6-21-2014

Deterministic Versus Randomized Kaczmarz Iterative Projection

Tim Wallace
Tennessee State University

Ali Sekmen
Tennessee State University

Recommended Citation

Tim Wallace, Ali Sekmen, "Deterministic Versus Randomized Kaczmarz Iterative Projection", arXiv:1407.5593 https://doi.org/10.48550/arXiv.1407.5593

Deterministic Versus Randomized Kaczmarz Iterative Projection

Tim Wallace and Ali Sekmen†

Department of Computer Science

Tennessee State University

Nashville, TN USA

Abstract

The Kaczmarz's alternating projection method has been widely used for solving a consistent (mostly over-determined) linear system of equations $A\boldsymbol{x} = \boldsymbol{b}$. 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. It was recently shown that picking a row at random, proportional with its norm, makes the iteration converge exponentially in expectation with a decay constant that depends on the scaled condition number of A and not the number of equations. 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 explain the link between them. It also propose new versions of randomization that may speed up convergence.

I. Introduction

Kaczmarz (in [1]) introduced an iterative algorithm for solving a consistent linear system of equations $A\boldsymbol{x} = \boldsymbol{b}$ with $A \in \mathbb{R}^{M\times N}$. This method projects the estimate $\boldsymbol{x}^j$ onto a subspace normal to the row a_i at step $j+1$ cyclically with $i = j \pmod M + 1$. The block Kaczmarz algorithm first groups the rows into matrices $A_1, A_2, \ldots, A_k$ and then it projects the estimate $\boldsymbol{x}^j$ onto the subspace normal to the subspace spanned by the rows of A_i at step $j+1$ cyclically with $i = j \pmod k + 1$. Obviously, the block Kaczmarz is equivalent to the simple Kaczmarz for $k = M$. The Kaczmarz method is a method of alternating projection (MAP)

†Email: asekmen@tnstate.edu

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 *et al.* [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 $A\boldsymbol{x} = \boldsymbol{b}$. Needell also developed a randomized Kaczmarz method that improves the incoherency for iteration [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. Galantai (in [9], [10]) provides convergence analysis for block Kaczmarz method by expanding the convergence analysis (based on subspace angles) of Deutsch [11]. Brezinski (in [12]) utilizes the work of Galantai for accelerating convergence of regular Kaczmarz method.

A. *Paper Contributions*

- Research on regular and randomized Kaczmarz methods appear disconnected in the literature. Even though convergence rates have been studied separately, the link between them has not been explored sufficiently.
- A new randomization technique based on subspace angles has been developed which indicates an advantage with coherent data measurements.
- 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.
- The effects of measurement coherence are observed in the literature and illustrated in our simulations with norm and angle based iteration randomization.
- A broader review and mathematical analysis of common methods is presented from both statistical and deterministic perspectives.

II. Convergence of Regular Block Kaczmarz Method

Let $\boldsymbol{x}^*$ be the solution of consistent $A\boldsymbol{x} = \boldsymbol{b}$ where $A \in \mathbb{R}^{M\times M}$ is full column rank. Let A be row-partitioned as $\{A_1, \ldots, A_k\}$ where $A_i \in \mathbb{R}^{M_i \times M}$. Then, the simple block Kaczmarz

update is as follows:

$$\boldsymbol{x}_{j+1} = \boldsymbol{x}_j + A_i^T(A_iA_i^T)^{-1}(\boldsymbol{b}_i - A_i\boldsymbol{x}_j) \;\; i = j \pmod{k} + 1 \tag{1}$$

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

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

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

$$\boldsymbol{x}_{j+1} - \boldsymbol{x}^* = \boldsymbol{x}_j - \boldsymbol{x}^* - P_{Sp(A_i^T)}(\boldsymbol{x}_j - \boldsymbol{x}^*) \tag{2}$$

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

$$\boldsymbol{x}_{j+1} - \boldsymbol{x}^* = P_{Sp^{\perp}(A_i^T)}(\boldsymbol{x}_j - \boldsymbol{x}^*). \tag{3}$$

For one cycle of the blocks,

$$\boldsymbol{x}_k - \boldsymbol{x}^* = P_{Sp^{\perp}(A_k^T)}P_{Sp^{\perp}(A_{k-1}^T)}\ldots P_{Sp^{\perp}(A_1^T)}(\boldsymbol{x}_0 - \boldsymbol{x}^*). \tag{4}$$

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:

$$\boldsymbol{x}_{j+1} = \boldsymbol{x}_j + A_i^{\dagger}(\boldsymbol{b}_i - A_i\boldsymbol{x}_j) = \boldsymbol{x}_j + A_i^{\dagger}A_i(\boldsymbol{x}^* - \boldsymbol{x}_j) \;\; i = j \pmod{k} + 1 \tag{5}$$

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 Equation (2), and subsequently we get Equation (4),

$$\boldsymbol{x}_{j+1} - \boldsymbol{x}^* = \boldsymbol{x}_j - \boldsymbol{x}^* - P_{Sp(A_i^T)}(\boldsymbol{x}_j - \boldsymbol{x}^*). \tag{6}$$

A. Exponential Convergence

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

Proof: By Equation (2) and orthogonal projection,

$$\left\|\boldsymbol{x}_{j+1}-\boldsymbol{x}^*\right\|_2^2 = \left\|\boldsymbol{x}_j-\boldsymbol{x}^*\right\|_2^2 - \left\|P_{Sp(A_i^T)}(\boldsymbol{x}_j-\boldsymbol{x}^*)\right\|_2^2. \tag{7}$$

So,

$$\left\|\boldsymbol{x}_{j+1}-\boldsymbol{x}^*\right\|_2^2 \le \left\|\boldsymbol{x}_j-\boldsymbol{x}^*\right\|_2^2, \tag{8}$$

$\boldsymbol{x}_j-\boldsymbol{x}^*$ depends on the initial condition $\tilde{\boldsymbol{x}}_0 = \boldsymbol{x}_0-\boldsymbol{x}^*$, and this dependence is scale-invariant. To see this, let $\boldsymbol{e}_j = \boldsymbol{x}_j-\boldsymbol{x}^*$ and consider $c\tilde{\boldsymbol{x}}_0$ where $c\in\mathbb{R}$. By Equation (3),

$$\begin{aligned}
\boldsymbol{e}_{j+1}(c\tilde{\boldsymbol{x}}_0) &= P_{Sp^\perp(A_{j+1}^T)}\boldsymbol{e}_j(c\tilde{\boldsymbol{x}}_0)\\
&= P_{Sp^\perp(A_{j+1}^T)}P_{Sp^\perp(A_j^T)}\ldots P_{Sp^\perp(A_1^T)}\boldsymbol{e}_0(c\tilde{\boldsymbol{x}}_0)\\
&= P_{Sp^\perp(A_{j+1}^T)}P_{Sp^\perp(A_j^T)}\ldots P_{Sp^\perp(A_1^T)}(c\tilde{\boldsymbol{x}}_0)\\
&= cP_{Sp^\perp(A_{j+1}^T)}P_{Sp^\perp(A_j^T)}\ldots P_{Sp^\perp(A_1^T)}\boldsymbol{e}_0(\tilde{\boldsymbol{x}}_0)\\
&= c\boldsymbol{e}_{j+1}(\tilde{\boldsymbol{x}}_0).
\end{aligned} \tag{9}$$

We will first show that if $\boldsymbol{x}_0\neq\boldsymbol{x}^*$, then $\|\boldsymbol{x}_k-\boldsymbol{x}^*\|_2 < \|\boldsymbol{x}_0-\boldsymbol{x}^*\|_2$. By the way of contradiction, assume that $\boldsymbol{x}_0\neq\boldsymbol{x}^*$ and $\|\boldsymbol{x}_k-\boldsymbol{x}^*\|_2 = \|\boldsymbol{x}_0-\boldsymbol{x}^*\|_2$. By Equation (8),

$$\left\|\boldsymbol{x}_k-\boldsymbol{x}^*\right\|_2 \le \left\|\boldsymbol{x}_{k-1}-\boldsymbol{x}^*\right\|_2 \ldots < \left\|\boldsymbol{x}_0-\boldsymbol{x}^*\right\|_2$$

and therefore $\|\boldsymbol{x}_l-\boldsymbol{x}^*\|_2 = \|\boldsymbol{x}_0-\boldsymbol{x}^*\|_2$ for all $1\le l\le k$. By Equation (2), $P_{Sp(A_l^T)}(\boldsymbol{x}_{l-1}-\boldsymbol{x}^*) = 0$ for all $1\le l\le k$. By Equation (7), we get $\boldsymbol{x}_l = \boldsymbol{x}_0$ for all $1\le l\le k$. This implies that

$P_{Sp(A_l^T)}(\boldsymbol{x}_0 - \boldsymbol{x}^*) = 0$ for all $1 \le l \le k$. So,

$$P_{Sp^{\perp}(A_k^T) \cap Sp^{\perp}(A_k^T) \ldots \cap Sp^{\perp}(A_1^T)}(\boldsymbol{x}_0 - \boldsymbol{x}^*) = 0$$

$$P_{Sp^{\perp}(A^T)}(\boldsymbol{x}_0 - \boldsymbol{x}^*) = 0.$$

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

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

$$\left\|\boldsymbol{x}_k - \boldsymbol{x}^*\right\|_2 \le 1 - \epsilon. \tag{10}$$

