**CONVERGENCE OF A RANDOMIZED DOUGLAS-RACHFORD METHOD FOR LINEAR SYSTEM**

**LEYU HU AND XINGJU CAI**  
School of Mathematical Sciences  
Key Laboratory for NSLSCS of Jiangsu Province  
Nanjing Normal University, Nanjing, 210023, China

**Abstract.** In this article, we propose a randomized Douglas-Rachford(DR) method for linear system. This algorithm is based on the cyclic DR method. We consider a linear system as a feasible problem of finding intersection of hyperplanes. In each iteration, the next iteration point is determined by a random DR operator. We prove the convergence of the iteration points based on expectation. And the variance of the iteration points declines to zero. The numerical experiment shows that the proposed algorithm performs better than the cyclic DR method.

1. **Introduction.** Consider large-scale systems of linear equations of the form  
\[ Ax = b, \quad \text{with} \quad A \in \mathbb{C}^{m \times n} \quad \text{and} \quad b \in \mathbb{C}^m \]  
with arbitrary shape of the matrix (tall, fat or square).

Assume the problem (1) is consistent, which means that there is at least one \( x \in \mathbb{C}^n \) such that \( Ax = b \). We also assume the rank of \( A \) is at least 2. We denote the set of rows of \( A \) by \( \{ A(i) | i = 1, \ldots, m \} \). Problem (1) can be rewritten as  
\[
\begin{align*}
\text{Find} & \quad x \\
\text{s.t.} & \quad x \in C_i, \\
& \quad C_i = \{ x \in \mathbb{C}^n | A(i)x = b(i) \}, i = 1, \ldots, m.
\end{align*}
\]

A famous type of method for (2) is the projection method. There are plenty of projection-type methods, such as von Neumann method[19, 4, 6], Douglas-Rachford (DR) method[11] and Dykstra’s method[12, 9]. For further reviews, readers can refer to [5, 3, 10].

When the convex sets are hyperplanes, Kaczmarz method[15] is popular, which is a kind of Dykstra’s method. In practice, randomized Kaczmarz method is widely used but the convergence is proved by Strohmer[18] until 2008. Soon, there comes many types of Kaczmarz method with greedy selection rules[17, 14, 2], and with randomized block projections[13, 16]. Kaczmarz method simply projects the iteration point onto the next chosen hyperplane. So the distance between iteration point and next hyperplane directly determine the descent of residual.

**2020 Mathematics Subject Classification.** 65F10, 65F20, 65K05, 90C25, 90C15, 15A06.  
**Key words and phrases.** System of linear equations, Douglas-Rachford method, randomized iteration, convergence.

The research of the second author is partially supported by the NSFC grants 11871279, 11571178.
This brings different properties for designing random criterion. A cyclic DR method scheme for general convex sets has been proposed by Borwein and Tam [8]. Cyclic DR method has similar structure as the cyclic version of Kaczmarz method which can benefit from randomization. It is natural to find out whether randomization is also work for cyclic DR method.

In this article, we propose a generic randomization scheme for DR method with a given immutable random criterion. We prove the convergence of this randomized version of DR in term of expectation. The variance of the iteration points is also proved to converge to 0. The numeric experiments illustrate randomized DR performs better than cyclic DR.

2. Preliminary.

2.1. Notations and problems. We denote \( \mathcal{M} = \{1, \ldots, m\} \) and define a family of 2-element subsets for \( \mathcal{M} \),

\[
S_{all} = \{(i_{j,1}, i_{j,2})| i_{j,1}, i_{j,2} \in \mathcal{M}, i_{j,1} \neq i_{j,2}, j = 1, 2, \ldots, N\}.
\]

\( S_{all} \) has \( N = m(m-1) \) elements and \( j \) is the index of the elements.

For the \( j \)th element \( (i_{j,1}, i_{j,2}) \in S_{all} \), we denote

\[
A_{(j)} = (A^{(i_{j,1})}^T, A^{(i_{j,2})}^T)^T, b_{(j)} = (b^{i_{j,1}}, b^{i_{j,2}})^T,
\]

\[
C_{(j)} = \{x \in \mathbb{C}^n|A_{(j)} x = b_{(j)}\}, j = 1, 2, \ldots, N,
\]

where "\(^T\)" is the transpose symbol. \( A^* \) is the conjugate transposition of \( A \). We denote by \( \bar{A} \) the conjugate of \( A \) and by row(\( A \)) the row space of \( A \).

