

# Remarks on Kaczmarz algorithm for solving consistent and inconsistent system of linear equations<sup>\*</sup>

Xinyin Huang<sup>1</sup>, Gang Liu<sup>2</sup>, and Qiang Niu<sup>2,3</sup>[0000-0002-6880-6874]

<sup>1</sup> Department of Mathematics, King's College of London, United Kingdom.

<sup>2</sup> Department of Mathematical Sciences, Xi'an Jiaotong-Liverpool University, Suzhou, Jiangsu 215123, P.R. China.

<sup>3</sup> XJTLU Laboratory for Intelligent computation and Financial Technology, Suzhou, Jiangsu 215123, P.R. China.  
{Gang.Liu,Qiang.Niu}@xjtlu.edu.cn

**Abstract.** In this paper we consider the classical Kaczmarz algorithm for solving system of linear equations. Based on the geometric relationship between the error vector and rows of the coefficient matrix, we derive the optimal strategy of selecting rows at each step of the algorithm for solving consistent system of linear equations. For solving perturbed system of linear equations, a new upper bound in the convergence rate of the randomized Kaczmarz algorithm is obtained.

**Keywords:** Iterative methods · Kaczmarz method · Convergence rate · Orthogonal projection · Linear systems.

## 1 Introduction

Kaczmarz algorithm [28] is an iterative method for solving system of linear equations of the form

$$Ax = b, \quad (1)$$

where  $A \in \mathcal{R}^{m \times n}$  has full column rank,  $m \geq n$  and  $b \in \mathcal{R}^m$ . In the consistent case, the solution of (1) can be regarded as the coordinate of the common point of hyperplanes defined by each single equation in (1):

$$\mathcal{P}_i = \{x | a_i^T x = b_i\}, \quad (2)$$

where  $a_i^T$ ,  $i = 1, 2, \dots, m$ , denotes the  $i$ th row of  $A$  and  $b_i$  is the  $i$ th element of vector  $b$ .

The idea of the Kaczmarz type algorithms is to exploit the geometric structure of the problem (1), and the using a sequential of projections to seek the

---

<sup>\*</sup> This work was supported by XJTLU research enhancement fund with no. REF-18-01-04 and Key Programme Special Fund (KSF) in XJTLU with nos. KSF-E-32, KSF-P-02 and KSF-E-21. Partial of the work was supported by the Qing Lan project of Jiangsu Province.

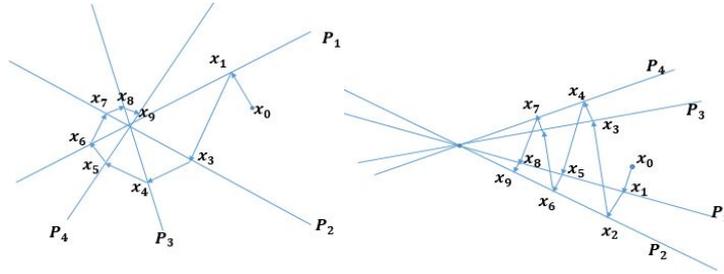
solution. The recursive process can be formulated as follows. Let  $x_0$  be an initial guess to the solution of (1), then the classical Kaczmarz algorithm iteratively generates a sequence of approximate solutions  $x_k$  by the recursive formula:

$$x_{k+1} = x_k + \frac{b_i - a_i^T x_k}{\|a_i\|_2^2} a_i, \quad (3)$$

where  $i = \text{mod}(k, m) + 1$ . For a given  $x_k$ , from (3) we can see that  $x_{k+1}$  satisfies the  $i$ th equation in (1), i.e.,  $a_i^T x_{k+1} = b_i$ . The updating formula (3) implicitly produces a solution to the following constraint optimization problem [21, 37]

$$\min_{\{x | a_i^T x_{k+1} = b_i\}} \|x - x_k\|_2,$$

which is equivalent to finding the projection of  $x_k$  from the hyperplane  $\mathcal{P}_i$ . Two geometric explanations of the above process can be illustrated by Fig 1:



**Fig. 1.** Geometric illustrations of the classical Kaczmarz iterations with  $m = 4$ .

By comparing the projection processes displayed in Fig. 1, it is natural to have the intuition that convergence of the classical Kaczmarz algorithm highly depends on the geometric positions of the associated hyperplanes. If the normal vectors of every two successive hyperplanes keep reasonably large angles, the convergence of the classical Kaczmarz algorithm will be fast, whereas two nearly parallel consecutive hyperplanes will make the convergence slow down. The Kaczmarz algorithm can be regarded as a special application of famous von Neumann's alternating projection [35] originally distributed in 1933. The fundamental idea can even trace the history back to Schwarz [38] in 1870s.