By Equations (9) and (10)

$$\left\|\boldsymbol{x}_k - \boldsymbol{x}^*\right\|_2 = \|\tilde{\boldsymbol{x}}_0\|_2 \, \boldsymbol{e}_k\left(\frac{\tilde{\boldsymbol{x}}_0}{\|\tilde{\boldsymbol{x}}_0\|_2}\right) \le (1-\epsilon)\|\tilde{\boldsymbol{x}}_0\|_2$$

$$\left\|\boldsymbol{x}_k - \boldsymbol{x}^*\right\|_2 \le (1-\epsilon)\left\|\boldsymbol{x}_0 - \boldsymbol{x}^*\right\|_2.$$

Now consider iteration for q cycles,

$$\left\|\boldsymbol{x}_{qk} - \boldsymbol{x}^*\right\|_2 \le (1-\epsilon)^q \left\|\boldsymbol{x}_0 - \boldsymbol{x}^*\right\|_2$$
$$\left\|\boldsymbol{x}_{qk} - \boldsymbol{x}^*\right\|_2 \le [(1-\epsilon)^{1/k}]^{qk} \left\|\boldsymbol{x}_0 - \boldsymbol{x}^*\right\|_2.$$

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 $\mathbb{E}\left\|\boldsymbol{x}_p - \boldsymbol{x}^*\right\|_2^2 \le (1-\kappa(A)^{-2})^p \|\boldsymbol{x}_0 - \boldsymbol{x}^*\|_2^2$, where $\kappa(A) = \|A\|_F \left\|A^{\dagger}\right\|_2$ is the scaled condition number of matrix A with $A^{\dagger}$ is the pseudo-inverse of A. ■

B. Iterative Subspace Projection Approach

We can use the following theorem (in [10], [11]) to show the convergence of regular block Kaczmarz method.

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

$$\lim_{q \to \infty} (P_{M_k} P_{M_{k-1}} \ldots P_{M_1})^q \boldsymbol{x} = P_M \boldsymbol{x}$$

where P_M is the orthogonal intersection projection.

The block Kaczmarz is an alternating projection method with $M_1 = Sp^{\perp}(A_1^T), \ldots, M_k = Sp^{\perp}(A_k^T)$. Also, $P_{M_1} = P_{Sp^{\perp}(A_1^T)}, \ldots, P_{M_k = Sp^{\perp}(A_k^T)}$ and $M = Sp^{\perp}(A_1^T) \cap \ldots \cap Sp^{\perp}(A_k^T) = Sp^{\perp}(A^T)$. Since A is full column rank, $Sp^{\perp}(A^T) = \{0\}$ and $P_M = \{0\}$. After q cycles,

$$\boldsymbol{x}_{qk} - \boldsymbol{x}^* = (P_{M_k} P_{M_{k-1}} \ldots P_{M_1})^q (\boldsymbol{x}_0 - \boldsymbol{x}*). \tag{11}$$

By Theorem 2, $\lim_{q \to \infty} \boldsymbol{x}_{qk} - \boldsymbol{x}^* = 0$ and $\lim_{q \to \infty} \boldsymbol{x}_{qk} = \boldsymbol{x}^*$. Galantai in [10] gives a bound for $\left\| \boldsymbol{x}_{qk} - \boldsymbol{x}^* \right\|_2$ in terms of principle angles between M_i's.

C. Bound for Block Kaczmarz in terms of Principle Angles

Smith, Salmon, and Wagner established the following convergence theorem for applying the alternating projection method in tomography [10], [13]:

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

$$\left\| (P_{M_k} P_{M_{k-1}} \ldots P_{M_1})^q \boldsymbol{x} - P_M \boldsymbol{x} \right\|_2^2 \leq (1 - \Pi_{j=1}^{k-1} \sin^2 \theta_j)^q \left\| \boldsymbol{x} - P_M \boldsymbol{x} \right\|_2^2$$

where P_M is the orthogonal intersection projection.

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

$$\left\| \boldsymbol{x}_{qk} - \boldsymbol{x}^* \right\|_2^2 = \left\| (P_{M_k} P_{M_{k-1}} \ldots P_{M_1})^q (\boldsymbol{x}_0 - \boldsymbol{x}^*) \right\|_2^2 \leq (1 - \Pi_{j=1}^{k-1} \sin^2 \theta_j)^q \left\| \boldsymbol{x}_o - \boldsymbol{x}^* \right\|_2^2 \tag{12}$$

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.

$$\left\|\boldsymbol{x}_{qk}-\boldsymbol{x}^*\right\|_2^2 \le [(1-\Pi_{j=1}^{k-1}\sin^2\theta_j)^{1/k}]^{qk}\left\|\boldsymbol{x}_o-\boldsymbol{x}^*\right\|_2^2 \tag{13}$$

Galantai in [10] 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 $\boldsymbol{x}^*$ be the solution of $A\boldsymbol{x}=\boldsymbol{b}$ for a consistent linear system with $A\in\mathbb{R}^{M\times M}$. Let A be row-partitioned as $\{A_1,\ldots,A_k\}$ where $A_i\in\mathbb{R}^{M_i\times N}$. Let $M_1=Sp^{\perp}(A_1^T),\ldots,M_k=Sp^{\perp}(A_k^T)$ and $A_iA_i^T=LL^T$ be the Cholesky decomposition of $A_iA_i^T$. Define $X_i=A_i^TL^{-T}$ and $X=[X_1,\ldots,X_k]$. Then for each $\boldsymbol{x}\in\mathbb{R}^N$ and integer $q\ge 1$,*

$$\left\|\boldsymbol{x}_{qk}-\boldsymbol{x}^*\right\|_2^2 \le [1-\det(X^TX)]^q\left\|\boldsymbol{x}_o-\boldsymbol{x}^*\right\|_2^2 = [(1-\det(X^TX))^{1/k}]^{qk}\left\|\boldsymbol{x}_o-\boldsymbol{x}^*\right\|_2^2$$

D. 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, $A_iA_i^T=\|\boldsymbol{a}_i\|_2^2=LL^T$. therefore, $L=\|\boldsymbol{a}_i\|_2$ and $L^{-T}=1/\|\boldsymbol{a}_i\|_2$. This implies that $X_i=[\frac{\boldsymbol{a}_i}{\|\boldsymbol{a}_i\|_2}]$. Then, $X\in\mathbb{R}^{M\times M}$ is defined as:

$$X=[\frac{\boldsymbol{a}_1}{\|\boldsymbol{a}_1\|_2},\ldots,\frac{\boldsymbol{a}_M}{\|\boldsymbol{a}_M\|_2}]. \tag{14}$$

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 $\boldsymbol{b}$ does not change the solution $\boldsymbol{x}$.

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

$$\left\|\boldsymbol{x}_{qM}-\boldsymbol{x}^*\right\|_2^2 \le [(1-\det(X^TX))^{1/M}]^{qM}\left\|\boldsymbol{x}_0-\boldsymbol{x}^*\right\|_2^2. \tag{15}$$

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

$$\left\|\boldsymbol{x}_{qM}-\boldsymbol{x}^*\right\|_2^2 \le [(1-\det(AA^T))^{1/M}]^{qM}\left\|\boldsymbol{x}_0-\boldsymbol{x}^*\right\|_2^2. \tag{16}$$

Algorithm 1 Randomized Kaczmarz (of [4])

Require: An over-determined linear set of consistent equations $A\boldsymbol{x} = \boldsymbol{b}$, where A is $M \times N$ matrix and $\boldsymbol{b} \in \mathbb{R}^M$. Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_M$ be the rows of A and b_j be the j^{th} element of $\boldsymbol{b}$.

1: Pick an arbitrary initial approximation $\boldsymbol{x}_0$.
2: Set $p = 0$.
3: **while** not converged **do**
4: Randomly choose $r(i)$ from $\{1, \ldots, M\}$ with probability proportional to $\left\|\boldsymbol{a}_{r(i)}\right\|_2^2$.
5: $\boldsymbol{x}_{p+1} = \boldsymbol{x}_p + \frac{b_{r(i)} - \langle \boldsymbol{a}_{r(i)}, \boldsymbol{x}_p \rangle}{\|\boldsymbol{a}_{r(i)}\|_2^2} \boldsymbol{a}_{r(i)}$
6: Set $p = p + 1$
7: **end while**

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

III. Randomized Kaczmarz Method

A. Randomization Based on Row ℓ_2 Norms

Strohmer *et al.* (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.* discusses (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 $\boldsymbol{x}^*$ be the solution of $A\boldsymbol{x} = \boldsymbol{b}$ Then, Algorithm 1 converges to $\boldsymbol{x}^*$ in expectation, with the average error*

$$\mathbb{E}\left\|\boldsymbol{x}_p - \boldsymbol{x}^*\right\|_2^2 \leq (1 - \kappa(A)^{-2})^p \left\|\boldsymbol{x}_0 - \boldsymbol{x}^*\right\|_2^2 \tag{17}$$

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

Note that A is a full column matrix ($A \in \mathbb{R}^{M \times N}$ with $rank(A) = N$) and therefore we define $A^\dagger$ as left pseudo-inverse 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 $\boldsymbol{a}_i$ and $\boldsymbol{a}_{i+1}$ 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 $A\boldsymbol{x} = \boldsymbol{b}$ be a consistent linear system of equations ($A \in \mathbb{R}^{M\times N}$) and let $\boldsymbol{x}_0$ be an arbitrary initial approximation to the solution of $A\boldsymbol{x} = \boldsymbol{b}$. For $k = 1,2,\ldots$ compute*

$$\boldsymbol{x}_{p+1} = \boldsymbol{x}_p + \frac{b_{r(i)} - \langle \boldsymbol{a}_{r(i)}, \boldsymbol{x}_p \rangle}{\left\| \boldsymbol{a}_{r(i)} \right\|_2^2} \boldsymbol{a}_{r(i)} \tag{18}$$

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

$$\mathbb{E} \left\| \boldsymbol{x}_p - \boldsymbol{x}^* \right\|_2^2 \leq (1 - \kappa(B)^{-2})^p \left\| \boldsymbol{x}_0 - \boldsymbol{x}^* \right\|_2^2 \tag{19}$$

where $\kappa(B) = \|B\|_F \left\| B^\dagger \right\|_2$ 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 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.

$$B = \left[\frac{\boldsymbol{a}_1}{\|\boldsymbol{a}_1\|_2}, \ldots, \frac{\boldsymbol{a}_M}{\|\boldsymbol{a}_M\|_2} \right]^T. \tag{20}$$

■

Algorithm 2 Randomized Kaczmarz Hyperplane Angles

Require: An over-determined linear set of consistent equations $A\boldsymbol{x} = \boldsymbol{b}$, where A is $M \times N$ matrix and $\boldsymbol{b} \in \mathbb{R}^M$. Let $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_M$ be the rows of A and $\boldsymbol{b}_j$ be the j^{th} element of $\boldsymbol{b}$.