We denote the projection and reflection operator of the hyperplane \( A^{(i)} x = b \) as \( P_i \) and \( R_i \) for \( i = 1, \ldots, m \), which is defined by

\[
P_i x = x + \frac{b^{(i)} - A^{(i)} x}{\| A^{(i)} \|_2^2} (A^{(i)})^*, \tag{6}
\]

\[
R_i x = (2P_i - I)x = x + 2\frac{b^{(i)} - A^{(i)} x}{\| A^{(i)} \|_2^2} (A^{(i)})^*, \tag{7}
\]

For simplicity, we define

\[
O_j = R_{i_{j,2}} R_{i_{j,1}}. \tag{8}
\]

**Lemma 2.1.** Let \( S \) be the subset of \( S_{all} \) that satisfies

\[
\bigcup_{(i_1, i_2) \in S} \{(i_1, i_2) | C_{i_1} \neq C_{i_2}\} \supseteq \mathcal{M}, \tag{9}
\]

where \( C_{i_1} \) is defined in (2). Define \( J_S \subseteq \{1, 2, \ldots, N\} \) as the index set of \( S \) so that

\[
J_S = \{j|(i_{j,1}, i_{j,2}) \in S\}. \tag{10}
\]

Then, vector \( x_* \in \mathbb{C}_n \) is the solution for problem (2) if and only if

\[
O_j x_* = x_*, \forall j \in J_S. \tag{11}
\]

**Proof.** Following from the definition of reflection operator, the necessity is obvious.

Conversely, assuming that \( O_j x_* = x_*, j \in J_S \), it can be proved by contradiction that \( R_{i_{j,2}} x_* = x_*, j \in J_S \). Assuming \( R_{i_{j,1}} x_* \neq x_* \) for some \( j \), let \( C_{i_{j,1}} \) and \( C_{i_{j,2}} \) be the hyperplanes related to reflection operators \( R_{i_{j,1}} \) and \( R_{i_{j,2}} \). We will prove they are the same. Following the definition of \( R_{i_{j,1}} \),

\[
P_{i_{j,1}} x_* = (R_{i_{j,1}} x_* + x_*)/2. \tag{12}
\]
Using the property of projection[7][Theorem 3.14], we know that for all \( y \in C_{i,1}, \)
\[
(x_\star - P_{i,1}x_\star, y - P_{i,1}x_\star) \leq 0. \tag{13}
\]
From (12) and (13), we know that \( x_\star - R_{i,1}x_\star \) is a normal vector of hyperplane \( C_{i,1}. \) For the same reason, \( R_{i,1}x_\star - O_jx_\star \) is a normal vector of \( C_{i,2}. \) We know that \( O_jx_\star = x_\star, \) which means \( R_{i,1}x_\star - x_\star \) is the normal vector of \( C_{i,2}. \) It yields that \( C_{i,1} \) and \( C_{i,2} \) are parallel. On the other hand, we have
\[
P_{i,2}R_{i,1}x_\star = (O_jx_\star + R_{i,1}x_\star)/2 = (x_\star + R_{i,1}x_\star)/2 = P_{i,2}x_\star. \tag{14}
\]
It means \( C_{i,1} \) and \( C_{i,2} \) share a common point. Parallel hyperplanes share a common point means they are the same which contradicts to the definition of \( S \) in (9).

So, \( R_{i,1}x_\star = R_{i,2}x_\star = x_\star, \) \( j \in J_S. \) Thus \( x_\star \in C_{i,1} \cap C_{i,2}; \) \( j \in J_S, \) which means \( x_\star \in C_i, \forall i \in M, \) i.e. \( x_\star \) is a solution to problem (2).

\[\square\]

**Remark 1.** Linear system (2) is based on hyperplanes \( C_i, i = 1, \ldots, m, \) so randomized Kaczmarz method need each hyperplane has a positive probability to be selected. However, DR operator relates to twins in \( S_{all}, \) which is an arrangement of \( \{1, 2, \ldots, m\}. \) The lemma here shows that by removing unnecessary twins in \( S_{all}, \) the solution to problem (2) doesn’t change. As a result, when designing random criterion for randomized DR, some of the twins in \( S_{all} \) don’t need to have positive probability. In the rest of this article, we assume that \( S \) satisfies (9) and \( J_S \) as the index set of \( S \) without further explanation.

For set \( S \) that satisfied (9), we can rewrite problem (2) as
\[
\begin{align*}
\text{Find } x \\
\text{s.t. } x \in C_{(j)}, \; j \in J_S.
\end{align*}
\tag{15}
\]
where \( C_{(j)} \) is defined in (5).

2.2. **Useful lemmas.** We denote by \( \text{tr}(\cdot) \) the trace of a matrix, which is defined as the sum of diagonal elements of a matrix. We denote by \( \text{vec}(\cdot) \) the vectorization of a matrix, i.e.
\[
\text{vec}(A) = [a_{11}, \ldots, a_{m1}, a_{12}, \ldots, a_{m2}, \ldots, a_{1n}, \ldots, a_{mn}]^T
\]
where \( a_{ij} \) represents the element in row \( i \) and column \( j \) of \( A. \) \( A \otimes B \) represents the Kronecker product of two matrices \( A, B; \) and \( \text{rank}(\cdot) \) indicates the rank of a matrix.

