AppliedMath

MDPI

Article

Kaczmarz Anomaly in Tomography Problems

Achiya Dax

Hydrological Service, P.O. Box 36118, Jerusalem 91360, Israel; dax20@water.gov.il

Abstract: The Kaczmarz method is an important tool for solving large sparse linear systems that arise in computerized tomography. The Kaczmarz anomaly phenomenon has been observed recently when solving certain types of random systems. This raises the question of whether a similar anomaly occurs in tomography problems. The aim of the paper is to answer this question, to examine the extent of the phenomenon and to explain its reasons. Another tested issue is the ability of random row shuffles to sharpen the anomaly and to accelerate the rate of convergence. The results add important insight into the nature of the Kaczmarz method.

Keywords: the smallest singular value anomaly; the condition number anomaly; Kaczmarz anomaly; tomography problems; random rows shuffles

check for updates

Citation: Dax, A. Kaczmarz Anomaly in Tomography Problems. *AppliedMath* **2022**, *2*, 196–211. https://doi.org/10.3390/appliedmath2020012

Received: 20 December 2021
Accepted: 14 April 2022
Published: 25 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

The Kaczmarz algorithm is an iterative method for solving large sparse linear systems of the form

$$A\mathbf{x} = \mathbf{b}, \tag{1}$$

where $A \in \mathbb{R}^{m\times n}$, $\mathbf{b} = (b_1, \ldots, b_m)^T \in \mathbb{R}^m$ and $\mathbf{x} \in \mathbb{R}^n$ denote the vectors of unknowns. Let the rows of A be denoted by the row vectors $\mathbf{a}_i^T$, $i = 1, \ldots, m$. Then, an equivalent way to write (1) is

$$\mathbf{a}_i^T\mathbf{x} = b_i \ \text{ for } \ i = 1, \ldots, m. \tag{2}$$

The idea of the Kaczmarz method is to handle one equation at a time. Let $\|\mathbf{a}_i\|_2 = (\mathbf{a}_i^T\mathbf{a}_i)^{1/2}$ denote the Euclidean vector norm and let w be a preassigned relaxation parameter that satisfies $0 < w < 2$. Then, the kth iteration of the Kaczmarz algorithm, $k = 1, 2, \ldots$, is composed of m steps. The ith step of the kth iteration, $i = 1, \ldots, m$, starts with the vector $\mathbf{x}_{k,i-1}$ and ends with the vector

$$\mathbf{x}_{k,i} = \mathbf{x}_{k,i-1} + w\mathbf{a}_i(b_i - \mathbf{a}_i^T\mathbf{x}_{k,i-1})/\|\mathbf{a}_i\|_2^2. \tag{3}$$

That is, the ith step uses only the ith equation. Observe that for $w = 1$, the point $\mathbf{x}_{k,i}$ is the projection of $\mathbf{x}_{k,i-1}$ on the hyperplane $\{\mathbf{x}|\mathbf{a}_i^T\mathbf{x} = b_i\}$. Note also that the kth iteration, $k = 1, 2, \ldots$, starts with the vector

$$\mathbf{x}_{k-1} = \mathbf{x}_{k-1,m} = \mathbf{x}_{k,0}, \tag{4}$$

and ends with

$$\mathbf{x}_k = \mathbf{x}_{k,m} = \mathbf{x}_{k+1,0}. \tag{5}$$

The starting point is denoted as $\mathbf{x}_0 = \mathbf{x}_{1,0}$.

The fact that the algorithm uses one row at a time makes it a popular tool for solving large sparse linear systems that arise in important applications, such as computerized tomography or digital signal processing. The literature on the Kaczmarz method is vast and covers various issues. See [1–26] and the references therein. Results on the theory behind the method and its rate of convergence can be found in [1,6,14,16,19,21,23,25], while efficient implementations and applications are considered in [3,4,10,12–15,19]. In

addition, there are several variants of the basic iteration, such as block versions and parallel computing techniques [4,14,19]. In particular, recently, there has been growing interest in Randomized Kaczmarz methods [5,20,22,24]. Some of the variants are easily described by restating the algorithm with the restriction that each iteration regards only one equation, which is chosen according to some rule. In the basic iteration (3), the rows are chosen in a sequential "cyclic" manner. In "Greedy" Kaczmarz, we select an equation that has a maximal residual, while in "Randomized" Kaczmarz, the equation's index i is selected at random, with probability proportional to $\|\mathbf{a}_i\|_2^2$, e.g., [5,20–22,24]. However, in the coming discussions, the terms "Kaczmarz method" and "Kaczmarz iteration" refer to (3). The original algorithm of Kaczmarz [16] is obtained from this framework when $m = n$ and $w = 1$.

The use of the relaxation parameter, w, is motivated by the close relation with the SOR method for solving the linear system

$$AA^T\mathbf{y} = \mathbf{b}, \tag{6}$$

where here, $\mathbf{y} \in \mathbb{R}^m$ denotes the vector of unknowns. Let $\mathbf{y}_k$ denote the current solution at the end of the kth iteration of this method, $k = 1, 2, \ldots$. Then, the following observation is well known, e.g., [1,6]. If the starting points satisfy

$$\mathbf{x}_0 = A^T\mathbf{y}_0, \tag{7}$$

then the equality

$$\mathbf{x}_k = A^T\mathbf{y}_k \tag{8}$$

holds for all k. This relation implies that several convergence properties of Kaczmarz method are inherited from those of the SOR method.

Let the linear system

$$\hat{\mathbf{a}}_i^T\mathbf{x} = \hat{b}_i, \quad i = 1, \ldots, m, \tag{9}$$

be obtained from (2) by "normalizing" the rows of A to have unit length. That is,

$$\hat{\mathbf{a}}_i = \mathbf{a}_i / \|\mathbf{a}_i\|_2 \quad \text{and} \quad \hat{b}_i = b_i / \|\mathbf{a}_i\|_2. \tag{10}$$

Then, it is easy to verify that applying the Kaczmarz method to solve (9) yields the same sequence as (3). Hence, when studying the convergence properties of Kaczmarz method, there is no loss of generality in assuming that the rows of A have unit length. (A similar remark applies to the related SOR method).

In this paper, we consider an interesting feature of Kaczmarz method. To illustrate this property, it is assumed that m is considerably larger than n. Let A_i and $\mathbf{b}_i$ be composed from the first i rows of A and $\mathbf{b}$, respectively. That is

$$A_i = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_i]^T \in \mathbb{R}^{i\times n} \tag{11}$$

and

$$\mathbf{b}_i = (b_1, b_2, \ldots, b_i)^T \in \mathbb{R}^i. \tag{12}$$

Then, the new feature is revealed when using Kaczmarz method for solving the linear systems

$$A_i\mathbf{x} = \mathbf{b}_i, \quad i = 1, 2, \ldots, m, \tag{13}$$

and watching how the number of rows, i, affects the rate of convergence. If A is an arbitrary matrix, we are not expecting a certain behavior. However, as we shall see, in some cases, the number of rows has a dramatic effect on the rate of convergence. Assume first that i is considerably smaller than n. In this case, the Kaczmarz method has a fast rate of convergence. Yet as i increases toward n, the rate of convergence slows down. That is, the more equations we have, the more iterations are needed to solve the system. In particular, as i approaches n, there is a dramatic increase in the number of iterations. The closer i and

n are, the slower the convergence. However, as i passes n, the situation is reversed. From now on, the more equations we have, the fewer iterations are needed. Finally, when i is considerably larger than n, the method returns to enjoy rapid convergence.

We call this behavior the **Kaczmarz anomaly**. One aim of this paper is to examine the presence of this phenomenon when solving tomography problems. The first report on the Kaczmarz anomaly appeared in [7], but it remained almost unnoticed. Recently, we have shown in [8] that it is likely to occur whenever the rows' directions $\hat{\mathbf{a}}_i = \mathbf{a}_i / \|\mathbf{a}_i\|_2$ scatter randomly over some portion of the unit sphere. This suggests that a random shuffle of the rows may improve their randomality and strengthen the anomaly. A second aim of the paper is to examine this idea.

