Greedy Randomized and Maximal Weighted Residual Kaczmarz Methods with Oblique Projection

Fang Wang, Weiguo Li*, Wendi Bao, Li Liu
China University of Petroleum
Qingdao, China
*liwg@upc.edu.cn
June 2021

Abstract
For solving large-scale consistent linear system, we combine two efficient row index selection strategies with Kaczmarz-type method with oblique projection, and propose a greedy randomized Kaczmarz method with oblique projection (GRKO) and the maximal weighted residual Kaczmarz method with oblique projection (MWRKO). Through those method, the number of iteration steps and running time can be reduced to a greater extent to find the least-norm solution, especially when the rows of matrix A are close to linear correlation. Theoretical proof and numerical results show that GRKO method and MWRKO method are more effective than greedy randomized Kaczmarz method and maximal weighted residual Kaczmarz method respectively.

Key words: oblique projection, convergence property, Kaczmarz method, correlation, large linear system.

1 Introduction
Consider to solve a large-scale consistent linear system

$$Ax = b,$$

where the matrix $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$. One of the solutions of the system (1) is $x^* = A^\dagger b$, which is the least Euclidean norm solution. Especially, when the coefficient matrix $A$ is full column rank, $x^*$ is the unique solution of the system (1).

There are many researches on solving the system (1) through iterative methods, among which the Kaczmarz method is a representative and efficient row-action method. The Kaczmarz method [25] selects the rows of the matrix $A$ by using the cyclic rule, and in each iteration, the current iteration point is orthogonally projected onto the corresponding hyperplane. Due to its simplicity and performance, the Kaczmarz method has been applied to many fields, such as computerized tomography [1,2], image reconstruction [3,4,5,6], distributed computing [7], and signal processing [8,11,2]; and so on [9,10,11,12]. Since the Kaczmarz method cycles through the rows of $A$, the performance may depend heavily on the ordering of these rows. A poor ordering may result in a very slow convergence rate. McCormick [13] proposed a maximal weighted residual Kaczmarz (MWRK) method, and proved its convergence. In recent work, a new theoretical convergence estimate was proposed for the MWRK method in [39]. Strohmer and Vershynin [37] proposed a randomized Kaczmarz (RK) method which selects a given row with proportional to the Euclidean norm of the rows of the coefficient matrix $A$, and proved its convergence. After the above work, research on the Kaczmarz-type methods...
A convergence is proved theoretically and numerically. We emphasize the efficiency of our proposed methods – the GRKO method and the MWRKO method, respectively, and their greedy versions – the GRK method and the MWRK method, we propose two new Kaczmarz-type methods based on oblique projection, which breaks the limitation of orthogonal projection. Therefore, in this paper, we propose a new descent direction based on the definition of oblique projection, which can guarantee the two entries of residual error to be zero during iteration, so as to accelerate convergence. Based on the row index selection rules of two representative randomized and non-randomized Kaczmarz-type methods – the GRK method and the MWRK method, we propose two new Kaczmarz-type methods with oblique projection (KO-type) – the GRKO method and the MWRKO method respectively, and their convergence is proved theoretically and numerically. We emphasize the efficiency of our proposed methods when the rows of the matrix $A$ are nearly linearly correlated, and find that Kaczmarz-type method based on orthogonal projection performed poorly when applied to this kind of matrices.

The organization of this paper is as follows. In Section 2, we introduce the KO-type method, and give its two lemmas. In Section 3, we propose the GRKO method and MWRKO method naturally and prove the convergence of the two methods. In Section 4, some numerical examples are provided to illustrate the efficiency of our new methods. Finally, some brief concluding remarks are described in Section 5.

In this paper, $\langle \cdot, \cdot \rangle$ stands for the scalar product. $\| x \|$ is the Euclidian norm of $x \in \mathbb{R}^n$. For a given matrix $G = (g_i) \in \mathbb{R}^{m \times n}$, $g_i^T$, $G^T$, $G^1$, $R(G)$, $N(G)$, $\| G \|_F$ and $\lambda_{\text{min}}(G)$, are used to denote the ith row, the transpose, the Moore-Penrose pseudoinverse [22], the range space, the null space, the Frobenius norm, and the smallest nonzero eigenvalue of $G$ respectively. $P_C(x)$ is the orthogonal projection of $x$ onto $C$, $\bar{x}$ is any solution of the system (1); $x^* = A^\dagger b$ is the least-norm solution of the system (1).

Let $E_k$ denote the expected value conditional on the first $k$ iterations, that is,

$$E_k[\cdot] = E[\cdot | j_0, j_1, ..., j_{k-1}],$$

where $j_s(s = 0, 1, ..., k - 1)$ is the column chosen at the $s$th iteration.

## 2 Kaczmarz-type Method with Oblique Projection and its Lemmas

The sets $H_i = \{ x \in \mathbb{R}^n, g_i^T x = b_i \}$ $(i = 1, 2, \cdots, m)$ are the hyperplanes which associated to the $i$th equation of the system (1). To project the current iteration point $x^{(k)}$ to one of the hyperplanes, the oblique projection [45, 34] can be expressed as follows:

$$x^{(k+1)} = P^d_{H_i} (x^{(k)}) = x^{(k)} - \langle a_i, x \rangle - b_i \frac{d}{\langle d, a_i \rangle} d,$$  \hfill (2)

where $d \in \mathbb{R}^n$ is a given direction. In Figure 1, $x^{(k+1)}$ is obtained by oblique projection of the current iteration point $x^{(k)}$ to the hyperplane $H_{i_{k+1}}$ along the direction $d$, i.e. $x^{(k+1)} = P^d_{H_{i_{k+1}}} (x^{(k)})$. $y^{(k+1)}$ is the iteration point obtained when the direction $d = a_{i_{k+1}}$, i.e. $y^{(k+1)} = P^0_{H_{i_{k+1}}} (x^{(k)})$. When the direction...
\(d = a_i (i = \text{mod}(m, k) + 1)\), it is the classic Kaczmarz method. However, when the hyperplanes are close to linear parallel, the Kaczmarz method based on orthogonal projection has a slow iteration speed. In this paper, we propose a new iteration direction \(d = w(i_k) = a_{i_{k+1}} - \frac{\langle a_{i_k}, a_{i_{k+1}} \rangle}{\|a_{i_k}\|^2} a_{i_k}\), to make the current iteration point approach to the intersection of two hyperplanes, i.e. \(z^{(k+1)} = P_{H_{i_{k+1}}} (x^{(k)})\).

The framework of KO-type mthod is given in Section 2.1.

**Algorithm 1** Kaczmarz-type Method with Oblique Projection

**Require:** \(A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, x^{(0)} \in \mathbb{R}^n, K, \epsilon > 0\)

1. For \(i = 1 : m, M(i) = \|a_i\|^2\)
2. Choose \(i_1\) based on a certain selection rule
3. Compute \(x^{(1)} = x^{(0)} + \frac{b_{i_1} - \langle a_{i_1}, x^{(0)} \rangle}{M(i_1)} a_{i_1}\)
4. for \(k = 1, 2, \cdots, K\) do
5. Choose \(i_{k+1}\) based on a certain selection rule
6. Compute \(r^{(i_k)} = b_{i_{k+1}} - \langle a_{i_{k+1}}, x^{(k)} \rangle\) and \(D_{i_k} = \frac{M(i_k)}{\|a_{i_k}\|^2} a_{i_k}\)
7. Compute \(w^{(i_k)} = a_{i_{k+1}} - \frac{D_{i_k}}{M(i_k)} a_{i_k}\) and \(h_{i_k} (= \|w^{(i_k)}\|^2) = M(i_{k+1}) - \frac{D_{i_k}}{M(i_k)} D_{i_k}\)
8. \(a^{(k)}_{i_k} = \frac{r^{(i_k)}}{h_{i_k}}\) and \(x^{(k+1)} = x^{(k)} + a^{(k)}_{i_k} w^{(i_k)}\)
9. end for
10. Output \(x^{(K+1)}\)

We will give two lemmas of KO-type method. The selection rule of its row index \(i_{k+1}\) does not affect the lemmas.