The following lemma is about the Kronecker product, vectorization, rank and trace:

**Lemma 2.2.** [1][Lemma 2.1] Let \( R, S, F \) and \( G \) be complex matrices of suitable dimensions. Then the following statements hold true:
\[
\begin{align*}
(i) \; \text{tr}(RS) = \text{tr}(SR) \; &\text{and} \; \text{tr}(R^*S) = \text{vec}(R)^* \text{vec}(S), \\
(ii) \; (R \otimes S)^* = R^* \otimes S^* \; &\text{and} \; \text{vec}(RSG^T) = (G \otimes R) \text{vec}(S), \\
(iii) \; (R \otimes S)(F \otimes G) = (RF) \otimes (SG), \\
(iv) \; \text{rank}(R \otimes S) = \text{rank}(R) \text{rank}(S).
\end{align*}
\]

**Definition 2.3.** [7][Definition 4.1] Let \( D \) be a non-empty subset of a Hilbert space \( \mathcal{H}, \) and let \( T : D \to \mathcal{H}. \) Then \( T \) is
\[
i \text{nonexpansive, if} \quad \|Tx - Ty\| \leq \|x - y\|, \forall x, y \in D;
\]
(ii) firmly nonexpansive, if
\[ \|Tx - Ty\|^2 + \|(I - T)x - (I - T)y\|^2 \leq \|x - y\|^2, \forall x, y \in D; \]

(iii) \(\alpha\)-averaged with constant \(\alpha \in (0, 1)\), if there exists a nonexpansive operator \(R : D \to H\) such that \(T = (1 - \alpha)I + \alpha R\).

Here, \(I\) is identity operator. It is obvious that a firmly nonexpansive operator is also nonexpansive and \(1/2\)-averaged.

**Lemma 2.4.** [7][Proposition 4.30] Let \(D\) be a nonempty subset of \(H\), \((T_i)_{i \in I}\) be a finite family of nonexpansive operators from \(D\) to \(H\), \((\omega_i)_{i \in I}\) be real numbers in \((0, 1]\) such that \(\sum_{i \in I} \omega_i = 1\), and \((\alpha_i)_{i \in I}\) be real numbers in \((0, 1]\) such that, for every \(i \in I\), \(T_i\) is \(\alpha_i\)-averaged, and set \(\alpha = \max_{i \in I} \alpha_i\). Then \(\sum_{i \in I} \omega_i T_i\) is \(\alpha\)-averaged.

The set of fixed points of an operator \(T : X \to X\) is denoted by \(\text{Fix } T\), i.e.,
\[ \text{Fix } T = \{ x \in X \mid Tx = x \}. \]

**Theorem 2.5.** Let \(\alpha \in (0, 1)\), and let \(T : H \to H\) be an \(\alpha\)-averaged operator such that \(\text{Fix } T \neq \emptyset\). Then for any \(x_0 \in H\), \(T^n x_0\) converge weakly to a point in \(\text{Fix } T\).

**Proof.** It is a direct result of [7][Proposition 5.15]. \(\square\)

**Remark 2.** In practical, \(H\) is a finite-dimensional normed vector space, so that \(T^n x_0\) converge strongly to a point in \(\text{Fix } T\).

3. **Randomized Douglas-Rachford method for linear systems.** We can use the DR method to find the intersection of \(C_{(j)}\) in problem (15). We define the DR operator as
\[ T_j = I + O_j. \]  

The randomized DR method is listed below.

**Algorithm 1** Randomized Douglas-Rachford method for linear systems

**Input:** Problem formed as (15) with set \(S\) which satisfies (9), probability \(\{p_j\}_{j \in J_S}\) for selecting DR operator at each iteration, maximum iteration number \(k_{\text{max}}\).

**Output:** A solution \(x_*\) for problem (15).

1: Randomly pick \(x_0 \in \text{row}(A)\) as an initial point,  
2: for \(k = 1, 2, \ldots, k_{\text{max}}\) do  
3: Pick \(j_k \in J_S\) randomly according to probability \(p_j\),  
4: Update \(x_k = T_{j_k} x_{k-1}\),  
5: end for  
6: Output \(x_* = x_{k_{\text{max}}}\).

Following the property of probability, we have
\[ p_j > 0, \forall j \in J_S, \]
\[ \sum_{j \in J_S} p_j = 1. \]

**Lemma 3.1.** Let \(P_i\), \(R_i\) and \(T_j\) defined as (6)/(7)/(16). Then \(R_i\) is nonexpansive and \(P_i, T_j\) are firmly nonexpansive.
Proof. See [7] [Proposition 4.8, Corollary 4.10, Remark 4.24].

Lemma 3.2. Let
\[ G = \sum_{j \in J_S} p_j T_j, \]  
(19)
then \( G \) is firmly nonexpansive.

Proof. It follows from (17) and (18) that \( G \) is a convex combination of \( T_j \); and from Lemma 3.1 that \( T_j \) is firmly nonexpansive for all \( j \in J_S \). As a result, \( G \) is firmly nonexpansive followed by Lemma 2.4.