The plan of the paper is as follows. The first two sections provide theoretical background that reveals the reasons behind the anomaly phenomenon. Section 2 overviews the condition number anomaly phenomenon, while Section 3 explains how the condition number of A affects the rate of convergence. Combining these features yields the Kaczmarz anomaly. The background is based on recent results by this author; see [8,9]. The use of random shuffles and related techniques is considered in Section 4. The paper ends with numerical experiments that illustrate the anomaly phenomena.

2. The Smallest Singular Value Anomaly and the Condition Number Anomaly

Let α_i denote the largest singular value of the matrix $A_i, i = 1, \ldots, m$. Then, α_i^2 is the largest eigenvalue of the cross-product matrix $A_i^T A_i$, and there exists a unit eigenvector, $\mathbf{u}_i$, that satisfies

$$A_i^T A_i \mathbf{u}_i = \alpha_i^2 \mathbf{u}_i \tag{14}$$

and

$$\alpha_i^2 = \mathbf{u}_i^T A_i^T A_i \mathbf{u}_i = \max \ \{\mathbf{x}^T A_i^T A_i \mathbf{x} \,|\, \mathbf{x} \in \mathbb{R}^n \ \text{ and } \ \|\mathbf{x}\|_2 = 1\}. \tag{15}$$

The first assertion characterizes the ascending behavior of the sequence $\alpha_1, \ldots, \alpha_m$. (For proofs of the coming theorems, see [8]).

Theorem 1. *For $i = 1, \ldots, m-1$, we have the inequalities*

$$\alpha_i^2 \le \alpha_i^2 + (\mathbf{u}_i^T \mathbf{a}_{i+1})^2 \le \alpha_{i+1}^2 \tag{16}$$

and

$$\alpha_{i+1}^2 \le \alpha_i^2 + (\mathbf{u}_{i+1}^T \mathbf{a}_{i+1})^2 \le \alpha_i^2 + \|\mathbf{a}_{i+1}\|_2^2. \tag{17}$$

Next, we explore the behavior of the smallest singular value, which is rather surprising. Let β_i denote the smallest singular value of the matrix A_i, $i = 1, \ldots, m$. Then, as the coming theorem shows, the first part of this sequence, $\beta_1, \ldots, \beta_n$, is descending.

Theorem 2. *Let $\hat{\mathbf{a}}_{i+1} = \mathbf{a}_{i+1} / \|\mathbf{a}_{i+1}\|_2$ be a unit vector in the direction of $\mathbf{a}_{i+1}$, and let the unit vector $\mathbf{v}_i \in \mathbb{R}^n$ be a right singular vector of A_i that corresponds to β_i. Then, for $i = 1, 2, \ldots, n-1$, we have the inequalities*

$$\begin{aligned} \beta_{i+1}^2 &\le \beta_i^2 [1 - (\mathbf{v}_i^T \hat{\mathbf{a}}_{i+1})^2] / [1 + \beta_i^2 (\mathbf{v}_i^T \hat{\mathbf{a}}_{i+1})^2 / \|\mathbf{a}_{i+1}\|_2^2] \\ &\le \beta_i^2 [1 - (\mathbf{v}_i^T \hat{\mathbf{a}}_{i+1})^2] \le \beta_i^2. \end{aligned} \tag{18}$$

The assumption $\beta_i > 0$ is not essential for the proof of Theorem 2, but it enables us to replace (18) with the inequality

$$\beta_{i+1}^2 / \beta_i^2 \le [1 - (\mathbf{v}_i^T \hat{\mathbf{a}}_{i+1})^2] / [1 + \beta_i^2 (\mathbf{v}_i^T \hat{\mathbf{a}}_{i+1})^2 / \|\mathbf{a}_{i+1}\|_2^2]. \tag{19}$$

This exposes the actual reasons that force β_{i+1}^2 to be smaller than β_i^2. One reason is the size of $\mathbf{a}_{i+1}$. We see that the smaller $\|\mathbf{a}_{i+1}\|_2^2$ is, the smaller is β_{i+1}^2. Another important factor is the size of the scalar product $\mathbf{v}_i^T \hat{\mathbf{a}}_{i+1}$. Since both $\mathbf{v}_i$ and $\hat{\mathbf{a}}_{i+1}$ are unit vectors, the

Cauchy–Schwartz inequality implies $(\mathbf{v}_i^T \hat{\mathbf{a}}_{i+1})^2 \leq 1$, and equality occurs if and only if $\hat{\mathbf{a}}_{i+1} = \pm \mathbf{v}_i$. Now, from (19), we see that the larger $(\mathbf{v}_i^T \hat{\mathbf{a}}_{i+1})^2$ is, the smaller is β_{i+1}^2.

In this paper, we concentrate on the behavior of Kaczmarz method, and for this purpose, it is possible to assume that all the rows of A have unit length. This assumption implies that $\beta_1 = 1$ and turns (19) into the form

$$\beta_{i+1}^2/\beta_i^2 \leq [1 - (\mathbf{v}_i^T \hat{\mathbf{a}}_{i+1})^2]/[1 + \beta_i^2 (\mathbf{v}_i^T \hat{\mathbf{a}}_{i+1})^2]. \tag{20}$$

A further simplification is allowed when β_i^2 becomes considerably smaller than one. In this case, the factor $1/[1 + \beta_i^2 (\mathbf{v}_i^T \hat{\mathbf{a}}_{i+1})^2]$ approaches one, and the bound

$$\beta_{i+1}^2/\beta_i^2 \leq 1 - (\mathbf{v}_i^T \hat{\mathbf{a}}_{i+1})^2 \tag{21}$$

is nearly as good as (20).

It is left to see how the second part of the sequence, $\beta_n, \beta_{n+1}, \ldots, \beta_m$, behaves. Below, we will show that this part is ascending. The proof is based on the observation that now, β_i^2 is the smallest eigenvalue of the cross-product matrix $A_i^T A_i$. Hence, for $i = n, n+1, \ldots, m$, there exists a unit eigenvector $\mathbf{v}_i$ such that

$$A_i^T A_i \mathbf{v}_i = \beta_i^2 \mathbf{v}_i \tag{22}$$

and

$$\mathbf{v}_i^T A_i^T A_i \mathbf{v}_i = \beta_i^2 = \min\{\mathbf{x}^T A_i^T A_i \mathbf{x} \,|\, \mathbf{x} \in \mathbb{R}^n \text{ and } \|\mathbf{x}\|_2 = 1\}. \tag{23}$$

Theorem 3. *For $i = n, n+1, \ldots, m-1$, we have the inequalities*

$$\beta_i^2 \leq \beta_i^2 + (\mathbf{v}_{i+1}^T \mathbf{a}_{i+1})^2 \leq \beta_{i+1}^2. \tag{24}$$

Assume for a moment that $\beta_i > 0$, which enables us to rewrite (24) in the form

$$\beta_{i+1}^2/\beta_i^2 \geq 1 + (\mathbf{v}_{i+1}^T \mathbf{a}_{i+1})^2/\beta_i^2. \tag{25}$$

Assume further that the rows of the matrix have unit length and random directions. Then, a small β_i^2 implies a large increase ratio, while a large β_i^2 means a slow increase. Consequently, when i is close to n, we expect a fast increase, but as i moves away from n, the rate of increase is likely to slow down.

Combining the results of Theorems 2 and 3 shows that the sequence $\beta_1, \ldots, \beta_n$ is decreasing, the sequence $\beta_n, \ldots, \beta_m$ is increasing, and β_n is the smallest number in the whole sequence. Moreover, in some cases, β_n can be considerably smaller than its neighbors. This behavior is called the **smallest singular value anomaly**.

Let k_i denote the condition number of A_i, $i = 1, \ldots, m$. In the rest of this section, we assume for simplicity that $\beta_i > 0$ for $i = 1, \ldots, m$, and that

$$k_i = \alpha_i/\beta_i. \tag{26}$$

(The discussion of the case when $\beta_i = 0$ is deferred to the next section. In this case, β_i is redefined as the smallest nonzero singular value of A_i). We have seen that the sequence $\alpha_1, \ldots, \alpha_n$ is ascending, while the sequence $\beta_1, \ldots, \beta_n$ is descending. This proves the following conclusion.

