

Research Article

Kaczmarz Iterative Projection and Nonuniform Sampling with Complexity Estimates

Tim Wallace and Ali Sekmen

Department of Computer Science, Tennessee State University, 3500 John A. Merritt Boulevard, Nashville, TN 37209-1500, USA

Correspondence should be addressed to Ali Sekmen; asekmen@tnstate.edu

Received 21 August 2014; Revised 26 October 2014; Accepted 27 October 2014; Published 15 December 2014

Academic Editor: Hengyong Yu

Copyright © 2014 T. Wallace and A. Sekmen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Kaczmarz's alternating projection method has been widely used for solving mostly over-determined linear system of equations $A\mathbf{x} = \mathbf{b}$ in various fields of engineering, medical imaging, and computational science. Because of its simple iterative nature with light computation, this method was successfully applied in computerized tomography. Since tomography generates a matrix A with highly coherent rows, randomized Kaczmarz algorithm is expected to provide faster convergence as it picks a row for each iteration at random, based on a certain probability distribution. Since Kaczmarz's method is a subspace projection method, the convergence rate for simple Kaczmarz algorithm was developed in terms of subspace angles. This paper provides analyses of simple and randomized Kaczmarz algorithms and explains the link between them. New versions of randomization are proposed that may speed up convergence in the presence of nonuniform sampling, which is common in tomography applications. It is anticipated that proper understanding of sampling and coherence with respect to convergence and noise can improve future systems to reduce the cumulative radiation exposures to the patient. Quantitative simulations of convergence rates and relative algorithm benchmarks have been produced to illustrate the effects of measurement coherency and algorithm performance, respectively, under various conditions in a real-time kernel.

1. Introduction

Kaczmarz (in [1]) introduced an iterative algorithm for solving a consistent linear system of equations $A\mathbf{x} = \mathbf{b}$ with $A \in \mathbb{R}^{M \times N}$. This method projects the estimate \mathbf{x}^j onto a subspace normal to the row a_i at step $j + 1$ cyclically with $i = j(\text{mod}M) + 1$. The block Kaczmarz algorithm first groups the rows into matrices A_1, A_2, \dots, A_k and then it projects the estimate \mathbf{x}^j onto the subspace normal to the subspace spanned by the rows of A_i at step $j + 1$ cyclically with $i = j(\text{mod}k) + 1$. Obviously, the block Kaczmarz is equivalent to the simple Kaczmarz for $k = M$. The Kaczmarz method is a method of alternating projection (MAP) and it has been widely used in medical imaging as an algebraic reconstruction technique (ART) [2, 3] due to its simplicity and light computation. Strohmer and Vershynin [4] proved that if a row for each iteration is picked in a random fashion with probability proportional with ℓ_2 norm of that row, then the algorithm converges in expectation exponentially with

a rate that depends on a scaled condition number of A (not on the number of equations). Needell (in [5]) extended the work of [4] for noisy linear systems and developed a bound for convergence to the least square solution for $A\mathbf{x} = \mathbf{b}$. Needell also developed a randomized Kaczmarz method that addresses coherence effects [6], and she analyzed the convergence of randomized block Kaczmarz method [7]. Chen and Powell (in [8]) consider a random measurement matrix A instead of random selection of measurements. Galántai (in [9, 10]) provides convergence analysis for block Kaczmarz method by expanding the convergence analysis (based on subspace angles) of Deutsch [11]. Brezinski and Redivo-Zaglia (in [12]) utilizes the work of Galántai for accelerating convergence of regular Kaczmarz method.

The work of this paper endeavors to make the following contributions.

- (i) Research on regular and randomized Kaczmarz methods appears disconnected in the literature. Even

though convergence rates have been studied separately, the link between them has not been explored sufficiently.

- (ii) A new randomization technique based on subspace angles has been developed which indicates an advantage with coherent data measurements.
- (iii) A further method is introduced which orthogonalizes the subspace blocks in order to mitigate the coherency. Convergence is consistent with statistical expectations from theory and simulations.
- (iv) The effects of measurement coherence are observed in the literature and illustrated in our simulations with norm and angle based iteration randomization.
- (v) A broader review and mathematical analysis of common methods is presented from both statistical and deterministic perspectives.
- (vi) Numerical simulations are provided to illustrate the typical effects of nonuniform sampling coherency upon convergence of Kaczmarz methods.
- (vii) Kaczmarz inversions versus matrix size were performed to allow comparison of the relative convergence rates of various well-known methods using typical hardware and software. The results show relative computational complexities of common methods for simple and randomized Kaczmarz, including the randomized Kaczmarz orthogonal block.

2. Methods and Materials

Data inversion and reconstruction in computed tomography is most often based upon the iterative Kaczmarz algorithm due to the $O(N^2)$ performance. First, in this section, given the number of methods currently in the literature, a broad but extensive overview of the mathematical theory for the more common methods is provided, such as simple, block, and randomized Kaczmarz. Where possible, the convergence results are compared from both random and deterministic perspectives to demonstrate similar results and convergence analysis methods. The concept of subspace projections is reviewed and the connection to iteration is noted.

Next, two new methods are proposed and analyzed in the context of coherent data measurements. These methods allow the algorithm to adapt to the changing environment of the sampling measurement system, in order to mitigate coherency.

Simulated methods for data acquisition under uniform and nonuniform X-ray beam measurements are included, and convergence results are computed comparing simple and random row selection methods.

Lastly, after the algorithm methods section, a brief overview of the methods used to obtain the complexity estimates is presented. Methods include using common software and hardware under dedicated kernel conditions. Simulations for Kaczmarz convergence and complexity were performed using Octave software.

2.1. Convergence of Regular Block Kaczmarz Method. Let \mathbf{x}^* be the solution of consistent $\mathbf{Ax} = \mathbf{b}$, where $A \in \mathbb{R}^{M \times M}$ is full column rank. Let A be row-partitioned as $\{A_1, \dots, A_k\}$ where $A_i \in \mathbb{R}^{M_i \times M}$. Then, the simple block Kaczmarz update is as follows:

$$\mathbf{x}_{j+1} = \mathbf{x}_j + A_i^T (A_i A_i^T)^{-1} (\mathbf{b}_i - A_i \mathbf{x}_j) \quad i = j \pmod{k} + 1, \quad (1)$$

where \mathbf{b}_i is the section of \mathbf{b} that corresponds to the rows of A_i . Note that since A_i is full row rank, $A_i^T (A_i A_i^T)^{-1}$ is the right pseudo-inverse of A_i . This is equivalent to

$$\mathbf{x}_{j+1} = \mathbf{x}_j + A_i^T (A_i A_i^T)^{-1} (A_i \mathbf{x}^* - A_i \mathbf{x}_j), \quad (2)$$

$$\mathbf{x}_{j+1} - \mathbf{x}^* = \mathbf{x}_j - \mathbf{x}^* - A_i^T (A_i A_i^T)^{-1} A_i (\mathbf{x}_j - \mathbf{x}^*).$$

Note that $A_i^T (A_i A_i^T)^{-1} A_i$ is the projection matrix for projection of the range of A_i^T :

$$\mathbf{x}_{j+1} - \mathbf{x}^* = \mathbf{x}_j - \mathbf{x}^* - P_{Sp(A_i^T)} (\mathbf{x}_j - \mathbf{x}^*), \quad (3)$$

$$\mathbf{x}_{j+1} - \mathbf{x}^* = (I - P_{Sp(A_i^T)}) (\mathbf{x}_j - \mathbf{x}^*), \quad (4)$$

$$\mathbf{x}_{j+1} - \mathbf{x}^* = P_{Sp^\perp(A_i^T)} (\mathbf{x}_j - \mathbf{x}^*).$$

For one cycle of the blocks,

$$\mathbf{x}_k - \mathbf{x}^* = P_{Sp^\perp(A_k^T)} P_{Sp^\perp(A_{k-1}^T)} \cdots P_{Sp^\perp(A_1^T)} (\mathbf{x}_0 - \mathbf{x}^*). \quad (5)$$

Note that if $A \in \mathbb{R}^{M \times N}$ is a full column rank with $M < N$, then the simple block Kaczmarz update is as follows:

$$\mathbf{x}_{j+1} = \mathbf{x}_j + A_i^\dagger (\mathbf{b}_i - A_i \mathbf{x}_j) = \mathbf{x}_j + A_i^\dagger A_i (\mathbf{x}^* - \mathbf{x}_j) \quad (6)$$

$$i = j \pmod{k} + 1,$$

where A_i^\dagger is the pseudo-inverse of A_i and $A_i^\dagger A_i$ is the orthogonal projection onto $Sp(A_i^T)$. Then, we get the same equation as (3), and subsequently we get (5),

$$\mathbf{x}_{j+1} - \mathbf{x}^* = \mathbf{x}_j - \mathbf{x}^* - P_{Sp(A_i^T)} (\mathbf{x}_j - \mathbf{x}^*). \quad (7)$$

2.2. Exponential Convergence

Theorem 1. *Let \mathbf{x}^* be the solution of consistent $\mathbf{Ax} = \mathbf{b}$, where $A \in \mathbb{R}^{M \times M}$ is full rank. Let A be row-partitioned as $\{A_1, \dots, A_k\}$ where $A_i \in \mathbb{R}^{M_i \times M}$ and $M_1 + M_2 + \dots + M_k = M$. Then, the simple block Kaczmarz converges exponentially and the convergence rate depends of the number of blocks.*

Proof. By (3) and orthogonal projection,

$$\|\mathbf{x}_{j+1} - \mathbf{x}^*\|_2^2 = \|\mathbf{x}_j - \mathbf{x}^*\|_2^2 - \|P_{Sp(A_i^T)} (\mathbf{x}_j - \mathbf{x}^*)\|_2^2. \quad (8)$$

So,

$$\|\mathbf{x}_{j+1} - \mathbf{x}^*\|_2^2 \leq \|\mathbf{x}_j - \mathbf{x}^*\|_2^2; \quad (9)$$

$\mathbf{x}_j - \mathbf{x}^*$ depends on the initial condition $\tilde{\mathbf{x}}_0 = \mathbf{x}_0 - \mathbf{x}^*$, and this dependence is scale-invariant. To see this, let $\mathbf{e}_j = \mathbf{x}_j - \mathbf{x}^*$ and consider $c\tilde{\mathbf{x}}_0$ where $c \in \mathbb{R}$. By (4),

$$\begin{aligned} \mathbf{e}_{j+1}(c\tilde{\mathbf{x}}_0) &= P_{Sp^+(A_{j+1}^T)} \mathbf{e}_j(c\tilde{\mathbf{x}}_0) \\ &= P_{Sp^+(A_{j+1}^T)} P_{Sp^+(A_j^T)} \cdots P_{Sp^+(A_1^T)} \mathbf{e}_0(c\tilde{\mathbf{x}}_0) \\ &= P_{Sp^+(A_{j+1}^T)} P_{Sp^+(A_j^T)} \cdots P_{Sp^+(A_1^T)} (c\tilde{\mathbf{x}}_0) \quad (10) \\ &= c P_{Sp^+(A_{j+1}^T)} P_{Sp^+(A_j^T)} \cdots P_{Sp^+(A_1^T)} \mathbf{e}_0(\tilde{\mathbf{x}}_0) \\ &= c \mathbf{e}_{j+1}(\tilde{\mathbf{x}}_0). \end{aligned}$$