Theorem 3.3. Assume \( \{ x_k \} \) is generated by Algorithm 1, then for any \( x_0 \in \text{row}(A) \), the expectation of \( \{ x_k \} \) converges to a solution of problem (15), i.e.
\[ \lim_{k \to \infty} E_{x_1, x_2, \ldots, x_k | x_0} x_k = x_*. \]

Proof. According to Algorithm 1, the \( k \)th iteration point is
\[ x_k = T_{j_k} x_{k-1}, \]
and the conditional expectation of \( x_k \) given the sequence \( \{ x_0, \ldots, x_{k-1} \} \) is
\[ E_{x_k \in J_S | x_0, \ldots, x_{k-1}} = E_{x_{k-1} \in J_S | x_0, \ldots, x_{k-2}} T_{j_k} x_{k-1} \]
\[ = E_{x_{k-1} \in J_S} T_{j_k} x_{k-1} \]
\[ = \sum_{j \in J_S} p_j T_j x_{k-1} \]
\[ = G x_{k-1}. \]
(20)
The selection of \( T_j \) is independent in each iteration, so
\[ E x_k = E \left[ E_{x_{k-1} \in J_S | x_0, \ldots, x_{k-2}} T_{k-1} x_{k-2} \right] \]
\[ = G E x_{k-1} \]
\[ = G^k x_0. \]
(21)

We use Theorem 2.5 here to finish the proof: according to Lemma 3.2, \( G \) is firmly nonexpansive; and by assumption, problem (15) is consistent, so \( \text{Fix} G \neq \emptyset \). By now we prove \( G^k x_0 \) converges to \( \text{Fix} G \). We still need to prove \( \text{Fix} G = X_* \).

Let \( x_* \) be a solution to problem (15), i.e.
\[ x_* \in C(j), \forall j \in J_S. \]
\( J_S \) satisfies (9), so
\[ b^{(i)} - A^{(i)} x_* = 0, \forall i = 1, \ldots, m. \]
Substituting it into (7) and (16), we have
\[ R_i x_* = T_j x_* = x_*, \forall i = 1, \ldots, m, j \in J_S. \]
According to (18)(19), we have
\[ G x_* = x_*, \]
(22)
which means \( X_* \subseteq \text{Fix} G \).

Conversely, we prove \( \text{Fix} G \subseteq X_* \Rightarrow \bar{X}_* \subseteq \overline{\text{Fix} G} \), where \( \overline{X} \) represents the complement of \( X \). Let \( x \) be an iteration point of the algorithm. For simplicity, we denote for any \( j \in J_S \) that \((j_1, j_2) = (i, j) \in S\) as the \( j \)th member of \( S \), then define
\[ O_j x = R_{j_2} R_{j_1} x, \]
and the corresponding iteration point is \( \frac{I + O_j}{2} x \).

Supposing that \( x \notin \mathcal{X}^* \), according to Lemma 2.1, there exists at least \( j' \in J_S \), such that
\[
O_{j'} x \neq x. \tag{23}
\]

Without loss of generality, we assume \( 0 \in \mathcal{X}^* \), so all the hyperplane associate with \( R_{j_2} \) and \( R_{j_1} \) are subspace. According to the properties of projection and reflection, \( R_{j_1} x - P_{j_1} x \perp P_{j_1} x \) and \( P_{j_1} x \perp P_{j_1} x \). Using Pythagoras theorem, for any \( i \in \mathcal{M} \),
\[
\| R_{j_1} x \|_2^2 = \| R_{j_1} x - P_{j_1} x + P_{j_1} x \|_2^2 = \| R_{j_1} x - P_{j_1} x \|_2^2 + \| P_{j_1} x \|_2^2 = \| P_{j_1} x - x \|_2^2 + \| P_{j_1} x \|_2^2 = \| x \|_2^2,
\]
which means
\[
\| x \|_2 = \| R_{j_1} x \|_2 = \| O_{j_1} x \|_2. \tag{24}
\]
In other words, \( x, R_{j_1} x \) and \( O_{j_1} x \) are on the 2-norm ball with a center of 0 and a radius of \( \| x \|_2 \).

(24) implies either
\[
\langle x, O_{j_1} x \rangle < \| O_{j_1} x \|_2 \| x \|_2 = \| x \|_2^2 \tag{25}
\]
or \( O_{j_1} x = x \). But (23) provides at least for \( j' \), \( O_{j'} x \neq x \), which means
\[
\langle x, O_{j'} x \rangle < \| x \|_2^2. \tag{26}
\]
For other \( j \in J_S \), it always holds that
\[
\langle x, O_{j} x \rangle \leq \| x \|_2^2. \tag{27}
\]