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

B. Randomization based on Subspace Angles

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 angles between the hyperplanes (represented by the rows of $A\boldsymbol{x} = \boldsymbol{b}$). Then, it randomly picks two hyperplanes using this probability distribution. This is followed by a two-step projection on these hyperplanes (see Algorithm 2).

C. P-Subspaces Approach

A new method has been developed which is intended to better accommodate the coherency of non-orthogonal 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.

The algorithm 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 non-orthogonal projections which may be present in the simple Kaczmarz.

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 is given in reference [16] and is 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 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:

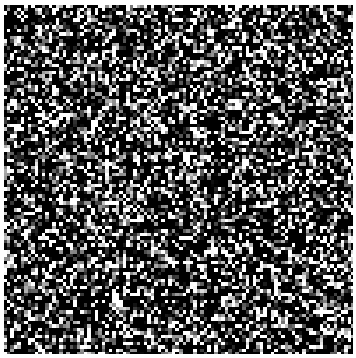
$$\boldsymbol{x}_{S_P} = \sum_{l=1}^{P} \langle \hat{\boldsymbol{u}}_l, \boldsymbol{x} \rangle \hat{\boldsymbol{u}}_l. \tag{21}$$

where $\boldsymbol{x}_{S_P}$ in $S_P \subseteq S_N$ subspace is the $P-$dimensional solution approximation which becomes exact for $S_{P=N}$ for $\boldsymbol{x}_{S_{P=N}} \in \mathbb{R}^N$ in the noiseless, self-consistent, case.[1]

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

[1]The u vector with the hat symbol $\hat{u}$ indicates unit ℓ^2-norm



(a) IID Gaussian Unit Vector Image

(b) CT Phantom Image

Fig. III.1: Representative test data

and incurring reasonable computational penalty. The algorithm is simply to add subspaces of larger dimensions. Let

$$\boldsymbol{x} - \boldsymbol{x}_{k+1} \quad = \quad \boldsymbol{x} - \boldsymbol{x}_k - \sum_{l=1}^{P} \langle \hat{\boldsymbol{u}}_l, \boldsymbol{x} - \boldsymbol{x}_k \rangle \hat{\boldsymbol{u}}_l. \tag{22}$$

It is convenient to make a substitution as follows:

$$\boldsymbol{z}_{k+1} = \boldsymbol{x} - \boldsymbol{x}_{k+1}. \tag{23}$$

Using above substitution and orthonormal condition[2] $\langle \hat{\boldsymbol{u}}_j, \hat{\boldsymbol{u}}_k \rangle = \delta_{j,k}$, where the Kronecker $\delta_{j,k} = \begin{cases} 0 & \text{if } j \neq k \\ 1 & \text{if } j = k, \end{cases}$, find the ℓ^2-norm squared of $\boldsymbol{z}_{k+1}$:

$$\|\boldsymbol{z}_{k+1}\|_2^2 = \|\boldsymbol{z}_k\|_2^2 - \sum_{l=1}^{P} |\langle \hat{\boldsymbol{u}}_l, \boldsymbol{z}_k \rangle|^2. \tag{24}$$

The ensemble average of the above Equation 24 yields the convergence result, which is

[2]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 Equation (24). 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 (24).

Algorithm 3 P-Subspace Kaczmarz Projections

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

1: Set $\boldsymbol{x}_0$ 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 $\{\boldsymbol{a}_{i,1},\ldots,\boldsymbol{a}_{i,P}\}$.
4: Perform Gram-Schmidt procedure on A_i to obtain the orthonormal set of columns $\{\boldsymbol{u}_{i,1},\ldots,\boldsymbol{u}_{i,P}\}$. Let $Q_i = \{\boldsymbol{u}_{i,1},\ldots,\boldsymbol{u}_{i,P}\} \in \mathbb{R}^{N\times P}$
5: Update $\boldsymbol{x}_i$ as follows:
$\boldsymbol{x}_i = \boldsymbol{x}_{i-1} + Proj_{Sp(Q_i)}(\boldsymbol{x}_{i-1} - \boldsymbol{x})$,
$\boldsymbol{x}_i = \boldsymbol{x}_{i-1} - Q_i Q_i^T(\boldsymbol{x} - \boldsymbol{x}_{i-1})$,
6: Compute $Q_i^T\boldsymbol{x}$ iteratively using $\{\boldsymbol{a}_{i,1},\ldots,\boldsymbol{a}_{i,P}\}$, $\{\boldsymbol{b}_{i,1},\ldots,\boldsymbol{b}_{i,P}\}$, $\{\boldsymbol{u}_{i,1},\ldots,\boldsymbol{u}_{i,P}\}$
7: Update $i = i + 1$
8: **end while**

the main topic of this section.

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 $\boldsymbol{x} \in \mathbb{R}^{N\times 1}$ onto a random subspace basis $U_P \in$ of dimension P,

$$\mathbb{E}[\|U_P^T\boldsymbol{x}\|_2^2],$$

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

$$\hat{\boldsymbol{u}}_j \;\rightarrow\; \hat{\boldsymbol{U}}_j = [U_{j,1},\ldots,U_{j,N}]\frac{1}{C_\sigma}, \tag{25}$$

$$= \frac{\boldsymbol{U}_j}{\|\boldsymbol{U}_j\|_2^2} \qquad \forall\, j \in [1,\ldots,P]. \tag{26}$$

where the upper case components0 $U_{j,i}$ represent the (j,i)-th IID random variable component, and normalization constant C_σ is to be determined.

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

$$\|U_P^T \boldsymbol{x}\|_2^2 = \boldsymbol{x}^T U_P U_P^T \boldsymbol{x}.$$

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

$$\mathbb{E}\left[\boldsymbol{x}^T \hat{\boldsymbol{U}}_j \hat{\boldsymbol{U}}_j^T \boldsymbol{x}\right] \; \forall j \in [1,\ldots,P].$$

b) Unit Vector : The deterministic identity for the magnitude of a unit vector is well known result for $\hat{\boldsymbol{u}} \in \mathbb{R}^N$,

$$\|\hat{\boldsymbol{u}}\|_2^2 = \sum_{i=1}^{N} \frac{u_i^2}{\|\boldsymbol{u}\|_2^2} = 1. \tag{27}$$

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

$$\begin{aligned} \mathbb{E}\left[\|\hat{\boldsymbol{U}}_j\|_2^2\right] &= \mathbb{E}\left[\hat{\boldsymbol{U}}_j^T \hat{\boldsymbol{U}}_j\right] = 1 \\ &= \mathbb{E}\left[U_{j,1}^2 + \cdots + U_{j,N}^2\right] \frac{1}{C_\sigma^2}. \end{aligned} \tag{28}$$

c) Normalization of Random Unit Vector : Denote $\hat{\boldsymbol{U}}_j$ as the j-th random variable unit-norm vector associated with a set of column vectors $\{\boldsymbol{U}_j\}_{j\in 1,\ldots,P}$ 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 Equation (28) for random vector $\boldsymbol{U}_j$ are found such that:

$$\mathbb{E}\sum_{i=1}^{N} \frac{U_{j,i}^2}{C_\sigma^2} = \sum_{i=1}^{N} \mathbb{E}\left[\frac{U_{j,i}^2}{C_\sigma^2}\right] = 1, \tag{29}$$

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

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

$$\mathbb{E}\left[U_{j,i}^2\right] = \sigma_{j,i}^2 = \frac{C_\sigma^2}{N} \qquad \forall\, U_{j,i\in 1,\ldots,N} \in f(U_{j,i}), \tag{30}$$

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 $\boldsymbol{x}$ 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 $\boldsymbol{\alpha} = U_P^T \boldsymbol{x}$ and find the ℓ^2-norm squared:

$$\begin{aligned} \|\boldsymbol{\alpha}\|_2^2 &= \alpha_1^2 + \alpha_2^2 + \ldots + \alpha_P^2 \\ &= \|U_P^T \boldsymbol{x}\|_2^2 = \boldsymbol{x}^T U_P U_P^T \boldsymbol{x}, \end{aligned} \tag{31}$$

where

$$\alpha_j^2 = \langle \hat{\boldsymbol{u}}_j, \boldsymbol{x} \rangle^2 \tag{32}$$

$$= \left(u_{j,1}\boldsymbol{x}_1 + \ldots + u_{j,N}\boldsymbol{x}_N\right)^2 \tag{33}$$

$$= \sum_{i,k}^{N,N} \frac{u_{j,k} u_{j,i} \boldsymbol{x}_k \boldsymbol{x}_i}{\|\boldsymbol{u_j}\|_2^2}. \tag{34}$$