Theorem 4. *The sequence $k_1, \ldots, k_n$ is ascending. That is,*

$$k_i \leq k_{i+1} \text{ for } i = 1, \ldots, n-1. \tag{27}$$

The behavior of the sequence $k_n, k_{n+1}, \ldots, k_m$ is not that straightforward. We know that the sequences $\alpha_n, \alpha_{n+1}, \ldots, \alpha_m$ and $\beta_n, \beta_{n+1}, \ldots, \beta_m$ are ascending, but this does not provide decisive information. Indeed, for $i \geq n$, one can find examples in which $k_{i+1} < k_i$ as well as examples with $k_{i+1} > k_i$, e.g., [8]. The **condition number anomaly** occurs when the sequence $k_n, \ldots, k_m$ is descending. That is, when

$$k_i \geq k_{i+1} \text{ for } \quad i = n, \ldots, m-1. \tag{28}$$

The reasons behind this behavior lie in the following observations.

Theorem 5. *Let* $\mathbf{u}_{i+1}$ *be as in Theorem 1, let* $\mathbf{v}_{i+1}$ *be as in Theorem 3, and consider the terms*

$$\eta_i^2 = (\mathbf{u}_{i+1}^T \mathbf{a}_{i+1})^2 / \alpha_i^2 \tag{29}$$

and

$$\nu_i^2 = (\mathbf{v}_{i+1}^T \mathbf{a}_{i+1})^2 / \beta_i^2. \tag{30}$$

Then, for $i = n, \ldots, m-1$,

$$k_{i+1}^2 \leq k_i^2 (1 + \eta_i^2) / (1 + \nu_i^2). \tag{31}$$

Proof. From (17), we see that

$$\alpha_{i+1}^2 \leq \alpha_i^2 + (\mathbf{u}_{i+1}^T \mathbf{a}_{i+1})^2 = \alpha_i^2 (1 + \eta_i^2) \tag{32}$$

while Theorem 3 gives

$$\beta_{i+1}^2 \geq \beta_i^2 + (\mathbf{v}_{i+1}^T \mathbf{a}_{i+1})^2 = \beta_i^2 (1 + \nu_i^2). \tag{33}$$

Hence, combining these inequalities yields (31). □

Corollary 1. *The inequality*

$$\nu_i^2 \geq \eta_i^2 \tag{34}$$

implies

$$k_i \geq k_{i+1}. \tag{35}$$

The last corollary is a key observation that indicates at which situations the condition number anomaly is likely to occur. Assume for example that the direction of $\mathbf{a}_{i+1}$ is chosen in some random way. Then, the scalar product terms $(\mathbf{u}_{i+1}^T \mathbf{a}_{i+1})^2$ and $(\mathbf{v}_{i+1}^T \mathbf{a}_{i+1})^2$ are likely to be about the same size. However, since β_i^2 is (considerably) smaller than α_i^2, the term $(\mathbf{v}_{i+1}^T \beta_{i+1})^2 / \beta_i^2$ is expected to be larger than $(\mathbf{u}_{i+1}^T \mathbf{a}_{i+1})^2 / \alpha_i^2$, which implies $k_i \geq k_{i+1}$. In other words, the condition number anomaly is likely to occur whenever the rows' directions scatter in some random way. This conclusion means that the phenomenon is shared by a wide range of matrices. See Tables 1–6 and the examples in [8].

3. The Rate of Convergence of the Kaczmarz Method

We have seen that the Kaczmarz method for solving (1) is closely related to the SOR method for solving the linear system

$$G\mathbf{y} = \mathbf{b}, \tag{36}$$

where

$$G = AA^T \in \mathbb{R}^{m \times m}. \tag{37}$$

Below, we will use this relation to obtain the iteration matrix of Kaczmarz method (3). As before, it is allowed to assume that the rows of A have unit length. That is

$$\|\mathbf{a}_i\|_2 = 1 \text{ for } \quad i = 1, \ldots, m. \tag{38}$$

This assumption implies that G has the form

$$G = I - L - L^T, \tag{39}$$

where I denotes the identity matrix, and L is a strictly lower triangular matrix. The SOR iteration splits G in the form

$$G = B_w - C_w, \tag{40}$$

where

$$B_w = (I - wL)/w \tag{41}$$

and

$$C_w = [(1 - w)I + wL^T]/w. \tag{42}$$

As before, w is a preassigned relaxation parameter that satisfies $0 < w < 2$. The kth SOR iteration, $k = 1, 2, \ldots$, starts with $\mathbf{y}_{k-1}$ and ends with $\mathbf{y}_k$, which is computed by solving the linear system

$$B_w \mathbf{y} = C_w \mathbf{y}_{k-1} + \mathbf{b}. \tag{43}$$

In other words, $\mathbf{y}_k$ is obtained from $\mathbf{y}_{k-1}$ by the rule

$$\mathbf{y}_k = H_w \mathbf{y}_{k-1} + \mathbf{d}_w, \tag{44}$$

where

$$H_w = B_w^{-1} C_w \tag{45}$$

is the related **iteration matrix**, and

$$\mathbf{d}_w = B_w^{-1} \mathbf{b}. \tag{46}$$

Observe that (40) enables us to express H_w in the form

$$H_w = I - B_w^{-1} G = I - B_w^{-1} A A^T. \tag{47}$$

Multiplying (44) by A^T gives

$$\mathbf{x}_k = A^T H_w \mathbf{y}_{k-1} + A^T \mathbf{d}_w, \tag{48}$$

while substituting $I - B_w^{-1} A A^T$ instead of H_w shows that

$$\mathbf{x}_k = (I - A^T B_w^{-1} A)\mathbf{x}_{k-1} + A^T \mathbf{d}_w. \tag{49}$$

This means that the **iteration matrix of the Kaczmarz method** has the form

$$F_w = I - A^T B_w^{-1} A. \tag{50}$$

Note that H_w is an $m \times m$ matrix, while F_w is an $n \times n$ matrix. However, as shown in [9], these matrices share the nonzero eigenvalues.

The theory of the Kaczmarz method tells us that the sequence $\mathbf{x}_k, k = 1, 2, \ldots$, converges for any choice of $\mathbf{b}$ and $\mathbf{x}_0$, e.g., [3,6,9,10,14,19,22,25]. Moreover, let $\hat{\mathbf{x}}$ denote the limit point of this sequence; then, the error vectors $\mathbf{x}_k - \hat{\mathbf{x}}$ satisfy

$$\mathbf{x}_k - \hat{\mathbf{x}} = (F_w)^k (\mathbf{x}_0 - \hat{\mathbf{x}}). \tag{51}$$

This shows that the rate of convergence depends on $\rho(F_w)$, the spectral radius of F_w. The smaller $\rho(F_w)$ is, the faster the convergence. It is interesting, therefore, to see which properties of A make $\rho(F_w)$ small. One answer is given by the following bound, which has recently been derived in [9]. (A second answer is given in the next section, in which we consider the effect of rows shuffling).

Let α denote the largest singular value of A, let $\beta > 0$ denote the smallest nonzero singular value of A, and let $k = \alpha/\beta$ denote the related condition number of A. Let ω_{opt} denote the optimal relaxation parameter of the Kaczmarz method. That is,

$$\rho(F_{w_{opt}}) = \min\left\{\rho(F_w)\,\middle|\, 0 < w < 2\right\}. \tag{52}$$

Then, it is proved in [9] that

$$\rho(F_{w_{opt}})^2 \leq 1 - 1/(k^2 c), \tag{53}$$

where c is a constant,

$$c = 1 + \log_2(2m). \tag{54}$$

The bound is not tight, and the actual rate of convergence (even for $w = 1$) is often faster than the implied rate. Recall that in many practical problems, w_{opt} is not known in advance, and its value is computed by repeated experiments; see Section 5. Yet the main consequence from this bound is that a small condition number forces fast convergence. Conversely, for large k, the bound tends to 1, which allows slow convergence. Indeed, as explained in [9], the existence of small nonzero singular values invites a slow rate of convergence.