Since \( x \) is chosen randomly, we know that for any \( x \notin \mathcal{X}^* \)
\[
\| Gx \|_2^2 \leq \left\| \sum_{j \in J_S} p_j \frac{I + O_j}{2} x \right\|_2^2 \leq \frac{1}{4} \left\| \sum_{j \in J_S} p_j O_j x \right\|_2^2 + \frac{1}{4} \left\| \sum_{j \in J_S} p_j x \right\|_2^2 + \frac{1}{2} \left\langle x, \sum_{j \in J_S} p_j O_j x \right\rangle \tag{28}
\]
\[
\leq \frac{1}{4} \| x \|_2^2 + \frac{1}{4} \sum_{j \in J_S} p_j \| O_j x \|_2^2 + \frac{1}{2} \sum_{j \in J_S} p_j \langle x, O_j x \rangle < \| x \|_2^2,
\]
so that \( Gx \neq x \), i.e. \( x \notin \text{Fix} G \), which means \( \text{Fix} G \subseteq \mathcal{X}^* \).

\[ \square \]

**Remark 3.** Note that for any \( j \in J_S \),
\[
T_j x = \frac{I + O_j}{2} x,
\]
where \( O_j \) is defined in (8). As a result, \( T_j x \) must be a linear combination of row vectors of \( A \) and \( x \). Let \( \text{row}(A) \) be the row space of \( A \), and \( x \in \text{row}(A) \), then we have \( T_j x \in \{ y | y = x + z, z \in \text{row}(A) \} \subseteq \text{row}(A) \). We will reach a conclusion that
\[
G^k x_0 \in \text{row}(A), \text{ if } x_0 \in \text{row}(A). \tag{29}
\]
So, if we begin with $x_0 \in \text{row}(A)$, $G^k x_0$ will converge to $X^* \cap \text{row}(A)$, which contains only the least norm least squares solution to the linear system according to Bai[2].

**Theorem 3.4.** Assume $\{x_k\}$ is generated by Algorithm 1, then the variance of $\{x_k\}$ converges to 0, i.e.

$$\lim_{k \to \infty} Dx_k = 0.$$  

**Proof.** In Theorem 3.3 we already knows that $\lim_{k \to \infty} Ex_k = x_*$, where $x_*$ is a solution point of problem (15). The component of $x_k$ is independent, so the variance of $x_k$ is a diagonal matrix and the $i$-row, $i$-column of $Dx$ is

$$(Dx)_{ii} = E(x^{(i)}_k)^2 - (Ex^{(i)}_k)^2,$$

where $x^{(i)}_k$ is the $i$th component of $x_k$. Without loss of generality, we assume $x_* = 0$, so that the 1-norm of vector $\text{vec}(Dx_k)$ satisfies

$$\| \text{vec}(Dx_k) \|_1 = \sum_{i=1}^n |(Dx)_{ii}| = \sum_{i=1}^n \left| E(x^{(i)}_k)^2 - \|x^{(i)}_*\|_2^2 \right| = E\|x_k\|_2^2. \quad (30)$$

Next, we show that $E\|x_k\|_2^2$ converges to 0. We have

$$x_k = T_{j_k} T_{j_{k-1}} \cdots T_{j_2} T_{j_1} x_0, \quad (31)$$

where we use $j_k$ to denote the index $j \in J_S$ selected in $k$th iteration. Using Lemma 2.2 we derive

$$\|x_k\|_2^2 = \|T_{j_k} T_{j_{k-1}} \cdots T_{j_2} T_{j_1} x_0\|_2^2$$

$$= \text{tr} (x_0^* T_{j_1}^* \cdots T_{j_k}^* T_{j_k} \cdots T_{j_2} T_{j_1} x_0)$$

$$= \text{tr} (T_{j_1}^* \cdots T_{j_k}^* T_{j_k} \cdots T_{j_2} T_{j_1} x_0 x_0^*)$$

$$= \text{vec}(T_{j_1}^* \cdots T_{j_k}^* T_{j_k} \cdots T_{j_2} T_{j_1}) \text{vec}(x_0 x_0^*)$$

$$= \text{vec}((I)^* (T_{j_1} \otimes T_{j_1})^* (T_{j_2} \otimes T_{j_2})^* \cdots (T_{j_k} \otimes T_{j_k})^* \text{vec}(x_0 x_0^*),$$

where $T$ represents the conjugate of $T$, $T^*$ represent the conjugate transpose of $T$.