In the past few years, the Kaczmarz algorithm has been interpreted as successive projection methods [4, 7, 8, 11–13], which are also known as projection onto convex sets (POCS) [9, 17, 18, 42–44] in the optimization community. Notice that each iteration of the Kaczmarz algorithm just need  $\mathcal{O}(n)$  flops and the cost is independent with the number of equations, this type of algorithms are well-suited to problems with  $m \gg n$ . Due to its simplicity and generality, Kaczmarz algorithms find viable applications in the area of image processing and

signal process [19, 20, 24–26, 30, 36] under the name of algebraic reconstruction techniques (ART). Since 1980s, relaxation variants [11, 25, 41]

$$x_{k+1} = x_k + \lambda_k \frac{b_i - a_i^T x_k}{\|a_i\|_2^2} a_i, \quad (4)$$

and the block versions [3, 33, 34]

$$x_{k+1} = x_k + A_\tau^\dagger (b_\tau - A_\tau^T x_k), \text{ with } A = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_M \end{pmatrix}, b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{pmatrix}, \tau \in \{1, 2, \dots, M\}, \quad (5)$$

of the Kaczmarz algorithm have been widely investigated, and some fruitful theoretical results have been obtained. In particular, for consistent linear systems, it is shown [5, 31, 21, 39] that the Kaczmarz iterations converges to the least square norm solution  $x = A^\dagger b$  with any starting vector  $x_0$  in the column space of  $A^T$ . For inconsistent linear systems, the cyclic subsequences generated by the Kaczmarz algorithm converges to a weighted least squares solution when the relaxation parameter  $\lambda_k$  goes to zero [12].

As indicated in Fig 1, convergence of the classical Kaczmarz algorithm depends on the sequence of successive projections, which relies upon the ordering of the rows in the matrix  $A$ . In some real applications, it is observed [25, 30] that instead of selecting rows of the matrix  $A$  sequentially at each step of the Kaczmarz algorithm, randomly selection can often improve its convergence. Recently, in the remarkable paper [39], Strohmer and Vershynin proved the rate of convergence for the following randomized Kaczmarz algorithm

$$x_{k+1} = x_k + \frac{b_{r(i)} - a_{r(i)}^T x_k}{\|a_{r(i)}\|_2^2} a_{r(i)}$$

where  $r(i)$  is chosen from  $\{1, 2, \dots, m\}$  with probabilities  $\frac{\|a_{r(i)}\|_2^2}{\|A\|_F^2}$ . In particular, the following bound on the expected rate of convergence for the randomized Kaczmarz method is proved

$$\mathbb{E}\|x_k - x\|_2^2 \leq \left(1 - \frac{1}{\kappa(A)^2}\right)^k \|x_0 - x\|_2^2, \quad (6)$$

where  $\kappa(A) = \|A\|_F \|A^{-1}\|_2$ , with  $\|A^{-1}\|_2 = \inf\{M : M\|Ax\|_2 \geq \|x\|_2\}$  be the scaled conditioned number of  $A$  introduced by J. Demmel [14]. Due to this pioneering work that characterized the convergence rate for the randomized Kaczmarz algorithms, the idea stimulated considerable interest in this area and various investigations [1, 2, 6, 10, 15] have been performed recently. In particular, some acceleration strategies have been proposed [6, 16, 22] and convergence analysis was performed in [21, 27, 23, 29, 31, 32]. See also [21, 23] for some comments on equivalent interpretations of the randomized Kaczmarz algorithms.