**Lemma 1.** For the Kaczmarz-type method with oblique projection, the residual satisfies the following equations:

\[ r^{(k)}_{i_k} = 0 \quad (\forall k > 0), \quad (3) \]

\[ r^{(k)}_{i_{k-1}} = 0 \quad (\forall k > 1). \quad (4) \]
Proof. From the definition of the KO-type method, for \( k = 1 \), we have

\[
\begin{align*}
r_{i_1}^{(1)} &= b_{i_1} - \langle a_{i_1}, x^{(1)} \rangle \\
&= b_{i_1} - \langle a_{i_1}, x^{(0)} \rangle + \frac{b_{i_1} - \langle a_{i_1}, x^{(0)} \rangle}{M(i_1)} a_{i_1} \\
&= 0.
\end{align*}
\]

For \( k > 1 \), we have

\[
\begin{align*}
r_{i_k}^{(k)} &= b_{i_k} - \langle a_{i_k}, x^{(k)} \rangle \\
&= b_{i_k} - \langle a_{i_k}, x^{(k-1)} \rangle + a_{i_{k-1}}^{(k-1)} w^{(i_{k-1})} \\
&= b_{i_k} - \langle a_{i_k}, x^{(k-1)} \rangle - a_{i_{k-1}}^{(k-1)} \langle a_{i_{k-1}}, w^{(i_{k-1})} \rangle \\
&= r_{i_k}^{(k-1)} - \frac{r_{i_k}^{(k-1)}}{h_{i_{k-1}}} (a_{i_{k-1}}, w^{(i_{k-1})}) \\
&= r_{i_k}^{(k-1)} - r_{i_k}^{(k-1)} \\
&= 0.
\end{align*}
\]

The fifth equality holds due to \( \langle a_{i_k}, w^{(i_{k-1})} \rangle = \|a_{i_k}\|^2 - \frac{(a_{i_{k-1}}, a_{i_k})^2}{\|a_{i_{k-1}}\|^2} = h_{i_{k-1}} \). Thus, the equation (3) holds. Since \( k > 1 \),

\[
x^{(k)} = x^{(k-1)} + a_{i_{k-1}}^{(k)} w^{(i_{k-1})}.
\]

We get

\[
(b - Ax^{(k)})_{i_{k-1}} = (b - Ax^{(k-1)})_{i_{k-2}} - (Aa_{i_{k-1}}^{(k)} w^{(i_{k-1})})_{i_{k-1}},
\]

that is,

\[
\begin{align*}
r_{i_{k-1}}^{(k-1)} &= r_{i_{k-1}}^{(k-1)} - a_{i_{k-1}}^{(k)} (a_{i_{k-1}}, w^{(i_{k-1})}) \\
&= a_{i_{k-1}}^{(k)} (a_{i_{k-1}}, a_{i_{k}}) - \frac{(a_{i_{k-1}}, a_{i_{k}})^2}{\|a_{i_{k-1}}\|^2} a_{i_{k-1}} \\
&= 0.
\end{align*}
\]

The second equality holds due to the equation (3). Thus, the equation (4) holds. \( \square \)

**Lemma 2.** The iteration sequence \( \{x^{(k)}\}_{k=0}^{\infty} \) generated by the Kaczmarz-type method with oblique projection, satisfies the following equations:

\[
\|x^{(k+1)} - \tilde{x}\|^2 = \|x^{(k)} - \tilde{x}\|^2 - \|x^{(k+1)} - x^{(k)}\|^2 \quad (\forall k \geq 0),
\]

where \( \tilde{x} \) is an arbitrary solution of the system (1). Especially, when \( P_{N(A)}(x^{(0)}) = P_{N(A)}(\tilde{x}) \), \( x^{(k)} \) is \( \tilde{x} \in R(A^T) \).

**Proof.** For \( k = 0 \), we have

\[
\begin{align*}
\langle a_{i_1}, x^{(1)} - \tilde{x} \rangle &= \langle a_{i_1}, x^{(0)} - \tilde{x} + \frac{b_{i_1} - \langle a_{i_1}, x^{(0)} \rangle}{M(i_1)} a_{i_1} \rangle \\
&= \langle a_{i_1}, x^{(0)} \rangle - b_{i_1} + \frac{b_{i_1} - \langle a_{i_1}, x^{(0)} \rangle}{M(i_1)} a_{i_1} \\
&= 0,
\end{align*}
\]
which shows that $x^{(1)} - \tilde{x}$ is orthogonal to $a_i$. Therefore, we know

$$(x^{(1)} - x^{(0)})^T(x^{(1)} - \tilde{x}) = 0.$$  

It follows that

$$||x^{(1)} - \tilde{x}||^2 = ||x^{(0)} - \tilde{x}||^2 - ||x^{(1)} - x^{(0)}||^2. \quad (6)$$

For $k > 0$, we have

$$\langle w^{(i_k)}, x^{(k+1)} - \tilde{x} \rangle = \langle w^{(i_k)}, x^{(k)} - \tilde{x} \rangle + \alpha_{i_k} M^{(i_k)} a_{i_k},$$

$$= \langle a_{i_k}, x^{(k)} - \tilde{x} \rangle + \alpha_{i_k} r^{(k)}_{i_k} + r^{(k)}_{i_k} a_{i_k},$$

$$= 0.$$  

The third and last equalities hold due to $h_{i_k} = \|w^{(i_k)}\|^2$, and the equation (3) respectively. Thus we get that $x^{(k+1)} - \tilde{x}$ is orthogonal to $w^{(i_k)}$. Therefore, we get that

$$(x^{(k+1)} - x^{(k)})^T(x^{(k+1)} - \tilde{x}) = 0.$$  

It follows that

$$||x^{(k+1)} - \tilde{x}||^2 = ||x^{(k)} - \tilde{x}||^2 - ||x^{(k+1)} - x^{(k)}||^2 \quad (\forall k > 0). \quad (7)$$

Thus, from the above proof, the equation (5) holds. According to the iterative formula

$$\begin{cases}
  x^{(1)} = x^{(0)} + b_{i_1} - \langle a_{i_1}, x^{(0)} \rangle M^{(i_1)} a_{i_1}, \\
  x^{(k+1)} = x^{(k)} + \alpha_{i_k} M^{(i_k)} a_{i_k} \quad (\forall k > 0),
\end{cases}$$

we can get $P_{N(A)}(x^{(k)}) = P_{N(A)}(x^{(k-1)}) = \ldots = P_{N(A)}(x^{(0)})$, and by the fact that $P_{N(A)}(x^{(0)}) = P_{N(A)}(\tilde{x})$, we can deduce that $x^{(k)} - \tilde{x} \in R(A^T)$.  

\[ \square \]

3 Greedy Randomized and Maximal Weighted Residual Kaczmarz methods with Oblique Projection

In this section, we combine the oblique projection with the GRK method [18] and the MWRK method [13] to obtain the GRKO method and the MWRKO method, and prove their convergence. Theoretical results show that the KO-type method can accelerate the convergence when there are suitable row index selection strategies.
3.1 Greedy Randomized Kaczmarz Method with Oblique projection

The core of the GRK method [18] is a new probability criterion, which can grasp the large items of the residual vector in each iteration, and randomly select the item with probability in proportion to the retained residual norm. Theories and experiments prove that it can speed up convergence speed. This paper uses its the row index selection rule in combination with the KO method to obtain the GRKO method, and the algorithm is as follows:

**Algorithm 2 Greedy Randomized Kaczmarz Method with Oblique Projection**

Require: \( A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, x(0) \in \mathbb{R}^n, \varepsilon > 0 \)
1: For \( i = 1 : m \), \( M(i) = \|a_i\|^2 \)
2: Randomly select \( i_0 \) and compute \( x(1) = x(0) + \frac{b_{i_0} - \langle a_{i_0}, x(0) \rangle}{M(i_0)} a_{i_0} \)
3: for \( k = 1, 2, \ldots, K - 1 \) do
4: Compute \( \varepsilon_k = \frac{1}{2} \left( \frac{1}{\|b - Ax(k)\|^2} \max_{1 \leq i \leq m} \left\{ \frac{\|b_{i_0} - \langle a_{i_0}, x(k) \rangle\|^2}{\|a_{i_0}\|^2} + \frac{1}{\|A_i\|^2} \right\} \right) \)
5: Determine the index set of positive integers \( \mathcal{I}_k = \{ i_{k+1} | b_{i_{k+1}} - \langle a_{i_{k+1}}, x(k) \rangle |^2 \geq \varepsilon_k \|b - Ax(k)\|^2 \|a_{i_{k+1}}\|^2 \} \)
6: Compute the th entry \( \hat{r}^{(k)}_i \) of the vector \( \hat{r}^{(k)} \) according to
   \[ \hat{r}^{(k)}_i = \begin{cases} b_i - \langle a_i, x^{(k)} \rangle, & \text{if } i \in \mathcal{I}_k \\ 0 & \text{otherwise} \end{cases} \]
7: Select \( i_{k+1} \in \mathcal{I}_k \) with probability \( Pr(\text{row } = i_{k+1}) = \frac{\| \hat{r}^{(k)}_i \|^2}{\| \hat{r}^{(k)} \|^2} \)
8: Compute \( D_i = \langle a_{i_{k+1}}, a_{i_{k+1}} \rangle \)
9: Compute \( w^{(i_{k+1})} = a_{i_{k+1}} - \frac{D_i}{M(i_{k+1})} D_i \) and
   \[ h_i(=\|w^{(i_{k+1})}\|^2) = M(i_{k+1}) - \frac{D_i}{M(i_{k+1})} D_i \]
10: \( \alpha^{(k)}_i = \frac{\| \hat{r}^{(k)}_i \|^2}{h_i} \left( = \frac{\| \hat{r}^{(k)}_i \|^2}{h_i} \right) \) and \( x^{(k+1)} = x^{(k)} + \alpha^{(k)}_i w^{(i_{k+1})} \)
11: end for
12: Output \( x^{(K)} \)

The convergence of the GRKO method is provided as follows.

**Theorem 1.** Consider the consistent linear system \( \{Ax = b\} \), where the coefficient matrix \( A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m \). Let \( x^{(0)} \in \mathbb{R}^n \) be an arbitrary initial approximation, \( x \) is a solution of system \( \{Ax = b\} \) such that \( P_{N(A)}(\tilde{x}) = P_{N(A)}(x^{(0)}) \). Then the iteration sequence \( \{x^{(k)}\}_{k=1}^\infty \) generated by the GRKO method obeys

\[
E\|x^{(k)} - \tilde{x}\|^2 \leq \prod_{s=0}^{k-1} \zeta_s \|x(0) - \tilde{x}\|^2.
\]

where \( \zeta_0 = 1 - \frac{\lambda_{\infty}(A^T A)}{m\|A\|^2}, \zeta_1 = 1 - \frac{1}{2} \left( 1 - \frac{1}{\gamma_1} \|A\|^2 + 1 \right) \frac{\lambda_{\infty}(A^T A)}{\Delta \|A\|^2}, \zeta_k = 1 - \frac{1}{2} \left( 1 - \frac{1}{\gamma_k} \|A\|^2 + 1 \right) \frac{\lambda_{\infty}(A^T A)}{\Delta \|A\|^2} \) \((\forall k > 1)\),

which

\[
\gamma_1 = \max_{1 \leq i \leq m} \sum_{s=1}^{m} \|a_i\|^2,
\]

and

\(
\left( \frac{\lambda_{\infty}(A^T A)}{\Delta \|A\|^2} \right) = \gamma_1 \).
\[ \gamma_2 = \max_{1 \leq i, j \leq m} \max_{i \neq j} \frac{|a_i|^2}{|a_j|^2}, \quad (10) \]

\[ \Delta = \max_{j \neq k} \sin(a_j, a_k)^2 \in (0, 1]. \quad (11) \]

In addition, if \( x^{(0)} \in R(A^T) \), the sequence \( \{x^{(k)}\}_{k=1}^\infty \) converges to the least-norm solution of the system \( \mathbb{I} \), i.e. \( \lim_{k \to \infty} x^{(k)} = x^* = A^T b \).

**Proof.** When \( k = 1 \), we can get

\[ \epsilon_1 ||A||_F^2 = \frac{\max_{1 \leq i, j \leq m} \left\{ \frac{|b_j - (a_j, x^{(1)})|^2}{|a_j|^2} \right\}}{2} + \frac{1}{2} \sum_{i=1}^m \left\{ \frac{|a_i|^2}{|A||a_j|^2} \right\} \]

\[ = \frac{\max_{1 \leq i, j \leq m} \left\{ \frac{|b_j - (a_j, x^{(1)})|^2}{|a_j|^2} \right\}}{2} + \frac{1}{2} \sum_{i=1}^m \left\{ \frac{|a_i|^2}{|A||a_j|^2} \right\} \]

\[ \geq \frac{1}{2} \left( \frac{1}{\gamma_1} ||A||_F^2 + 1 \right). \]

The second equality holds due to the equation (3).

When \( k > 1 \), we get

\[ \epsilon_k ||A||_F^2 = \frac{\max_{1 \leq i, j \leq m} \left\{ \frac{|b_{j, k+1} - (a_{j, k+1}, x^{(k)})|^2}{|a_{j, k+1}|^2} \right\}}{2} + \frac{1}{2} \sum_{i=1}^m \left\{ \frac{|a_{i, k+1}|^2}{|A||a_{j, k+1}|^2} \right\} \]

\[ = \frac{\max_{1 \leq i, j \leq m} \left\{ \frac{|b_{j, k+1} - (a_{j, k+1}, x^{(k)})|^2}{|a_{j, k+1}|^2} \right\}}{2} + \frac{1}{2} \sum_{i=1}^m \left\{ \frac{|a_{i, k+1}|^2}{|A||a_{j, k+1}|^2} \right\} \]

\[ \geq \frac{1}{2} \left( \frac{1}{\gamma_2} ||A||_F^2 + 1 \right). \]
The second equality holds due to the equation (5) and the equation (4).

Under the GRKO method, Lemma 2.2 still holds, so we can take the full expectation on both sides of the equation (5), and get that for $k = 0$,

$$E[\|x^{(1)} - \bar{x}\|^2] = \|x^{(0)} - \bar{x}\|^2 - E[\|x^{(1)} - x^{(0)}\|^2]$$

$$= \|x^{(0)} - \bar{x}\|^2 - \frac{1}{m} \sum_{i_{1}, i_{2} \in \Psi_{k}} \frac{|b_{i_{2}} - \langle a_{i_{1}}, x^{(0)} \rangle|}{M(i_{1})} a_{i_{1}} ||^2$$

$$\leq \|x^{(0)} - \bar{x}\|^2 - \frac{1}{m} \|b - Ax^{(0)}\|^2$$

$$\leq \left(1 - \frac{\lambda_{\text{min}}(A^{T}A)}{m||A||_{F}^{2}}\right) \|x^{(0)} - \bar{x}\|^2$$

and for $k > 0$,

$$E_{k}[\|x^{(k+1)} - \bar{x}\|^2] = \|x^{(k)} - \bar{x}\|^2 - E_{k}[\|x^{(k+1)} - x^{(k)}\|^2]$$

$$= \|x^{(k)} - \bar{x}\|^2 - \frac{1}{m} \sum_{i_{1}, i_{2} \in \Psi_{k}} \frac{|b_{i_{2}} - \langle a_{i_{1}}, x^{(k)} \rangle|}{M(i_{1})} a_{i_{1}} ||^2$$

$$\leq \|x^{(k)} - \bar{x}\|^2 - \frac{1}{m} \|b - Ax^{(k)}\|^2$$

$$\leq \|x^{(k)} - \bar{x}\|^2 - \frac{\epsilon_{k}}{\Delta} \|A(\bar{x} - x^{(k)})\|^2$$

$$\leq (1 - \frac{\epsilon_{k} \lambda_{\text{min}}(A^{T}A)}{\Delta}) \|x^{(k)} - \bar{x}\|^2.$$