If we add an expectation symbol before the equation, since every selection in the algorithm is independent, the expectation symbol can be allocate to each factor:

$$E\|x_k\|_2^2 = \text{vec}(I)^* E((T_{j_1} \otimes T_{j_1})^* (T_{j_2} \otimes T_{j_2})^* \cdots (T_{j_k} \otimes T_{j_k})^* \text{vec}(x_0 x_0^*)) \quad (32)$$

Since the probability $p_j$ does not change during iterations, all the expectation above is the same. We denote $H = E(T_{j_k} \otimes T_{j_k})$ for all $k$. Thus the variance becomes

$$Dx_k = \text{vec}(I)^* (H^k)^* \text{vec}(x_0 x_0^*). \quad (33)$$

We will show that $H$ is an average operator. According to the definition of $H$ we can state it as

$$H = \sum_{j \in J_S} p_j (T_j \otimes T_j) \quad (34)$$

$$= \sum_{j \in J_S} p_j \left( \frac{T+O_j}{2} \otimes \frac{I+O_j}{2} \right). \quad (35)$$
Consider a matrix $V \in \mathbb{C}^{n \times n}$. We write $v = \text{vec}(V)$ for its vector form. Then, by using Lemma 2.2, we have

$$Hv = \sum_{j \in J_S} p_j \left( \frac{I + O_j}{2} \otimes \frac{I + O_j}{2} \right) \text{vec}(V)$$

$$= \sum_{j \in J_S} p_j \text{vec} \left( \frac{I + O_j^* V I + O_j}{2} \right)$$

$$= \sum_{j \in J_S} p_j \text{vec} \left( \frac{V + O_j^* V + VO_j + O_j^* VO_j}{4} \right)$$

$$= \sum_{j \in J_S} p_j \left( \frac{1}{4} \text{vec}(V) + \frac{3}{4} \left( \text{vec}(O_j^* V + VO_j + O_j^* VO_j) \right) \right)$$

We define operator $U_j : \mathbb{C}^n \to \mathbb{C}^n$ as

$$U_j v = \text{vec} \left( \frac{O_j^* V + VO_j + O_j^* VO_j}{3} \right).$$

Since we have assumed $0 = x_* \in \mathcal{X}_*$, for all $x \in \mathbb{C}^n$, (24) holds. Thus, we have

$$\|v\|_2 = \|\text{vec}(V)\|_2 = \|V\|_F = \|O_j^* V\|_F = \|\text{vec}(O_j^* V)\|_2.$$ (36)

Here, $\|\cdot\|_F$ means the Frobenius norm. So that,

$$\|v\|_2 = \|V\|_F = \|\text{vec}(V^*)\|_2$$

$$\quad = \|\text{vec}(O_j^* V^*)\|_2 = \|\text{vec}(VO_j)\|_2 = \|\text{vec}(O_j^* VO_j)\|_2.$$ (37)

Through (36),(37), we can prove $U_j$ is nonexpansive. For any $A, B \in \mathbb{C}^{n \times n}$, we write any $a = \text{vec}(A), b = \text{vec}(B) \in \mathbb{C}^{n^2}$. Considering $A - B$ as $V$ in (36),(37), we have

$$\|U_j a - U_j b\|_2 = \left\| \text{vec} \left( \frac{O_j^* (A-B) + (A-B) O_j + O_j^* (A-B) O_j}{3} \right) \right\|_2$$

$$\leq \|\text{vec}(O_j^* (A-B))\|_2 + \|\text{vec}((A-B) O_j)\|_2 + \|\text{vec}(O_j^* (A-B) O_j)\|_2$$ (38)

$$= \|A - B\|_F = \|a - b\|_2.$$

Note that

$$H = \sum_{j \in J_S} p_j \left( \frac{1}{4} I + \frac{3}{4} U_j \right),$$

which means $H$ is a $\frac{1}{4}$-averaged operator, according to Lemma 2.4.

We will then find out the fixed points of $H$. Suppose that $x$ is a solution to problem (15). Let $v = \tilde{x} \otimes x$. It follows from Lemma 2.1 that

$$Hv = \sum_{j \in J_S} p_j (T_j \otimes T_j)(\tilde{x} \otimes x) = v.$$ (39)

So $\{v | v = \tilde{x} \otimes x\} \subseteq \text{Fix} H$.

Suppose the opposite, $x$ is not a solution to problem (15). Let $v = \tilde{x} \otimes x$. And there exist at least a $j' \in J_S$, such that $U_{j'} v \neq v$, using the same technique as (26) and (28), we have

$$\langle v, U_{j'} v \rangle < \|U_{j'} v\|_2 \|v\|_2 = \|v\|_2^2.$$ (40)
Since \( \langle v, U_j v \rangle \leq \|U_j v\|_2 \|v\|_2 \) is always true for any \( j \),

\[
\|Hv\|_2^2 = \left\| \sum_{j \in J_S} p_j \left( \frac{1}{4} I + \frac{3}{4} U_j \right) v \right\|_2^2 \\
= \frac{1}{16} \|v\|_2^2 + \frac{9}{16} \left\| \sum_{j \in J_S} p_j U_j v \right\|_2^2 + \frac{6}{16} \langle v, U_j v \rangle
\]  

(39)

This implies \( v / \notin \text{Fix } H \), which means \( \text{Fix } H \subseteq \{v|v = \bar{x} \otimes x\} \), and thus \( \text{Fix } H = \{v|v = \bar{x} \otimes x\} \neq \emptyset \).