## 2 Optimal row selecting strategy of the Kaczmarz algorithm for solving consistent system of linear equations

In this section, we consider the case that system of linear equations (1) is consistent and  $x$  is a solution. If the  $i$ th row is selected at the  $(k+1)$ th iteration of the Kaczmarz algorithm, i.e.,

$$x_{k+1} = x_k + \frac{b_i - a_i^T x_k}{\|a_i\|_2^2} a_i,$$

then  $x_{k+1}$  can be reformulated as

$$\begin{aligned} x_{k+1} &= x_k + \frac{b_i - a_i^T x_k}{\|a_i\|_2^2} a_i \\ &= x_k + \frac{b_i}{\|a_i\|_2^2} a_i - \frac{x_k^T a_i}{\|a_i\|_2^2} a_i \\ &= x_k + \frac{a_i^T x}{\|a_i\|_2^2} a_i - \frac{a_i^T x_k}{\|a_i\|_2^2} a_i \\ &= x_k + \frac{a_i^T (x - x_k)}{\|a_i\|_2^2} a_i \\ &= x_k + \frac{a_i a_i^T}{\|a_i\|_2^2} (x - x_k). \end{aligned}$$

It follows that

$$\begin{aligned} x - x_{k+1} &= x - x_k - \frac{a_i a_i^T}{\|a_i\|_2^2} (x - x_k) \\ &= (I - \frac{a_i a_i^T}{\|a_i\|_2^2}) (x - x_k) \end{aligned} \quad (7)$$

and thus

$$x_{k+1} - x_k = \frac{a_i a_i^T}{\|a_i\|_2^2} (x - x_k). \quad (8)$$

From (7) and (8), we can see that

$$x - x_{k+1} \perp x_{k+1} - x_k, \quad (9)$$

i.e.,

$$x - x_{k+1} \perp a_i. \quad (10)$$

To this end, let us make the following orthogonal direct sum decomposition  $x - x_k$ ,

$$x - x_k = \alpha \hat{a}_i + \beta \hat{a}_i^\perp, \quad (11)$$

where  $\hat{a}_i = \frac{a_i}{\|a_i\|_2}$  and  $\hat{a}_i^\perp$  is a normalized vector orthogonal to  $a_i$ . Then coefficients  $\alpha$  and  $\beta$  can be written as

$$\alpha = \|x - x_k\|_2 \cos \theta_{k_i},$$

$$\beta = \|x - x_k\|_2 \sin \theta_{k_i},$$

where  $\theta_{k_i} = \angle(x - x_k, a_i)$  is the angle between the vectors  $(x - x_k)$  and  $a_i$ .

Substituting the above decomposition (11) into (7) gives

$$\begin{aligned} x - x_{k+1} &= (I - \frac{a_i a_i^T}{\|a_i\|_2^2})(\alpha \hat{a}_i + \beta \hat{a}_i^\perp) \\ &= \beta \hat{a}_i^\perp \\ &= \|x - x_k\|_2 \sin \theta_{k_i} \hat{a}_i^\perp. \end{aligned} \tag{12}$$

It follows that

$$\|x - x_{k+1}\|_2 = \|x - x_k\|_2 \cdot |\sin \theta_{k_i}|. \tag{13}$$

From (13) we can see that the error norms generated by the Kaczmarz algorithm are monotonically nonincreasing. Moreover, the convergence can be optimized if  $|\sin \theta_{k_i}|$  is minimized at every iteration, which is equivalent to selecting the row  $a_i$  that solves the optimization problem

$$|\sin \angle(x - x_k, a_i)| = \min_j |\sin \angle(x - x_k, a_j)|.$$

As  $x$  is the unknown solution, the above minimization problems seems unsolvable. However, noting that consistent linear system (1) implies

$$a_j^T x = b_j, \quad j = 1, 2, \dots, m$$

and  $x_k$  is fixed at the  $(k + 1)$ th iteration. The minimization problem can be tackled by maximizing  $|\cos \angle(x - x_k, a_j)|$ , i.e.,

$$\begin{aligned} |\cos \angle(x - x_k, a_j)| &= \frac{|a_j^T(x - x_k)|}{\|x - x_k\|_2 \|a_j\|_2} \\ &= \frac{|b_j - a_j^T x_k|}{\|x - x_k\|_2 \|a_j\|_2} \\ &= \frac{|r_k(j)|}{\|x - x_k\|_2 \|a_j\|_2}, \end{aligned} \tag{14}$$

where  $r_k = b - Ax_k = (r_k(1), r_k(2), \dots, r_k(m))^T$ .