The first inequality of the equation (14) is achieved with the use of the fact that $|b_{i_{2}}| + |b_{i_{1}}| \geq |b_{i_{2}} - |b_{i_{1}}|$, (if $|a_{1}| > 0, |a_{2}| > 0$), and the first inequality of the equation (15) is achieved with the use of the fact that $||w_{i}||^2 = ||a_{i_{1}}||^2 - \frac{(a_{i_{1}}a_{i_{2}})^{2}}{||a_{i_{2}}||^2} = \sin(a_{i_{1}}, a_{i_{2}})^2 ||a_{i_{2}}||^2 \leq \Delta \cdot ||a_{i_{1}}||^2$, and the second inequality of the equation (15) is achieved with the use of the definition of $\Psi_{k}$ which lead to

$$|b_{i_{2}} - \langle a_{i_{1}}, x^{(k)} \rangle|^2 \geq \epsilon_{k} ||b - Ax^{(k)}||^2 ||a_{i_{1}}||^2, \forall i_{k+1} \in \Psi_{k}.$$

Here in the last inequalities of the equation (14) and (15), we have used the estimate $||Au||^2 \geq \lambda_{\text{min}}(A^{T}A)||u||^2$, which holds true for any $u \in C^{k}$ belonging to the column space of $A^{T}$. According to the lemma 2.2, it holds.

By making use of the equation (12), (13) and (15), we get

$$E_{1}[\|x^{(2)} - \bar{x}\|^2] \leq \left[1 - \frac{1}{2} \left(\frac{1}{\gamma_{1}} + \frac{\lambda_{\text{min}}(A^{T}A)}{\Delta \cdot ||A||_{F}^{2}}\right)\right] \|x^{(1)} - \bar{x}\|^2$$

$$= \zeta_{1} \|x^{(1)} - \bar{x}\|^2,$
\[ E_k \|x^{(k+1)} - \tilde{x}\|^2 \leq \left[ 1 - \frac{1}{2\gamma_2} \|A\|^2 + 1 \right] \frac{\lambda_{\min}(A^T A)}{\Delta \cdot |A|} \|x^{(k)} - \tilde{x}\|^2 \]

= \zeta_k \|x^{(k)} - \tilde{x}\|^2 \quad (\forall k > 1).

Finally, by recursion and taking the full expectation, the equation (8) holds.

**Remark 1.** In the GRKO method, \( h_{i_k} \) is not zero. Suppose \( h_{i_k} = 0 \), which means \( \exists \lambda > 0 \), \( \lambda a_{i_k} = a_{i_{k+1}} \). Due to the system is consistent, it holds \( \langle a_{i_{k+1}}, x^* \rangle = \lambda \langle a_{i_k}, x^* \rangle = \lambda b_{i_k} = b_{i_{k+1}} \). According to the equation (8), it holds \( i_k^{(k)} = i_{k+1}^{(k)} = 0 \). From step 5 of Algorithm 3.1, we can know that such index \( i_{k+1} \) will not be selected.

**Remark 2.** Set \( \tilde{\zeta}_k = 1 - \frac{1}{2} \left( \frac{\gamma_2}{\|A\|^2} + 1 \right) \left( \frac{\lambda_{\min}(A^T A)}{\|A\|^2} \right) \) \((\forall k > 0)\), and the convergence of GRK method in [32] meets:

\[ E_k \|x^{(k+1)} - x^*\|^2 \leq \tilde{\zeta}_k \|x^{(k)} - x^*\|^2. \]

Obviously, \( \zeta_1 \leq \tilde{\zeta}_1, \tilde{\zeta}_k < \zeta_k \) \((\forall k > 1)\) is satisfied, so the convergence speed of GRKO method is faster than GRK method.

### 3.2 Maximal Weighted Residual Kaczmarz Method with Oblique Projection

The selection strategy for the index \( i_k \) used in the maximal weighted residual Kaczmarz (MWRK) method [13] is: Set

\[ i_k = \arg \max_{i \in \{1, 2, \ldots, m\}} \left| a_i^T x^{(k)} - b_i \right| / \|a_i\|. \]

McCormick proved the exponential convergence of the MWRKO method. In [32], a new convergence conclusion of the MWRKO method is given. We use its row index selection rule combined with KO-type method to obtain MWRKO method, and the algorithm is as follows:

**Algorithm 3** Maximal Weighted Residual Kaczmarz Method with Oblique Projection (MWRKO)

**Require:** \( A \in R^{m \times n}, b \in R^m, x^{(0)} \in R^n, K, \varepsilon > 0 \)

1. For \( i = 1 : m \), \( M(i) = \|a_i\|^2 \)
2. Compute \( i_1 = \arg \max_{i \in \{1, 2, \ldots, m\}} \left| a_i^T x^{(0)} - b_i \right| / \|a_i\| \), and \( x^{(1)} = x^{(0)} + \frac{b_i - \langle a_i, x^{(0)} \rangle}{M(i)} a_i \)
3. for \( k = 1, 2, \ldots, K \) do
4. Compute \( i_{k+1} = \arg \max_{i \in \{1, 2, \ldots, m\}} \left| a_i^T x^{(k)} - b_i \right| / \|a_i\| \)
5. Compute \( D_i = \langle a_{i_k}, a_{i_{k+1}} \rangle \) and \( r_{i_{k+1}} = b_{i_{k+1}} - \langle a_{i_{k+1}}, x^{(k)} \rangle \)
6. Compute \( w^{(i)} = a_{i_k} - \frac{D_i}{M(i)} a_i \) and \( h_i = \|w^{(i)}\|^2 = M(i_{k+1}) - \frac{D_i}{M(i)} D_i \)
7. \( a_{i_{k+1}} = \frac{h_{i_{k+1}}}{h_i} a_i \) and \( x^{(k+1)} = x^{(k)} + a_{i_{k+1}} w^{(i)} \)
8. end for
9. Output \( x^{(K+1)} \)

The convergence of the MWRKO method is provided as follows.

**Theorem 2.** Consider the consistent linear system (1), where the coefficient matrix \( A \in R^{m \times n}, b \in R^m \). Let \( x^{(0)} \in R^n \) be an arbitrary initial approximation, \( x \) is a solution of system (1) such that \( P_{h(A)}(\tilde{x}) = P_{N(A)}(x^{(0)}) \). Then the iteration sequence \( \{x^{(k)}\}_{k=1}^{\infty} \) generated by the MWRKO method obeys

\[ \|x^{(k)} - \tilde{x}\|^2 \leq \prod_{s=0}^{k-1} \rho_s \|x^{(0)} - \tilde{x}\|^2, \quad (16) \]
where \( \rho_0 = 1 - \frac{\lambda_{\text{min}}(A^T A)}{\Delta \gamma_1} \), \( \rho_1 = 1 - \frac{\lambda_{\text{min}}(A^T A)}{\Delta \gamma_2} \), \( \rho_k = 1 - \frac{\lambda_{\text{min}}(A^T A)}{\Delta \gamma_2} \) (\( \forall k > 1 \)), which \( \gamma_1, \gamma_2 \) and \( \Delta \) are defined by equations (9), (10) and (11) respectively.