Let upper case $U_{j,k}$ denote the k-th IID element random[3] variable of the j-th column vector $\boldsymbol{U}_j$ associated with column vector $\boldsymbol{u_j}$; let $\boldsymbol{x}$ 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:

$$\mathbb{E}\left[\alpha_j^2\right] = \mathbb{E}\left[\sum_{i,k}^{N,N} \frac{U_{j,k} U_{j,i} \boldsymbol{x}_k \boldsymbol{x}_i}{C_\sigma^2}\right] = \sum_{i,k}^{N,N} \mathbb{E}\left[\frac{U_{j,k} U_{j,i} \boldsymbol{x}_k \boldsymbol{x}_i}{C_\sigma^2}\right] \tag{35}$$

[3]This is not the same k-variable as the Kaczmarz iteration variable

$$= \sum_{k=1}^{N} \mathbb{E}\left[\frac{U_k^2 \boldsymbol{x}_k^2}{C_\sigma^2}\right] = \sum_{k}^{N} \mathbb{E}\left[\frac{U_{j,k}^2}{C_\sigma^2}\right] \boldsymbol{x}_k^2$$

$$= \mathbb{E}\left[\frac{U_{j,k}^2}{C_\sigma^2}\right] \sum_{k}^{N} \boldsymbol{x}_k^2 = \mathbb{E}\left[\frac{U_{j,k}^2}{C_\sigma^2}\right] \|\boldsymbol{x}\|_2^2$$

$$= \frac{1}{C_\sigma^2} \frac{C_\sigma^2}{N} \|\boldsymbol{x}\|_2^2$$

$$= \frac{1}{N} \|\boldsymbol{x}\|_2^2.$$

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

$$\mathbb{E}\left[\|\boldsymbol{\alpha}\|_2^2\right] = \mathbb{E}\left[\sum_{j=1}^{P} \alpha_j^2\right] = \frac{P}{N} \|\boldsymbol{x}\|_2^2 \tag{36}$$

subject to IID constraint on $\hat{\boldsymbol{U}}_j$ where it is further noted that $\sigma^2 N = C_\sigma^2$ in Equation (30).

e) Error per Iteration: For a given k-th Kaczmarz iteration, the expectation of the projection of fixed vector $\boldsymbol{x}$ 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[4] with Equation (37):

$$\|\boldsymbol{z}_{k+1}\|_2^2 = \|\boldsymbol{z}_k\|_2^2 - \sum_{l=1}^{P} |\langle \boldsymbol{z}_k, \hat{\boldsymbol{u}}_l \rangle|^2 \tag{37}$$

$$\mathbb{E}_{\{k+1|\boldsymbol{z}_0, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_k\}}\left[\|\boldsymbol{z}_{k+1}\|_2^2\right] = \tag{38}$$

$$= \mathbb{E}_{\{k+1|\boldsymbol{z}_0, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_k\}}\left[\|\boldsymbol{z}_k\|_2^2 - \sum_{l=1}^{P} |\langle \boldsymbol{z}_k, \hat{\boldsymbol{u}}_l \rangle|^2\right]$$

$$= \mathbb{E}_{\{k+1|\boldsymbol{z}_0, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_k\}}\left[\|\boldsymbol{z}_k\|_2^2\right] - \mathbb{E}_{\{k+1|\boldsymbol{z}_0, \boldsymbol{z}_1, \ldots, \boldsymbol{z}_k\}}\left[\sum_{l=1}^{P} |\langle \boldsymbol{z}_k, \hat{\boldsymbol{u}}_l \rangle|^2\right].$$

[4]Recall that derivation of this equation (37) requires orthogonality among the $\hat{u}_l$ subspace basis vectors.

We identify the term on the right as:

$$\mathbb{E}_{\{k+1|\boldsymbol{z}_0,\boldsymbol{z}_1,\ldots,\boldsymbol{z}_k\}}\left[\sum_{l=1}^{P}|\langle \boldsymbol{z}_k,\hat{\boldsymbol{u}}_l\rangle|^2\right]=\mathbb{E}_{\{k+1|\boldsymbol{z}_0,\boldsymbol{z}_1,\ldots,\boldsymbol{z}_k\}}\left[\|U_P\boldsymbol{z}_k\|_2^2\right]$$

$$=\frac{P}{N}\times\mathbb{E}_{\{k+1|\boldsymbol{z}_0,\boldsymbol{z}_1,\ldots,\boldsymbol{z}_k\}}\left[\|\boldsymbol{z}_k\|_2^2\right]. \tag{39}$$

The results from the two equations ((39) and (38)) above may then be combined to obtain,

$$\mathbb{E}_{\{k+1|\boldsymbol{z}_0,\boldsymbol{z}_1,\ldots,\boldsymbol{z}_k\}}\left[\|\boldsymbol{z}_{k+1}\|_2^2\right]=$$

$$\left(1-\frac{P}{N}\right)\times\mathbb{E}_{\{k|\boldsymbol{z}_0,\boldsymbol{z}_1,\ldots,\boldsymbol{z}_k-1\}}\left[\|\boldsymbol{z}_k\|_2^2\right],$$

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 $\boldsymbol{z_0}\equiv\boldsymbol{x}-\boldsymbol{x}_0$:

$$\mathbb{E}_{\{\beta+1|\boldsymbol{z}_0\}}\left[\|\boldsymbol{z}_{\beta+1}\|_2^2\right]=\left(1-\frac{P}{N}\right)^{\beta}\|\boldsymbol{z}_0\|_2^2\;\forall\;\beta\in 1,2,3,\ldots. \tag{40}$$

f) Asymptotic Convergence: The statistical ensemble average of the above Equation (24) for the β-th iteration yields the convergence result given in Equation (40). These results assume random variables identically and independently distributed, but compare well to others in the literature, such as the convergence result in Strohmer [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 (41) has an asymptotic form:

$$\frac{\mathbb{E}_{\{\beta+1|\boldsymbol{z}_0\}}}{\left[\|\boldsymbol{z}_{\beta+1}\|_2^2\right]}\|\boldsymbol{z}_0\|_2^2\quad= \tag{41}$$

$$\lim_{\beta\rightarrow\infty}\left[1-\frac{P}{N}\right]^{\beta}\simeq e^{-\beta P/N}$$

$$P=\dim(S_P),\;\beta\gg 1,2,3,\ldots\rightarrow k\in P,2P,3P,\ldots.$$

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

$$\frac{\mathbb{E}_{\{k+1|\boldsymbol{z}_0\}}}{\left[\|\boldsymbol{z}_{k+1}\|_2^2\right]}\|\boldsymbol{z}_0\|_2^2 = \left[1-\frac{1}{N}\right]^k \tag{42}$$

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

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

g) Theory and Simulation: Simulations in reference [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 $\boldsymbol{x}$ is chosen as a random point on the unit sphere - which is illustrated in Figure III.1a. In a second problem, a measurement of the standard phantom using parallel beam measurements is included, which contains coherent measurements.

3) QR Representation: An alternative method for finding the expected convergence of the RKOS iterative block Kaczmarz method used to solve $A\boldsymbol{x}^* = \boldsymbol{b}^*$ for $(\boldsymbol{x}^*;\boldsymbol{b}^*) \in \mathbb{R}^N$, and $A \in \mathbb{R}^{N\times N}$ is considered below. The formalism is slightly more rigorous and contemporary, allows direct computation of matrix quantities (instead of recursive GS), but is consistent with the former method of finding the orthogonal projection subspaces U_i.

The method includes sufficient algebra to allow representation of the Kaczmarz orthogonal block iterative process subject to the Smith Solmon Wagner [13] inequality, by incorporating the subspace projection concepts from Galantai *et a l* [21].

In this work, it is assumed that measurement matrix $A \in \mathbb{R}^{N\times N}$ is square full row rank, however, the results may be extended to cases where $M \geq N$ with proper modification.

a) Approach: The i-th block iteration of the RKOS selects blocks of M_i-rows of matrix A to form A_i. In general, the blocks may be selected to allow overlapping rows or unique row selections per cycle, in natural row order or via random *a priori* partitioning into the set $\{M_1, M_2, \ldots, M_k\}$ of row blocks comprising $A_i \in \mathbb{R}^{M_i\times N}$. However, in the following

analysis, we assume set is subject to

$$\sum_{i=1}^{k} M_i = N \tag{43}$$

which applies to the case in which rows are selected uniquely without replacement for each cycle.

Let $\mathbb{H}$ be a Hilbert space having a defined inner product and finite norm. Let the measurement matrix $A \in \mathbb{R}^{N\times N}$ be full row rank in $\mathbb{H}$ and segmented into k-blocks according to

$$I_N = [E_1,\ldots,E_k] \left(E_i \in \mathbb{R}^{N\times M_i},\, i = 1,\ldots,k\right)$$

where E_i is a set of M_i-column index vectors (which may be non-contiguous) of the identity matrix I_{NxN} to form $A_i^T = A^T E_i$.[5]

The segmentation of the blocks and the order of blocks is stationary with respect to iteration number in this treatment.

b) QR and Gram-Schmidt: In the RKOS algorithm, the process of decomposing A_i^T into the QR [17], [19] factorization performs the Gram-Schmidt process for orthogonalization. Algorithm (3) recursively solves for the orthonormal set and allows recursive computation of the projections of exact solution $\boldsymbol{x}^*$ onto the the orthogonal basis in terms measurements $\boldsymbol{b}_i$.