It is clear from (14) that the optimal updating strategy for the Kaczmarz algorithm is to select the row  $\hat{i}$  that satisfies

$$|b_{\hat{i}} - a_{\hat{i}}^T x_k| = \max_j |b_j - a_j^T x_k| = \|b - Ax_k\|_\infty,$$

i.e., the index where  $r_k$  has the largest entry in absolute value. We refer to the above row selection method as the *optimal selecting strategy*, and call the Kaczmarz algorithm with the optimal selecting strategy as the *optimal Kaczmarz algorithm*.

Next, we analyze the convergence of the optimal Kaczmarz algorithm for solving consistent system of linear equations. To simplify the analysis, we introduce two notations

$$\theta_k^{\hat{i}} = \min_j \angle(x_k - x, a_j),$$

and

$$\theta_p^{\hat{i}} = \max_k \theta_k^{\hat{i}},$$

where  $1 \leq \hat{i} \leq m$  and  $1 \leq p \leq k$ .

Based on (13), the  $(k+1)$ th error can be bounded as follows

$$\begin{aligned} \|x - x_{k+1}\|_2 &= \|x - x_k\|_2 \cdot |\sin \theta_k^{\hat{i}}| \\ &= \|x - x_0\|_2 \cdot |\sin \theta_k^{\hat{i}}| \cdot |\sin \theta_{k-1}^{\hat{i}}| \cdots |\sin \theta_0^{\hat{i}}| \\ &\leq \|x - x_0\|_2 \cdot |\sin \theta_p^{\hat{i}}|^k, \end{aligned} \quad (15)$$

where  $1 \leq p \leq k$ .

Notice that

$$0 \leq \sin \theta_p^{\hat{i}} \leq 1,$$

we can theoretically divide the convergence history of the Kaczmarz algorithm into two periods:

- when  $\sin \theta_p^{\hat{i}} < 1$ , the algorithm converge exponentially,
- when  $\sin \theta_p^{\hat{i}} = 1$ , we have

$$\max_j a_j^T (x_p - x) = 0$$

and thus,

$$a_j^T (x_p - x) = 0, \quad j = 1, 2, \dots, m.$$

This implies that  $Ax_p = b$ , i.e.,  $x_p$  solves the system of linear equation (1).

In summary, for solving consistent system of linear equations (1), there exists a theoretical optimal selecting strategy or optimal randomization strategy for Kaczmarz algorithm. With the strategy, the algorithm converges exponentially and will achieve convergence when

$$\max_k \min_{1 \leq j \leq m} \angle(x_k - x, a_j) = \frac{\pi}{2}.$$

### 3 Randomized Kaczmarz algorithm for solving inconsistent system of linear equations

Suppose (1) is a consistent system of linear equations and its right hand side is perturbed with a noise vector  $r$  as follows:

$$Ax \simeq b + r, \quad (16)$$

where (16) can be either consistent or inconsistent. In this section, we give some remarks on the convergence of randomized Kaczmarz algorithm for solving (16), which was investigated by D. Needell [32].

Firstly, we recall the Lemma 2.2 in [32].

**Lemma 1.** Let  $H_i$  be the affine subspaces of  $\mathcal{R}^n$  consisting of the solutions to unperturbed equations,  $H_i = \{x \mid \langle a_i, x \rangle = b_i\}$ . Let  $\tilde{H}_i$  be the solution spaces of the noisy equations,  $\tilde{H}_i = \{x \mid \langle a_i, x \rangle = b_i + r_i\}$ . Then

$$\tilde{H}_i = \{w + \alpha_i a_i \mid w \in H_i\}$$

where  $\alpha_i = \frac{r_i}{\|a_i\|_2^2}$ .

**Remarks:** If the Lemma 1 is used to interpret the Kaczmarz algorithm for solving the perturbed and unperturbed equations, we need to introduce a vector  $a_i^\perp$  in the as the orthogonal complement of the vector  $a_i$ , and write  $\tilde{x}_i \in \tilde{H}_i$  as

$$\tilde{x}_i = x_i + \alpha_i a_i + \beta v_i$$

where  $x_i$  is a solution generated by Kaczmarz algorithm for solving the unperturbed equations, and  $v_i$  is a vector in the orthogonal complement of  $a_i$ .

**Example 1.** Consider the  $2 \times 2$  system of linear equations

$$\begin{cases} x_1 + x_2 = 1, \\ x_1 - x_2 = 1, \end{cases}$$

and the perturbed equations

$$\begin{cases} x_1 + x_2 = 1.5, \\ x_1 - x_2 = 1.5, \end{cases}$$

i.e.,  $A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ ,  $b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$  and  $r = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$ .