In addition, if \( x^{(k)} \in \mathbb{R}(A^T) \), the sequence \( \{x^{(k)}\}_{k=1}^{\infty} \) converges to the least-norm solution of the system (1), i.e. \( \lim_{k \to \infty} x^{(k)} = x^- = A^T b \).

**Proof.** Under the MWRKO method, Lemma 2.2 still holds. For \( k = 1 \), we have

\[
\|x^{(1)} - \bar{x}\|^2 = \|x^{(0)} - \bar{x}\|^2 - \|x^{(1)} - x^{(0)}\|^2
\]
\[
= \|x^{(0)} - \bar{x}\|^2 - \left| \frac{b_{i_1} - \langle a_{i_1}, x^{(0)} \rangle}{M(i_1)} \right|^2
\]
\[
= \|x^{(0)} - \bar{x}\|^2 - \frac{\|A\|^2_F}{M(i_1)} \sum_{i=1}^{m} \frac{|b_{i} - \langle a_{i}, x^{(0)} \rangle|^2}{M(i)} \cdot M(i)
\]
\[
\leq \|x^{(0)} - \bar{x}\|^2 - \frac{\|A(x - x^{(0)})\|^2}{\|A\|^2_F}
\]
\[
\leq \|x^{(0)} - \bar{x}\|^2 - \frac{\lambda_{\text{min}}(A^T A)}{\|A\|^2_F} \|x^{(0)} - \bar{x}\|^2
\]
\[
= \left(1 - \frac{\lambda_{\text{min}}(A^T A)}{\|A\|^2_F}\right) \|x^{(0)} - \bar{x}\|^2
\]
\[
\leq \rho_0 \|x^{(0)} - \bar{x}\|^2.
\]

For \( k = 1 \), we have

\[
\|x^{(2)} - \bar{x}\|^2 = \|x^{(1)} - \bar{x}\|^2 - \|x^{(2)} - x^{(1)}\|^2
\]
\[
= \|x^{(1)} - \bar{x}\|^2 - \left| \frac{b_{i_2} - \langle a_{i_2}, x^{(1)} \rangle}{M(i_2)} \right|^2
\]
\[
\leq \|x^{(1)} - \bar{x}\|^2 - \frac{\|A\|^2_F}{\Delta \cdot M(i_2)} \cdot \sum_{i=1;j \neq i}^{m} \frac{|b_{i} - \langle a_{i}, x^{(1)} \rangle|^2}{M(i)} \cdot M(i)
\]
\[
\leq \|x^{(1)} - \bar{x}\|^2 - \frac{\|A(x - x^{(1)})\|^2}{\Delta \cdot \gamma_1}
\]
\[
\leq \|x^{(1)} - \bar{x}\|^2 - \frac{\lambda_{\text{min}}(A^T A)}{\Delta \cdot \gamma_1} \|x^{(1)} - \bar{x}\|^2
\]
\[
= \left(1 - \frac{\lambda_{\text{min}}(A^T A)}{\Delta \cdot \gamma_1}\right) \|x^{(1)} - \bar{x}\|^2
\]
\[
= \rho_1 \|x^{(1)} - \bar{x}\|^2.
\]
For $k > 1$, we have
\[
\|x^{(k+1)} - \tilde{x}\|^2 = \|x^{(k)} - \tilde{x}\|^2 - \|x^{(k+1)} - x^{(k)}\|^2 \\
= \|x^{(k)} - \tilde{x}\|^2 - \frac{|b_{i_{k+1}} - (a_{i_{k+1}} x^{(k)})|^2}{\|w^{(k)}\|^2} \\
\leq \|x^{(k)} - \tilde{x}\|^2 - \frac{|b_{i_{k+1}} - (a_{i_{k+1}} x^{(k)})|^2}{\Delta \cdot M(i_{k+1})} \\
\leq \|x^{(k)} - \tilde{x}\|^2 - \frac{\lambda_{\min}(A^T A)\|x^{(k)} - \tilde{x}\|^2}{\Delta \cdot \gamma_2} \\
= \left(1 - \frac{\lambda_{\min}(A^T A)}{\Delta \cdot \gamma_2}\right)\|x^{(k)} - \tilde{x}\|^2 \\
= \rho_k\|x^{(k)} - \tilde{x}\|^2.
\]

(19)

Here in the last inequalities of the equation (17), (18) and (19), we have used the estimate
\[
\|Au\|^2 \geq \lambda_{\min}(A^T A)|u|^2,
\]
which holds true for any $u \in C^n$ belonging to the column space of $A^T$. According to the lemma 2.2, it holds. From the equation (17), (18) and (19), the equation (16) holds.

Remark 3. When multiple indicators $i_{k+1}$ are met in Step 2 of Algorithm 3.2 in the iterative process, we randomly select any one of them.

Remark 4. In the MWRKO method, the reason of $h_{i_k} \neq 0$ is similar to Remark 1.

Remark 5. Set $\tilde{\rho}_0 = 1 - \frac{\lambda_{\min}(A^T A)}{\|A\|^2} - \frac{\lambda_{\min}(A^T A)}{\gamma_1}$, $(\forall k > 0)$, and the convergence of MWRK method in [39] meets:
\[
\|x^{(k)} - x^*\|^2 \leq \sum_{i=0}^{k-1} \tilde{\rho}_i\|x^{(0)} - x^*\|^2.
\]

Obviously, $\rho_k < \tilde{\rho}_k$ $(\forall k > 1)$, $\rho_1 \leq \tilde{\rho}_1$ and $\rho_0 = \tilde{\rho}_0$, so the convergence speed of MWRKO method is faster than MWRK method. Note that $\tilde{\rho}_k < \tilde{\zeta}_k$, $\rho_k < \zeta_k$ $(\forall k > 0)$, that is $V_{MWRK} < V_{MWRKO}$, $V_{GRK} < V_{GRKO}$, $V_{GRK} < V_{MWRK}$, $V_{GRKO} < V_{MWRKO}$, where $V$ represents the convergence speed.

4 Numerical Experiments

In this section, some numerical examples are provided to illustrate the effectiveness of the greedy randomized Kaczmarz (GRK) method, the greedy randomized Kaczmarz method with oblique projection (GRKO), the maximal weighted residual Kaczmarz method (MWRK), and the maximal weighted residual Kaczmarz method (MWRKO). All experiments are carried out using MATLAB (version R2019b) on a personal computer with 1.60 GHz central processing unit (Intel(R) Core(TM)i5-10210U CPU), 8.00 GB memory, and Windows operating system (64 bit Windows 10).

In our implementations, the right vector $b = Ax^*$ such that the exact solution $x^* \in R^n$ is a vector generated by the rand function. Define the relative residual error (RRE) at the $k$th iteration as follows:
\[
RRE = \frac{\|b - Ax^{(k)}\|^2}{\|b\|^2}.
\]

11
The initial point \( x^{(0)} \in \mathbb{R}^n \) is set to be a zero vector, and the iterations are terminated once the relative solution error satisfies \( \text{RRE} < \omega \) or the number of iteration steps exceeds 100,000. If the number of iteration steps exceeds 100,000, it is denoted as "-".

We will compare the numerical performance of these methods in terms of the number of iteration steps (denoted as 'IT') and the computing time in seconds (denoted as 'CPU'). Here the CPU and IT mean the arithmetical averages of the elapsed running times and the required iteration steps with respect to 50 trials repeated runs of the corresponding method.

4.1 Experiments for Random Matrix Collection in \([0, 1]\)

The random matrix collection in \([0, 1]\) is randomly generated by using the MATLAB function \( \text{rand} \), and the numerical results are reported in Tables 1-2 and Figures 2-3. In this subsection, we let \( \omega = 0.5 \times 10^{-8} \). According to the characteristics of the matrix generated by MATLAB function \( \text{rand} \), Table 1 and Table 2 are the experiments for the overdetermined consistent linear systems, underdetermined consistent linear systems respectively. Under the premise of convergence, all methods can find the unique least Euclidean norm solution \( x^* \).