The relation between the condition number and the rate of convergence suggests that the Kaczmarz anomaly phenomenon is caused by the condition number anomaly. We have seen that the last phenomenon is expected to occur whenever the rows' directions scatter randomly. This raises the question of whether tomography problems possess these properties. The next sections attempt to answer this question.

4. From Random Shuffles to Optimal Ordering

The Kaczmarz anomaly phenomenon is observed by watching how the number of rows affects the rate of convergence. Another property that affects the rate of convergence is row ordering. The initial ordering in tomography problems is often rather poor, in the sense that it yields a slow rate of convergence. A typical tomography matrix is composed from several blocks of rows, where each block is generated by one "view". The views (and the blocks) are ordered according to the size of the view angle, which is the natural geometric order, e.g., [15], p. 602. Yet this natural order minimizes the angle between adjacent views, which is a property that retards convergence. A possible remedy for this difficulty is to apply a random shuffle of rows before starting the Kaczmarz process. The shuffle is aimed at achieving a faster rate of convergence (see below). Yet, at the same time, it improves the randomality of the rows' directions, which sharpens the anomaly phenomena.

The term "random shuffle" means that the rows of the linear system (1) are reordered by applying a random permutation. This converts (1) into the form

$$\mathbf{P}A\mathbf{x} = \mathbf{P}\mathbf{b}, \tag{55}$$

where $\mathbf{P}$ is a random permutation matrix. To simplify the coming discussions and experiments, we assume that the random permutation is chosen by MATLAB's command "randperm (m)", and the matrix $\mathbf{P}A$ is generated by the command "shuffle (A)", which uses the "randperm (m)" command.

The reordering of the rows is expected to change the iteration matrix of the Kaczmarz method as well as its rate of convergence. It is easy to verify this assertion by using the relation with the SOR method for solving the system

$$\mathbf{P}AA^T\mathbf{P}^T\mathbf{y} = \mathbf{P}\mathbf{b}. \tag{56}$$

Now, it is easy to see that the SOR iteration for solving (56) differs from that of (6). Indeed, the observation that the reordering of rows changes the rate of convergence of the SOR method is not new. See, for example, [27] and the references therein.

As mentioned above, it has been observed by several authors that when solving tomography problems, a random row shuffle may improve the rate of convergence of

the Kaczmarz method. See, for example, [14,15,19,21,22,24] and the references therein. A possible explanation of this phenomenon comes from geometric interpretation of the basic step (3) when $i > 1$ and $w = 1$. Let $0 \leq \theta_i \leq \pi/2$ denote the angle between $\mathbf{a}_{i-1}$ and $\mathbf{a}_i$. Then, in two-dimensional space, the distance to the solution point is reduced by the factor $\cos\theta_i$. That is, a small angle yields a small reduction while a large angle implies a large reduction. When moving to larger dimensions, the situation is not that simple, but it is still true that a small θ_i forces a small step toward the solution, while a large θ_i allows larger steps. (Recall that a large θ_i means that $\mathbf{a}_{i-1}$ and $\mathbf{a}_i$ are nearly orthogonal). These considerations suggest that a random shuffle may improve the rate of convergence if it improves orthogonality between adjacent rows.

Similar arguments have motivated Herman and Meyer [15] to propose an optimal ordering of rows that takes advantage of the special structure of tomography problems to maximize orthogonality between adjacent rows. Further optimal ordering schemes that follow this approach are described in [11,17,18].

The observation that a random ordering of rows may improve the rate of convergence has motivated the **Randomized Kaczmarz** algorithm of Strohmer and Vershynin [24]. In this algorithm, the basic step treats one equation whose index is selected at random with probability proportional to $\|\mathbf{a}_i\|_2^2$. Thus, when all the rows have unit length, all the indices have equal probability. (In our experiments, the row index is obtained by the "randi(m)" command.)

The use of a random shuffle has recently been considered by Oswald and Zhou [21,22], who proposed an improved randomized method, the **Shuffled Kaczmarz** algorithm. In this method, each iteration is preceded by a random shuffle of the rows. This formulation has two advantages. First, as in the Kaczmarz method, each iteration treats all the equations. Second, since all the shuffled matrices have the same singular values as A, the bound on the rate of convergence is the same as in the Kaczmarz method.

It is interesting to compare the above randomized methods with the **Initial Shuffle** method, which uses one random shuffle before starting the Kaczmarz algorithm. Both approaches share the same motivation. If the given system has bad ordering, then a random shuffle is likely to provide a better ordering. Furthermore, basically, we are not expecting large differences in the quality of the generated random shuffles. Hence, in practice, the initial shuffle method is likely to run at the same speed as the randomized Kaczmarz methods. Yet, since it uses only one shuffle, there is a tiny probability to obtain bad ordering, while randomized algorithms avoid this possibility.

5. Numerical Experiments

The experiments examine the behavior of the Kaczmarz method (3) when solving tomography test problems. The test problems are generated by using MATLAB's functions from "AIR tools", which is a MATLAB package of algebraic reconstruction iterative methods prepared by P.C. Hansen and others [12,13]. The test problems imitate the scanning of an $N \times N$ array of square cells. This generates a linear system with $n = N^2$ unknowns. (The unknowns present the densities of the cells while equations describe rays.) **In our experiments, $N = 20$, and all the test matrices have $n = N^2 = 400$ columns**. The number of rows depends on the nature of the scanning device and the specific details of the experiment.

The **Parallel beam** tomography problems are generated by using MATLAB's function

$$\text{paralleltomo}\ (N, \theta, p) \tag{57}$$

with $N = 20$, theta = 1:1:180, and $p = \text{round}\ (\sqrt{2}N) = 28$. (The vector theta contains the angles of the views, while p denotes the number of parallel rays for each view.) This results in a linear system with $n = N^2 = 400$ unknowns and $m = 180 \times 28 = 5040$ equations. However, if a ray passes outside the array, it generates a null row. So, we use MATLAB's function

$$rzr(A, \mathbf{b}) \tag{58}$$

to remove zero rows, which yields a linear system with $n = 400$ unknowns and $m = 4340$ rows.

In **Fan beam** tomography, each angle (each view) is related to a "fan" of p rays, and the problem is generated by using the function

$$\text{fanbeamtomo}\ (N) \tag{59}$$

with $N = 20$ and the default values theta = 0:1:359 and $p = \text{round}\ (\sqrt{2}N) = 28$. This builds a linear system with $n = N^2 = 400$ unknowns and $m = 28 \times 360$ = 10,080 equations. Then, after removing zero rows, we remain with $m = 9520$ equations.

In **Seismic** tomography problems, the linear system is generated by applying the function

$$\text{seismictomo}\ (N, s, p) \tag{60}$$

with $N = 20$, $s = 2N$ sources, and $p = 4N$ receivers. This setting builds a linear system with $n = N^2 = 400$ unknowns and $m = s \times p = 3200$ equations. (In this case, there are no zero rows.)

The experiments were carried out as follows. At first, we have generated an $m \times n$ linear system, $A\mathbf{x} = \mathbf{b}$, as described above. Together with A and $\mathbf{b}$, we are given a prescribed solution $\mathbf{x}^* \in \mathbb{R}^n$, which is the one that has been used to build A and $\mathbf{b}$. Then, in the second stage, the rows of A are normalized to have a unit norm. Thus, for $i = 1, \ldots, m$, the ith row of A is redefined as $\mathbf{a}_i = \mathbf{a}_i / \|\mathbf{a}_i\|_2$, while $\mathbf{b}$ is updated as $\mathbf{b} = A\mathbf{x}^*$. Finally, after the normalization, the Kaczmarz method was applied to solve partial systems of the form (13). The starting point in our runs is always $\mathbf{x}_0 = \mathbf{0}$, and the iterative process was terminated after 666 iterations.