Let

$$H_i \doteq \{x \mid \langle a_i, x \rangle = b_i\}$$

and

$$\tilde{H}_i \doteq \{\tilde{x} \mid \langle a_i, \tilde{x} \rangle = b_i + r_i\}.$$

If we use  $x_0 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$  as the same initial guess for the perturbed and unperturbed linear system, then

$$H_1 = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \xi \begin{pmatrix} -1 \\ 1 \end{pmatrix} \mid \xi \in \mathcal{R} \right\}$$

and

$$\tilde{H}_1 = \left\{ \begin{pmatrix} 1.5 \\ 0 \end{pmatrix} + \xi \begin{pmatrix} -1 \\ 1 \end{pmatrix} \mid \xi \in \mathcal{R} \right\}$$

Note that  $a_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ ,  $\|a_1\|_2^2 = 2$  and  $r_1 = \frac{1}{2}$ . We have

$$\begin{pmatrix} 1.5 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \frac{1}{4} \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{1}{4} \begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

i.e.,

$$\tilde{x}_1 = x_1 + \frac{r_1}{\|a_1\|_2^2} a_1 + \frac{r_1}{\|a_1\|_2^2} a_1^\perp.$$

In order to derive the convergence rate of randomized Kaczmarz algorithm for solving the perturbed linear equations (16), we need to make use of the established convergence results [39] for the unperturbed linear system (1), together with the relationship between the approximate solutions generated by the Kaczmarz algorithm [39] for perturbed and unperturbed linear equations. In [32], D. Needell analyzed the convergence rate and error bound of the randomized Kaczmarz algorithm for solving the perturbed linear equations, in which the author take the approximate solution to the perturbed linear equations as the guess for the unperturbed system, which make the derivation process simplified. However, the approximate solutions generated by applying the randomized Kaczmarz algorithm to the perturbed linear system may not converge to the solution of the unperturbed linear system.

In what follows, we will consider the convergence rate of the randomized Kaczmarz algorithm for solving (16) from a different perspective. We try to bound the difference between the solution for the unperturbed linear system (1) and approximate solutions generated by applying the randomized Kaczmarz algorithm to the perturbed linear system.

In the following discussion, we use  $x_k$  and  $\tilde{x}_k$  to denote the approximate solutions generated by applying the randomized Kaczmarz algorithm to (1) and (16), respectively. The recursive formulas can be written as

$$x_{k+1} = x_k + \frac{b_{i_k} - x_k^T a_{i_k}}{\|a_{i_k}\|_2^2} a_{i_k} \quad (17)$$

and

$$\tilde{x}_{k+1} = \tilde{x}_k + \frac{b_{i_k} + r_{i_k} - \tilde{x}_k^T a_{i_k}}{\|a_{i_k}\|_2^2} a_{i_k}, \quad (18)$$

where the subscript  $i_k \in \{1, 2, \dots, m\}$  is used to denote that the  $i_k$ th row is selected with probability  $\frac{\|a_{i_k}\|_2^2}{\|A\|_F^2}$  at the  $k$ th iteration.

Suppose the same initial guess  $x_0 = \tilde{x}_0$  is used as the starting vector. Then

$$\tilde{x}_1 = \tilde{x}_0 + \frac{b_{i_0} + r_{i_0} - \tilde{x}_0^T a_{i_0}}{\|a_{i_0}\|_2^2} a_{i_0}$$