We will first show that if $\mathbf{x}_0 \neq \mathbf{x}^*$, then $\|\mathbf{x}_k - \mathbf{x}^*\|_2 < \|\mathbf{x}_0 - \mathbf{x}^*\|_2$. By the way of contradiction, assume that $\mathbf{x}_0 \neq \mathbf{x}^*$ and $\|\mathbf{x}_k - \mathbf{x}^*\|_2 = \|\mathbf{x}_0 - \mathbf{x}^*\|_2$. By (9),

$$\|\mathbf{x}_k - \mathbf{x}^*\|_2 \leq \|\mathbf{x}_{k-1} - \mathbf{x}^*\|_2 \cdots < \|\mathbf{x}_0 - \mathbf{x}^*\|_2 \quad (11)$$

and therefore $\|\mathbf{x}_l - \mathbf{x}^*\|_2 = \|\mathbf{x}_0 - \mathbf{x}^*\|_2$ for all $1 \leq l \leq k$. By (3), $P_{Sp^+(A_l^T)}(\mathbf{x}_{l-1} - \mathbf{x}^*) = 0$ for all $1 \leq l \leq k$. By (8), we get $\mathbf{x}_l = \mathbf{x}_0$ for all $1 \leq l \leq k$. This implies that $P_{Sp^+(A_l^T)}(\mathbf{x}_0 - \mathbf{x}^*) = 0$ for all $1 \leq l \leq k$. So,

$$\begin{aligned} P_{Sp^+(A_k^T) \cap Sp^+(A_{k-1}^T) \cdots \cap Sp^+(A_1^T)}(\mathbf{x}_0 - \mathbf{x}^*) &= 0, \\ P_{Sp^+(A^T)}(\mathbf{x}_0 - \mathbf{x}^*) &= 0. \end{aligned} \quad (12)$$

Since A is full column rank, we get $\mathbf{x}_0 = \mathbf{x}^*$, which is a contradiction. So we know that $\|\mathbf{x}_k - \mathbf{x}^*\|_2 < \|\mathbf{x}_0 - \mathbf{x}^*\|_2$ (for one full cycle of k -iterations).

By compactness, there exists an $\epsilon \in (0, 1)$ such that, for all $\tilde{\mathbf{x}}_0 = \mathbf{x}_0 - \mathbf{x}^* \in S^{N-1}$,

$$\|\mathbf{x}_k - \mathbf{x}^*\|_2 \leq (1 - \epsilon). \quad (13)$$

By (10) and (13),

$$\|\mathbf{x}_k - \mathbf{x}^*\|_2 = \|\tilde{\mathbf{x}}_0\|_2 \mathbf{e}_k \left(\frac{\tilde{\mathbf{x}}_0}{\|\tilde{\mathbf{x}}_0\|_2} \right) \leq (1 - \epsilon) \|\tilde{\mathbf{x}}_0\|_2, \quad (14)$$

$$\|\mathbf{x}_k - \mathbf{x}^*\|_2 \leq (1 - \epsilon) \|\mathbf{x}_0 - \mathbf{x}^*\|_2.$$

Now consider iteration for q cycles:

$$\begin{aligned} \|\mathbf{x}_{qk} - \mathbf{x}^*\|_2 &\leq (1 - \epsilon)^q \|\mathbf{x}_0 - \mathbf{x}^*\|_2, \\ \|\mathbf{x}_{qk} - \mathbf{x}^*\|_2 &\leq [(1 - \epsilon)^{1/k}]^{qk} \|\mathbf{x}_0 - \mathbf{x}^*\|_2. \end{aligned} \quad (15)$$

Therefore, we conclude that the exponential decay depends on the number of blocks k . Note that $k = M$ for regular simple Kaczmarz and the exponential decay depends on the number of rows in this case. The randomized Kaczmarz algorithm proposed by Strohmer and Vershynin [4] avoids this and it converges in expectation as $\mathbb{E}\|\mathbf{x}_p - \mathbf{x}^*\|_2^2 \leq (1 - \kappa(A)^{-2})^p \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2$, where $\kappa(A) = \|A\|_F \|A^\dagger\|_2$ is the scaled condition number of matrix A with A^\dagger is the pseudoinverse of A . \square

2.3. Iterative Subspace Projection Approach. We can use the following theorem (in [9, 11]) to show the convergence of regular block Kaczmarz method.

Theorem 2. Let M_1, M_2, \dots, M_k be closed subspaces of the real Hilbert space \mathbb{H} . Let $M = \bigcap_{i=1}^k M_i$ and P_{M_i} ($i = 1, \dots, k$) be orthogonal projection on M_i . Then, for each $\mathbf{x} \in \mathbb{H}$,

$$\lim_{q \rightarrow \infty} (P_{M_k} P_{M_{k-1}} \cdots P_{M_1})^q \mathbf{x} = P_M \mathbf{x}, \quad (16)$$

where P_M is the orthogonal intersection projection.

The block Kaczmarz is an alternating projection method with $M_1 = Sp^+(A_1^T), \dots, M_k = Sp^+(A_k^T)$. Also,

$$P_{M_1} = P_{Sp^+(A_1^T)}, \dots, P_{M_k} = P_{Sp^+(A_k^T)}, \quad (17)$$

$$M = Sp^+(A_1^T) \cap \cdots \cap Sp^+(A_k^T) = Sp^+(A^T).$$

Since A is full column rank, $Sp^+(A^T) = \{0\}$ and $P_M = \{0\}$. After q cycles,

$$\mathbf{x}_{qk} - \mathbf{x}^* = (P_{M_k} P_{M_{k-1}} \cdots P_{M_1})^q (\mathbf{x}_0 - \mathbf{x}^*). \quad (18)$$

By Theorem 2, $\lim_{q \rightarrow \infty} \mathbf{x}_{qk} - \mathbf{x}^* = 0$ and $\lim_{q \rightarrow \infty} \mathbf{x}_{qk} = \mathbf{x}^*$. Galántai in [9] gives a bound for $\|\mathbf{x}_{qk} - \mathbf{x}^*\|_2$ in terms of principle angles between M_i 's.

2.4. Bound for Block Kaczmarz in Terms of Principle Angles. Smith et al. established the following convergence theorem for applying the alternating projection method in tomography [9, 13].

Theorem 3. Let M_1, M_2, \dots, M_k be closed subspaces of the real Hilbert space \mathbb{H} . Let $M = \bigcap_{i=1}^k M_i$ and P_{M_i} ($i = 1, \dots, k$) be orthogonal projection on M_i (P_M is the orthogonal intersection projection). Let $\theta_j = \alpha(M_j, \bigcap_{i=j+1}^k M_i)$; then, for each $\mathbf{x} \in \mathbb{H}$ and integer $q \geq 1$,

$$\begin{aligned} \left\| (P_{M_k} P_{M_{k-1}} \cdots P_{M_1})^q \mathbf{x} - P_M \mathbf{x} \right\|_2^2 \\ \leq (1 - \prod_{j=1}^{k-1} \sin^2 \theta_j)^q \|\mathbf{x} - P_M \mathbf{x}\|_2^2, \end{aligned} \quad (19)$$

where P_M is the orthogonal intersection projection.

In the special case of the block Kaczmarz, we have $\mathbb{H} = \mathbb{R}^N$, $M_1 = Sp^+(A_1^T), \dots, M_k = Sp^+(A_k^T)$. Also, $P_{M_1} = P_{Sp^+(A_1^T)}, \dots, P_{M_k} = P_{Sp^+(A_k^T)}$ and $M = Sp^+(A_1^T) \cap \cdots \cap Sp^+(A_k^T) = Sp^+(A^T)$. Since A is full column rank, $Sp^+(A^T) = \{0\}$ and $P_M = \{0\}$. Therefore, after q cycles,

$$\begin{aligned} \|\mathbf{x}_{qk} - \mathbf{x}^*\|_2^2 &= \left\| (P_{M_k} P_{M_{k-1}} \cdots P_{M_1})^q (\mathbf{x}_0 - \mathbf{x}^*) \right\|_2^2 \\ &\leq (1 - \prod_{j=1}^{k-1} \sin^2 \theta_j)^q \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2, \end{aligned} \quad (20)$$

where θ_j is as defined in Theorem 3. Note that the exponential decay rate depends on the number of blocks k as shown below:

$$\|\mathbf{x}_{qk} - \mathbf{x}^*\|_2^2 \leq \left[(1 - \prod_{j=1}^{k-1} \sin^2 \theta_j)^{1/k} \right]^{qk} \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2. \quad (21)$$

Require: An over-determined linear set of consistent equations $\mathbf{Ax} = \mathbf{b}$, where A is $M \times N$ matrix and $\mathbf{b} \in \mathbb{R}^M$.
 Let $\mathbf{a}_1, \dots, \mathbf{a}_M$ be the rows of A and b_j be the j th element of \mathbf{b} .
 (1) Pick an arbitrary initial approximation \mathbf{x}_0 .
 (2) Set $p = 0$.
 (3) **while** not converged **do**
 (4) Randomly choose $r(i)$ from $\{1, \dots, M\}$ with probability proportional to $\|\mathbf{a}_{r(i)}\|_2^2$.
 (5) $\mathbf{x}_{p+1} = \mathbf{x}_p + \frac{b_{r(i)} - \langle \mathbf{a}_{r(i)}, \mathbf{x}_p \rangle}{\|\mathbf{a}_{r(i)}\|_2^2} \mathbf{a}_{r(i)}$
 (6) Set $p = p + 1$
 (7) **end while**

ALGORITHM 1: Randomized Kaczmarz (of [4]).

Galántai in [9] developed another bound (for $A \in \mathbb{R}^{M \times M}$) by defining a new matrix X_i for each block A_i as follows.