From Table 1 and Figure 2, we can see that when the linear system is overdetermined, with the increase of \( m \), the IT of all methods decreases, but the CPU shows an increasing trend. Our new methods – the GRKO method and the MWRKO method, perform better than the GRK method and the MWRK method respectively in both iteration steps and running time. Among the four methods, the MWRKO method performs best. From Table 2 and Figure 3, we can see that in the case of underdetermined linear system, with the increase of \( m \), the IT and CPU of all methods decrease.

In this group of experiments, whether it is an overdetermined or underdetermined linear system, whether in terms of the IT or CPU, the GRKO method and the MWRKO method perform very well compared with the GRK method and the MWRK method. These experimental phenomena are consistent with the theoretical convergence conclusions we got.

| \( m \) | \( 1000 \) | \( 2000 \) | \( 3000 \) | \( 4000 \) | \( 5000 \) | \( 6000 \) | \( 7000 \) | \( 8000 \) | \( 9000 \) |
|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| IT    | 12072  | 4726   | 3362   | 2663   | 2398   | 2100   | 1970   | 1861   | 1750   |
| CPU   | 1.2824 | 1.4792 | 1.7550 | 1.9415 | 2.4134 | 2.6235 | 2.6019 | 3.1035 | 3.0223 |

Table 1: IT and CPU of GRK, GRKO, MWRK and MWRKO for \( m \times n \) matrices \( A \) with \( n = 500 \) and different \( m \) when the consistent linear system is overdetermined.
Figure 2: Matrix $A \in \mathbb{R}^{m \times 500}$ is generated by the $\text{rand}$ function in the interval $[0, 1]$. (a): IT of the GRK, GRKO, MWRK, MWRKO methods changes with $m$. (b): CPU of the GRK, GRKO, MWRK, MWRKO methods changes with $m$.

Table 2: IT and CPU of GRK, GRKO, MWRK and MWRKO for $m \times n$ matrices $A$ with $n = 2000$ and different $m$ when the consistent linear system is underdetermined.

| m   | IT    | CPU    |
|-----|-------|--------|
|     | GRK   | GRKO   | MWRK   | MWRKO  | GRK   | GRKO   | MWRK   | MWRKO  |
| 100 | 802   | 286    | 848    | 272    | 0.0496| 0.0223 | 0.0258 | 0.0165 |
| 200 | 1968  | 523    | 1948   | 481    | 0.1648| 0.0496 | 0.0831 | 0.0276 |
| 300 | 3104  | 759    | 3148   | 709    | 0.3982| 0.1090 | 0.2404 | 0.0664 |
| 400 | 4586  | 1002   | 4612   | 930    | 1.0539| 0.2594 | 0.8433 | 0.1920 |
| 500 | 6233  | 1250   | 6336   | 1215   | 1.9528| 0.4409 | 1.6836 | 0.3576 |
| 600 | 8671  | 1576   | 8882   | 1497   | 3.6363| 0.7493 | 3.1625 | 0.5957 |
| 700 | 11895 | 2063   | 11575  | 1879   | 5.8642| 1.1078 | 5.0029 | 0.9087 |
| 800 | 14758 | 2451   | 14888  | 2394   | 8.4280| 1.5350 | 7.7007 | 1.6405 |
| 900 | 18223 | 3250   | 18608  | 2945   | 12.0469| 2.2765 | 10.9511| 1.8884 |
Figure 3: Matrix $A \in \mathbb{R}^{m \times 2000}$ is generated by the $\text{rand}$ function in the interval $[0, 1]$. (a): IT of the GRK, GRKO, MWRK, MWRKO methods changes with $m$. (b): CPU of the GRK, GRKO, MWRK, MWRKO methods changes with $m$.

### 4.2 Experiments for Random Matrix Collection in $[c, 1]$

In this subsection, the entries of our coefficient matrix are randomly generated in the interval $[c, 1]$. This set of experiments was also done in [30] and [46], and pointed out that when the value of $c$ is close to 1, the rows of matrix $A$ is closer to linear correlation. Theorem 3.1 and theorem 3.2 have shown the effectiveness of the GRKO method and the MWRKO method in this case. In order to verify this phenomenon, we construct several $1000 \times 500$ and $500 \times 1000$ matrices $A$, which entries is independent identically distributed uniform random variables on some interval $[c, 1]$. Note that there is nothing special about this interval, and other intervals yield the same results when the interval length remains the same. In the experiment of this subsection, we take $\omega = 0.5 \times 10^{-8}$.

From Table 3 and Figure 4, it can be seen that when the linear system is overdetermined, with $c$ getting closer to 1, the GRK method and the MWRK method have a significant increase in the number of iterations and running time. When $c$ increases to 0.7, the GRK method and the MWRK method exceeds the maximum number of iterations. But the IT and CPU of the GRKO method and the MWRKO method have decreasing trends. From Table 4 and Figure 5, we can get that the numerical experiment of the coefficient matrix $A$ in the underdetermined case has similar laws to the numerical experiment in the overdetermined case.

In this group of experiments, it can be observed that when the rows of the matrix are close to linear correlation, the GRKO method and the MWRKO method can find the least Euclidean norm solution more quickly than the GRK method and the MWRK method.
Table 3: IT and CPU of GRK, GRKO, MWRK and MWRKO for matrices $A \in \mathbb{R}^{1000 \times 500}$ generated by the rand function in the interval $[c, 1]$

| c   | IT  |          |          | CPU  |          |          |
|-----|-----|----------|----------|------|----------|----------|
|     | GRK | GRKO     | MWRK     | MWRKO| GRK      | GRKO     | MWRK     | MWRKO |
| 0.1 | 14757 | 2036 | 14594 | 1830  | 1.5811 | 0.2180 | 0.9419 | 0.0969 |
| 0.2 | 21103 | 1840 | 20717 | 1714  | 2.1684 | 0.2287 | 1.1828 | 0.1003 |
| 0.3 | 27375 | 1708 | 26986 | 1569  | 3.5926 | 0.1789 | 1.5865 | 0.1195 |
| 0.4 | 36293 | 1708 | 35595 | 1394  | 3.6751 | 0.1802 | 2.0682 | 0.0885 |
| 0.5 | 53485 | 1428 | 52853 | 1310  | 5.3642 | 0.1486 | 3.0024 | 0.0847 |
| 0.6 | 84204 | 1353 | 81647 | 1185  | 9.0879 | 0.1388 | 4.5468 | 0.0767 |
| 0.7 | -    | 1227 | -    | 1036  | -      | 0.1298 | -      | 0.0564 |
| 0.8 | -    | 1080 | -    | 926   | -      | 0.1107 | -      | 0.0580 |
| 0.9 | -    | 715  | -    | 583   | -      | 0.0707 | -      | 0.0324 |

Figure 4: Matrix $A \in \mathbb{R}^{1000 \times 500}$ is generated by the rand function in the interval $[c, 1]$. (a): IT of the GRK, GRKO, MWRK, MWRKO methods changes with $c$. (b): CPU of the GRK, GRKO, MWRK, MWRKO methods changes with $c$.