and potentially

$$x_1 = x_0 + \frac{b_{i_0} - x_0^T a_{i_0}}{\|a_{i_0}\|_2^2} a_{i_0}.$$

It follows that

$$\tilde{x}_1 = x_1 + \frac{r_{i_0} a_{i_0}}{\|a_{i_0}\|_2^2}. \quad (19)$$

In the next iteration, we have

$$\begin{aligned}
 \tilde{x}_2 &= \tilde{x}_1 + \frac{b_{i_1} + r_{i_1} - \tilde{x}_1^T a_{i_1}}{\|a_{i_1}\|_2^2} a_{i_1} \\
 &= x_1 + \frac{r_{i_0} a_{i_0}}{\|a_{i_0}\|_2^2} + \frac{b_{i_1} - (x_1 + \frac{r_{i_0} a_{i_0}}{\|a_{i_0}\|_2^2})^T a_{i_1}}{\|a_{i_1}\|_2^2} a_{i_1} + \frac{r_{i_1} a_{i_1}}{\|a_{r_1}\|_2^2} \\
 &= x_1 + \underbrace{\frac{b_{i_1} - x_1^T a_{i_1}}{\|a_{i_1}\|_2^2} a_{i_1}}_{x_2} + \underbrace{\frac{r_{i_1} a_{i_1}}{\|a_{i_1}\|_2^2} + (I - \frac{a_{i_1} a_{i_1}^T}{\|a_{i_1}\|_2^2}) \frac{r_{i_0} a_{i_0}}{\|a_{i_0}\|_2^2}}_{a_{i_1}^\perp} \\
 &= x_2 + \frac{r_{i_1} a_{i_1}}{\|a_{i_1}\|_2^2} + v_{i_1}
 \end{aligned}$$

where  $v_{i_1} = (I - \frac{a_{i_1} a_{i_1}^T}{\|a_{i_1}\|_2^2}) \frac{r_{i_0} a_{i_0}}{\|a_{i_0}\|_2^2} \in \text{span}\{a_{i_1}\}^\perp$  with  $\|v_{i_1}\|_2 = \frac{|r_{i_0}|}{\|a_{i_0}\|_2}$ .

Continue the above process, we have

$$\tilde{x}_k = x_k + \frac{r_{i_{k-1}}}{\|a_{i_{k-1}}\|_2^2} a_{i_{k-1}} + \sum_{j=1}^{k-2} v_{i_j}, \tag{20}$$

where  $v_{i_j} = (I - \frac{a_{i_j} a_{i_j}^T}{\|a_{i_j}\|_2^2}) \frac{r_{i_{j-1}} a_{i_{j-1}}}{\|a_{i_{j-1}}\|_2^2} \in \text{span}\{a_{i_j}\}^\perp$  and  $\|v_{i_j}\|_2 = \frac{|r_{i_{j-1}}|}{\|a_{i_{j-1}}\|_2}$ .

Subtracting  $x$  on both sides of (20) gives

$$\tilde{x}_k - x = x_k - x + \frac{r_{i_{k-1}} a_{i_{k-1}}}{\|a_{i_{k-1}}\|_2^2} + \sum_{j=1}^{k-2} v_{i_j}. \tag{21}$$

Based on Jensen's inequality and (6), we have

$$\mathbb{E}\|x_k - x\|_2 \leq (1 - \frac{1}{\kappa(A)^2})^{\frac{k}{2}} \|x_0 - x\|_2, \tag{22}$$

where  $\kappa(A) = \|A\|_F \|A^{-1}\|_2$ , with  $\|A^{-1}\|_2 = \inf\{M : M \|Ax\|_2 \geq \|x\|_2\}$ .

Taking norm on both sides of (21) and using triangle inequality, we have

$$\begin{aligned}
 \mathbb{E}(\|\tilde{x}_k - x\|_2) &\leq \mathbb{E}(\|x_k - x\|_2) + \mathbb{E}\|\frac{r_{i_{k-1}} a_{i_{k-1}}}{\|a_{i_{k-1}}\|_2^2}\|_2 + \sum_{j=1}^{k-2} \|v_{i_j}\|_2 \\
 &\leq (1 - \frac{1}{\kappa(A)^2})^{\frac{k}{2}} \|x_0 - x\|_2 + \sum_{j=1}^{k-1} \|v_{i_j}\|_2 \\
 &= (1 - \frac{1}{\kappa(A)^2})^{\frac{k}{2}} \|x_0 - x\|_2 + \sum_{j=1}^{k-1} \frac{|r_{i_j}|}{\|a_{i_j}\|_2} \\
 &= (1 - \frac{1}{\kappa(A)^2})^{\frac{k}{2}} \|x_0 - x\|_2 + (k-1)\gamma
 \end{aligned}$$

where  $\gamma = \max_{1 \leq i \leq m} \frac{|r_i|}{\|a_i\|_2}$ .

In conclusion, we have derived the following theorem.

