b}_{\mathcal{R}(\mathbf{A})}\|_2^2}{\left(\|\mathbf{A}\|_F - \frac{\gamma}{\|\mathbf{A}\|_F}\right)^2} \quad (\text{by Ineq.}(3.3))\\
&\leq \cdots \leq \beta^{k^*}\|\mathbf{x}^{(0)}-\mathbf{x}^*\|_2^2 + \sum_{\ell=0}^{k^*-2}\beta^\ell \frac{\|\mathbf{b}_{\mathcal{R}(\mathbf{A})}\|_2^2}{\left(\|\mathbf{A}\|_F - \frac{\gamma}{\|\mathbf{A}\|_F}\right)^2}\\
&\leq \|\mathbf{x}^*\|_2^2 + \sum_{\ell=0}^{k^*-2}\beta^\ell \frac{\|\mathbf{b}_{\mathcal{R}(\mathbf{A})}\|_2^2}{\left(\|\mathbf{A}\|_F - \frac{\gamma}{\|\mathbf{A}\|_F}\right)^2},
\end{aligned}$$

since $\beta<1$ and $\mathbf{x}^{(0)}=0$. Simplifying the right hand side using the fact that

$$\sum_{\ell=0}^{k^*-2}\beta^\ell = \frac{1}{1-\beta} = \frac{2\|A\|_{\mathbf{F}}^2}{\left(\frac{1}{\gamma}\|\mathbf{A}\|_{\mathbf{F}}^2+1\right)\sigma_{\min}^2(\mathbf{A})},$$

it follows

$$\mathbf{E}\|\mathbf{x}^{(k^*)}-\mathbf{x}^*\|_2^2 \leq \|\mathbf{x}^*\|_2^2 + \frac{2\|\mathbf{b}_{\mathcal{R}(\mathbf{A})}\|_2^2\|\mathbf{A}\|_{\mathbf{F}}^2}{\left(\|\mathbf{A}\|_F - \frac{\gamma}{\|\mathbf{A}\|_F}\right)^2\left(\frac{1}{\gamma}\|\mathbf{A}\|_{\mathbf{F}}^2+1\right)\sigma_{\min}^2(\mathbf{A})}.$$

Moreover, observe that for every $\ell \geq 0$, we have

$$\mathbf{E}\|\mathbf{z}^{(\ell+k^*)}-\mathbf{b}_{\mathcal{R}(\mathbf{A})^\perp}\|_2^2 \leq \alpha^{\ell+k^*}\|\mathbf{b}_{\mathcal{R}(\mathbf{A})}\|_2^2 \leq \alpha^{k^*}\|\mathbf{b}_{\mathcal{R}(\mathbf{A})}\|_2^2.$$

Now for any $k>0$

$$\begin{aligned}
&\mathbf{E}\|\mathbf{x}^{(k^*+k)}-\mathbf{x}^*\|_2^2\\
&\leq \beta\mathbf{E}\|\mathbf{x}^{(k^*+k-1)}-\mathbf{x}^*\|_2^2 + \frac{\mathbf{E}\|\mathbf{b}_{\mathcal{R}(\mathbf{A})^\perp}-\mathbf{z}^{(k^*+k-1)}\|_2^2}{\left(\|\mathbf{A}\|_F - \frac{\gamma}{\|\mathbf{A}\|_F}\right)^2}\\
&\leq \cdots \leq \beta^k\mathbf{E}\|\mathbf{x}^{(k^*)}-\mathbf{x}^*\|_2^2 + \sum_{\ell=0}^{k-1}\beta^{(k-1)-\ell}\cdot\frac{\mathbf{E}\|\mathbf{b}_{\mathcal{R}(\mathbf{A})^\perp}-\mathbf{z}^{(\ell+k^*)}\|_2^2}{\left(\|\mathbf{A}\|_F - \frac{\gamma}{\|\mathbf{A}\|_F}\right)^2}
\end{aligned}$$