Theorem 4. Let \mathbf{x}^* be the solution of $\mathbf{Ax} = \mathbf{b}$ for a consistent linear system with $A \in \mathbb{R}^{M \times M}$. Let A be row-partitioned as $\{A_1, \dots, A_k\}$ where $A_i \in \mathbb{R}^{M_i \times N}$. Let $M_1 = Sp^+(A_1^T), \dots, M_k = Sp^+(A_k^T)$ and $A_i A_i^T = LL^T$ be the Cholesky decomposition of $A_i A_i^T$. Define $X_i = A_i^T L^{-T}$ and $X = [X_1, \dots, X_k]$. Then, for each $\mathbf{x} \in \mathbb{R}^N$ and integer $q \geq 1$,

$$\begin{aligned} \|\mathbf{x}_{qk} - \mathbf{x}^*\|_2^2 &\leq [1 - \det(X^T X)]^q \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2 \\ &= \left[(1 - \det(X^T X))^{1/k} \right]^{qk} \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2. \end{aligned} \quad (22)$$

2.5. Special Case: Simple Kaczmarz for $A \in \mathbb{R}^{M \times M}$. Note that this section assumes that $A \in \mathbb{R}^{M \times M}$. The block Kaczmarz algorithm is equivalent to the simple Kaczmarz algorithm if the number of blocks k is equal to the number of rows M . In this case, $A_i A_i^T = \|\mathbf{a}_i\|_2^2 = LL^T$. Therefore, $L = \|\mathbf{a}_i\|_2$ and $L^{-T} = 1/\|\mathbf{a}_i\|_2$. This implies that $X_i = [\mathbf{a}_i/\|\mathbf{a}_i\|_2]$. Then, $X \in \mathbb{R}^{M \times M}$ is defined as

$$X = \left[\begin{array}{c} \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|_2}, \dots, \frac{\mathbf{a}_M}{\|\mathbf{a}_M\|_2} \end{array} \right]. \quad (23)$$

Assume the matrix A has normalized rows and we pick a row at each iteration uniformly randomly. Note that this assumption is feasible as scaling a row of A and the corresponding measurement in \mathbf{b} does not change the solution \mathbf{x} .

X is the Gram matrix with $0 \leq \det(X^T X) \leq \|\mathbf{x}_1\|_2^2 \|\mathbf{x}_2\|_2^2 \dots \|\mathbf{x}_M\|_2^2$. Since $\|\mathbf{x}_i\|_2 = 1$ and X is full rank, we have $0 < \det(X^T X) \leq 1$. Using Theorem 4, we get the following deterministic bound:

$$\|\mathbf{x}_{qM} - \mathbf{x}^*\|_2^2 \leq \left[(1 - \det(X^T X))^{1/M} \right]^{qM} \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2. \quad (24)$$

Since A is normalized, we get $X = A^T$ and therefore

$$\|\mathbf{x}_{qM} - \mathbf{x}^*\|_2^2 \leq \left[(1 - \det(AA^T))^{1/M} \right]^{qM} \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2. \quad (25)$$

Bai and Liu (in [14]) uses the Meany Inequality to develop a general form of this inequality.

2.6. Randomized Kaczmarz Method. Several methods of randomized Kaczmarz are discussed in this section.

2.7. Randomization Based on Row ℓ_2 Norms Method. Strohmer and Vershynin (in [4]) developed a randomized Kaczmarz algorithm that picks a row of A in a random fashion with probability proportional with ℓ_2 norm of that row. They proved that this method has exponential expected convergence rate. Since the rows are picked based on a probability distribution generated by the ℓ_2 norms of the rows of A , it is clear that scaling some of the equations does not change the solution set. However, it may drastically change the order of the rows picked at each iteration. Censor et al. discuss (in [15]) that this should not be better than the simple Kaczmarz as picking a row based on its ℓ_2 norm does not change the geometry of the problem. Theorem 5 is from [4].

Theorem 5. Let \mathbf{x}^* be the solution of $\mathbf{Ax} = \mathbf{b}$ Then, Algorithm 1 converges to \mathbf{x}^* in expectation, with the average error

$$\mathbb{E} \|\mathbf{x}_p - \mathbf{x}^*\|_2^2 \leq (1 - \kappa(A)^{-2})^p \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2, \quad (26)$$

where $\kappa(A) = \|A\|_F \|A^\dagger\|_2$ is the scaled condition number of matrix A with A^\dagger is the left pseudoinverse of A .

Note that A is a full column matrix ($A \in \mathbb{R}^{M \times N}$ with $\text{rank}(A) = N$) and therefore we define A^\dagger as left pseudoinverse of A . We observe that the randomization should work better than the simple (cyclic) Kaczmarz algorithm for matrices with highly coherent rows (e.g., matrices generated by the computerized tomography). Since the Kaczmarz algorithm is based on projections, the convergence will be slow if the consecutive rows selected are highly coherent (i.e., the angle between \mathbf{a}_i and \mathbf{a}_{i+1} is small). Picking rows randomly (not necessarily based on the ℓ_2 norms) makes picking more incoherent rows possible in each iteration. Therefore, the randomization may be useful for certain applications such as medical imaging. Note that matrix A generated by computerized tomography has coherent and sparse rows due to physical nature of data collection. In fact, using Theorem 5, we can develop the following proposition.

Proposition 6. Let $\mathbf{Ax} = \mathbf{b}$ be a consistent linear system of equations ($A \in \mathbb{R}^{M \times N}$) and let \mathbf{x}_0 be an arbitrary initial

Require: An over-determined linear set of consistent equations $\mathbf{Ax} = \mathbf{b}$, where A is $M \times N$ matrix and $\mathbf{b} \in \mathbb{R}^M$.

Let $\mathbf{a}_1, \dots, \mathbf{a}_M$ be the rows of A and \mathbf{b}_j be the j th element of \mathbf{b} .

- (1) Pick an arbitrary initial approximation \mathbf{x}_0 .
- (2) Set $k = 0$.
- (3) Randomly choose $f(i)$ from $\{1, 2, \dots, M\}$ with a uniform distribution.
- (4) **while** not converged **do**
- (5) Randomly choose $g(i)$ from $\{1, \dots, M\}$ with probability proportional to $1 - (\langle \mathbf{a}_{f(i)}, \mathbf{a}_{g(i)} \rangle^2 / \|\mathbf{a}_{f(i)}\|_2^2 \|\mathbf{a}_{g(i)}\|_2^2) = 1 - \cos^2 \theta_{f(i), g(i)}$
- (6) Compute $\mathbf{x}_{k+1} = \mathbf{x}_k + ((b_{f(i)} - \langle \mathbf{a}_{f(i)}, \mathbf{x}_k \rangle) / \|\mathbf{a}_{f(i)}\|_2^2) \mathbf{a}_{f(i)}$
- (7) Compute $\mathbf{x}_{k+2} = \mathbf{x}_{k+1} + ((b_{g(i)} - \langle \mathbf{a}_{g(i)}, \mathbf{x}_{k+1} \rangle) / \|\mathbf{a}_{g(i)}\|_2^2) \mathbf{a}_{g(i)}$
- (8) Set $f(i) = g(i)$
- (9) Set $k = k + 2$
- (10) **end while**

ALGORITHM 2: Randomized Kaczmarz Hyperplane Angles.

approximation to the solution of $\mathbf{Ax} = \mathbf{b}$. For $k = 1, 2, \dots$, compute

$$\mathbf{x}_{p+1} = \mathbf{x}_p + \frac{b_{r(i)} - \langle \mathbf{a}_{r(i)}, \mathbf{x}_p \rangle}{\|\mathbf{a}_{r(i)}\|_2^2} \mathbf{a}_{r(i)}, \quad (27)$$

where $r(i)$ is chosen from the set $\{1, 2, \dots, M\}$ at random, with "any probability distribution." Let \mathbf{x}^* be the solution of $\mathbf{Ax} = \mathbf{b}$. Then,

$$\mathbb{E} \|\mathbf{x}_p - \mathbf{x}^*\|_2^2 \leq (1 - \kappa(B)^{-2})^p \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2, \quad (28)$$

where $\kappa(B) = \|B\|_F \|B^\dagger\|_2$ is the scaled condition number of a matrix B that is obtained by some row-scaling of A .

Proof. This is due to the fact that row-scaling of A (with scaling of the corresponding \mathbf{b}) does not change the geometry of the problem and we can scale the rows to generate any probability distribution. In other words, we can obtain another matrix B from A by scaling its rows in such a way that picking the rows of B based on the ℓ_2 norms of the rows will be equivalent to picking the rows of A based on the chosen probability distribution. Therefore, clearly, any randomization of the row selection will have exponential convergence; however, the rate will depend on the condition number of another matrix. For example, if we use uniform distribution, we can then normalize each row to have matrix B as follows and then pick the rows at random with probability proportional to the norms of the rows:

$$B = \left[\frac{\mathbf{a}_1}{\|\mathbf{a}_1\|_2}, \dots, \frac{\mathbf{a}_M}{\|\mathbf{a}_M\|_2} \right]^T. \quad (29)$$

□

2.8. Randomization Based on Subspace Angles Method. Our approach iterates through the rows of A based on a probability distribution using the hyperplane (subspace) angles. Therefore, it is immune to scaling or normalization. This approach first generates a probability distribution based on the central angle(s) $\{\theta_{i,j}\}$ between the hyperplanes (represented by the rows of $\mathbf{Ax} = \mathbf{b}$). Then, it randomly picks

two hyperplanes using this probability distribution. This is followed by a two-step projection on these hyperplanes (see Algorithm 2).

2.9. P-Subspaces Method. A new method has been developed which is intended to better accommodate the coherency of nonorthogonal data measurements. This next section makes contributions towards proving the statistical convergence of the randomized Kaczmarz orthogonal subspace (RKOS) algorithm. As described in [16], the RKOS initially uses ℓ^2 -norm random hyperplane selection and subsequent projection into a constructed P -dimensional orthogonal subspace S_P comprised of an additional $P - 1$ hyperplanes selected uniformly at random.

Algorithm 3 uses a recursive method to solve for the projections into the orthogonal subspace which is constructed using Gram-Schmidt (GS) procedure. However, a second approach demonstrates an alternate method of arriving at similar results, based upon an a closed form matrix for QR decomposition [17] of projection blocks.

In each of the above cases, vector operations inside the orthogonal subspace preserve the ℓ^2 -norm and reduce errors that would normally be induced for coherent nonorthogonal projections which may be present in the simple Kaczmarz.

2.9.1. Orthogonal Subspaces. A statistical convergence analysis for Randomized Kaczmarz Orthogonal Subspace (RKOS) method is developed assuming identically and independently distributed (IID) random variables as vector components of each row of the measurement matrix A .

(a) Orthogonal Construction. In many problems, $M \gg N$ and fast but optimal solutions are needed, often in noisy environments. In most cases, orthogonal data projection sampling is not feasible due to the constraints of the measurement system. The algorithm and procedure for the RKOS method are given in reference [16] and are intended to construct orthogonal measurements subspaces (see Algorithm 3).

The general technique is to solve using a constructed orthogonal basis from a full rank set of linearly independent

Require: Matrix $A \in \mathbb{R}^{M \times N}$ full-rank consistent measurements subject to $A\mathbf{x} = \mathbf{b}$, for $\mathbf{b} \in \mathbb{R}^M$.