Since we already proved that \( H \) is \( \frac{1}{4} \)-averaged, we conclude by Theorem 2.5 that \( H^k \text{vec}(x_0 x_0^*) \) converge to \( \text{vec}(x_\infty x_\infty^*) \) as \( k \to \infty \), where \( x_\infty \in X_* \) is a solution to problem (15). So,

\[
\lim_{k \to \infty} Dx_k = \text{vec}(I)^* \text{vec}(x_\infty x_\infty^*) = \|x_\infty\|_2^2.
\]

At last, we will show that \( x_\infty = x_* = 0 \), since \( x_\infty \) and \( x_* \) could both be the solution point but not equal to each other at the same time. In fact, similar to Remark 3, \( O_j x \in \{y|y = x + z, z \in \text{row}(A)\} \). So

\[
U_j (x \otimes x^*) \in \{y \otimes y^* | y = x + z, z \in \text{row}(A)\}.
\]

Considering the fact that \( H \) is a linear combination of \( U_j \) and identity operator, we reach the similar conclusion as (29) that

\[
H^k x_0 \in \{y \otimes y^* | y = x_0 + z, z \in \text{row}(A)\},
\]

(40)

which means that

\[
Dx_k^2 \in \{\|y\|_2^2 | y = x_0 + z, z \in \text{row}(A)\}.
\]

(41)

Since there is only one point in the intersection \( \{y|y = x_0 + z, z \in \text{row}(A)\} \cap X_* \) because of the fundamental theorem of algebra, and that \( x_*, x_\infty \) are both in this intersection according to (29) and (41), it must be true that \( x_\infty = x_* = 0 \), which means \( \lim_{k \to \infty} Dx_k = 0 \).

**4. Experimental results.** In this section, we will implement the randomized DR method, the cyclic DR method[8], and the randomized Kaczmarz method[18]. We will show that randomized DR and randomized Kaczmarz perform almost the same on problem 1, and better than their cyclic version. Kaczmarz method is a kind of projection method for solving linear systems. Randomized Kaczmarz is a well-known randomized method for solving linear systems.

Our experiment uses randomly generated matrices by MATLAB function \texttt{randn}, which produces independent standard normal entries for matrix \( A \in \mathbb{C}^{m \times n} \), i.e., \( A \sim \mathcal{N}(0, 1) \). We generate the solution vector \( x_* \) using \texttt{randn}. The right-hand side term \( b \in \mathbb{C}^m \) is taken as \( Ax_* \). All algorithm starts from an initial point \( x_0 = 0 \).

Before the simulation, we need to point out that the time cost of randomized algorithms can be affected by the efficiency of \texttt{rand} function. In a process of a randomized algorithm, we perform a random selection at every iteration. Although we only need an integer ranging from 1 to \( m \) (the total number of the rows of matrix \( A \)), the \texttt{rand} function still generate a floating point number or a large integer
Fig. 1. Comparison of number of projections/reflections between randomized Kaczmarz (RK), cyclic DR (CDR) and randomized DR (RDR) when the size of $A$ and $b$ is $m = 1000$ and $n = 100$.

depending on its data type. It is not a problem on most platform, since a `rand` only cost 1 FLOPS. However, MATLAB gives a paralleled optimization for matrix multiplication while `rand` function is not optimized. Furthermore, there are plenty of tricks that have no relationship with the algorithm. In order to get rid of these tricks and concentrate on the algorithm itself, we uses the number of projection-reflections (NPR) to illustrate the cost of different algorithms.

According to Strohmer[18], randomized Kaczmarz is linearly convergent. Fig 1 shows that randomized DR and cyclic DR are both linearly convergent and randomized DR has the similar converge rate as randomized Kaczmarz.

**Table 1.** Number of Projections/Reflections (NPR) for $n = 50$

|   | 1000 | 2000 | 3000 | 4000 | 5000 |
|---|------|------|------|------|------|
| CDR | 1264 | 1404 | 1282 | 1488 | 1390 |
| NPR RK | 725.3 | 701.1 | 696.1 | 671.3 | 683.8 |
| RDR | 705.2 | 683.8 | 696 | 669.6 | 698.8 |

**Table 2.** Number of Projections/Reflections (NPR) for $n = 100$

|   | 1000 | 2000 | 3000 | 4000 | 5000 |
|---|------|------|------|------|------|
| CDR | 2620 | 2604 | 2692 | 2770 | 2770 |
| NPR RK | 1508.4 | 1419 | 1425.5 | 1422.5 | 1395.1 |
| RDR | 1523.2 | 1425.8 | 1409.6 | 1421.4 | 1381.8 |
Table 3. Number of Projections/Reflections (NPR) for $n = 150$

| $m$     | 1000   | 2000   | 3000   | 4000   | 5000   |
|---------|--------|--------|--------|--------|--------|
| CDR     | 4272   | 4286   | 4104   | 4096   | 3854   |
| NPR     | 2466.9 | 2194.4 | 2175.4 | 2092.1 | 2152   |
| RK      | 2574.2 | 2227.8 | 2160.8 | 2133   | 2113.8 |
| RDR     | 2152   | 2092.1 | 2133   | 2113.8 |        |

We perform these algorithms on different sizes of matrices. For each size of matrix, we generate a matrix and repeat every algorithm on the same matrix for 10 times and average the results. We made the fastest data on bold for convenience. From Table 1-3 we see that randomized Kaczmarz and randomized DR perform almost the same in terms of NPR, while randomized DR always outperforms cyclic DR. It shows that the acceleration of randomization works not only for projection method like Kaczmarz, but also for DR method.