The shuffled test problems are obtained by reordering the rows of A, using a random permutation. The actual reordering is carried out by applying MATLAB's function

$$\text{shuffle}\ (A). \tag{61}$$

After the shuffling, the vector $\mathbf{b}$ is redefined as $\mathbf{b} = A\mathbf{x}^*$ where $\mathbf{x}^*$ denotes the known solution. (The shuffling takes place after the normalization but before starting the solution of the partial linear systems.)

The rows in Tables 1–6 describe the use of the Kaczmarz method to solve partial linear systems of the form

$$A_i\mathbf{x} = \mathbf{b}_i. \tag{62}$$

Recall that A_i is a $i \times n$ submatrix of A which is composed from the first i rows of A, and $\mathbf{b}_i = (b_1, \ldots, b_i)^T \in \mathbb{R}^i$ is composed from the first i entries of $\mathbf{b}$. The results for the linear system (62) start with the number of rows, i, and the number of zero singular values of A_i. Then, we provide the values of α_i, β_i, and k_i, as well as the related residual values. As noted in the tables' headlines, α_i is the largest singular value of A_i, β_i is the smallest nonzero singular value of A_i, and $k_i = \alpha_i / b_i$ is the condition number of A_i. The residual values are defined as

$$\|A_i\mathbf{x}_{666} - \mathbf{b}_i\|_2 / \|A_i\mathbf{x}_0 - \mathbf{b}_i\|_2 = \|A_i\mathbf{x}_{666} - \mathbf{b}_i\|_2 / \|\mathbf{b}_i\|_2 \tag{63}$$

where $\|\cdot\|_2$ denotes the Euclidean vector norm and $\mathbf{x}_{666}$ denotes the computed solution after 666 iterations. The Kaczmarz method uses a relaxation parameter, w, and the residual values are given for $w = 1$ and $w = w_{opt}$. The value of w_{opt} was obtained by running the Kaczmarz method with values of w from the set

$$\{0.1, 0.2, \ldots, 1.8, 1.9\} \bigcup \{1.95\} \tag{64}$$

and taking a value of w that yields the smallest residual.

Table 1. Parallel beam tomography.

Number of Rows i	Number of Zero Singular Values	Largest Singular Value α_i	Smallest Nonzero Singular Value β_i	Condition Number k_i	Residual after 666 Iterations $w = 1$	$w = w_{opt}$	w_{opt}
50	6	1.723	6.46×10^{-2}	2.67× 10	1.91×10^{-8}	7.95×10^{-13}	1.4
100	8	2.213	9.77×10^{-3}	2.26×10^{2}	1.49×10^{-4}	1.12×10^{-4}	1.5
200	8	3.133	7.50×10^{-4}	4.17× 10^{3}	3.83×10^{-4}	2.77×10^{-4}	1.2
300	20	3.740	5.18×10^{-4}	7.22× 10^{3}	1.38×10^{-3}	1.27×10^{-3}	1.1
360	27	4.105	2.91×10^{-5}	1.41×10^{5}	2.04×10^{-3}	2.02×10^{-3}	1.1
380	31	4.218	2.67×10^{-5}	1.58×10^{5}	2.37×10^{-3}	2.37×10^{-3}	1.0
400	37	4.307	2.12×10^{-5}	2.03×10^{5}	2.84×10^{-3}	2.84×10^{-3}	1.0
420	22	4.363	5.09×10^{-6}	8.58×10^{5}	3.10×10^{-3}	3.09×10^{-3}	0.9
440	16	4.468	7.07×10^{-5}	6.32×10^{4}	3.13×10^{-3}	3.13×10^{-3}	1.0
500	11	4.774	2.15×10^{-4}	2.22×10^{4}	2.82×10^{-3}	2.82×10^{-3}	1.0
1000	0	6.504	2.47×10^{-3}	2.64× 10^{3}	3.11×10^{-3}	2.83×10^{-3}	0.6
2000	0	9.111	2.29×10^{-2}	3.97×10^{2}	1.48×10^{-3}	1.24×10^{-3}	0.6
3000	0	11.246	6.87×10^{-2}	1.64×10^{2}	6.82×10^{-4}	3.64×10^{-4}	0.2
4000	0	12.898	1.01×10^{-1}	1.28×10^{2}	4.66×10^{-4}	2.14×10^{-4}	0.2
4340	0	13.498	1.15×10^{-1}	1.18×10^{2}	3.76×10^{-4}	1.76×10^{-4}	0.2

Table 2. Parallel beam with initial shuffle.

Number of Rows i	Number of Zero Singular Values	Largest Singular Value α_i	Smallest Nonzero Singular Value β_i	Condition Number k_i	Residual after 666 Iterations $w = 1$	$w = w_{opt}$	w_{opt}
40	1	1.786	8.94×10^{-2}	2.00× 10	5.79×10^{-9}	$2.08 \times 10^{-1}6$	1.5
50	1	1.966	6.61×10^{-2}	2.97× 10	4.15×10^{-6}	$5.31 \times 10^{-1}5$	1.6
60	1	1.999	6.57×10^{-2}	3.04× 10	6.41×10^{-6}	$9.79 \times 10^{-1}5$	1.6
100	1	2.600	6.41×10^{-2}	3.52× 10	1.43×10^{-5}	$1.48 \times 10^{-1}3$	1.6
200	5	3.064	1.31×10^{-2}	2.34×10^{2}	3.14×10^{-4}	8.42×10^{-6}	1.6
300	9	3.694	7.37×10^{-5}	5.01×10^{4}	8.55×10^{-4}	7.55×10^{-4}	1.4
360	15	4.026	1.41×10^{-3}	2.85× 10^{3}	1.64×10^{-3}	1.60×10^{-3}	0.9
380	18	4.119	3.94×10^{-4}	1.05×10^{4}	1.95×10^{-3}	1.81×10^{-3}	0.8
400	22	4.200	4.41×10^{-4}	9.53× 10^{3}	1.72×10^{-3}	1.59×10^{-3}	0.8
420	5	4.304	1.09×10^{-4}	3.96×10^{4}	1.86×10^{-3}	1.80×10^{-3}	0.9
440	0	4.407	6.25×10^{-4}	7.05× 10^{3}	1.73×10^{-3}	1.71×10^{-3}	0.9
500	0	4.678	4.77×10^{-3}	9.80×10^{2}	1.59×10^{-3}	1.58×10^{-3}	1.1
1000	0	6.534	3.86×10^{-2}	1.69×10^{2}	1.18×10^{-4}	1.40×10^{-5}	1.95
2000	0	9.216	7.22×10^{-2}	1.28×10^{2}	7.45×10^{-6}	4.89×10^{-7}	1.8
3000	0	11.243	9.15×10^{-2}	1.23×10^{2}	1.03×10^{-6}	2.00×10^{-8}	1.8
4000	0	12.965	1.10×10^{-1}	1.18×10^{2}	9.63×10^{-8}	$1.07 \times 10^{-1}0$	1.95
4340	0	13.498	1.15×10^{-1}	1.18×10^{2}	4.36×10^{-8}	$9.80 \times 10^{-1}1$	1.8

Table 3. Fan beam tomography.