**Theorem 1.** *Let  $A$  be a matrix full column rank and assume the system  $Ax = b$  is consistent. Let  $\tilde{x}_k$  be the  $k$ th iterate of the noisy randomized Kaczmarz method run with  $Ax \simeq b + r$ , and let  $a_1, \dots, a_m$  denote the rows of  $A$ . Then we have*

$$\mathbb{E}\|\tilde{x}_k - x\|_2 \leq \left(1 - \frac{1}{\kappa(A)^2}\right)^{\frac{k}{2}}\|x_0 - x\|_2 + (k-1)\gamma,$$

where  $\kappa(A) = \|A\|_F \|A^{-1}\|_2$  and  $\gamma = \max_{1 \leq i \leq m} \frac{|r_i|}{\|a_i\|_2}$ .

## 4 Conclusions

In this paper, we provide a new look at the Kaczmarz algorithm for solving system of linear equations. The optimal row selecting strategy of the Kaczmarz algorithm for solving consistent system of linear equations is derived. The convergence of the randomized Kaczmarz algorithm for solving perturbed system of linear equations is analyzed and a new bound of the convergence rate is obtained from a new perspective.

## References

1. A. Agaskar, C. Wang, and Y. M. Lu, *Randomized Kaczmarz algorithms: Exact MSE analysis and optimal sampling probabilities*, Proceedings of the 2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Atlanta, GA, December, 3-5 (2014), 389-393.
2. A. A. Ivanov and A.I. Zhdanov, *Kaczmarz algorithm for Tikhonov regularization problem*, Appl. Math. E-Notes, 13 (2013), 270-276.
3. A. A. Ivanov and A. I. Zhdanov, *The block Kaczmarz algorithm based on solving linear systems with arrowhead matrices*, Appl. Math. E-Notes., 17 (2017), 142-156.
4. Z.-Z. Bai and X.-G. Liu, *On the Meany inequality with applications to convergence analysis of several row-action methods*, Numer. Math., 124 (2013), 215-236.
5. Z.-Z. Bai and W.-T. Wu, *On greedy randomized Kaczmarz method for solving large sparse linear systems*, SIAM J. Sci. Comput., 40 (2018), A592-A606.
6. Z.-Z. Bai and W.-T. Wu, *On relaxed greedy randomized Kaczmarz methods for solving large sparse linear systems*, Appl. Math. Letters, 83 (2018), 21-26.
7. M. Benzi and C. D. Meyer, *A direct projection method for sparse linear systems*, SIAM J. Sci. Comput., 16 (1995), 1159-1176.
8. C. Brenzinski, *Projection methods for systems of equations*, Elsevier Science B. V., Amsterdam, 1997.
9. M. Benzi and C. D. Meyer, *The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming*, USSR Comput. Math. Math. Physics., 7 (1967), 200-217.
10. J. Cai and Y. Tang, *A new randomized Kaczmarz based kernel canonical correlation analysis algorithm with applications to information retrieval*, Neural Networks., 98 (2018), 178-191.
11. Y. Censor, *Row-action methods for huge and sparse systems and their applications*, SIAM Rev., 23 (1981), 444-466.