Table 4: IT and CPU of GRK, GRKO, MWRK and MWRKO for matrices $A \in \mathbb{R}^{500 \times 1000}$ generated by the rand function in the interval $[c, 1]$

| c   | IT  |          |          | CPU  |          |          |
|-----|-----|----------|----------|------|----------|----------|
|     | GRK | GRKO     | MWRK     | MWRKO| GRK      | GRKO     | MWRK     | MWRKO |
| 0.1 | 16828 | 1968 | 16913 | 1795  | 1.7612 | 0.2103 | 0.9333 | 0.1083 |
| 0.2 | 23518 | 2003 | 23234 | 1857  | 2.3037 | 0.2066 | 1.3119 | 0.1230 |
| 0.3 | 30875 | 1661 | 31017 | 1688  | 2.9310 | 0.1635 | 1.7373 | 0.0997 |
| 0.4 | 41242 | 1511 | 40986 | 1515  | 4.3004 | 0.1726 | 2.2899 | 0.1025 |
| 0.5 | 60000 | 1399 | 59750 | 1349  | 5.4754 | 0.1252 | 2.8920 | 0.0727 |
| 0.6 | 97045 | 1270 | 95969 | 1264  | 8.5229 | 0.1173 | 4.8380 | 0.0688 |
| 0.7 | -    | 1082 | -    | 1022  | -      | 0.1168 | -      | 0.0646 |
| 0.8 | -    | 858  | -    | 863   | -      | 0.0960 | -      | 0.0585 |
| 0.9 | -    | 549  | -    | 598   | -      | 0.0582 | -      | 0.0353 |
Figure 5: Matrix $A \in \mathbb{R}^{500 \times 1000}$ is generated by the rand function in the interval $[c, 1]$. (a): IT of the GRK, GRKO, MWRK, MWRKO methods changes with $c$. (b): CPU of the GRK, GRKO, MWRK, MWRKO methods changes with $c$.

### 4.3 Experiments for Sparse Matrix

In this subsection, we will give three examples to illustrate the effectiveness of our new methods applied to sparse matrix. The coefficient matrices $A$ of these three examples are the practical problems from [44] and the two test problems from [43]. We uniformly take $\omega = 0.5 \times 10^{-5}$ in these three numerical examples.

**Example 1.** We solve the problem (1) with the coefficient matrix $A \in \mathbb{R}^{m \times n}$ chosen from the University of Florida sparse matrix collection [44]. the matrices are divorce, photogrammetry, Ragusa18, Trec8, Stranke94, and well1033. In Table 5, we list some properties of these matrices, where density is defined as follows:

$$\text{density} = \frac{\text{number of nonzeros of } m \times n \text{ matrix}}{mn}.$$

In order to solve Example 4.1, we list the IT, CPU and historical convergence of the GRK, GRKO, MWRK, and MWRKO methods in Figure 6 and Table 6, respectively. It can be seen that MWRKO’s IT and CPU are the least. Although the GRKO method is not faster than the MWRK method for most of the experiments in Table 6, it is always faster than the GRK method.

| $m \times n$     | divorce | photogrammetry | Ragusa18 | Trec8  | Stranke94 | well1033 |
|------------------|---------|----------------|----------|--------|-----------|----------|
| rank             | 9       | 390            | 23×23    | 23×84  | 10×10     | 1033×320 |
| cond(A)          | 19.3908 | 4.35e+8        | 3.48e+35 | 26.8949| 51.7330   | 166.1333 |
| density          | 50.00%  | 2.18%          | 12.10%   | 28.42% | 90.00%    | 1.43%    |
Table 6: IT and CPU of GRK, GRKO, MWRK and MWRKO for different sparse matrices

| A                | GRK | GRKO | MWRK | MWRKO | IT         | CPU         |
|------------------|-----|------|------|-------|------------|-------------|
| divorce          | 51  | 28   | 54   | 22    | 0.0053     | 0.0037      |
| photogrammetry    | 85938 | 48933 | 90480 | 27084 | 9.9917     | 8.0424      |
| Ragusa18         | 744 | 262  | 727  | 280   | 0.0577     | 0.0270      |
| Trec8            | 465 | 152  | 538  | 139   | 0.0382     | 0.0168      |
| Stranke94        | 1513 | 197  | 1453 | 181   | 0.1291     | 0.0187      |
| well1033         | 22924 | 9825 | 25250 | 8655  | 2.4278     | 1.5112      |

Example 2. We consider $f_{ancurvedtomo}(N, \theta, P)$ test problem from the MATLAB package AIR Tools [43], which generates sparse matrix $A$, an exact solution $x^*$ and $b = Ax^*$. We set $N = 60, \theta = 0 : 0.5 : 179.5^\circ, P = 50$, then resulting matrix is of size $32400 \times 3600$. We test RRE every 10 iterations and run these four methods until $\text{RRE} < \omega$ is satisfied, where $\omega = 0.5 \times 10^{-5}$.

We first remove the rows of $A$ where the entries are all 0, and perform row unitization processing on $A$ and $b$. We emphasized that this will not cause a change in $x^*$. In Figure 7, we give $60 \times 60$ images of the exact phantom and the approximate solutions obatined by the GRK, GRKO, MWRK, MWRKO methods. In Figure 7, these four methods can basically restore the original image, but in the subgraph (f) of Figure 7, we can see that the MWRKO methods needs the least iterative steps, and the GRKO method has less iterative steps than GRK method. It can be observed from Table 7 that the MWRKO method is the best in terms of IT and CPU.
Figure 7: Performance of GRK, GRKO, MERK, MWRKO methods for an curved tomo test problem. (a): Exact phantom. (b): GRK. (c): GRKO. (d): MWRK. (e): MERKO. (f): Convergence history of GRK, GRKO, MWRK, MWRKO methods.

Table 7: IT and CPU of GRK, GRKO, MWRK and MWRKO for an curved tomo test problem

| Method   | IT   | CPU   |
|----------|------|-------|
| GRK      | 13550| 581.17|
| GRKO     | 12750| 538.82|
| MWRK     | 12050| 504.83|
| MWRKO    | 10790| 452.86|

Example 3. We use an example from 2D seismic travel-time tomography reconstruction, implemented in the function `seismictomo(N, s, p)` in the MATLAB package AIR Tools [43], which generates sparse matrix $A$, an exact solution $x^*$ and $b = Ax^*$. We set $N = 12$, $s = 24$, $p = 35$, then resulting matrix is of size $840 \times 144$. We run these four methods until $\text{RRE} < \omega$ is satisfied, where $\omega = 0.5 \times 10^{-5}$.

We first remove the rows of $A$ where the entries are all 0, and perform row unitization processing on $A$ and $b$. In Figure 8, we give $12 \times 12$ images of the exact phantom and the approximate solutions obtained by the GRK, GRKO, MWRK, MWRKO methods. From the subgraph (f) of Figure 8 and Table 8, we can see that the MRKO method, the GRKO method, and the MWRK method perform similarly in the number of iteration steps, and are twice as small as the number of iteration steps of the GRK method. It can be observed from Table 8 that MWRKO method is the best in terms of IT and CPU.
Figure 8: Performance of GRK, GRKO, MERK, MWRKO methods for *seismictomo* test problem. (a): Exact phantom. (b): GRK. (c): GRKO. (d): MWRK. (e): MERKO. (f): Convergence history of GRK, GRKO, MERK, MWRKO methods.

Table 8: IT and CPU of GRK, GRKO, MWRK and MWRKO for *seismictomo* test problem

| method  | IT   | CPU      |
|---------|------|----------|
| GRK     | 831  | 0.0443   |
| GRKO    | 452  | 0.0273   |
| MWRK    | 447  | 0.0125   |
| MWRKO   | 420  | 0.0108   |

5 Conclusion

Combined with the representative randomized and non-randomized row index selection strategies, two Kaczamrz-type methods with oblique projection for solving large-scale consistent linear systems are proposed, namely the GRKO method and the MWRKO method. The exponential convergence of the GRKO method and the MWRKO method are deduced. Theoretical and experimental results show that the convergence rates of the GRKO method and the MWRKO method are better than GRK method and the MWRK method respectively. Numerical experiments show the effectiveness of these two methods, especially when the rows of the coefficient matrix $A$ are close to linear correlation.
Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities [grant number 19CX05003A-20], the National Key Research and Development Program of China [grant number 2019YFC1408400], and the Science and Technology Support Plan for Youth Innovation of University in Shandong Province [No.YCX2021151].