$$
\begin{aligned}
&\le \beta^k \mathbf{E}\|\mathbf{x}^{(k^*)} - \mathbf{x}^*\|_2^2 + \frac{\alpha^{k^*}\|\mathbf{b}_{\mathcal{R}(\mathbf{A})}\|_2^2}{\left(\|\mathbf{A}\|_F - \frac{\gamma}{\|\mathbf{A}\|_F}\right)^2}\sum_{\ell=0}^{k-1}\beta^\ell \\
&\le \beta^k\left(\|\mathbf{x}^*\|_2^2 + \frac{2\|\mathbf{b}_{\mathcal{R}(\mathbf{A})}\|_2^2}{\left(1-\frac{\gamma}{\|\mathbf{A}\|_\mathbf{F}^2}\right)^2\left(\frac{1}{\gamma}\|\mathbf{A}\|_\mathbf{F}^2+1\right)\sigma_{\min}^2(\mathbf{A})}\right) \\
&\qquad + \frac{2\alpha^{k^*}\|\mathbf{b}_{\mathcal{R}(\mathbf{A})}\|_2^2}{\left(\|1-\frac{\gamma}{\|\mathbf{A}\|_\mathbf{F}^2}\right)^2\left(\frac{1}{\gamma}\|\mathbf{A}\|_\mathbf{F}^2+1\right)\sigma_{\min}^2(\mathbf{A})} \\
&= \beta^k\|\mathbf{x}^*\|_2^2 + (\beta^k+\alpha^{k^*})\frac{2\|\mathbf{b}_{\mathcal{R}(\mathbf{A})}\|_2^2}{\left(1-\frac{\gamma}{\|\mathbf{A}\|_\mathbf{F}^2}\right)^2\left(\frac{1}{\gamma}\|\mathbf{A}\|_\mathbf{F}^2+1\right)\sigma_{\min}^2(\mathbf{A})} \\
&\le \beta^k\|\mathbf{x}^*\|_2^2 + (\beta^k+\alpha^{k^*})\frac{2}{\left(1-\frac{\gamma}{\|\mathbf{A}\|_\mathbf{F}^2}\right)^2\left(\frac{1}{\gamma}\|\mathbf{A}\|_\mathbf{F}^2+1\right)}\kappa^2(\mathbf{A})\|\mathbf{x}^*\|_2^2,
\end{aligned}
$$

the last inequality is due to the fact that $\|\mathbf{b}_{\mathcal{R}(\mathbf{A})}\|_2 \le \sigma_{\max}\|\mathbf{x}^*\|_2$ and $\kappa^2(\mathbf{A}) = \sigma_{\max}^2(\mathbf{A})/\sigma_{\min}^2(\mathbf{A})$. Now, consider two cases, if T is even, set $k = k^*$, otherwise set $k = k^*+1$, then $\beta^k \le \beta^{k^*}$ and

$$
\mathbf{E}\|\mathbf{x}^{(k^*+k)} - \mathbf{x}^*\|_2^2 \le \alpha^{k^*}\left(1 + \frac{2}{\left(1-\frac{\gamma}{\|\mathbf{A}\|_\mathbf{F}^2}\right)^2\left(\frac{1}{\gamma}\|\mathbf{A}\|_\mathbf{F}^2+1\right)}\kappa^2(\mathbf{A})\right)\|\mathbf{x}^*\|_2^2, \tag{3.5}
$$

since $\beta < \alpha$. □

Remark 3.1. The greedy randomized extended Kaczmarz algorithm converges faster than the randomized extended Kaczmarz algorithm because it adopts a probability criterion that is essentially determined by the largest entry of the residual with respect to the current iterate. In fact, the convergence factor of the proposed algorithm is smaller than that of randomized extended Kaczmarz algorithm in [16] since

$$
1 + \frac{2}{\left(1-\frac{\gamma}{\|\mathbf{A}\|_\mathbf{F}^2}\right)^2\left(\frac{1}{\gamma}\|\mathbf{A}\|_\mathbf{F}^2+1\right)}\kappa^2(\mathbf{A}) < 1 + 2\kappa^2(\mathbf{A}),
$$

and also we have magnified the factor $\left(\beta^k + (\beta^k+\alpha^{k^*})\frac{2}{\left(1-\frac{\gamma}{\|\mathbf{A}\|_\mathbf{F}^2}\right)^2\left(\frac{1}{\gamma}\|\mathbf{A}\|_\mathbf{F}^2+1\right)}\kappa^2(\mathbf{A})\right)\|\mathbf{x}^*\|_2^2$ by $\beta^k \le \beta^{k^*} \le \alpha^{k^*}$ in inequality (3.5).