- (1) Set \mathbf{x}_0 to initial approximation, $i = 1$
- (2) **while** not converged **do**
- (3) Select $\dim(S_p) = P < N$ distinct linearly independent rows of A relative to random rule. Construct block matrix $A_i \in \mathbb{R}^{P \times N}$ comprised of rows $\{\mathbf{a}_{i,1}, \dots, \mathbf{a}_{i,P}\}$.
- (4) Perform Gram-Schmidt procedure on A_i to obtain the orthonormal set of columns $\{\mathbf{u}_{i,1}, \dots, \mathbf{u}_{i,P}\}$.
Let $Q_i = \{\mathbf{u}_{i,1}, \dots, \mathbf{u}_{i,P}\} \in \mathbb{R}^{N \times P}$
- (5) Update \mathbf{x}_i as follows:
 $\mathbf{x}_i = \mathbf{x}_{i-1} + \text{Proj}_{S_p(Q_i)}(\mathbf{x}_{i-1} - \mathbf{x})$,
 $\mathbf{x}_i = \mathbf{x}_{i-1} - Q_i Q_i^T (\mathbf{x} - \mathbf{x}_{i-1})$,
- (6) Compute $Q_i^T \mathbf{x}$ iteratively using $\{\mathbf{a}_{i,1}, \dots, \mathbf{a}_{i,P}\}$, $\{\mathbf{b}_{i,1}, \dots, \mathbf{b}_{i,P}\}$, $\{\mathbf{u}_{i,1}, \dots, \mathbf{u}_{i,P}\}$
- (7) Update $i = i + 1$
- (8) **end while**

ALGORITHM 3: P -Subspace Kaczmarz Projections.

measurements in for each subspace in Gram-Schmidt fashion [18, 19].

The subspace estimation may be computed as P -dimensional subspace projection into the subspace orthonormal vector basis:

$$\mathbf{x}_{S_p} = \sum_{l=1}^P \langle \hat{\mathbf{u}}_l, \mathbf{x} \rangle \hat{\mathbf{u}}_l, \quad (30)$$

where \mathbf{x}_{S_p} in $S_p \subseteq S_N$ subspace is the P -dimensional solution approximation which becomes exact for $S_{P=N}$ for $\mathbf{x}_{S_{P=N}} \in \mathbb{R}^N$ in the noiseless, self-consistent, case (the u vector with the hat symbol \hat{u} indicates unit ℓ^2 -norm).

(b) *Modified Kaczmarz.* The standard Kaczmarz equation is essentially iterative projections into a single subspace of dimension one; based upon the sampling hyperplanes, these projections are often oblique, especially in highly-coherent sampling.

The approach herein is motivated towards constructing an iterative algorithm based upon Kaczmarz which may be accelerated while controlling the potential projection errors and incurring reasonable computational penalty. The algorithm is simply to add subspaces of larger dimensions. Let

$$\mathbf{x} - \mathbf{x}_{k+1} = \mathbf{x} - \mathbf{x}_k - \sum_{l=1}^P \langle \hat{\mathbf{u}}_l, \mathbf{x} - \mathbf{x}_k \rangle \hat{\mathbf{u}}_l. \quad (31)$$

It is convenient to make a substitution as follows:

$$\mathbf{z}_{k+1} = \mathbf{x} - \mathbf{x}_{k+1}. \quad (32)$$

Using above substitution and orthonormal condition (It is worthwhile to note that in the problem setup, a fixed vector is projected into a randomized P -dimensional subspace, where algebraic orthogonality was used to obtain (34). In the this statistical treatment of the same equation, the expectation of two random unit vectors vanishes for independent uncorrelated zero mean probability distribution functions,

providing the statistical orthogonality on average satisfying (34). $\langle \hat{\mathbf{u}}_j, \hat{\mathbf{u}}_k \rangle = \delta_{j,k}$, where the Kronecker

$$\delta_{j,k} = \begin{cases} 0, & \text{if } j \neq k, \\ 1, & \text{if } j = k, \end{cases} \quad (33)$$

find the ℓ^2 -norm squared of \mathbf{z}_{k+1} :

$$\|\mathbf{z}_{k+1}\|_2^2 = \|\mathbf{z}_k\|_2^2 - \sum_{l=1}^P |\langle \hat{\mathbf{u}}_l, \mathbf{z}_k \rangle|^2. \quad (34)$$

The ensemble average of the above equation (34) yields the convergence result, which is the main topic of this section.

2.9.2. *Convergence for IID Measurement Matrix.* Firstly, the expectation of a single random projection is computed. In the second step, the terms are summed for the P -dimensional subspace. Experimental results are included in a latter section.

(a) *Expectation of IID Projections.* Consider the expectation of the ℓ^2 -norm squared of the projection of fixed vector $\mathbf{x} \in \mathbb{R}^{N \times 1}$ onto a random subspace basis $U_p \in$ of dimension P :

$$\mathbb{E} \left[\|U_p^T \mathbf{x}\|_2^2 \right], \quad (35)$$

where the matrix basis $U_p \in \mathbb{R}^{N \times P}$ is comprised of P -columns of unit vectors $\hat{\mathbf{u}}_j \in \mathbb{R}^N$ in a constructed orthogonal basis for

$$\begin{aligned} \hat{\mathbf{u}}_j &\longrightarrow \hat{\mathbf{U}}_j = [U_{j,1}, \dots, U_{j,N}] \frac{1}{C_\sigma} \\ &= \frac{U_j}{\|U_j\|_2^2} \quad \forall j \in [1, \dots, P], \end{aligned} \quad (36)$$

where the upper case components $U_{j,i}$ represent the (j, i) th IID random variable component, and normalization constant C_σ is to be determined.

Further noting that complex conjugate $(\cdot)^*$ reduces to transpose $(\cdot)^T$ for real components, the ℓ^2 -norm squared of the projection expands to

$$\|U_P^T \mathbf{x}\|_2^2 = \mathbf{x}^T U_P U_P^T \mathbf{x}. \quad (37)$$

In the next section, the goal is to find the expected value for outer product of the projection:

$$\mathbb{E} [\mathbf{x}^T \widehat{\mathbf{U}}_j \widehat{\mathbf{U}}_j^T \mathbf{x}] \quad \forall j \in [1, \dots, P]. \quad (38)$$

(b) *Unit Vector.* The deterministic identity for the magnitude of a unit vector is well known result for $\widehat{\mathbf{u}} \in \mathbb{R}^N$:

$$\|\widehat{\mathbf{u}}\|_2^2 = \sum_{i=1}^N \frac{u_i^2}{\|\mathbf{u}\|_2^2} = 1. \quad (39)$$

The following statistical result must apply for the j th column unit vector:

$$\begin{aligned} \mathbb{E} [\|\widehat{\mathbf{U}}_j\|_2^2] &= \mathbb{E} [\widehat{\mathbf{U}}_j^T \widehat{\mathbf{U}}_j] = 1 \\ &= \mathbb{E} [U_{j,1}^2 + \dots + U_{j,N}^2] \frac{1}{C_\sigma^2}. \end{aligned} \quad (40)$$

(c) *Normalization of Random Unit Vector.* Denote $\widehat{\mathbf{U}}_j$ as the j th random variable unit-norm vector associated with a set of column vectors $\{\mathbf{U}_j\}_{j \in 1, \dots, P}$ comprising a random subspace matrix $U_{N \times P}$ having IID random variable components $U_{j,i}$. However, no additional assumptions on the distribution of the random variables are made at this time, other than IID. The expectation of both sides of (40) for random vector \mathbf{U}_j are found such that

$$\mathbb{E} \sum_{i=1}^N \frac{U_{j,i}^2}{C_\sigma^2} = \sum_{i=1}^N \mathbb{E} \left[\frac{U_{j,i}^2}{C_\sigma^2} \right] = 1, \quad (41)$$

$$N \times \frac{\mathbb{E} [U_{j,i}^2]}{C_\sigma^2} = 1.$$

Solving above for each unit vector component in this treatment implies a random variable $U_{j,i}$ with zero mean and variance as follows:

$$\mathbb{E} [U_{j,i}^2] = \sigma_{j,i}^2 = \frac{C_\sigma^2}{N} \quad \forall U_{j,i \in 1, \dots, N} \in f(U_{j,i}), \quad (42)$$

where $f(U_{i,j})$ is the associated IID probability distribution.

(d) *P-Dimensional Random Projection.* The next step is to compute the expectation of the magnitude of the projection of fixed vector \mathbf{x} onto random P -dimensional orthonormal subspace U_P projection term by term. Let $\boldsymbol{\alpha} \in \mathbb{R}^P$ be a column vector defined as $\boldsymbol{\alpha} = U_P^T \mathbf{x}$ and find the ℓ^2 -norm squared:

$$\begin{aligned} \|\boldsymbol{\alpha}\|_2^2 &= \alpha_1^2 + \alpha_2^2 + \dots + \alpha_P^2 \\ &= \|U_P^T \mathbf{x}\|_2^2 = \mathbf{x}^T U_P U_P^T \mathbf{x}, \end{aligned} \quad (43)$$

where

$$\begin{aligned} \alpha_j^2 &= \langle \widehat{\mathbf{u}}_j, \mathbf{x} \rangle^2 \\ &= (u_{j,1} \mathbf{x}_1 + \dots + u_{j,N} \mathbf{x}_N)^2 \\ &= \sum_{i,k}^{N,N} \frac{u_{j,k} u_{j,i} \mathbf{x}_k \mathbf{x}_i}{\|\mathbf{u}_j\|_2^2}. \end{aligned} \quad (44)$$

Let upper case $U_{j,k}$ denote the k th IID element random (this is not the same k -variable as the Kaczmarz iteration variable) variable of the j th column vector \mathbf{U}_j associated with column vector \mathbf{u}_j ; let \mathbf{x} vector denote a fixed point. Next, take the expectation of the term over the possible outcomes of $U_{j,k}$ random variables. Using the IID assumption, the expected value for a single projection component preserves terms squared as follows:

$$\begin{aligned} \mathbb{E} [\alpha_j^2] &= \mathbb{E} \left[\sum_{i,k}^{N,N} \frac{U_{j,k} U_{j,i} \mathbf{x}_k \mathbf{x}_i}{C_\sigma^2} \right] = \sum_{i,k}^{N,N} \mathbb{E} \left[\frac{U_{j,k} U_{j,i} \mathbf{x}_k \mathbf{x}_i}{C_\sigma^2} \right] \\ &= \sum_{k=1}^N \mathbb{E} \left[\frac{U_{j,k}^2 \mathbf{x}_k^2}{C_\sigma^2} \right] = \sum_k^N \mathbb{E} \left[\frac{U_{j,k}^2}{C_\sigma^2} \right] \mathbf{x}_k^2 \\ &= \mathbb{E} \left[\frac{U_{j,k}^2}{C_\sigma^2} \right] \sum_k^N \mathbf{x}_k^2 = \mathbb{E} \left[\frac{U_{j,k}^2}{C_\sigma^2} \right] \|\mathbf{x}\|_2^2 \\ &= \frac{1}{C_\sigma^2} \frac{C_\sigma^2}{N} \|\mathbf{x}\|_2^2 \\ &= \frac{1}{N} \|\mathbf{x}\|_2^2. \end{aligned} \quad (45)$$

It is now possible to determine the expectation for P -terms of the projection as

$$\mathbb{E} [\|\boldsymbol{\alpha}\|_2^2] = \mathbb{E} \left[\sum_{j=1}^P \alpha_j^2 \right] = \frac{P}{N} \|\mathbf{x}\|_2^2 \quad (46)$$

subject to IID constraint on $\widehat{\mathbf{U}}_j$ where it is further noted that $\sigma^2 N = C_\sigma^2$ in (42).