References

[1] W. Guo, H. Chen, W. Geng, L. Li, A Modified Kaczmarz Algorithm for Computerized Tomographic Image Reconstruction, International Conference on Biomedical Engineering and Informatics IEEE, (2009) 1-4.

[2] S. Lee, H. J. Kim, Noise properties of reconstructed images in a kilo-voltage on-board imaging system with iterative reconstruction techniques: A phantom study, Physica Medica, 30(2014), 365-373.

[3] D. Carmona-Ballester, J. M. Trujillo-Sevilla, Bonaque-González, et al, Weighted nonnegative tensor factorization for atmospheric tomography reconstruction, Astronomy and Astrophysics, 614 (2018) A41.

[4] T. Li, D. Isaacson, J. C. Newell, et al, Adaptive techniques in electrical impedance tomography reconstruction, Physiological Measurement, 35 (2014) 1111-1124.

[5] R. Ramlau, M. Rosensteiner, An efficient solution to the atmospheric turbulence tomography problem using Kaczmarz iteration, Inverse Problems, 28 (2012) 095004-1-095004–3.

[6] G. Thoppe, V. S. Borkar, D. Manjunath, A stochastic Kaczmarz algorithm for network tomography, Automatica, 50 (2014) 910-914.

[7] A. Hefny, D. Needell, A. Ramdas, Rows versus Columns: Randomized Kaczmarz or Gauss-Seidel for Ridge Regression, SIAM Journal on Scientific Computing, 39 (2017) S528-S542.

[8] V. Borkar, N. Karamchandani, S. Mirani, Randomized Kaczmarz for rank aggregation from pairwise comparisons, IEEE Information Theory Workshop (ITW), Cambridge, (2016) 389-393.

[9] J. Loera, J. Haddock, D. Needell, A sampling Kaczmarz-Motzkin algorithm for linear feasibility, SIAM Journal on Scientific Computing, 39 (2017) S66-S87.

[10] H.Q. Guan, R. Gordon, A projection access order for speedy convergence of ART (algebraic reconstruction technique): a multilevel scheme for computed tomography, Physics in Medicine and Biology, 39 (1994) 2005-2022.

[11] X. Intes, V. Ntziachristos, J. P. Culver, et al, Projection access order in algebraic reconstruction technique for diffuse optical tomography, Physics in Medicine and Biology, 47 (2002) N1-N10.

[12] X. L. Xu, J. S. Liow, S. C. Strother, Iterative algebraic reconstruction algorithms for emission computed tomography: A unified framework and its application to positron emission tomography, Medical Physics, 20 (1993) 1675-1684.

[13] S. F. McCormick, The methods of Kaczmarz and row orthogonalization for solving linear equations and least squares problems in Hilbert space, Indiana Univ. Math. J., 26 (1977) 1137-1150.
[14] D. Needell J. A. Tropp, Paved with good intentions: analysis of a randomized block Kaczmarz method. Linear Algebra Appl., 441 (2014) 199-221.

[15] D. Needell, R. Zhao, and A. Zouzias, Randomized block Kaczmarz method with projection for solving least squares. Linear Algebra Appl., 484 (2015) 322-343.

[16] I. Necoara. Faster randomized block kaczmarz algorithms. SIAM J. Matrix Anal. Appl., (2019) 1425-145.

[17] J. Liu, S. J. Wright, An accelerated randomized Kaczmarz algorithm, Math. Comp., 85 (2016) 153-178.

[18] Z.Z. Bai, W.T. Wu, On greedy randomized Kaczmarz method for solving large sparse linear systems, SIAM J. Sci. Comput., 40 (2018) A592-A606. 17M1137747.

[19] Z.Z. Bai, W.T. Wu, On relaxed greedy randomized Kaczmarz methods for solving large sparse linear systems, Appl. Math. Lett., 83 (2018) 21-26.

[20] 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) 1-15.

[21] Z.Z. Bai, W.T. Wu, On partially randomized extended Kaczmarz method for solving large sparse overdetermined inconsistent linear systems, Linear Algebra Appl. 578 (2019) 225-250.

[22] A. Ben-Israel, Generalized inverses: Theory and applications, Pure Appl. Math. 139 (1974) 125-147.

[23] K.W. Chang, C.J. Hsieh, C.J. Lin, Coordinate descent method for large-scale l2-loss linear support vector machines, J. Mach. Learn. Res 9 (2008) 1369-1398.

[24] G. Golub, C.V. Loan, Matrix Computations, Johns Hopkins University Press, 1996.

[25] S. Kaczmarz, Angenäherte auflösung von systemen linearer gleichungen, Bull. Internat. A-cad. Polon.Sci. Lettres A 29 (1937) 335-357.

[26] D. Leventhal, A. Lewis, Randomized methods for linear constraints: convergence rates and conditioning, Math. Oper. Res. 35 (2010) 641-654.

[27] Z. Lu, L. Xiao, On the complexity analysis of randomized block-coordinate descent methods, Math. Program. 152 (2015) 615-642.

[28] 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.

[29] I. Necoara, Y. Nesterov, F. Glineur, Random block coordinate descent methods for linearly constrained optimization over networks, J. Optim. Theory Appl. 173 (2017) 227-254.

[30] D. Needell, R. Ward, Two-subspace projection method for coherent overdetermined systems, J. Fourier Anal. Appl. 19 (2013) 256-269.

[31] X. Yang, A geometric probability randomized Kaczmarz method for large scale linear systems, Appl. Numer. Math. 164 (2021) 139-160.

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

21
[33] Y. Nesterov, S. Stich, Efficiency of the accelerated coordinate descent method on structured optimization problems, SIAM J. Optim. 27 (2017) 110-123.

[34] C. PoPa, T. Preclik, H. Köstler, U. Rüde, On Kaczmarz's projection iteration as a direct solver for linear least squares problems, Linear. Algebra Appl. 436 (2012) 389-404.

[35] C.G. Kang, H. Zhou, The extensions of convergence rates of Kaczmarz-type methods, J. Comput. Appl. Math. 382 (2021) 113577.

[36] P. Richtárik, M. Takáč, Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function, Math. Program. 144 (2014) 1-38.

[37] T. Strohmer, R. Vershynin, A randomized Kaczmarz algorithm with exponential convergence, J. Fourier Anal. Appl. 15 (2009) 262-278.

[38] Y. Liu, C.Q. Gu, Variant of greedy randomized Kaczmarz for ridge regression, Appl. Numer. Math. 143 (2019) 223-246.

[39] K. Du, H. Gao, A new theoretical estimate for the convergence rate of the maximal residual Kaczmarz algorithm, Numer. Math. Theor. Meth. Appl. 12 (2019) 627-639.

[40] Y.J. Guan, W.G. Li, L.L. Xing and T.T. Qiao, A note on convergence rate of randomized Kaczmarz method, Calcolo 57 (2020).

[41] S. Wright, Coordinate descent algorithms, Math. Program. 151 (2015) 3-34.

[42] J.H. Zhang, J.H. Guo, On relaxed greedy randomized coordinate descent methods for solving large linear least-squares problems, Appl. Numer. Math 157 (2020) 372-384.

[43] P.C. Hansen, J.S. Jorgensen, AIR tools II: algebraic iterative reconstruction methods, improved implementation, Numer. Algor. 79 (2018) 107-137.

[44] T.A. Davis, Y. Hu, The University of Florida sparse matrix collection, ACM Trans. Math. Software 38 (2011) 1-25.

[45] C. Popa, Projection algorithms - classical results and developments, Lap Lambert Academic Publishing, 2012.

[46] W.T. Wu, On two-subspace randomized extended Kaczmarz method for solving large linear least-squares problems. Numer. Algor. (2021).