Number of Rows i	Number of Zero Singular Values	Largest Singular Value α_i	Smallest Nonzero Singular Value β_i	Condition Number k_i	Residual after 666 Iterations $w = 1$	$w = w_{opt}$	w_{opt}
40	3	1.601	8.98×10^{-2}	1.784× 10	$2.29 \times 10^{-1}0$	$6.29 \times 10^{-1}6$	1.4
50	3	1.601	8.98×10^{-2}	1.784× 10	$2.07 \times 10^{-1}0$	$5.64 \times 10^{-1}6$	1.5
100	11	2.275	1.37×10^{-3}	1.658× 10^3	1.51×10^{-5}	6.36×10^{-6}	1.2
200	31	3.165	2.59×10^{-4}	1.220×10^4	3.37×10^{-4}	2.94×10^{-4}	1.2
300	55	3.812	7.03×10^{-4}	5.418× 10^3	1.40×10^{-3}	1.37×10^{-3}	1.1
360	64	4.068	5.17×10^{-4}	7.875× 10^3	1.93×10^{-3}	1.89×10^{-3}	0.9
380	68	4.213	4.70×10^{-4}	8.969× 10^3	2.11×10^{-3}	2.09×10^{-3}	0.9
400	74	4.346	9.89×10^{-5}	4.393×10^4	2.25×10^{-3}	2.24×10^{-3}	0.9
420	61	4.452	8.22×10^{-5}	5.416×10^4	2.77×10^{-3}	2.72×10^{-3}	0.9
440	44	4.531	3.81×10^{-5}	1.190×10^5	2.95×10^{-3}	2.93×10^{-3}	0.9
500	12	4.810	3.13×10^{-5}	1.537×10^5	3.05×10^{-3}	2.85×10^{-3}	0.7
1000	0	6.640	3.60×10^{-3}	1.845× 10^3	3.12×10^{-3}	2.84×10^{-3}	0.5
2000	0	9.317	1.37×10^{-2}	6.792×10^2	1.79×10^{-3}	1.68×10^{-3}	0.6
3000	0	11.435	4.13×10^{-2}	2.767×10^2	3.10×10^{-4}	2.16×10^{-4}	0.4
4000	0	13.141	5.93×10^{-2}	2.217×10^2	3.13×10^{-4}	2.28×10^{-4}	0.4
8000	0	18.460	8.99×10^{-2}	2.053×10^2	1.90×10^{-4}	1.32×10^{-4}	0.2
9520	0	20.106	9.76×10^{-2}	2.06×10^2	1.05×10^{-4}	7.77×10^{-5}	0.4

Table 4. Fan beam with initial shuffle.

Number of Rows i	Number of Zero Singular Values	Largest Singular Value α_i	Smallest Nonzero Singular Value β_i	Condition Number k_i	Residual after 666 Iterations $w = 1$	$w = w_{opt}$	w_{opt}
40	0	1.589	2.63×10^{-1}	6.04	1.08×10^{-16}	4.90×10^{-17}	0.7
50	0	1.764	2.53×10^{-1}	6.96	1.35×10^{-16}	5.91×10^{-17}	0.9
100	2	2.254	4.08×10^{-3}	5.52×10^2	1.37×10^{-4}	1.23×10^{-4}	0.9
200	5	3.056	2.75×10^{-3}	1.11× 10^3	3.89×10^{-4}	3.88×10^{-4}	1.1
300	8	3.700	1.50×10^{-3}	2.47× 10^3	1.23×10^{-3}	1.22×10^{-3}	0.9
360	10	4.041	4.04×10^{-4}	1.00×10^4	1.25×10^{-3}	1.25×10^{-3}	1.0
380	13	4.139	2.32×10^{-4}	1.78×10^4	1.28×10^{-3}	1.26×10^{-3}	1.2
400	17	4.240	3.17×10^{-4}	1.34×10^4	1.28×10^{-3}	1.26×10^{-3}	0.9
420	1	4.332	1.13×10^{-6}	3.83×10^6	1.71×10^{-3}	1.52×10^{-3}	0.7
440	0	4.430	7.70×10^{-4}	5.75× 10^3	1.49×10^{-3}	1.47×10^{-3}	0.9
500	0	4.699	4.19×10^{-3}	1.12× 10^3	7.96×10^{-4}	7.92×10^{-4}	1.1
1000	0	6.620	2.60×10^{-2}	2.54×10^2	9.46×10^{-5}	9.46×10^{-5}	1.0
2000	0	9.288	4.15×10^{-2}	2.24×10^2	5.13×10^{-5}	2.98×10^{-5}	1.7
3000	0	11.291	5.08×10^{-2}	2.22×10^2	2.65×10^{-5}	4.51×10^{-6}	1.95
4000	0	13.005	5.96×10^{-2}	2.18×10^2	1.45×10^{-5}	4.27×10^{-6}	1.8
8000	0	18.420	9.01×10^{-2}	2.04×10^2	8.79×10^{-7}	1.13×10^{-8}	1.9
9520	0	20.106	9.76×10^{-2}	2.06×10^2	3.23×10^{-7}	9.81×10^{-10}	1.95

Table 5. Seismic tomography problems.

Number of Rows i	Number of Zero Singular Values	Largest Singular Value α_i	Smallest Nonzero Singular Value β_i	Condition Number k_i	Residual after 666 Iterations $w = 1$	$w = w_{opt}$	w_{opt}
40	1	2.929	4.73×10^{-2}	6.19×10	9.55×10^{-5}	4.61×10^{-7}	1.6
50	1	3.103	2.93×10^{-2}	1.06×10^{2}	7.95×10^{-5}	2.75×10^{-5}	1.6
60	1	3.199	2.92×10^{-2}	1.09×10^{2}	9.29×10^{-5}	1.95×10^{-5}	1.6
100	5	3.803	2.92×10^{-2}	1.30×10^{2}	6.44×10^{-5}	1.30×10^{-5}	1.6
200	11	5.159	1.06×10^{-3}	4.89×10^{3}	4.40×10^{-3}	3.81×10^{-3}	0.5
300	24	6.154	4.33×10^{-4}	1.42×10^{4}	3.92×10^{-3}	3.88×10^{-3}	0.6
360	33	6.600	1.22×10^{-4}	5.42×10^{4}	5.06×10^{-3}	4.25×10^{-3}	0.4
380	34	6.773	1.06×10^{-5}	6.41×10^{5}	4.47×10^{-3}	4.19×10^{-3}	0.5
400	45	6.859	2.40×10^{-6}	2.86×10^{6}	4.82×10^{-3}	4.40×10^{-3}	0.5
420	34	6.985	2.41×10^{-6}	2.90×10^{6}	4.88×10^{-3}	4.44×10^{-3}	0.5
440	32	7.173	2.43×10^{-6}	2.95×10^{6}	4.27×10^{-3}	3.79×10^{-3}	0.6
500	26	7.488	1.06×10^{-5}	7.03×10^{5}	4.72×10^{-3}	4.27×10^{-3}	0.5
1000	6	9.807	1.89×10^{-3}	5.18×10^{3}	2.61×10^{-3}	2.43×10^{-3}	0.5
2000	4	12.513	9.91×10^{-3}	1.26×10^{3}	6.54×10^{-4}	6.45×10^{-4}	1.2
3000	4	14.474	2.79×10^{-2}	5.18×10^{2}	4.56×10^{-5}	4.56×10^{-5}	1.0
3200	4	14.902	2.87×10^{-2}	5.19×10^{2}	4.51×10^{-5}	4.51×10^{-5}	1.0

Table 6. Seismic tomography problems with initial shuffle.

Number of Rows i	Number of Zero Singular Values	Largest Singular Value α_i	Smallest Nonzero Singular Value β_i	Condition Number k_i	Residual after 666 Iterations $w = 1$	$w = w_{opt}$	w_{opt}
40	0	1.940	3.46×10^{-1}	5.61	1.04×10^{-16}	1.91×10^{-17}	0.4
50	0	2.099	1.70×10^{-1}	1.23×10	1.75×10^{-16}	1.12×10^{-16}	1.2
60	0	2.341	1.17×10^{-1}	2.01×10	6.16×10^{-11}	1.30×10^{-16}	1.3
100	1	2.876	8.21×10^{-2}	3.50×10	1.69×10^{-7}	1.38×10^{-8}	1.1
200	2	3.818	4.19×10^{-3}	9.11×10^{2}	4.00×10^{-4}	3.94×10^{-4}	0.9
300	8	4.719	2.22×10^{-7}	2.12×10^{7}	4.24×10^{-4}	3.67×10^{-4}	0.7
360	21	5.075	5.56×10^{-7}	9.13×10^{6}	3.37×10^{-4}	3.28×10^{-4}	1.2
380	32	5.225	6.33×10^{-7}	8.25×10^{6}	2.87×10^{-4}	2.87×10^{-4}	1.0
400	40	5.365	3.28×10^{-8}	1.64×10^{8}	2.09×10^{-4}	2.09×10^{-4}	1.1
420	32	5.514	5.29×10^{-8}	1.04×10^{8}	2.21×10^{-4}	2.20×10^{-4}	1.1
440	24	5.656	1.15×10^{-7}	4.93×10^{7}	2.05×10^{-4}	2.05×10^{-4}	0.9
500	21	6.034	6.29×10^{-5}	9.59×10^{4}	1.95×10^{-4}	1.94×10^{-4}	0.9
1000	10	8.370	4.37×10^{-3}	1.92×10^{3}	6.94×10^{-5}	6.94×10^{-5}	1.0
2000	6	11.857	2.09×10^{-2}	5.67×10^{2}	3.16×10^{-5}	2.52×10^{-5}	1.5
3000	4	14.462	1.96×10^{-2}	7.37×10^{2}	1.90×10^{-5}	1.36×10^{-5}	1.6
3200	4	14.902	2.87×10^{-2}	5.19×10^{2}	1.56×10^{-5}	1.08×10^{-5}	1.6