(e) *Error per Iteration.* For a given k th Kaczmarz iteration, the expectation of the projection of fixed vector \mathbf{x} onto the random P -dimensional subspace U_P is known from above. The total convergence expectation may then be computed, using a method similar to Strohmer's, starting (recall that derivation of (47) requires orthogonality among the $\widehat{\mathbf{u}}_l$ subspace basis vectors) with (47):

$$\|\mathbf{z}_{k+1}\|_2^2 = \|\mathbf{z}_k\|_2^2 - \sum_{l=1}^P |\langle \mathbf{z}_k, \widehat{\mathbf{u}}_l \rangle|^2 \quad (47)$$

$$\begin{aligned}
& \mathbb{E}_{\{k+1|z_0, z_1, \dots, z_k\}} \left[\|\mathbf{z}_{k+1}\|_2^2 \right] \\
&= \mathbb{E}_{\{k+1|z_0, z_1, \dots, z_k\}} \left[\|\mathbf{z}_k\|_2^2 - \sum_{l=1}^P |\langle \mathbf{z}_k, \hat{\mathbf{u}}_l \rangle|^2 \right] \\
&= \mathbb{E}_{\{k+1|z_0, z_1, \dots, z_k\}} \left[\|\mathbf{z}_k\|_2^2 \right] - \mathbb{E}_{\{k+1|z_0, z_1, \dots, z_k\}} \left[\sum_{l=1}^P |\langle \mathbf{z}_k, \hat{\mathbf{u}}_l \rangle|^2 \right]. \tag{48}
\end{aligned}$$

We identify the term on the right as

$$\begin{aligned}
\mathbb{E}_{\{k+1|z_0, z_1, \dots, z_k\}} \left[\sum_{l=1}^P |\langle \mathbf{z}_k, \hat{\mathbf{u}}_l \rangle|^2 \right] &= \mathbb{E}_{\{k+1|z_0, z_1, \dots, z_k\}} \left[\|U_P \mathbf{z}_k\|_2^2 \right] \\
&= \frac{P}{N} \times \mathbb{E}_{\{k+1|z_0, z_1, \dots, z_k\}} \left[\|\mathbf{z}_k\|_2^2 \right]. \tag{49}
\end{aligned}$$

The results from the two equations ((49) and (48)) above may then be combined to obtain

$$\begin{aligned}
& \mathbb{E}_{\{k+1|z_0, z_1, \dots, z_k\}} \left[\|\mathbf{z}_{k+1}\|_2^2 \right] \\
&= \left(1 - \frac{P}{N} \right) \times \mathbb{E}_{\{k|z_0, z_1, \dots, z_{k-1}\}} \left[\|\mathbf{z}_k\|_2^2 \right], \tag{50}
\end{aligned}$$

where the expectation on the right hand side includes $k+1 \rightarrow k$ accounting for the previous iteration. Next, apply induction to arrive at the expectation for the whole iterative sequence up to the β th iteration given that $\mathbf{z}_0 \equiv \mathbf{x} - \mathbf{x}_0$:

$$\mathbb{E}_{\{\beta+1|z_0\}} \left[\|\mathbf{z}_{\beta+1}\|_2^2 \right] = \left(1 - \frac{P}{N} \right)^\beta \|\mathbf{z}_0\|_2^2 \quad \forall \beta \in 1, 2, 3, \dots \tag{51}$$

(f) *Asymptotic Convergence.* The statistical ensemble average of the above equation (34) for the β th iteration yields the convergence result given in (51). These results assume random variables identically and independently distributed but compare well to others in the literature, such as the convergence result in Strohmer and Vershynin [20].

The theoretical convergence iterative limit for uniform random IID sampling was compared to numerical simulations using random solution vector point on a unit sphere. Equation (52) has an asymptotic form:

$$\begin{aligned}
& \frac{\mathbb{E}_{\{\beta+1|z_0\}} \left[\|\mathbf{z}_{\beta+1}\|_2^2 \right]}{\|\mathbf{z}_0\|_2^2} \\
&= \lim_{\beta \rightarrow \infty} \left[1 - \frac{P}{N} \right]^\beta \simeq e^{-\beta P/N}, \tag{52}
\end{aligned}$$

$$P = \dim(S_P), \quad \beta \gg 1, 2, 3, \dots \rightarrow k \in P, 2P, 3P, \dots$$

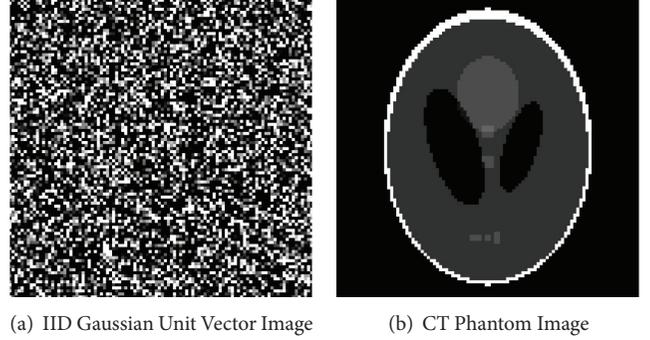


FIGURE 1: Representative test data.

For comparison, recall the convergence for RK method of Strohmer for IID measurements with $R = N$ which is approximately

$$\frac{\mathbb{E}_{\{k+1|z_0\}} \left[\|\mathbf{z}_{k+1}\|_2^2 \right]}{\|\mathbf{z}_0\|_2^2} = \left[1 - \frac{1}{N} \right]^k \tag{53}$$

$$\lim_{k \rightarrow \infty} \left[1 - \frac{1}{N} \right]^k \simeq e^{-k/N} \quad \forall k \gg 1, 2, 3, \dots$$

Estimated noise bound convergence complexity to ϵ error is $O(N^2)$. Since the value of \mathbf{z}_0 is given, the expectation is known to be the same.

(g) *Theory and Simulation.* Simulations in [16] compare theory to Gaussian IID with noise variance added to the measurements with magnitude $\beta = 0.05$ (about five percent) and iteration termination at $\beta = 0.05/4 = 0.0125$. In the first problem, the exact solution \mathbf{x} is chosen as a random point on the unit sphere—which is illustrated in Figure 1(a). In a second problem, a measurement of the standard phantom using parallel beam measurements is included, which contains coherent measurements.

2.10. *Regular Versus Randomized Kaczmarz Method.* The randomized Kaczmarz's algorithm developed by Strohmer and Vershynin in [4] has the following convergence in expectation:

$$\mathbb{E} \left\| \mathbf{x}_{qM} - \mathbf{x}^* \right\|_2^2 \leq \left(1 - \frac{1}{\kappa(A)^2} \right)^{qM} \left\| \mathbf{x}_0 - \mathbf{x}^* \right\|_2^2, \tag{54}$$

where $\kappa(A) = \|A\|_F \|A^\dagger\|_2$ is the scaled condition number of matrix A with A^\dagger as the left pseudoinverse of A . The bound for regular Kaczmarz is given in (25). Note that we assume $A \in \mathbb{R}^{M \times M}$. Now, we need to compare $(1 - (1/\|A\|_F^2 \|A^\dagger\|_2^2))$ and $(1 - \det(AA^T))^{1/M}$ to assess which bound is tighter. Let $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_M > 0$ be ordered singular values of A . Then,

$$\begin{aligned}
\|A^\dagger\|_2^2 &= \frac{1}{\sigma_N^2}, \\
\|A\|_F^2 &= \sum_{i=1}^M \sigma_i^2. \tag{55}
\end{aligned}$$

Also, note that

$$AA^T = \begin{bmatrix} 1 & \cos \theta_{12} & \cdots & \cos \theta_{1M} \\ \cos \theta_{21} & 1 & \cdots & \cos \theta_{2M} \\ \vdots & \vdots & \cdots & \vdots \\ \cos \theta_{M1} & \cos \theta_{M2} & \cdots & 1 \end{bmatrix}, \quad (56)$$

where θ_{ij} denotes the angles between the rows a_i and a_j of A . Then,

$$\det(AA^T) \leq \prod_{i=1}^M \sum_{j=1}^M \cos^2 \theta_{ij}. \quad (57)$$

Note that

$$\prod_{i=1}^M \sigma_i^2(A) = \prod_{i=1}^M \lambda_i(A^T A) = \det(A^T A) = \det(AA^T); \quad (58)$$

therefore,

$$[1 - \det(AA^T)]^{1/M} = \left(1 - \prod_{i=1}^M \sigma_i^2\right)^{1/M}. \quad (59)$$

Now, (54) and (25) become

$$\begin{aligned} \mathbb{E} \|\mathbf{x}_{qM} - \mathbf{x}^*\|_2^2 &\leq \left(1 - \frac{\sigma_M^2}{\sum_{i=1}^M \sigma_i^2}\right)^{qM} \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2, \\ \|\mathbf{x}_{qM} - \mathbf{x}^*\|_2^2 &\leq \left[\left(1 - \prod_{i=1}^M \sigma_i^2\right)^{1/M}\right]^{qM} \|\mathbf{x}_0 - \mathbf{x}^*\|_2^2. \end{aligned} \quad (60)$$

2.11. Complexity Measurement Methods. In order to visualize and assess the relative performance of the well-known common methods of Kaczmarz, a simple routine was written in Octave [21] to record the total central processing unit (CPU) times for each method and variable matrix sizes of interest. The Linux (tm) kernel was modified to include the real-time patches for the x86/64 CPU architecture, and the process was run on a (see the *cset* in the *cpuset* package) shielded CPU set to single processor unit CPU7 to avoid process contention and interrupts. All system functions were locked to other CPU instances and not allowed to execute on CPU7.

Execution times were measured with *rusage()* calls before and after calls to each method in Octave. The elapsed time has microsecond resolution from the *rusage()* function. In addition, the interrupt balance kernel function was disabled. The status of the CPU cores and interrupts may be observed in `/proc/interrupts`. The CPU clock frequency was locked to a fixed value of 800 Mhz for the x86_64 system. Virtual memory was dropped prior to execution to ensure maximum physical memory availability.