Moreover, the most time-consuming step in Algorithm 2 is the calculating of the residual vector r_k, but we observe that

$$
\begin{aligned}
r_{k+1} &= \mathbf{b} - \mathbf{z}^{(k+1)} - \mathbf{A}\mathbf{x}^{(k+1)} \\
&= \mathbf{b} - \mathbf{z}^{(k)} + \frac{\langle \mathbf{A}_{(j_k)}, \mathbf{z}^{(k)}\rangle}{\|\mathbf{A}_{(j_k)}\|_2^2}\mathbf{A}_{(j_k)} - \mathbf{A}\mathbf{x}^{(k)} - \frac{\mathbf{b}^{(i_k)} - \mathbf{z}_{i_k}^{(k)} - \langle \mathbf{x}^{(k)}, \mathbf{A}^{(i_k)}\rangle}{\|\mathbf{A}^{(i_k)}\|_2^2}\mathbf{A}(\mathbf{A}^{(i_k)})^\top \\
&= r_k + \frac{\langle \mathbf{A}_{(j_k)}, \mathbf{z}^{(k)}\rangle}{\|\mathbf{A}_{(j_k)}\|_2^2}\mathbf{A}_{(j_k)} - \frac{\mathbf{b}^{(i_k)} - \mathbf{z}_{i_k}^{(k)} - \langle \mathbf{x}^{(k)}, \mathbf{A}^{(i_k)}\rangle}{\|\mathbf{A}^{(i_k)}\|_2^2}\mathbf{B}_{(ik)}
\end{aligned} \tag{3.6}
$$

where $\mathbf{B} = \mathbf{A}\mathbf{A}^T$.

The recursive formula given in (3.6) indicate that the algorithm could be faster if we have $\mathbf{B} = \mathbf{A}\mathbf{A}^T$ at the beginning. Meanwhile, the second and the third part of (3.6) are exactly the amount to be calculated in step 8 and step 9 of the greedy randomized extended Kaczmarz algorithm.

4. Experimental results

In this section, we illustrate the performance of the proposed greedy randomized extended Kaczmarz(GREK) algorithm versus the randomized extended Kaczmarz(REK) algorithm and the greedy randomized Kaczmarz(GRK) algorithm in there experiments. The results indicate that the former is numerically better than the latter in terms of the number of iteration steps(IT), the computing time in seconds(CPU) and the relative standard error(RSE). The results in the experiment are the average of 50 repeated runs of the correspongding method.

In the experiment, we set $\mathbf{x}^{(0)} = \mathbf{0}$ and $\mathbf{z}^{(0)} = \mathbf{b}$, the RSE is defined as follow

$$RSE = \frac{\|\mathbf{x}^{(k)} - \mathbf{x}^*\|_2^2}{\|\mathbf{x}^*\|_2^2}$$

and the iteration terminates when the RSE less than 10^{-5}. The coefficient matrix $\mathbf{A} \in \mathbf{R}^{m\times n}$ and the solution vector $\mathbf{x}^* \in \mathbf{R}^n$ are randomly generated. The vector $\mathbf{y} \in \mathbf{R}^m$ is generated by $\mathbf{y} = \mathbf{A}\mathbf{x}^*$. Then we set $\mathbf{w} = noiselev \cdot \frac{\|\mathbf{y}\|_2}{\|\mathbf{e}\|_2} \cdot \mathbf{e}$, where the level of the noise is in $[0,1]$ and the vector $\mathbf{e}$ is randomly generated. Thus $\mathbf{b} = \mathbf{A}\mathbf{x}^* + \mathbf{w}$. We apply the greedy randomized extended Kaczmarz algorithm and the randomized extended Kaczmarz algorithm to the linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$.

Firstly, we implement the GREK and REK for overdetermined and underdetermined systems without noise. Table 1, Table 2 and Table 3 show the results for the underdetermined systems, Table 4, Table 5 and Table 6 shows the results for the overdetermined systems. In both cases, the results demonstrate that the GREK method outperforms the REK method in terms of both the numbers of iteration steps and CPU times.

Figure 1 shows the relative standard errors for GREK and REK with $m = 1000, 2000, 3000$ and $n = 150$. Figure 2 shows the relative standard errors for GREK and REK with $m = 400$ and $n = 3000, 4000, 5000$. Note that the GREK method performs consistently well over all trials, and the convergence rate of the GREK method is much faster than the REK method.

Then, in order to further verify the effectiveness of the GREK method, we compare the REK, GRK and the proposed algorithm for noisy linear systems, including overdetermined systems and underdetermined systems. Figure 3 shows the results for GREK, GRK and REK with noisy linear systems. The results indicate that the convergence rate of the GREK method is much faster than the REK method, and the GRK method fails with the increase of the noise level.