The reading of the tables is simple. Consider for example Table 2 when the number of rows equals 400. In this case, the related 400×400 matrix has 22 zero singular values, $\alpha_i = 4.200, \beta_i = 4.41 \times 10^{-4}$, $k_i = 9.53 \times 10^3$, and the residual values are 1.72×10^{-3} for $w = 1$, and 1.59×10^{-3} for $w_{opt} = 0.8$.

The experiments reveal interesting features of the anomaly phenomena. First, note the slow increase of the sequence $\alpha_1, \ldots, \alpha_m$. We see that α_i is considerably smaller than i. Moreover, the larger i is, the smaller the ratio α_i / i. This behavior is due to the fact that the rows have unit length and random directions; see (17).

The second remark is about the smallest singular value anomaly and the related condition number anomaly. The derivation of these properties relies on the assumption that the submatrices A_i, $i = 1, \ldots, m$, do not have zero singular values. Yet, as our tables show, several submatrices have zero singular values. Consequently, in some cases, we can see a slight violation of the anomaly behavior.

The third point is about the use of an initial random shuffle. Note that the shuffle reduces the number of zero singular values in the submatrices. In addition, as expected, the shuffled systems enjoy sharper anomaly. In particular, we see that for highly overdetermined (underdetermined) linear systems, the use of a shuffle improves the rate of convergence!

Tables 7 and 8 display experiments with randomized Kaczmarz methods. In **Shuffled Kaczmarz**, each iteration starts with a random shuffle of the linear system that is solved. In **Randomized Kaczmarz**, each iteration is composed from m steps, where each step treats one randomly chosen equation. Thus, in both methods, the computational effort per iteration is slightly larger than that of Kaczmarz iteration. Consider for example Table 7, which describes experiments with the Shuffled Kaczmarz method. Now, let us inspect the solution of the Fan beam tomography problem of the form (62) with $i = 1000$ rows. In this case, the related residual values are 1.64×10^{-3} and 8.54×10^{-5}, where the smaller value is due to initial shuffle.

The results of Tables 7 and 8 are quite interesting. First, note that the two randomized methods behave in a similar way. In particular, both methods possess the anomaly phenomenon, and the use of an initial shuffle sharpens the anomaly. However, when solving highly overdetermined systems, the use of initial shuffle has a smaller effect, since now, each iteration includes an internal shuffle. Moreover, comparing Tables 7 and 8 with Tables 1–6 indicates that the randomized methods are not faster than the Kaczmarz method with initial shuffle. That is, one shuffle is enough!

Summarizing our experiments, we see that the asymptotic rate of convergence of the Kaczmarz method can be rather slow. Yet, the rate of convergence is considerably affected by a number of factors, such as the number of rows (the Kaczmarz anomaly phenomenon), the value of w, and rows ordering.

Table 7. Solving tomography problems with the shuffled Kaczmarz method.

Number of Rows	Residuals after 666 Iterations					
	No Initial Shuffle			With Initial Shuffle		
i	Parallel Beam	Fan Beam	Seismic	Parallel Beam	Fan Beam	Seismic
40	7.90×10^{-8}	5.03×10^{-9}	2.21×10^{-4}	6.65×10^{-7}	8.22×10^{-17}	1.31×10^{-16}
50	3.27×10^{-6}	4.85×10^{-9}	3.40×10^{-4}	3.47×10^{-5}	5.32×10^{-17}	7.12×10^{-15}
100	3.01×10^{-4}	4.53×10^{-5}	2.91×10^{-4}	9.92×10^{-5}	9.18×10^{-5}	1.24×10^{-5}
200	1.01×10^{-3}	6.90×10^{-4}	2.30×10^{-3}	5.28×10^{-4}	4.86×10^{-4}	2.89×10^{-4}
300	2.21×10^{-3}	1.95×10^{-3}	2.72×10^{-3}	9.34×10^{-4}	1.46×10^{-3}	3.73×10^{-4}
360	2.65×10^{-3}	1.80×10^{-3}	3.26×10^{-3}	1.74×10^{-3}	1.26×10^{-3}	3.29×10^{-4}
380	3.07×10^{-3}	2.43×10^{-3}	3.14×10^{-3}	1.80×10^{-3}	1.36×10^{-3}	2.79×10^{-4}
400	2.92×10^{-3}	2.67×10^{-3}	3.11×10^{-3}	1.52×10^{-3}	1.14×10^{-3}	2.31×10^{-4}
420	2.66×10^{-3}	2.85×10^{-3}	3.18×10^{-3}	2.03×10^{-3}	1.25×10^{-3}	1.93×10^{-4}
440	3.11×10^{-3}	2.42×10^{-3}	3.46×10^{-3}	1.80×10^{-3}	1.51×10^{-3}	1.75×10^{-4}
500	2.99×10^{-3}	2.44×10^{-3}	3.01×10^{-3}	1.58×10^{-3}	8.18×10^{-4}	1.81×10^{-4}
1000	1.87×10^{-3}	1.64×10^{-3}	1.17×10^{-3}	1.17×10^{-4}	8.54×10^{-5}	7.62×10^{-5}
2000	3.85×10^{-4}	6.66×10^{-4}	4.06×10^{-4}	7.60×10^{-6}	5.85×10^{-5}	3.73×10^{-5}
3000	1.26×10^{-5}	3.74×10^{-5}	1.65×10^{-5}	9.43×10^{-7}	3.23×10^{-5}	1.80×10^{-5}
4000	1.90×10^{-7}	1.48×10^{-5}		8.06×10^{-8}	1.45×10^{-5}	
8000		7.21×10^{-7}			6.95×10^{-7}	
9520		3.30×10^{-7}			2.96×10^{-7}	

Table 8. Solving tomography problems with the randomized Kaczmarz method.