5. Conclusions and future works. In this article, we propose a randomized DR method for linear system. We prove that the method is convergent by expectation and the variance is converges to 0 when adopt any random criterion that is fix during iteration. When we use a uniform random criterion, the method has the same amount of computation as the Randomized Kaczmarz method which is also a randomized method for linear system that uses a uniform random criterion.

The Randomized Kaczmarz method performs a simple projection in every iteration, so most refinement of the algorithm is about finding the next hyperplane with the biggest resident. However, the descent of DR method relates with the angle between the two hyperplane chosen in every iteration that can be calculate before the algorithm in parallel. This may bring new ways to optimize randomized algorithm for linear systems.

REFERENCES

[1] Z.-Z. Bai and W.-T. Wu, On convergence rate of the randomized kaczmarz method, *Linear Algebra and Its Applications*, 553 (2018), 252–269.
[2] Z.-Z. Bai and W.-T. Wu, On greedy randomized kaczmarz method for solving large sparse linear systems, *SIAM Journal on Scientific Computing*, 40 (2018), A592–A606.
[3] H. H. Bauschke, Projection algorithms: results and open problems, in *Studies in Computational Mathematics*, Elsevier, 8 (2001), 11–22.
[4] H. H. Bauschke and J. M. Borwein, On the convergence of von neumann’s alternating projection algorithm for two sets, *Set-Valued Analysis*, 1 (1993), 185–212.
[5] H. H. Bauschke and J. M. Borwein, On projection algorithms for solving convex feasibility problems, *SIAM Review*, 38 (1996), 367–426.
[6] H. H. Bauschke, J. M. Borwein and A. S. Lewis, The method of cyclic projections for closed convex sets in hilbert space, *Contemporary Mathematics*, 204 (1997), 1–38.
[7] H. H. Bauschke, P. L. Combettes et al., *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, vol. 408, Springer, 2011.
[8] J. M. Borwein and M. K. Tam, A cyclic douglas–rachford iteration scheme, *Journal of Optimization Theory and Applications*, 160 (2014), 1–29.
[9] J. P. Boyle and R. L. Dykstra, A method for finding projections onto the intersection of convex sets in hilbert spaces, in *Advances in order restricted statistical inference*, Springer, (1986), 28–47.
[10] C. Brezinski, Projection methods for linear systems, *Journal of Computational and Applied Mathematics*, 77 (1997), 35–51.
[11] J. Douglas and H. H. Rachford, On the numerical solution of heat conduction problems in two and three space variables, Transactions of the American mathematical Society, 82 (1956), 421–439.

[12] R. L. Dykstra, An algorithm for restricted least squares regression, Journal of the American Statistical Association, 78 (1983), 837–842.

[13] T. Elfving, Block-iterative methods for consistent and inconsistent linear equations, Numerische Mathematik, 35 (1980), 1–12.

[14] M. Griebel and P. Oswald, Greedy and randomized versions of the multiplicative schwarz method, Linear Algebra and Its Applications, 437 (2012), 1596–1610.

[15] S. Kaczmarz, Angenäherte aufflösung von systemen linearer gleichungen, Bulletin International de l’Académie Polonaise des Sciences et des Lettres. Classe des Sciences Mathématiques et Naturelles. Série A, Sciences Mathématiques, 35 (1937), 355–357.

[16] D. Needell, R. Zhao and A. Zouzias, Randomized block kaczmarz method with projection for solving least squares, Linear Algebra and Its Applications, 484 (2015), 322–343.

[17] J. Nutini, B. Sepehry, I. Laradji, M. Schmidt, H. Koepke and A. Virani, Convergence rates for greedy kaczmarz algorithms, and faster randomized kaczmarz rules using the orthogonality graph, in Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence, AUAI Press, (2016), 547–556.

[18] T. Strohmer and R. Vershynin, A randomized kaczmarz algorithm with exponential convergence, Journal of Fourier Analysis and Applications, 15 (2008), 262.

[19] J. Von Neumann, The geometry of orthogonal spaces, Functional operators-vol. II, Annals of Mathematics Studies, 22 (1950), This is a reprint of mimeographed lecture notes, first distributed in 1933.

Received March 2020; 1st revision August 2020; Final revision September 2020.

E-mail address: kz.david@outlook.com
E-mail address: caixingju@njnu.edu.cn