Direct QR decomposition for row block A_i is noted to be

$$A_i^T = A^T E_i = Q_i R_i = U_i R_i \tag{44}$$

is equivalent to GS and may be directly computed[6], where $U_i \in \mathbb{R}^{N\times M_i}$ is the i-th orthonormal basis (columns) constructed from the M_i-rows randomly selected from matrix A, and

[5]To understand the sampling vector E_i, consider the following example.
Let $I_{6,6}$ be the identity matrix and select non-continguous sampling set $M_i = \{3,5,6\}$ and form E_i as $E_i(3,5,6) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and $E_i(3,5,6)^T = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$.

[6]The transpose is needed since the columns of U_i are the rows of A_i block

$R_i \in \mathbb{R}^{M_i \times M_i}$ is upper triangular matrix. It is important to note that matrix $Q_i = U_i$ in the RKOS algorithm (3).

For $\boldsymbol{x} \in \mathbb{H}$, define the j-th iterative error estimate as, $\boldsymbol{z}_j \equiv \boldsymbol{x}_i^* - \boldsymbol{x}_j$ and $\widetilde{\boldsymbol{z}}_j \equiv \boldsymbol{x}_i^* + \boldsymbol{\epsilon}_x(i) - \boldsymbol{x}_j$ respectively without and with noise, where $\boldsymbol{x}_j$ is the j-th iterative estimate for the i-th block projection of k-blocks per cycle; $\boldsymbol{x}^*$ is the desired noise-free solution to $A\boldsymbol{x}^* = \boldsymbol{b}^*$; $\boldsymbol{x}_i^*$ is the i-th block estimate of the noise free solution; and $\boldsymbol{\epsilon}_x(i)$ is the i-th propagated measurement noise vector in the current basis.[7]

The simple block Kaczmarz's equation (without noise) using the orthogonal projection matrix U_i may be written as

$$\begin{aligned} \boldsymbol{x}_{j+1} &= \boldsymbol{x}_j + U_i U_i^T (\boldsymbol{x}^* - \boldsymbol{x}_j) \\ \boldsymbol{x}^* - \boldsymbol{x}_{j+1} &= \boldsymbol{x}^* - \boldsymbol{x}_j - U_i U_i^T (\boldsymbol{x}^* - \boldsymbol{x}_j) \\ \boldsymbol{z}_{j+1} &= \boldsymbol{z}_j - U_i U_i^T \boldsymbol{z}_j = (I - U_i U_i^T)\boldsymbol{z}_j \\ & \quad (i \equiv j(\mod k) + 1). \end{aligned} \tag{45}$$

In above, notice that U_i is orthonormal column matrix, i.e. $U_i^T U_i = I$ under contraction, but on projection, $U_i U_i^T = P_i$ acts to preserves components within the subspace U_i. The following relations are noted:

$$A_i A_i^T = (U_i R)_i^T (U_i R_i) = R_i^T U_i^T U_i R_i = R_i^T R_i \tag{46}$$

$$A_i^T A_i = (U_i R_i)(U_i R_i)^T = U_i R_i R_i^T U_i^T.$$

To find the new basis, use definition in Equation (44) solve to find

$$U_i = A_i^T R_i^T (R_i R_i^T)^{-1}, \tag{47}$$

$$U_i^T U_i R_i = R_i = U_i^T A_i^T. \tag{48}$$

c) Block Equations: Next, consider that measurement vector $\boldsymbol{b}$ is comprised of (a) $\boldsymbol{b}^*$ the self-consistent error free measurement vector solution of $A\boldsymbol{x}^* = \boldsymbol{b}^*$, and (b) the

[7]It should be noted that the noise terms are generally not separable in practice, but are explicitly shown here in order to facilitate the analysis.

measurement noise term, $\boldsymbol{\epsilon}_b$. Therefore, $A\boldsymbol{x}^* = \boldsymbol{b}^*$, $\boldsymbol{x} = \boldsymbol{x}^* + \boldsymbol{\epsilon}_x$, $\boldsymbol{b} = \boldsymbol{b}^* + \boldsymbol{\epsilon_b}$. Then we may find,

$$A\boldsymbol{x} = \boldsymbol{b} = \boldsymbol{b}^* + \boldsymbol{\epsilon}_b, \tag{49}$$

$$A_i\boldsymbol{x} = (U_iR_i)^T\boldsymbol{x} = R_i^TU_i^T\boldsymbol{x} = \boldsymbol{b}_i = \boldsymbol{b}_i^* + \boldsymbol{\epsilon}_b(i),$$

where $\boldsymbol{b}_i = E_i\boldsymbol{b}$ to obtain the i-th under-determined block estimate for the solution,

$$\boldsymbol{x}_i = \left(A_i^TA_i\right)^{-1}A_i^T\boldsymbol{b}_i = \left(A_i^TA_i\right)^{-1}A_i^T(\boldsymbol{b}_i^* + \boldsymbol{\epsilon}_b(i)). \tag{50}$$

The next objective is to find the result in the new basis. First, substitute from Equation (44) and multiply both sides by R_i as follows:

$$A_i\boldsymbol{x} = (U_iR_i)^T\boldsymbol{x} = R_i^TU_i^T\boldsymbol{x} = \boldsymbol{b}_i, \tag{51}$$

$$R_iR_i^TU_i^T\boldsymbol{x} = R_i\boldsymbol{b}_i,$$

$$U_i^T[\boldsymbol{x}] = (R_iR_i^T)^{-1}R_i\boldsymbol{b}_i = (R_iR_i^T)^{-1}R_i(\boldsymbol{b}_i^* + \boldsymbol{\epsilon}_b(i)), \tag{52}$$

which has been converted to terms of U_i and R_i. Using the orthogonality of U_i, Equation (52) may be solved for $\boldsymbol{x}$ in terms of U_i, R_i as follows:

$$\begin{aligned} U_iU_i^T\boldsymbol{x} &= U_iI(R_iR_i^T)^{-1}R_i\boldsymbol{b}_i \\ \left(U_iU_i^T\right)\boldsymbol{x} &= \left(U_iU_i^T\right)U_i(R_iR_i^T)^{-1}R_i\boldsymbol{b}_i \\ \rightarrow \boldsymbol{x}_i &= \boldsymbol{x}_i^* + \epsilon_x(i) = U_i(R_iR_i^T)^{-1}R_i\boldsymbol{b}_i \end{aligned} \tag{53}$$

where $\boldsymbol{b}_i = \boldsymbol{b}_i^* + \boldsymbol{\epsilon}_b(i)$ and the contraction of the orthonormal matrix $I = U_i^TU_i$ is used on the right hand side, $U_iU_i^T$ is non-singular, $\boldsymbol{x} = \boldsymbol{x}^* + \boldsymbol{\epsilon}_x$, and the i-th block estimate $\boldsymbol{x}_i = \boldsymbol{x}_i^* + \epsilon_x(i)$. The result $\boldsymbol{x}_i = U_i(R_iR_i^T)^{-1}R_i\boldsymbol{b}_i$ may be verified by

$$U_i^T\boldsymbol{x}_i = U_i^TU_i(R_iR_i^T)^{-1}R_i\boldsymbol{b}_i = (R_iR_i^T)^{-1}R_i\left(\boldsymbol{b}_i^* + \boldsymbol{\epsilon}_b(i)\right)$$

which is equation (52) as expected.

d) Block Iteration and Noise: Making the substitutions for the consistent noise free solution $\boldsymbol{x}^*$ and the measurement noise $\boldsymbol{\epsilon}_b$, the j-th error difference vector terms are as

follows:

$$\begin{aligned} \widetilde{\boldsymbol{z}}_j \quad &\equiv \quad \boldsymbol{z}_j + \boldsymbol{\epsilon}_x = \boldsymbol{x}^* + \boldsymbol{\epsilon}_x - \boldsymbol{x}_j, \\ \rightarrow \quad \widetilde{\boldsymbol{z}}_j &= \boldsymbol{x}^* \quad + U_i (R_i R_i^T)^{-1} R_i \boldsymbol{\epsilon}_b(i) - \boldsymbol{x}_j. \end{aligned} \tag{54}$$

The orthogonal block Kaczmarz Equation (45) for $\boldsymbol{z}_{j+1} \equiv \boldsymbol{x}^* - \boldsymbol{x}_{j+1}$ may be written as follows:

$$\begin{aligned} \boldsymbol{z}_{j+1} + \boldsymbol{\epsilon}_x(i) = \quad & \left(\boldsymbol{z}_j + U_i (R_i R_i^T)^{-1} R_i \boldsymbol{\epsilon}_b(i)\right) \\ & - U_i U_i^T \left(\boldsymbol{z}_j + U_i (R_i R_i^T)^{-1} R_i \boldsymbol{\epsilon}_b(i)\right) \\ = \quad & (I - U_i U_i^T) \left[\boldsymbol{z}_j + U_i (R_i R_i^T)^{-1} R_i \boldsymbol{\epsilon}_b(i)\right] \end{aligned} \tag{55}$$

or,

$$\widetilde{\boldsymbol{z}}_{j+1} \quad = \quad (I - U_i U_i^T) \widetilde{\boldsymbol{z}}_j \; (i \equiv j (mod\, k) + 1) \tag{56}$$

where

$$\widetilde{\boldsymbol{z}}_{j+1} \quad = \quad \boldsymbol{z}_{j+1} + \boldsymbol{\epsilon}_x(i) = \boldsymbol{x}^* + \boldsymbol{\epsilon}_x(i) - \boldsymbol{x}_{j+1}, \tag{57}$$

and the estimated noise component in the block-row basis is $\boldsymbol{\epsilon}_x(i) = U_i (R_i R_i^T)^{-1} R_i \boldsymbol{\epsilon}_b(i)$. In actual practice, the projected component in the new orthogonal subspace basis is computed as $U_i U_i^T \boldsymbol{x} = U_i I (R_i R_i^T)^{-1} R_i \boldsymbol{b}_i$ from the right hand side, where the value of the under-determined solution vector $\boldsymbol{x}$ for the block estimate is not explicitly realized.

e) Cyclical Projections: In the notation of Halperin [22] and Galantai [10], $A_i^T = A^T E_i = U_i R_i$, and the projection operator, null subspace, and orthonormal condition may be identified as follows:

$P_{M_j} = I - U_j U_j^T$, $M_j = \mathscr{R}^{\perp}(U_j^T)$, $U_j \in \mathbb{R}^{N \times M_i}$ $U_j^T U_j = I_{M_i \times M_i}$ where during the first cycle, observe that $j = i$ for $j = 1, \ldots, k$.