We also implement the GREK and REK for some practical application matrices. The coefficient matrices comes from https://sparse.tamu.edu/. The variables $\mathbf{y}, \mathbf{x}^*, \mathbf{b}, \mathbf{w}$ are generated as before. The results given in Table 7 still show that the proposed GREK algorithm is effective.

Table 1. IT of GREK and REK for matrices **A** with $m = 50$ and different n

	n	1000	2000	3000	4000	5000
REK	IT	683	642	714	592	602
	CPU	0.238	0.234	0.268	0.23	0.244
GREK	IT	**582**	**542**	**607**	**516**	**504**
	CPU	**0.216**	**0.212**	**0.24**	**0.207**	**0.208**

Table 2. IT of GREK and REK for matrices **A** with $m = 1000$ and different n

	n	1000	2000	3000	4000	5000
REK	IT	1671	1517	1458	1450	1469
	CPU	0.574	0.554	0.554	0.567	0.588
GREK	IT	**1375**	**1303**	**1262**	**1263**	**1311**
	CPU	**0.514**	**0.51**	**0.519**	**0.537**	**0.582**

Table 3. IT of GREK and REK for matrices **A** with $m = 1000$ and different n

	n	1000	2000	3000	4000	5000
REK	IT	2906	2479	2311	2209	2146
	CPU	1.034	0.91	0.87	0.853	0.86
GREK	IT	**2317**	**2065**	**1965**	**1861**	**1845**
	CPU	**0.909**	**0.837**	**0.847**	**0.848**	**0.851**

Table 4. IT of GREK and REK for matrices **A** with $n = 50$ and different m

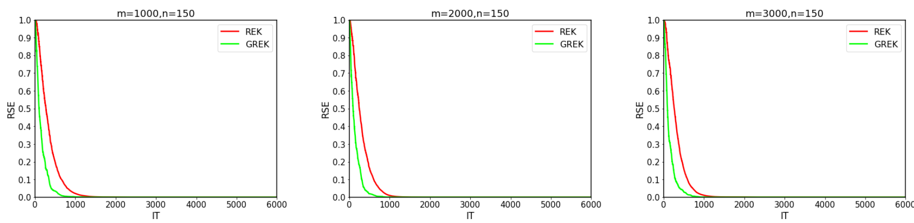
	m	1000	2000	3000	4000	5000
REK	IT	731	681	677	675	647
	CPU	0.201	0.189	0.19	0.201	0.228
GREK	IT	**467**	**423**	**412**	**399**	**383**
	CPU	**0.181**	**0.172**	**0.178**	**0.184**	**0.182**

Table 5. IT of GREK and REK for matrices **A** with $n = 100$ and different m

	m	1000	2000	3000	4000	5000
REK	IT	1627	1441	1404	1366	1370
	CPU	0.554	0.511	0.507	0.505	0.54
GREK	IT	**1196**	**986**	**967**	**917**	**881**
	CPU	**0.462**	**0.414**	**0.431**	**0.432**	**0.448**

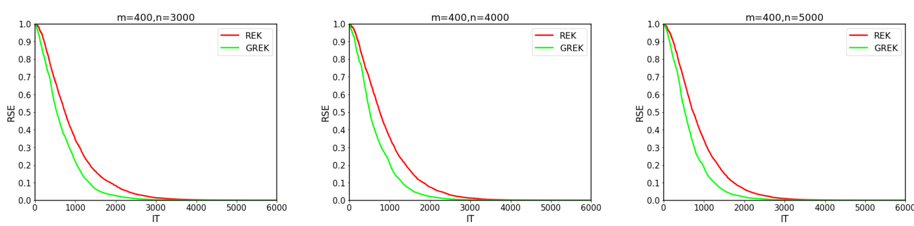
Table 6. IT of GREK and REK for matrices **A** with $n = 150$ and different m

	m	1000	2000	3000	4000	5000
REK	IT	2740	2281	2162	2164	2092
	CPU	0.959	0.82	0.799	0.828	0.832
GREK	IT	**2024**	**1626**	**1486**	**1455**	**1416**
	CPU	**0.815**	**0.711**	**0.699**	**0.732**	**0.748**



(a) (b) (c)

Figure 1. RSE versus IT for GREK and REK with $m = 1000, 2000, 3000$ and $n = 150$.