12. Y. Censor, P. P. B. Eggermont and D. Gordon, *Strong underrelaxation in Kaczmarz's method for inconsistent systems*, Numer. Math., 41 (1983), 83-92.
13. Y. Censor, G. T. Herman and M. Jiang, *A note on the behaviour of the randomized Kaczmarz algorithm of Strohmer and Vershynin*, J. Fourier Anal. Appl., 15 (2009), 431-436.
14. J. Demmel, *The probability that a numerical analysis problem is difficult*, Math. Comp., 50 (1988), 449-480.
15. J. A. De Loera, J. Haddock and D. Needell, *A sampling Kaczmarz-Motzkin algorithm for linear feasibility*, SIAM J. Sci. Comput., 39 (2017), S66-S87.
16. Y. C. Eldar and D. Needell, *Acceleration of randomized Kaczmarz method via the Johnson-Lindenstrauss lemma*, Numer. Algorithms., 58 (2011), 163-177.
17. H. G. Feichtinger, C. Cenkner, M. Mayer, H. Steier and T. Strohmer, *New variants of the POCS method using affine subspaces of finite codimension with applications to irregular sampling*, SPIE: VCIP, (1992), 299-310,
18. A. Galántai, *On the rate of convergence of the alternating projection method in finite dimensional spaces* J. Math. Anal. Appl., 310 (2005), 30-44,
19. R. Gordon, R. Bender and G. T. Herman, *Algebraic Reconstruction Techniques (ART) for threedimensional electron microscopy and x-ray photography*, J. Theor. Biol., 29 (1970), 471-481.
20. R. Gordon, G. T. Herman and S. A. Johnson, *Image reconstruction from projections*, Sci. Am., 233. (1975), 56-71.
21. R. M. Gower and P. Richtárik, *Randomized iterative methods for linear systems*, SIAM J. Matrix Anal. 36 (2015), 1660-1690.
22. M. Hanke and W. Niethammer, *On the acceleration of Kaczmarz's method for inconsistent linear systems*, Lin. Alg. Appl., 130 (1990), 83-98.
23. A. Hefny, D. Needell and Aaditya Ramdas, *Rows vs. Columns: Randomized Kaczmarz or Gauss-Seidel for Ridge Regression*, SIAM J. Sci. Comput., 39 (2016), S528-S542.
24. G. T. Herman, *Image reconstruction from projection, the fundamentals of the computerized tomography*, New York: Academic Press, 1980.
25. G. T. Herman, A. Lent and P. H. Lutz, *Relaxation methods for image reconstruction*, Commun. Assoc. Comput. Mach., 21 (1978), 152-158.
26. G. T. Herman and L. B. Meyer, *Algebraic reconstruction techniques can be made computationally efficient*, IEEE Trans. Medical Imaging., 12 (1993), 600-609.
27. Y.-L. Jiao, B.-T. Jin and X.-L. Lu, *Preasymptotic convergence of randomized Kaczmarz method*, Inverse Problems., 33 (2017), 125012.
28. S. Kaczmarz, *Angenäherte auflösung von systemen linearer gleichungen*, Bull. Acad. Polon. Sci. Lett., A35 (1933), 355-357.
29. L. Leventhal and A. S. Lewis, *Randomized methods for linear constraints: convergence rates and conditioning*, Math. Oper. Res., 35 (2010), 641-654.
30. F. Natterer. *The mathematics of computerized tomography*, SIAM, 2001.
31. A. Ma, D. Needell and A. Ramdas, *Convergence properties of the randomized extended Gauss-Seidel and Kaczmarz methods*, SIAM J. Matrix Anal. Appl., 36 (2015), 1590-1604.
32. D. Needell, *Randomized Kaczmarz solver for noisy linear systems*. BIT., 50 (2010), 395-403.
33. D. Needell, R. Zhao, A. Zouzias, *Randomized block Kaczmarz method with projection for solving least squares*, Linear Algebra. Appl., 484 (2015), 322-343.
34. D. Needell and J. A. Tropp, *Paved with good intentions: Analysis of a randomized block Kaczmarz method*, Lin. Alg. Appl., 441 (2014), 199-221.

35. J. von Neumann, *The geometry of orthogonal spaces*, Volume 2. Front Cover. John Von Neumann. Princeton University Press, 1950. This is a mimeographed lecture notes, first distributed in 1933.
36. 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*, UAI, 2016.
37. M. Schmidt, *Notes on randomized Kaczmarz*, Lecture notes, April 9, 2015.
38. H. A. Schwarz, *Ueber einen Grenz?ubergang durch alternirendes Verfahren*, Vierteljahrsschrift der Naturforschenden Gessellschaft in Zurich 15 (1870), 272-286.
39. T. Strohmer and R. Vershynin, *A randomized Kaczmarz algorithm with exponential convergence*, Journal of Fourier Analysis and Applications., 15 (2009), 262-278.
40. T. Strohmer, *Comments on the randomized Kaczmarz method*, unpublished manuscript, 2009.
41. K. Tanabe, *Projection method for solving a singular system of linear equations and its applications*, Numer. Math., 17 (1971), 203-214.
42. H. Trussell and M. Civanlar, *Signal deconvolution by projection onto convex sets*, Acoustics, Speech, and Signal Processing, IEEE International Conference on ICAS-SP'84., 9 (1984), 496-499.
43. D. C. Youla and H. Webb, *Image restoration by the method of convex projections: part 1 theory.*, IEEE Trans Med Imaging., 1 (1982), 81-94.
44. D. C. Youla and H. Webb, *Image restoration by the method of convex projections: part 2 applications and numerical results.*, IEEE Trans Med Imaging., 1 (1982), 95-101.
45. A. Zouzias and N. M. Freris, *Randomized extended Kaczmarz for solving least squares*, SIAM J. Matrix Anal., 34 (2013), 773-793.