It is further noted that the cumulative projection and null space intersection for the k-th iteration block are as follows:

$$\Omega = P_k, \ldots, P_2 P_1 = (I - U_k U_k^T), \ldots, (I - U_1 U_1^T), \tag{58}$$

$$M = \bigcap_{j=1}^{k} \mathscr{R}^{\perp}(U_j) = \mathscr{R}^{\perp}([U_1, \ldots, U_k]) = \mathscr{R}^{\perp}(U), \tag{59}$$

respectively, with $P_M = P_{\mathscr{R}^{\perp}(U)} = I - P_{\mathscr{R}(U)}$. The Smith Solmon Wagner [13] referenced in Theorem 4 of Galantai [10], has the form

$$\left\| \left[(I - U_k U_k^T), \ldots, (I - U_1 U_1^T) \right]^N \boldsymbol{z}_0 - P_M \boldsymbol{z}_0 \right\| \tag{60}$$

$$\leq c_{SSW}^N \left\| \boldsymbol{z}_0 - P_M \boldsymbol{z}_0 \right\|$$

where $c_{SSW} = \left(\prod_{j=1}^{k-1} \sin^2 \theta_j \right)^{1/2}$ and angle

$$\theta_j = \alpha \left(M_j, \bigcap_{i=j+1}^{k} M_i \right) = \alpha \left(\mathscr{R}^{\perp}(U_j), \mathscr{R}^{\perp}([U_{j+1}, \ldots, U_k]) \right). \tag{61}$$

The above result provides a bound for convergence using linear block projections

f) Gram-Schmidt and QR Summary : The expected statistical convergence method described using Gram-Schmidt (GS) shows good agreement to experimental simulations. The results are consistent with Strohmer for $P = 1$. The P-dimensional orthogonal subspace method based upon QR gives similar convergence result, and the deterministic bounds are consistent with the results of Galantai. In both of the above cases, i.e. Gram-Schmidt and QR decomposition, the proofs of convergence were based upon IID probability distribution of the measurement noise and the measurement sampling vectors.

The propagation of measurement noise is seen to be dependent upon the iterative convergence and general iterative process. An additional study may be worthwhile to determine a possible method for noise minimization and feasibility.

4) Convergence for Almost Any Probability Distribution: Although the former methods for RKOS Gram Schmidt and QR assumed IID random variables, it is noted that application of Theorem (2) to Equation (60) in section (III-C3e) yields convergence regardless of the

distribution[8] of the sampling and IID variates as follows:

$$\lim_{q\to\infty}\left[(I-U_kU_k^T),\ldots,(I-U_1U_1^T)\right]^q \boldsymbol{z}_0 = P_M\boldsymbol{z}_0. \tag{62}$$

As noted before, the block Kaczmarz is an alternating projection method with $M_1 = Sp^{\perp}(U_1^T),\ldots,M_k = Sp^{\perp}(U_k^T)$. Also, $P_{M_1} = P_{Sp^{\perp}(U_1^T)},\ldots,P_{M_k=Sp^{\perp}(U_k^T)}$ and $M = Sp^{\perp}(U_1^T)\cap\ldots\cap Sp^{\perp}(U_k^T) = Sp^{\perp}(A^T)$. Since A is full column rank, $Sp^{\perp}(A^T) = \{0\}$ and $P_M = \{0\}$. After q cycles,

$$\boldsymbol{z}_{qk} = \boldsymbol{x}_{qk} - \boldsymbol{x}^* = (P_{M_k}P_{M_{k-1}}\ldots P_{M_1})^q(\boldsymbol{x}_0 - \boldsymbol{x}^*). \tag{63}$$

By Theorem 3, $\lim_{q\to\infty}\boldsymbol{x}_{qk} - \boldsymbol{x}^* = 0$ and $\lim_{q\to\infty}\boldsymbol{x}_{qk} = \boldsymbol{x}^*$. Here, it should be noted that orthogonality of U_k is consistent with Galantai.

IV. Regular versus Randomized Kaczmarz

The randomized Kaczmarz's algorithm developed by Strohmer in [4] has the following convergence in expectation:

$$\mathbb{E}\left\|\boldsymbol{x}_{qM} - \boldsymbol{x}^*\right\|_2^2 \le (1 - \frac{1}{\kappa(A)^2})^{qM}\left\|\boldsymbol{x}_0 - \boldsymbol{x}^*\right\|_2^2 \tag{64}$$

where $\kappa(A) = \|A\|_F\left\|A^{\dagger}\right\|_2$ is the scaled condition number of matrix A with $A^{\dagger}$ is the left pseudo-inverse of A. The bound for regular Kacmarz is given in Equation (16). Note that we assume $A \in \mathbb{R}^{M\times M}$. Now, we need to compare $(1 - \frac{1}{\|A\|_F^2\left\|A^{\dagger}\right\|_2^2})$ and $(1 - \det(AA^T))^{1/M}$ to assess which bound is tighter. Let $\sigma_1 \ge \sigma_2 \ge \ldots \ge \sigma_M > 0$ be ordered singular values of A. Then,

$$\left\|A^{\dagger}\right\|_2^2 = 1/\sigma_N^2 \tag{65}$$

$$\|A\|_F^2 = \sum_{i=1}^{M}\sigma_i^2. \tag{66}$$

[8]Note that the span of the solution space must be completely sampled with non-zero probability

Also, note that

$$AA^T = \begin{bmatrix} 1 & \cos\theta_{12} & \dots & \cos\theta_{1M} \\ \cos\theta_{21} & 1 & \dots & \cos\theta_{2M} \\ \vdots & \vdots & \dots & \vdots \\ \cos\theta_{M1} & \cos\theta_{M2} & \dots & 1 \end{bmatrix} \tag{67}$$

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

$$\det(AA^T) \leq \prod_{i=1}^{M} \sum_{j=1}^{M} \cos^2\theta_{ij}. \tag{68}$$

Note that

$$\prod_{i=1}^{M} \sigma_i^2(A) = \prod_{i=1}^{M} \lambda_i(A^T A) = \det(A^T A) = \det(AA^T) \tag{69}$$

therefore

$$[1 - \det(AA^T)]^{1/M} = (1 - \prod_{i=1}^{M} \sigma_i^2)^{1/M}. \tag{70}$$

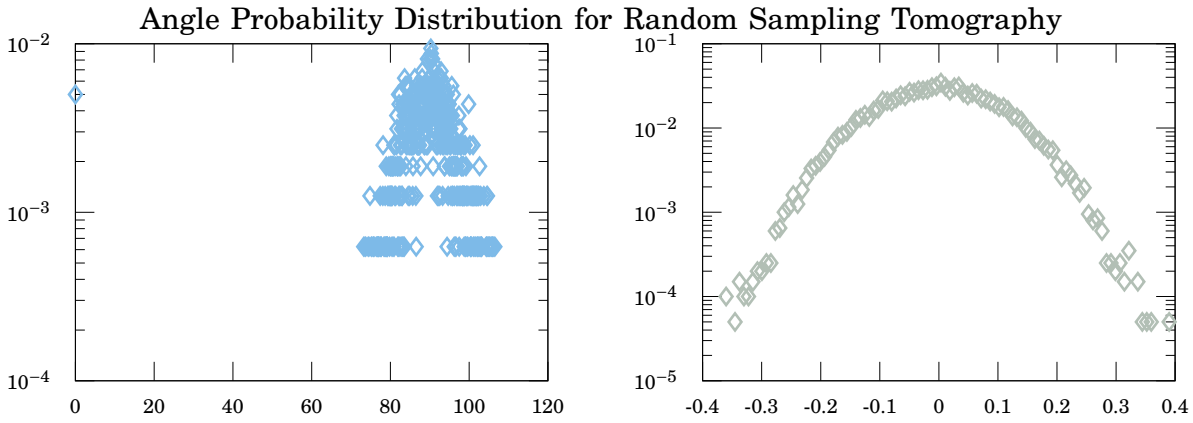
Now, Equations 64 and 16 become:

$$\mathbb{E} \left\| \boldsymbol{x}_{qM} - \boldsymbol{x}^* \right\|_2^2 \leq (1 - \frac{\sigma_M^2}{\sum_{i=1}^{M} \sigma_i^2})^{qM} \left\| \boldsymbol{x}_0 - \boldsymbol{x}^* \right\|_2^2, \tag{71}$$

$$\left\| \boldsymbol{x}_{qM} - \boldsymbol{x}^* \right\|_2^2 \leq [(1 - \prod_{i=1}^{M} \sigma_i^2)^{1/M}]^{qM} \left\| \boldsymbol{x}_0 - \boldsymbol{x}^* \right\|_2^2. \tag{72}$$