Number of Rows	Residuals after 666 Iterations					
	No Initial Shuffle			With Initial Shuffle		
i	Parallel Beam	Fan Beam	Seismic	Parallel Beam	Fan Beam	Seismic
40	1.34×10^{-6}	6.32×10^{-8}	5.12×10^{-4}	9.04×10^{-6}	1.31×10^{-16}	3.30×10^{-17}
50	2.57×10^{-5}	6.65×10^{-8}	5.78×10^{-4}	8.31×10^{-5}	6.11×10^{-17}	5.72×10^{-11}
100	3.02×10^{-4}	9.44×10^{-5}	3.93×10^{-4}	2.63×10^{-4}	3.11×10^{-4}	6.47×10^{-5}
200	1.46×10^{-3}	8.01×10^{-4}	2.51×10^{-3}	7.35×10^{-4}	6.37×10^{-4}	3.98×10^{-4}
300	2.36×10^{-3}	2.74×10^{-3}	2.92×10^{-3}	1.43×10^{-3}	1.36×10^{-3}	3.99×10^{-4}
360	2.62×10^{-3}	2.22×10^{-3}	3.21×10^{-3}	2.01×10^{-3}	1.60×10^{-3}	3.73×10^{-4}
380	2.92×10^{-3}	2.34×10^{-3}	3.09×10^{-3}	2.20×10^{-3}	1.64×10^{-3}	3.48×10^{-4}
400	3.39×10^{-3}	2.52×10^{-3}	3.26×10^{-3}	1.56×10^{-3}	1.65×10^{-3}	2.58×10^{-4}
420	3.26×10^{-3}	2.83×10^{-3}	3.74×10^{-3}	2.36×10^{-3}	1.29×10^{-3}	2.61×10^{-4}
440	3.63×10^{-3}	2.60×10^{-3}	3.38×10^{-3}	2.22×10^{-3}	1.62×10^{-3}	2.04×10^{-4}
500	3.17×10^{-3}	3.14×10^{-3}	3.10×10^{-3}	2.08×10^{-3}	9.14×10^{-4}	2.02×10^{-4}
1000	2.19×10^{-3}	2.09×10^{-3}	1.29×10^{-3}	1.58×10^{-4}	1.14×10^{-4}	7.45×10^{-5}
2000	4.40×10^{-4}	8.05×10^{-4}	3.62×10^{-4}	1.57×10^{-5}	6.40×10^{-5}	4.38×10^{-5}
3000	2.06×10^{-5}	4.36×10^{-5}	1.94×10^{-5}	1.95×10^{-6}	4.87×10^{-5}	2.07×10^{-5}
4000	4.26×10^{-7}	1.88×10^{-5}		1.84×10^{-7}	2.03×10^{-5}	
8000		9.80×10^{-7}			1.07×10^{-6}	
9520		3.85×10^{-7}			3.60×10^{-7}	

6. Concluding Remarks

Although Kaczmarz method has been well known for many years, the Kaczmarz anomaly phenomenon was observed only recently. This is, perhaps, because it requires a certain randomness of the rows' directions. A major application of Kaczmarz method is to solve large sparse linear systems that arise in computerized tomography. Hence, it is important to expose the extent of the phenomenon when solving such problems. The theory presented in the paper explains the reasons behind the anomaly, while the experiments display its nature.

The Kaczmarz anomaly phenomenon is observed by watching how the number of rows changes the asymptotic rate of convergence. Another property that affects the rate of convergence is row ordering. The initial ordering of tomography problems is often rather poor, which yields a slow rate of convergence. A common remedy that helps to overcome this difficulty is an initial random shuffle. The shuffle is likely to improve the randomality of the rows' directions and, therefore, to sharpen the anomaly phenomenon. The experiments that we have done illustrate this feature.

Repeating the use of a random shuffle at each iteration gives rise to a new randomized algorithm, the Shuffled Kaczmarz method of Oswald and Zhou [21,22], which is not inferior to the celebrated Randomized Kaczmarz method of Strohmer and Vershynin [24]. However, one consequence of our experiments is that randomized methods are not faster than Kaczmarz method with one initial random shuffle.

In our experiments, the random shuffle is based on a random permutations generator. Yet, following Herman and Meyer [15], it is possible to construct an improved initial shuffle that takes advantage of the special structure of tomography problems. The idea is to seek a permutation that improves the orthogonality between adjacent rows. In general, there is no easy way to achieve this task, but the special structure of tomography problems enables effective solutions of this problem, e.g., [11,15,17,18]. As with random shuffles, the use of optimal ordering is expected to sharpen the anomaly phenomenon. However, the testing of this issue is left to future research.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declare no conflict of interest.

References

1. Björk, A.; Elfving, T. Accelerated projection methods for computing pseudoinverse solutions of systems of linear equation. *BIT* **1979**, *19*, 145–163. [CrossRef]
2. Cegielski, A.; Censor, Y. Projection methods: An annotated bibliography of books and reviews. *Optimization* **2015**, *64*, 2343–2358.
3. Censor, Y. Row-action methods for huge and sparse systems and their applications. *SIAM Rev.* **1981**, *23*, 444–466. [CrossRef]
4. Censor, Y.; Zenios, S. *Parallel Optimization: Theory, Algorithms and Applications*; Oxford University Press: Oxford, UK, 1997.
5. Censor, Y.; Herman, G.T.; Jiang, M. A note on the behavior of the randomized Kaczmarz algorithm of Strohmer and Vershynin. *J. Fourier Anal. Appl.* **2009**, *15*, 431–436. [CrossRef]
6. Dax, A. The convergence of linear stationary iterative processes for solving singular unstructured systems of linear equations. *SIAM Rev.* **1990**, *32*, 611–635. [CrossRef]
7. Dax, A. The adventures of a simple algorithm. *Linear Alg. Its Appl.* **2003**, *361*, 41–61. [CrossRef]
8. Dax, A. The smallest singular value anomaly and the condition number anomaly. *Axioms* **2022**, *11*, 99. [CrossRef]
9. Dax, A. The rate of convergence of the SOR method in the positive semidefinite case. *Comput. Math. Methods* **2022**, *2022*, 6143444. [CrossRef]
10. Gordon, R.; Bender, R.; Herman, G.T. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *J. Theor. Biol.* **1970**, *29*, 471–481. [CrossRef]
11. Guan, H.; Gordon, R. A projection access order for speedy convergence of ART (algebraic reconstruction technique): A multilevel scheme for computed tomography. *Phy. Med. Biol.* **1994**, *39*, 2005–2022. [CrossRef]

12. Hansen, P.C.; Saxild-Hansen, M. AIR Tools—A MATLAB package of algebraic iterative reconstruction methods. *J. Comp. Appl. Math.* **2012**, *236*, 2167–2178. [CrossRef]
13. Hansen, P.C.; Jørgensen, J.S. AIR Tools II: Algebraic iterative reconstruction methods, improved implementation. *Numer. Algorithms* **2018**, *79*, 107–137. [CrossRef]
14. Herman, G.T. *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*; Academic Press: New York, NY, USA, 1980.
15. Herman, G.T.; Meyer, L.B. Algebraic reconstruction techniques can be made computationally efficient. *IEEE Trans. Med. Imaging* **1993**, *12*, 600–609. [CrossRef]
16. Kaczmarz, S. Angenäherte Auflösung von Systemen linearer Gleichungen. *Bull. l'Académie Pol. Sci. Lett.* **1937**, *A35*, 355–357.
17. Kazantsev, I.G.; Matej, S.; Lewitt, M. Optimal ordering of projections using permutation matrices and angles between projection subspaces. *Electron. Notes Discret. Math.* **2005**, *20*, 205–216. [CrossRef]
18. Mueller, K.; Yagel, R.; Cornhill, J.F. The weighted distance scheme: A globally optimizing projection ordering method for ART. *IEEE Trans. Med. Imaging* **1997**, *16*, 223–230. [CrossRef]
19. Natterer, E. *The Mathematics of Computerized Tomography*; Classics in Applied Mathematics; SIAM: Philadelphia, PA, USA, 2001.
20. Needell, D. Randomized Kaczmarz solver for noisy linear systems. *BIT* **2010**, *50*, 395–403. [CrossRef]
21. Oswald, P.; Zhou, W. Convergence analysis for Kaczmarz-type methods in a Hilbert space framework. *Linear Algebra Its Appl.* **2015**, *478*, 131–161. [CrossRef]
22. Oswald, P.; Zhou, W. Random reordering in SOR-type methods. *Numer. Math.* **2017**, *135*, 1207–1220. [CrossRef]
23. Popa, C. *Projection Algorithms—Classical Results and Developments*; Applications to image reconstruction; Lambert Academic Publishing: Chisinau, Republic of Moldova, 2012.
24. Strohmer, T.; Vershynin, R. A randomized Kaczmarz algorithm with exponential convergence. *J. Fourier Anal. Appl.* **2009**, *15*, 262–278. [CrossRef]
25. Tanabe, K. Projection method for solving a singular system of linear equations and its applications. *Numer. Math.* **1971**, *17*, 203–214. [CrossRef]
26. van Dijke, M.C. Iterative Methods in Image Reconstruction. Ph.D. Thesis, Utrecht University, Utrecht, The Netherlands, 1992.
27. Varga, S. Orderings of the successive overrelaxation scheme. *Pac. J. Math.* **1959**, *9*, 925–939. [CrossRef]