In addition to the 64-bit platform, an embedded 32-bit ARM architecture was also configured for computational reference. Due to time constraints, no real-time kernel was implemented for the ARM. The CPU clock frequency was locked to a fixed value of 840 Mhz for the embedded ARM system.

In both of the above benchmark cases, all noncritical network and file system processes and daemons were stopped prior to code execution. In both cases, program execution was monitored and noted to reside in virtual (nonswaped) memory mapped to physical memory.

2.11.1. Methods. The *kaczmarz()*, *randkaczmarz()*, and *cimmino()* methods were called directly from the AIR toolkit [22]. The functions for the block method *chenko()* were implemented from Vasilchenko and Svetlakov [23] of the form

$$\mathbf{x}_{j+1} = \mathbf{x}_j + A_i^T (A_i A_i^T)^{-1} (b_i - A_i \mathbf{x}_j) \quad i = j \pmod{k} + 1 \quad (61)$$

for $A_i = E_i^T A$.

The *rkos()* was implemented from [16] algorithm. The function for least squares *ls()* was computed from the equation

$$\mathbf{x}^* = (A^T A)^{-1} A^T \mathbf{b}. \quad (62)$$

The function for singular value decomposition *svd()* used the decomposition of the sampling matrix $A = U\Sigma V^T$, and the solution was computed as follows:

$$\mathbf{x}^* = V\Sigma^{-1}U^T A \mathbf{x}^* = V\Sigma^{-1}U^T \mathbf{b}. \quad (63)$$

The resulting plots are intended to illustrate how the methods scale with increasing measurement matrix size and not directly used in absolute terms, since each method has various dependencies in hardware and software.

A simple baseline Kaczmarz code was written to ensure a baseline reference was available. The Octave source code for a typical method is included in the inset Listing 1.

2.11.2. Notes Regarding Simulations. The sizes of the matrix blocks equal matrix dimensions until constant block size of sixteen rows. The legend key is as follows:

- (1) least squares: *LS()*,
- (2) singular value decomposition: *SVD()*,
- (3) *Kaczmarz()*,
- (4) randomized Kaczmarz: *RK()*,
- (5) block method randomized Kaczmarz orthogonal subspaces: *RKOS()*,
- (6) *Cimmino()*,
- (7) block method *Vasilchenko()*,
- (8) reference P1: simple matrix multiply with known $O(N^2)$ complexity,
- (9) reference P2: N times P1 for known $O(N^3)$ complexity,
- (10) reference Kaczmarz as shown in Listing 1.

```

(1) function [X info] = kazbase (A, b)
    % note that matrix A is assumed to have normalized rows in below.
    load x_solution.data x_sol
    x_exact = x_sol
    load convergence_limit.data
(6) convergence_limit
    myKmax = ceil(-log(convergence_limit))
    ntimes = myKmax
    [M, N] = size (A)
    jj = 0
(11) xk0 = zeros (M, 1);
    x0 = xk0;
    for n = 1: ntimes
        for kn = 1: M
            jj = jj + 1;
(16) %%%%%%%%%%%%%%%
            delta_xk_plus_10 = (eye (N) - A(kn, :)’ * A(kn, :)) * (x_exact - xk0);
            xk0 = x_exact - delta_xk_plus_10;
            %error0 (jj) = norm(delta_xk_plus_10, 2)/norm(x_exact - x0, 2);
            %%%%%%%%%%%%%%%
(21) end
    end
    X = xk0

```

LISTING 1: Reference Kaczmarz Baseline.

3. Results and Discussion

Here, we compare our angle-based randomization with norm-based randomization of Strohmer and Vershynin [4] in the context of uniform and nonuniform measurement methods. In particular, Shepp-Logan and *sinc2d()* phantom images were used as the solutions in four different simulation experiments [3]. Figure 3(d) shows that our approach (angle-based randomization) provides a better convergence rate over the randomized Kaczmarz (norm-based randomization) in the case of fan-beam sampling of the *sinc2d()* phantom. However, our method is computationally more complex, and therefore we devised the P -subspace algorithm (presented in the previous section).

In the experiments which follow, three factors of interest are the sampling angular distribution of the measurements (affects coherence), the algorithms’ method of iteration through the sampling hyperplanes, and the rate of convergence of the solution. The simulations were computed in parallel for each of the methods: Kaczmarz (K), randomized Kaczmarz hyperplane angles (RKHA), and randomized Kaczmarz (RK). Iteration is terminated at stopping point defined as the condition when at least one of the three methods attains a normalized error of 10% or less. Estimates for the signal to noise ratio are computed at same common stopping point.

3.1. Sampling Distributions. The following experiments compare Kaczmarz (K), randomized Kaczmarz (RK), and randomized Kaczmarz hyperplane angles (RKHA) via simulations. The objectives are to illustrate the effect of

row randomization upon the convergence and observe the dependency upon the sampling methods.

3.1.1. Angular Distribution of Sampling Hyperplanes. A comparison of the distribution of hyperplane sampling angles in computed tomography (CT) was performed to investigate the convergence rate versus measurement strategy. Example results are presented for iterative convergence of methods K, RK, and RKHA under conditions of random, fan, and parallel beam sampling strategies using the Shepp-Logan phantom (see Figure 1(b)) [22], *paralleltomo.m* and *fanbeamtomo.m* from the AIRtools distribution [22] and *randn()* from the built-in function method [24]. The typical results for the angular distributions are provided in Figures 2(a) and 2(b) and discussed in more detail below.

3.1.2. Sampling Coherence. In linear algebra, the coherence or mutual coherence [25] of a row measurement matrix A is defined as the maximum absolute value of the cross-correlations between the normalized rows of A .

Formally, let $\{\mathbf{a}_1, \dots, \mathbf{a}_M\} \in \mathbb{R}^N$ be the set of row vectors of the matrix $A \in \mathbb{R}^{M \times N}$ normalized such that $\langle \mathbf{a}_i, \mathbf{a}_i \rangle = \mathbf{a}_i^H \mathbf{a}_i = 1$ where $(\cdot)^H$ is the Hermitian conjugate and where $M > N$. Let the mutual coherence of A be defined as

$$\phi_{i,j} = \max_{1 \leq i \neq j \leq M} |\mathbf{a}_i^H \mathbf{a}_j|. \quad (64)$$

A lower bound was derived as $\phi \geq (M - N)/N(M - 1)$ in Welch [26].

In the distributions of Figures 2(a) and 2(b), it should be noted that the random sampling is concentrated near

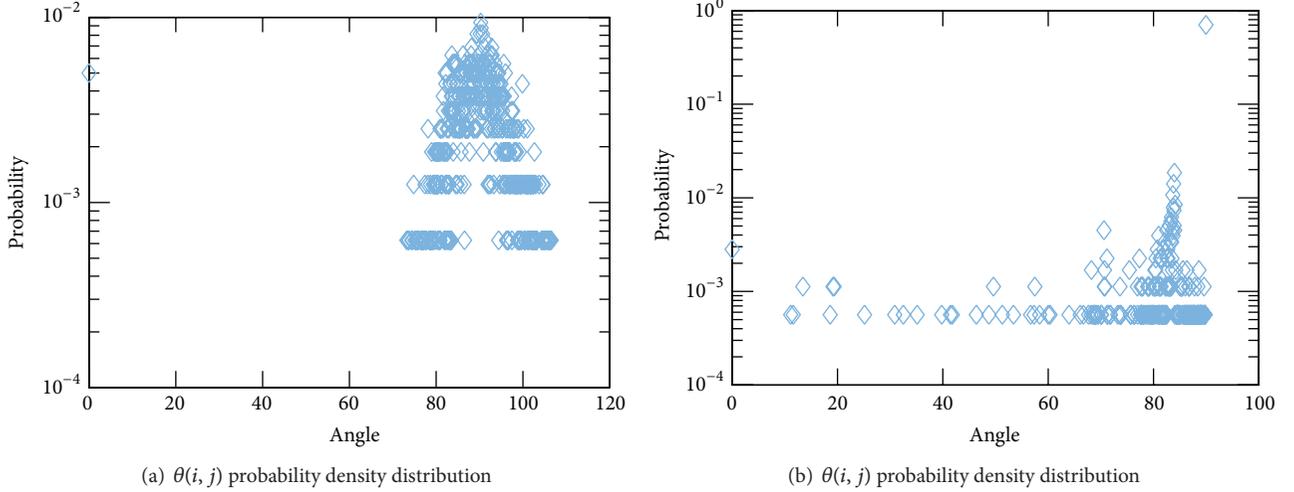


FIGURE 2: Example angles distribution (y -axis) from AA^T where $\theta_{i,j} = \cos^{-1}(\langle \hat{\mathbf{a}}_i, \hat{\mathbf{a}}_j \rangle) \forall i, j \in \{1, \dots, M\}$ versus angles (x -axis) degrees using (a) random and (b) fan-beam data acquisition strategy where the unit norm vectors $\hat{\mathbf{a}}_i, \hat{\mathbf{a}}_j$ are selected rows of matrix A .

90 degrees probability but fan sampling is less concentrated across the interval $[0, 90]$ degrees.

3.1.3. Observations on Effects of Sampling Distribution, Algorithm, and Convergence. Firstly, the convergence rates of K, RK, and RKHA are noted to be closely correlated for the case of random data sampling of the phantom in Figures 3(a) and 3(b). This is consistent with the mean values of coherence near zero for random sampling.

The cases for fan and parallel sampling have increasingly higher coherence and increasing numbers of iterations required to meet the 10% error stopping condition. Both test images for fan-beam sampling converge in example Figures 3(c) and 3(d) and show slight benefit from the methods which minimize the coherence, such as RK, RKHA, and RKOS. Example quantitative results for the two cases of fan sampling are shown in (a) Shepp-Logan Table 3 and Figure 4 and (b) $\text{sinc}2d()$ Table 4 and Figure 5.

Comparison of convergence results to the estimated coherence for the three cases given in Table 5 suggest consistent interpretation. Comparison of percent error and SNR of Tables 1, 2, 3, and 4 indicate the randomization methods have slight advantage under coherent fan sampling, providing increased SNR and lower percent error.

3.1.4. Potential Applications. Since the iterative methods utilize projections, the angles between the optical lines-of-sight (LOS) forming the measurement hyperplanes are of considerable interest in terms of data acquisition and system design. Most methods of computed tomography do not reasonably allow random or orthogonal data sampling of the object of interest.

Therefore, these systems which acquire coherent data may benefit from use of the randomized methods RK, RKHA, or RKOS in data inversions. Typical examples of such systems may include computed tomography in medical and

TABLE 1: Random sampling of Shepp-Logan SNR and percent error at stopping.

Performance estimates	Kaczmarz	RKHA	RK
SNR	10.1	5.60	4.5
% error	9.9	19.9	22.1

TABLE 2: Random sampling of $\text{sinc}2d()$ SNR and percent error at stopping.