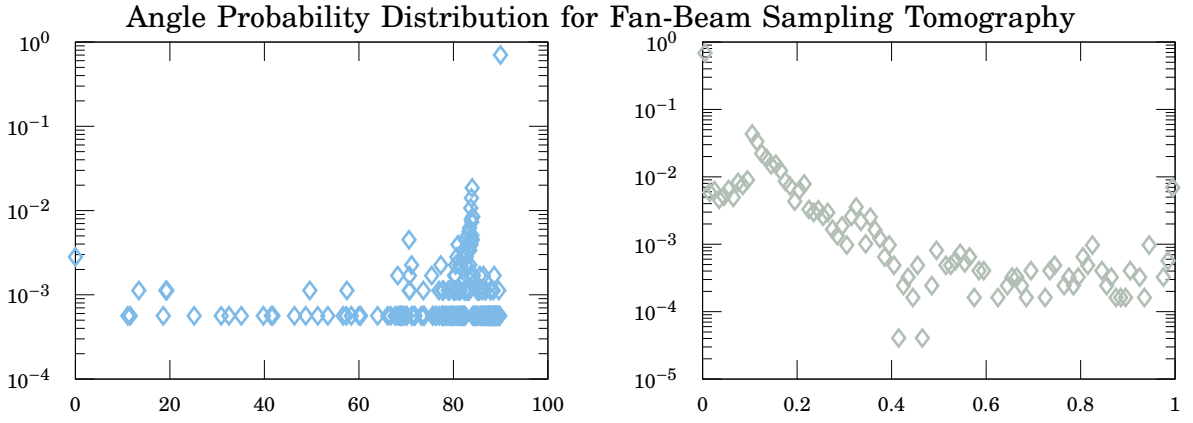
V. Experimental Results

Here, we compare our angle-based randomization with norm-based randomization of Strohmer [4] in the context of measurement methods. In particular, a phantom image was used as the solution in simulation experiments [3]. Figure V.5 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. However, our method is



(a) $\theta(i,j)$ probability density distribution (b) Gramian $G(i,j)$ density distribution

Fig. V.1: (a) Example angles distribution (y-axis) from AA^T where $\theta_{i,j} = \cos^{-1}(\langle \hat{\boldsymbol{a}}_i, \hat{\boldsymbol{a}}_j \rangle)\ \forall\ i,j \in \{1,\ldots,M\}$ vs angles (x-axis) degrees using random data acquisition strategy, (b) Gramian matrix $\langle \hat{\boldsymbol{a}}_i, \hat{\boldsymbol{a}}_j \rangle$ distribution



(a) $\theta(i,j)$ probability density distribution (b) Gramian $G(i,j)$ density distribution

Fig. V.2: (a) Example normalized angles distribution (y-axis) for the first eight columns of AA^T where $\theta_{i,j} = \cos^{-1}(\langle \hat{\boldsymbol{a}}_i, \hat{\boldsymbol{a}}_j \rangle)\ \forall\ i,j \in \{1,\ldots,M\}$ vs angles (x-axis) degrees using fan-beam tomographic data acquisition strategy, (b) Gramian matrix $\langle \hat{\boldsymbol{a}}_i, \hat{\boldsymbol{a}}_j \rangle$ distribution

computationally more complex, and therefore we devised another algorithm (explained in the next following section) that addresses this issue.

The following experiments compare Kaczmarz (K), randomized Kaczmarz (RK), and randomized Kaczmarz hyperplane angles (RKHA) via simulations. The objective is to illustrate the effect of row randomization upon the convergence and observe the dependence upon the sampling methods.

A. Angular Distribution of 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 (III.1b))[9], *paralleltomo.m* and *fanbeamtomo.m* from the AIRtools distribution [23], and *randn()* from the built-in function method [24].

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

$$\phi_{i,j}=\max_{1\le i\ne j\le M}\left|\boldsymbol{a}_i^H\boldsymbol{a}_j\right|. \tag{73}$$

A lower bound was derived as $\phi\ge\frac{M-N}{N(M-1)}$ in reference Welch [26].

It is noted that the statistical expectation[10] of the non-diagonal Gramian matrix elements $G_{i,j}=\langle\hat{\boldsymbol{a}}_i,\hat{\boldsymbol{a}}_j\rangle(1\le i\ne j\le M)$ for normalized random unit vectors $\{\hat{\boldsymbol{a}}_i,\hat{\boldsymbol{a}}_j\}$ would be zero for two independent random IID row vectors, $1/N$ for the case of a single dependent vector component (one variable in N variables), and the maximum expected value occurs when two unit row vectors are parallel, which gives a value of unity. Estimated numerical results for the three sampling methods are shown in Table (I) along with values for the mean of the Gramian.

Computations of the Gramian and angular density distributions are shown in Figures (V.1), (V.2), and (V.3). It should be noted that the random sampling is concentrated near 90

[9]Shepp-Logan phantom was generated from AIRtools/paralleltomo.m with non-uniform coherent parallel tomographic CT sampling, P. C. Hansen and M. Saxild-Hansen, AIR Tools - A MATLAB (tm) Package of Algebraic Iterative Reconstruction Methods, Journal of Computational and Applied Mathematics, 236 (2012), pp. 2167-2178

[10]A more formal treatment of the expectation of random IID vectors is given in section (III-C2a).

Coherence vs Measurement Method[11]	Random	Fan	Parallel
coherence Eq. (73)	.4	1.0	1.0
average value of $G_{i,j} = \langle \hat{\boldsymbol{a}}_i, \hat{\boldsymbol{a}}_j \rangle (1 \le i \ne j \le M)$	-.0013	.06	.18
median value of $G_{i,j} = \langle \hat{\boldsymbol{a}}_i, \hat{\boldsymbol{a}}_j \rangle (1 \le i \ne j \le M)$	-0.0009	0	.12

TABLE I: Typical coherence estimates for $N = 100$, $M = 200$ for random *randn()* and $N = 100$, $M = 222$ for fan *fanbeamtomo()* and parallel *paralleltomo()*

degrees probability and zero for the Gramian, but parallel sampling is spread out across the interval $[0, 90]$ degrees.

C. Distribution of Measurement Angles for K, RK, and RKHA for Shepp-Logan Versus Measurement Method

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. This is consistent with the mean values of coherence near zero for random sampling.

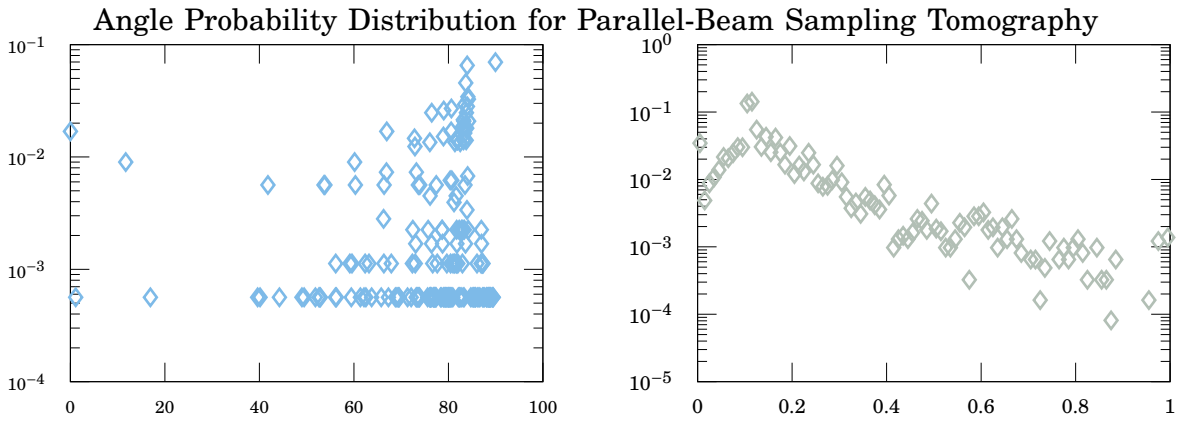
The cases for fan and parallel sampling have increasingly higher coherence, and generally benefit from methods which minimize the coherence, such as RK, RKHA, and RKOS.

Representative results for convergence are shown in Figures (V.4), (V.5), and (V.6). Comparison of convergence results to the estimated coherence for the three cases given in Table (I) suggest consistent interpretation.

Since the iterative methods utilize projections, the angles between the optical lines of sight (LOS) forming the measurement hyperplanes is of considerable interest. The figures also show example computations of distribution of measurement hyperplane angles relative to a hyperplane reference as given by $\theta_{i,j} = \cos^{-1}(\langle \hat{\boldsymbol{a}}_i, \hat{\boldsymbol{a}}_j \rangle) \; \forall \, \hat{\boldsymbol{a}}_i, \hat{\boldsymbol{a}}_j \in A_i \; \forall i, j \in \{1, \ldots, M\}$ where the unit norm vectors $\hat{\boldsymbol{a}}_i, \hat{\boldsymbol{a}}_j$ are selected rows of A.