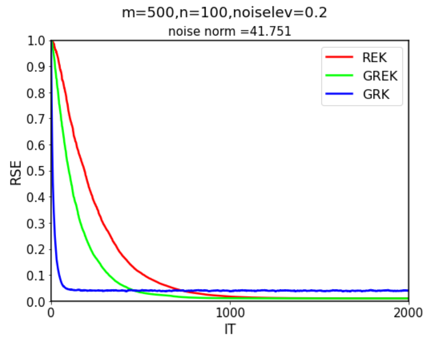


(a) (b) (c)

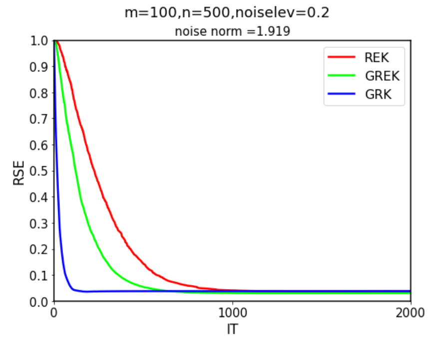
Figure 2. RSE versus IT for GREK and REK with $m = 400$ and $n = 3000, 4000, 5000$.

Table 7. IT of GREK and REK for practical application matrices **A**

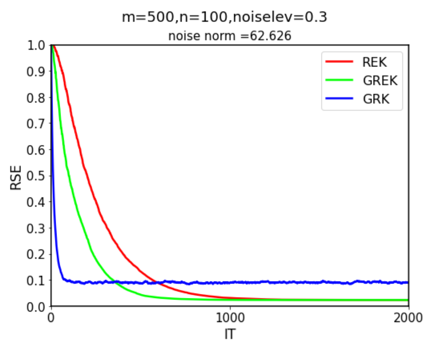
matrix A	$m * n$	cond(A)	REK(IT(CPU))	GREK(IT(CPU))
Stranke94	10*10	5.17	45330(11.916)	**21119(5.908)**
divorce	50*9	19.39	3660(1.005)	**2391(0.704)**
ash291	219*85	3.02	2236(0.776)	**1570(0.63)**
ash958	958*292	3.2	7035(2.629)	**4956(2.012)**
ash608	608*188	3.37	4537(1.587)	**3170(1.516)**
cari	400*1200	3.31	5747(2.14)	**4098(1.639)**



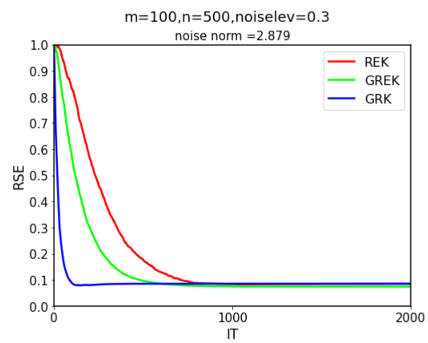
(a)



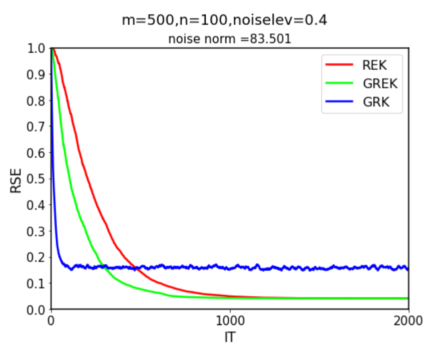
(b)



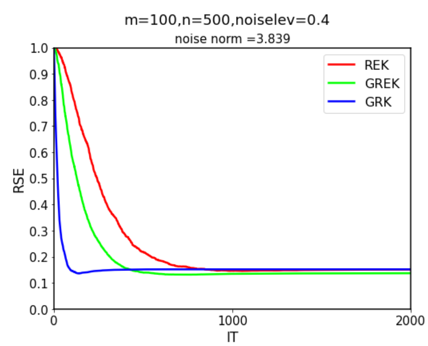
(c)



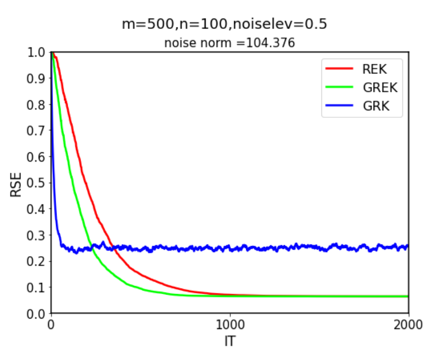
(d)



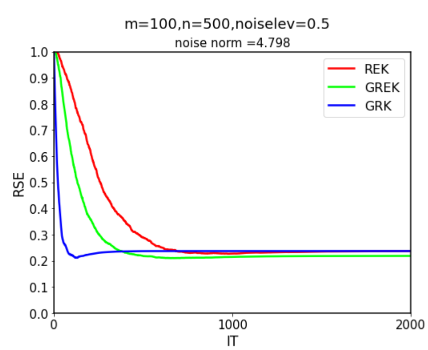
(e)



(f)



(g)



(h)

Figure 3. RSE versus IT for GREK, GRK and REK with noisy linear systems with noisy level 0.2 (Row 1), 0.3 (Row 2), 0.4 (Row 3) and 0.5 (Row 4). Left: The results for overdetermined systems. Right: The results for underdetermined systems.