Performance estimates	Kaczmarz	RKHA	RK
SNR	10.0	5.6	5.2
% error	9.9	17.6	19.1

TABLE 3: Fan beam sampling of Shepp Logan, SNR, and percent error at stopping.

Performance estimates	Kaczmarz	RKHA	RK
SNR	10.0	9.95	9.5
% error	9.9	10.0	10.5

TABLE 4: Fan beam sampling of $\text{sinc}2d()$ SNR and percent error at stopping.

Performance estimates	Kaczmarz	RKHA	RK
SNR	7.1	10.0	9.2
% error	14.0	9.9	10.7

nonmedical X-ray, transmission ultrasound [27, 28], and resonance optical absorption and molecular fluorescence *in-vivo* imaging [29]. Each of the aforementioned systems are potentially feasible applications of RKHA or RKOS, since in each case, the measurements are path integrated along a mostly nonorthogonal set of electromagnetic LOS's and

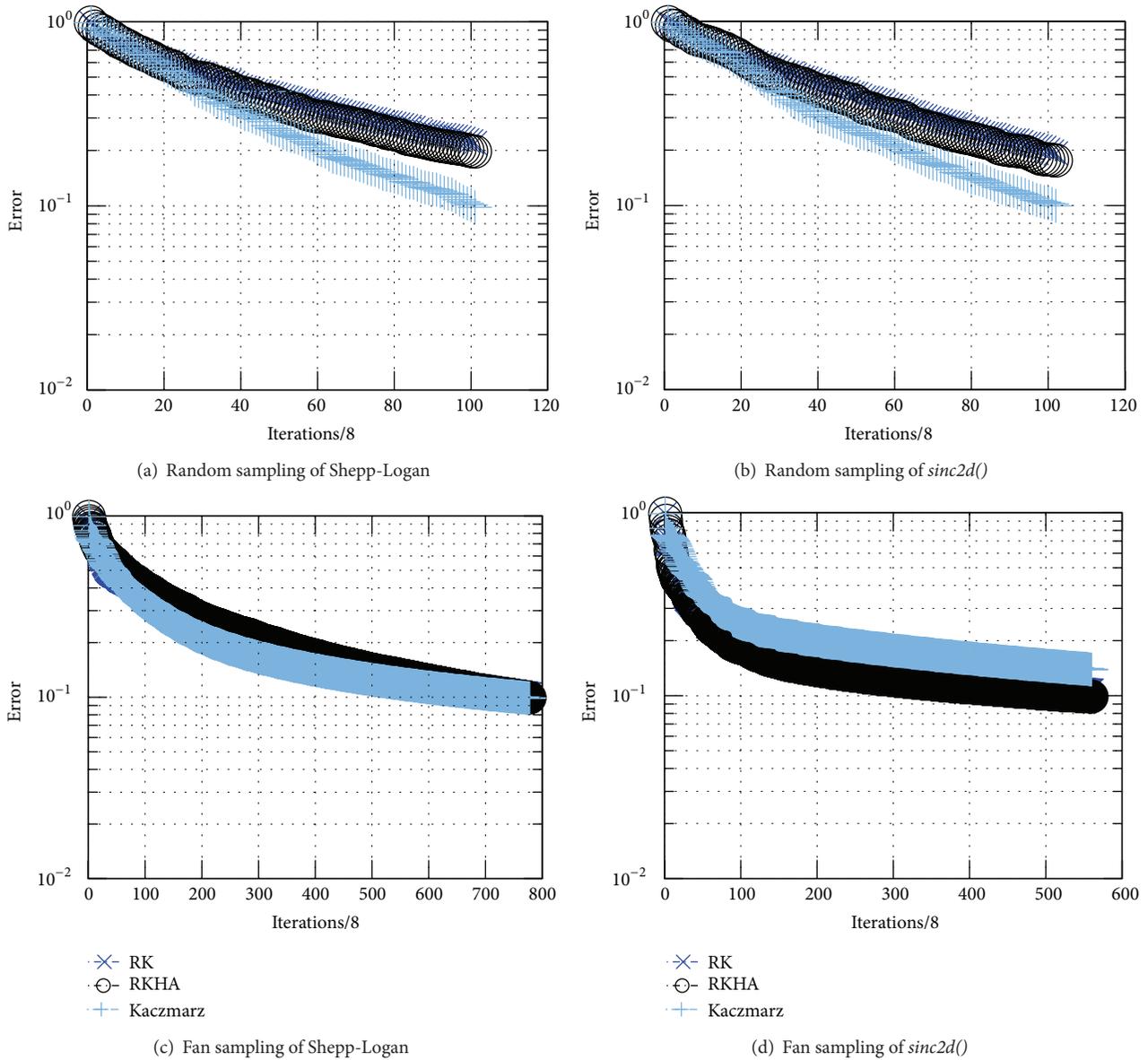


FIGURE 3: Semilog (y -axis) plot example of normalized convergence error results for K, RK, and RKHA on Shepp-Logan phantom using random and fan tomographic data acquisition sampling and stopping at 10 percent error.



FIGURE 4: Reconstructed test images for Shepp-Logan with fan sampling with error from left to right: exact, K (9.9% error), RKHA, and RK Shepp Logan.

generally require inversion to obtain the parameters of interest, such mole-fraction or species density.

3.2. Experimental Results for Convergence of K, RK, and RKHA. Iterative simulations were performed to estimate the

relative convergence rates of methods K, RK, RKHA for the data examples above: random and fan beam sampling. Representative results are shown in Figures 3(a), 3(b), 3(c), and 3(d) for noiseless data measurement scenarios of the standard Shepp-Logan phantom and $\text{sinc}2d()$ solutions under random

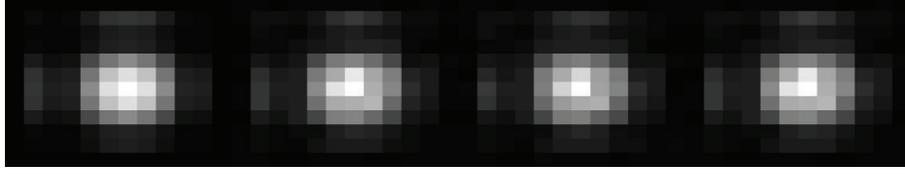


FIGURE 5: Reconstructed test images for $\text{sinc2d}()$ with fan sampling with error from left to right: exact, K, RKHA (9.9%), and RK.

TABLE 5: Typical coherence estimates for $N = 100, M = 200$ for random $\text{randn}()$ and $N = 100, M = 222$ for fan $\text{fanbeamtomo}()$ and parallel $\text{paralleltomo}()$.

Coherence versus measurement method	Random	Fan	Parallel
Coherence (64)	.4	1.0	1.0
Average value of $G_{i,j} = \langle \hat{\mathbf{a}}_i, \hat{\mathbf{a}}_j \rangle$ ($1 \leq i \neq j \leq M$)	-.0013	.06	.18
Median value of $G_{i,j} = \langle \hat{\mathbf{a}}_i, \hat{\mathbf{a}}_j \rangle$ ($1 \leq i \neq j \leq M$)	-0.0009	0	.12

and fan-beam sampling simulations. Parallel beam sampling and $\text{rect_sinc2d}()$ solution vectors were also simulated. The results were consistent with observations reported herein but not included in this report.

3.2.1. Test Images for Algorithms. The Shepp Logan image was generated from the phantom code in reference [22], which is also the source for the fan-beam sampling algorithm. The $\text{sinc2d}()$ function is defined as

$$\text{sinc2d}(x, y) = \frac{\sin(\pi x)}{\pi x} \frac{\sin(\pi y)}{\pi y} \quad (65)$$

and was computed as the outer-product of two independent vectors constructed from the included code `sinc.m` in the signal processing package of Octave [21]. The rectangular `rect_sinc2d.m` function was used from source (https://engineering.purdue.edu/VISE/ee438/demos/2D_signals_systems/rect_sinc2d.m) which computes the two-dimensional $\text{sinc}()$ function from the Fourier transform of the two-dimensional rectangle function.

3.2.2. Image Reconstruction Error. A comparison of the toy phantoms was performed based upon the estimated signal to noise ratio and total percent error versus iteration. Percent error and SNR were estimated for the Shepp-Logan phantom and the artifact based upon first method to obtain stopping error at 10% percent normalized error.

The normalized error is defined as

$$\epsilon_k = \frac{\|\mathbf{x} - \mathbf{x}_k\|_2}{\|\mathbf{x} - \mathbf{x}_0\|_2}, \quad (66)$$

and estimate the signal to noise ratio (SNR) is estimated as $1/\epsilon_k$ with the \mathbf{x}_0 vector set to the the zero vector.

TABLE 6: Numerical estimates for computational complexity using naive codes.

Computational complexity	LS	SVD	Kaczmarz	RK	RKOS
Theoretical	$O(N^3)$	$O(N^3)$	$O(N^2)$	$O(N^2)$	$O(N^2)$
Numerical estimates	2.7	3.1	1.1	1.1	1.1

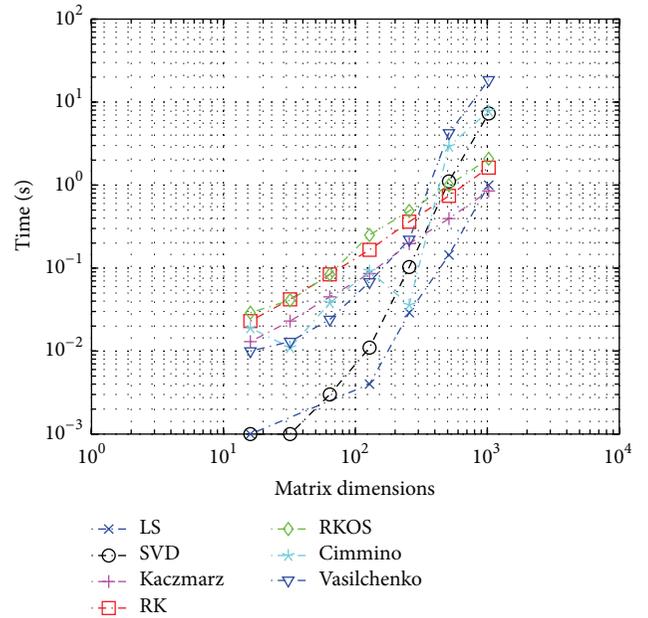


FIGURE 6: Relative computational time using dedicated custom real-time Linux kernel 3.12.5 patched versus square matrix dimensions $N = 2^q \forall q \in [4, \dots, 10]$; CPU is shielded and frequency locked for single process and single core (one of eight) 800 MHz Intel 64-bit i7.

3.3. Experimental Results for Complexity Estimates. Timings and plots of the methods were computed for a range of matrix sizes and plotted on log-log plots as shown in Figures 6 and 7, with measured run-times for 64-bit and 32-bit, respectively. For each method, the size of A matrix was modified and average time to complete was measured. The data matrix was chosen to be $\text{hadamard}()$ function for ease of implementation and reference. The solution vector was chosen as a random point in the binomial distribution with $p = 0.5$ using the $\text{binornd}()$ function.