D. Convergence of K, RK, and RKHA vs Measurement Method

Iterative simulations were performed to estimate the relative convergence rates of methods K, RK, RKHA for the data examples above, random, parallel, and fan beam sampling. Representative results are shown in Figures (V.4), (V.5), and (V.6) for noiseless data measurement scenarios of the standard Shepp-Logan phantom.



(a) $\theta(i,j)$ probability density distribution

(b) Gramian $G(i,j)$ density distribution

Fig. V.3: (a) Example normalized angles distribution (y-axis) from AA^T where $\theta_{i,j} = \cos^{-1}(\langle \hat{\boldsymbol{a}}_i, \hat{\boldsymbol{a}}_j \rangle)\ \forall\ i,j \in \{1,\ldots,M\}$ vs angles (x-axis) degrees using parallel-beam tomographic data acquisition strategy, (b) Gramian matrix $\langle \hat{\boldsymbol{a}}_i, \hat{\boldsymbol{a}}_j \rangle$ distribution

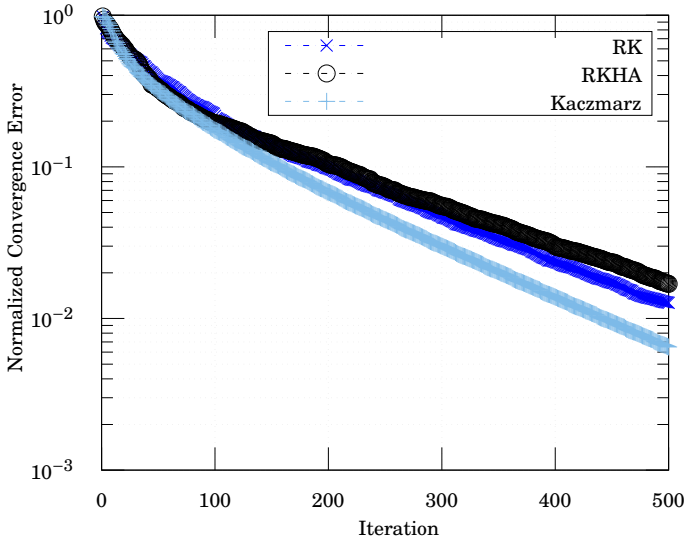


Fig. V.4: Semilog (y-axis) plot example convergence result for K, RK, RKHA on Shepp-Logan phantom using IID random tomographic data acquisition for 10 cycles of iteration (x-axis). Note that randomization tends to equalize convergence rates and diminish advantage of a particular method.

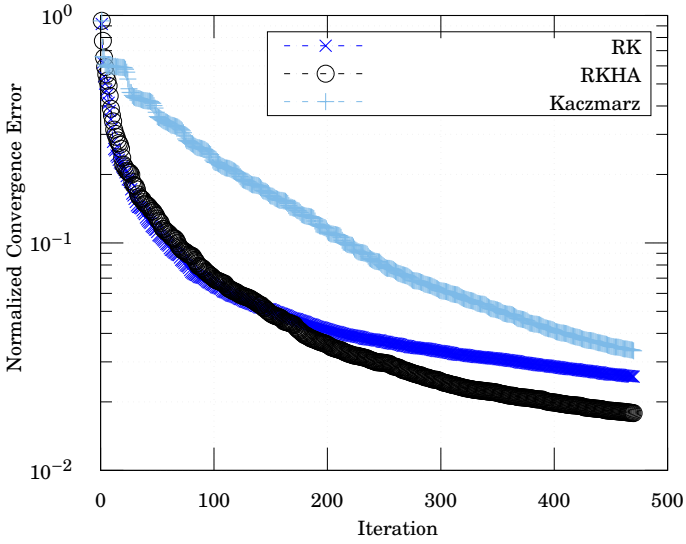


Fig. V.5: Semilog (y-axis) plot example convergence result for K, RK, RKHA on Shepp-Logan phantom using fan tomographic data acquisition for 10 cycles of iteration (x-axis). Note that both RK and RKHA appear to have advantage since each method utilizes randomization which improves avoidance of coherent neighbors, but simple Kaczmarz is too naive.

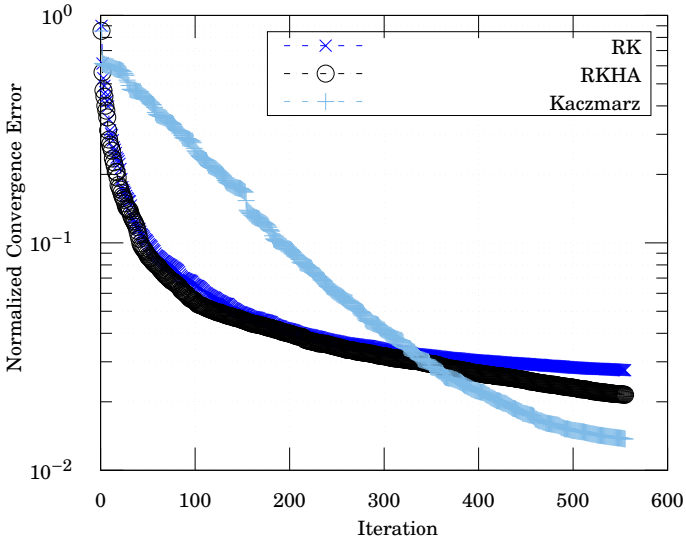


Fig. V.6: Semilog (y-axis) plot example convergence result for K, RK, RKHA on Shepp-Logan phantom using parallel tomographic data acquisition for 10 cycles of iteration (x-axis). Note that initially, both RK and RKHA have similar advantage, but simple Kaczmarz eventually improves.

VI. Conclusions

A new iterative selection rule based upon the relative central angle 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 incoherence, 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 convergence of above subspace methods was demonstrated using statistical IID assumptions. But, the more generalized approach based upon cyclical projections using the formalism of Galantai also prove convergence, without the statistical argument.

It is worthwhile to note that an additional method to prove the convergence rate for a given angular probability distribution function is currently underway and is considered an essential task towards validation of the RKHA results.

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, p. 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, p. 471–481, 1970.

[3] G. T. Herman, *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*, 2nd ed. Springer Publishing Company, Incorporated, 2009.

[4] T. Strohmer and R. Vershynin, "A randomized kaczmarz algorithm with exponential convergence," *Journal of Fourier Analysis and Applications*, vol. 15, no. 2, p. 262–278, 2009.

[5] D. Needell, "Randomized kaczmarz solver for noisy linear systems," *BIT Numerical Mathematics*, vol. 50, no. 2, p. 395–403, 2010.

[6] D. Needell and R. Ward, "Two-subspace projection method for coherent overdetermined systems," *Journal of Fourier Analysis and Applications*, vol. 19, no. 2, p. 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, no. 0, p. 199–221, 2014.

[8] X. Chen and A. Powell, "Almost sure convergence of the kaczmarz algorithm with random measurements," *Journal of Fourier Analysis and Applications*, vol. 18, no. 6, p. 1195–1214, 2012.

[9] A. Galantai, *Projectors and projection methods*, ser. Advances in mathematics. Boston, Dordrecht, London: Kluwer Academic, 2004.

[10] Galantai, "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. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0022247X04010716

[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, p. 381–405, 1997.

[12] C. Brezinski and M. Redivo-Zaglia, "Convergence acceleration of kaczmarz's method," *Journal of Engineering Mathematics*, p. 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, 11 1977. [Online]. Available: http://projecteuclid.org/euclid.bams/1183539851

[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, p. 215–236, 2013. [Online]. Available: http://dx.doi.org/10.1007/s00211-012-0512-6

[15] Y. Censor, G. 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, pp. 431–436, 2009.

[16] T. Wallace and A. Sekmen, "Acceleration of Kaczmarz Using Subspace Orthogonal Projections,T. Wallace and A. Sekmen, IEEE Biomedical Science and Engineering Conference (BSEC) at Oak Ridge National Laboratory, 2013," *IEEE Biomedical Science and Engineering Conference (BSEC) at Oak Ridge National Laboratory, 2013*, 2013.

[17] G. H. Golub and C. F. Van Loan, *Matrix Computations (Johns Hopkins Studies in Mathematical Sciences)(3rd Edition)*, 3rd ed. The Johns Hopkins University Press, Oct. 1996. [Online]. Available: http://www.worldcat.org/isbn/0801854148

[18] H. Yanai, K. Takeuchi, and Y. Takane, *Projection Matrices, Generalized Inverse Matrices, and Singular Value Decomposition*, ser. Statistics for Social and Behavioral Sciences. Dordrecht: Springer, 2011.

[19] Meyer, *Matrix analysis and applied linear algebra*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2000.

[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] A. Galántai, *Projectors and Projection Methods*, ser. Advances in Mathematics (Kluwer Academic Publishers), V. 6. Springer, 2003. [Online]. Available: http://books.google.com/books?id=n1ZpS0aTtjwC

[22] I. Halperin, "The product of projection operators," *Acta Sci. Math. (Szeged) 23, 96-99.*, 1962.

[23] 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, inverse Problems: Computation and Applications. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0377042711005188

[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, 10 2000. [Online]. Available: http://www.jstatsoft.org/v05/i08

[25] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *Information Theory, IEEE Transactions on*, vol. 52, no. 1, pp. 6–18, Jan 2006.

[26] L. Welch, "Lower bounds on the maximum cross correlation of signals (corresp.)," *Information Theory, IEEE Transactions on*, vol. 20, no. 3, pp. 397–399, May 1974.