5. Conclusions

In this paper, we investigate the problem of solving the noisy linear system of equations. We first prove that the estimate of the greedy randomized Kaczmarz algorithm for noisy linear system is within a fixed distance from the solution, and the distance is proportional to the norm of the noise vector. Then we propose a new greedy randomized extended Kaczmarz algorithm by introducing an effective greedy criterion for selecting the working rows and a randomized orthogonal projection for reducing the influence of the noise term. Theoretical results demonstrate that the convergence rate of the proposed greedy randomized Kaczmarz algorithm is faster than the randomized extended Kaczmarz algorithm. Numerical experiments also illustrate the superiority of the proposed algorithm.

References

[1] Z. Bai and W. Wu, *On greedy randomized Kaczmarz method for solving large sparse linear systems*, SIAM J. Sci. Comput., 2018, 40(1), 592–606.

[2] Z. Bai and W. Wu, *On convergence rate of the randomized Kaczmarz method*, Linear Algebra Appl., 2018, 553, 252–269.

[3] Z. Bai and W. Wu, *On relaxed greedy randomized Kaczmarz methods for solving large sparse linear systems*, Appl. Math. Lett., 2018, 83, 21–26.

[4] S. Kaczmarz, *Angenäherte Auflösung von systemen linearer Gleichungen*, Bull. Int. Acad. Pol. Sci. Lett. A, 1937, 35, 355–357.

[5] D. A. Lorenz, S. Wenger, F. Schöofer and M. Magnor, *A sparse Kaczmarz solver and a linearized Bregman method for online compressed sensing*, IEEE International Conference on Image Processing (ICIP), Paris, France, 2014, 1347–1351.

[6] J. Lin and D. Zhou, *Learning theory of randomized Kaczmarz method*, J. Machine Learning Res., 2015, 16(103), 3341–3365.

[7] Y. Lei and D. Zhou, *Learning theory of randomized sparse Kaczmarz method*, SIAM J. Imaging Sci., 2018, 11(1), 547–574.

[8] A. Ma, D. Needell and A. Ramdas, *Convergence properties of the randomized extended Gauss-Seidel and Kaczmarz methods*, SIAM J. Matrix Anal. Appl., 2015, 36(4), 1590–1604.

[9] D. Needell, N. Srebro and R. Ward, *Stochastic gradient descent, weighted sampling, and the randomized Kaczmarz algorithm*, Math. Program., Ser. A, 2016, 155(1), 549–573.

[10] D. Needell, *Randomized Kaczmarz solver for noisy linear systems*, BIT Numerical Mathematics, 2010, 50(2), 395–403.

[11] C. Popa and R. Zdunek, *Kaczmarz extended algorithm for tomographic image reconstruction from limited data*, Math. Comput. Simulation, 2004, 65(6), 579–598.

[12] E. Rebrova and D. Needell, *On block Gaussian sketching for the Kaczmarz method*, Numer. Algorithms, 2021, 86(1), 443–473.

[13] F. Schöofer and D. A. Lorenz, *Linear convergence of the randomized sparse Kaczmarz method*, Math. Program., Ser. A, 2019, 173(1–2), 509–536.

[14] T. Strohmer and R. Vershynin, *A randomized Kaczmarz algorithm with exponential convergence*, J. Fourier Anal. Appl., 2009, 15(2), 262–278.

[15] O. Wang, W. Li, W. Bao and X. Gao, *Nonlinear Kaczmarz algorithms and their convergence*, J. Comput. Appl. Math., 2022, 399, 223720.

[16] A. Zouzias and N. M. Freris, *Randomized extended Kaczmarz for solving least squares*, SIAM J. Matrix Anal. Appl., 2013, 34(2), 773–793.

[17] J. Zhang, *A new greedy Kaczmarz algorithm for the solution of very large linear systems*, Appl. Math. Lett., 2019, 91, 207–212.