It is notable that the theoretical complexity of noise limited Kaczmarz is $O(N^2)$ but the slope of the timings is closer to 1.1 than 2.0. This is attributed to relatively small matrix

TABLE 7: Numerical estimates for computational complexity using naive codes.

Computational complexity	Cimmino	Vasilchenko	Ref. P1	Ref. P2	Ref. Kaczmarz
Theoretical	$O(N^2)$	$O(N^2)$	$O(N^2)$	$O(N^3)$	$O(N^2)$
Numerical estimates	1.6	2.1	2	3.2	1.1

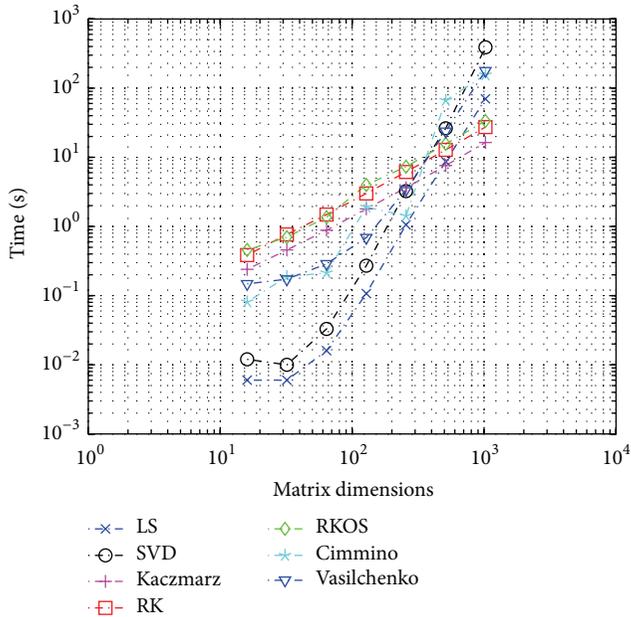


FIGURE 7: Relative computational time for embedded microprocessor dedicated task custom Linux kernel 3.12.7 ARMv7 relative computational time versus matrix dimensions $N = 2^q \forall q \in [4, \dots, 10]$; CPU ARM frequency locked kernel only process stack 840 MHz 32-bit.

sizes, and it is expected that the complexity asymptotically approaches ~ 2.0 for large N .

The theoretical and numerical estimates for complexity are shown in Tables 6 and 7.

4. Conclusions

A new iterative selection rule based upon the relative central angle (RKHA) shows enhanced convergence in measurements which contain coherence. However, the method requires a computational penalty related to the dot-products of all to all rows, which may be overcome by *a priori* determination. A new block method using constructed orthogonal subspace projections provides enhanced tolerance to measurement coherence, but may be affected by noise at least as much as simple Kaczmarz. The exponential convergence is accelerated by the P/N term and is computationally feasible for small P relative to N .

The theoretical convergence rates of above subspace methods were demonstrated using statistical IID assumptions or cyclical projections using the formalism of Galantai. Numerical results were presented from simulations of algorithm convergence under measurement distributions for

fan and parallel X-ray beams, as well as random uniform sampling.

Relative performance benchmarks for complexity were obtained for typical hardware and software for various methods of well-known Kaczmarz algorithms. As expected, the Kaczmarz methods out-perform other methods, such as least-squares and singular value decomposition. The performance of embedded 32-bit ARM CPU architecture was sufficient to demonstrate functional capability over a range of low-power application environments such as mobile medicine platforms.

The new angle-based method (RKHA) and orthogonal block (RKOS) inversion method demonstrated herein showed quantitative convergence improvement consistent with increasing orthogonality and decreasing coherency of measurements. Future designs for tomography should consider optimization of angular sampling distributions in addition to other factors, such signal to noise ratio, as important system parameters, since these criteria ultimately affect the spatial-temporal resolution and uncertainty for a given number of samples per unit volume.

Conflict of Interests

The authors declare that there is no known conflict of interests regarding the publication of this paper.

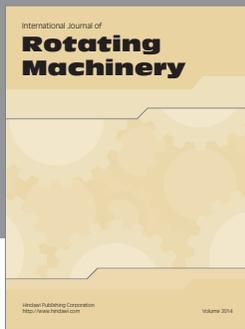
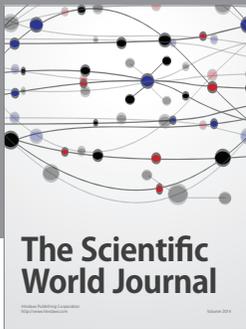
Acknowledgments

The authors thank Akram Aldroubi and Alex Powell for their invaluable feedback. The research of Ali Sekmen is supported in part by NASA Grant NNX12AI14A.

References

- [1] S. Kaczmarz, "Approximate solution of systems of linear equations," *International Journal of Control*, vol. 57, no. 6, pp. 1269–1271, 1993.
- [2] R. Gordon, R. Bender, and G. T. Herman, "Algebraic reconstruction techniques (art) for three-dimensional electron microscopy and x-ray photography," *Journal of Theoretical Biology*, vol. 29, no. 3, pp. 471–481, 1970.
- [3] G. T. Herman, *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*, Springer, 2nd edition, 2009.
- [4] T. Strohmer and R. Vershynin, "A randomized Kaczmarz algorithm with exponential convergence," *The Journal of Fourier Analysis and Applications*, vol. 15, no. 2, pp. 262–278, 2009.
- [5] D. Needell, "Randomized Kaczmarz solver for noisy linear systems," *BIT Numerical Mathematics*, vol. 50, no. 2, pp. 395–403, 2010.

- [6] D. Needell and R. Ward, "Two-subspace projection method for coherent overdetermined systems," *The Journal of Fourier Analysis and Applications*, vol. 19, no. 2, pp. 256–269, 2013.
- [7] D. Needell and J. A. Tropp, "Paved with good intentions: analysis of a randomized block Kaczmarz method," *Linear Algebra and its Applications*, vol. 441, pp. 199–221, 2014.
- [8] X. Chen and A. M. Powell, "Almost sure convergence of the Kaczmarz algorithm with random measurements," *Journal of Fourier Analysis and Applications*, vol. 18, no. 6, pp. 1195–1214, 2012.
- [9] A. Galántai, "On the rate of convergence of the alternating projection method in finite dimensional spaces," *Journal of Mathematical Analysis and Applications*, vol. 310, no. 1, pp. 30–44, 2005.
- [10] A. Galántai, *Projectors and Projection Methods*, Advances in Mathematics, Kluwer Academic, London, UK, 2004.
- [11] F. Deutsch and H. Hundal, "The rate of convergence for the method of alternating projections, II," *Journal of Mathematical Analysis and Applications*, vol. 205, no. 2, pp. 381–405, 1997.
- [12] C. Brezinski and M. Redivo-Zaglia, "Convergence acceleration of Kaczmarz's method," *Journal of Engineering Mathematics*, pp. 1–17, 2013.
- [13] K. T. Smith, D. C. Solmon, and S. L. Wagner, "Practical and mathematical aspects of the problem of reconstructing objects from radiographs," *Bulletin of the American Mathematical Society*, vol. 83, no. 6, pp. 1227–1270, 1977.
- [14] Z.-Z. Bai and X.-G. Liu, "On the Meany inequality with applications to convergence analysis of several row-action iteration methods," *Numerische Mathematik*, vol. 124, no. 2, pp. 215–236, 2013.
- [15] Y. Censor, G. T. Herman, and M. Jiang, "A note on the behavior of the randomized Kaczmarz algorithm of Strohmer and Vershynin," *Journal of Fourier Analysis and Applications*, vol. 15, no. 4, pp. 431–436, 2009.
- [16] T. Wallace and A. Sekmen, "Acceleration of Kaczmarz using subspace orthogonal projections," in *IEEE Biomedical Science and Engineering Conference (BSEC) at Oak Ridge National Laboratory*, 2013.
- [17] G. H. Golub and C. F. van Loan, *Matrix Computations*, Johns Hopkins Studies in Mathematical Sciences, The Johns Hopkins University Press, 3rd edition, 1996.
- [18] C. Meyer, *Matrix Analysis and Applied Linear Algebra*, Society for Industrial and Applied Mathematics, Philadelphia, Pa, USA, 2000.
- [19] H. Yanai, K. Takeuchi, and Y. Takane, *Projection Matrices, Generalized Inverse Matrices, and Singular Value Decomposition*, Statistics for Social and Behavioral Sciences, Springer, Dordrecht, The Netherlands, 2011.
- [20] T. Strohmer and R. Vershynin, "A randomized Kaczmarz algorithm with exponential convergence," *Journal of Fourier Analysis and Applications*, vol. 15, no. 2, pp. 262–278, 2009.
- [21] J. Eaton, W. Rik, D. Bateman, and S. Hauberg, *GNU Octave Version 3.8.1 Manual: A High-Level Interactive Language for Numerical Computations*, CreateSpace Independent Publishing Platform, 2014.
- [22] P. C. Hansen and M. Saxild-Hansen, "AIR-tools—a MATLAB package of algebraic iterative reconstruction methods," *Journal of Computational and Applied Mathematics*, vol. 236, no. 8, pp. 2167–2178, 2012.
- [23] G. P. Vasil'chenko and A. A. Svetlakov, "A projection algorithm for solving systems of linear algebraic equations of high dimensionality," *USSR Computational Mathematics and Mathematical Physics*, vol. 20, no. 1, pp. 1–8, 1980.
- [24] G. Marsaglia and W. W. Tsang, "The ziggurat method for generating random variables," *Journal of Statistical Software*, vol. 5, no. 8, pp. 1–7, 2000.
- [25] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 6–18, 2006.
- [26] L. R. Welch, "Lower bounds on the maximum cross correlation of signals (corresp.)," *IEEE Transactions on Information Theory*, vol. 20, no. 3, pp. 397–399, 1974.
- [27] M. Birk, R. Dapp, N. V. Ruiter, and J. Becker, "GPU-based iterative transmission reconstruction in 3D ultrasound computer tomography," *Journal of Parallel and Distributed Computing*, vol. 74, no. 1, pp. 1730–1743, 2014.
- [28] I. Peterlik, N. Ruiter, and R. Stotzka, "Algebraic reconstruction technique for ultrasound transmission tomography," in *Proceedings of the IEEE International Conference on Information Technology Applications in Biomedicine (ITAB 2006)*, Ioannina, Greece, October 2006.
- [29] V. Ntziachristos, "Fluorescence molecular imaging," *Annual Review of Biomedical Engineering*, vol. 8, no. 1, pp. 1–33, 2006.



Hindawi
Submit your manuscripts at
<http://www.hindawi.com>

