1-1-2012

Two-subspace Projection Method for Coherent Overdetermined Systems

Deanna Needell
Claremont McKenna College

Rachel Ward
University of Texas at Austin

TWO-SUBSPACE PROJECTION METHOD FOR COHERENT OVERDETERMINED SYSTEMS

DEANNA NEEDELL AND RACHEL WARD

ABSTRACT. We present a Projection onto Convex Sets (POCS) type algorithm for solving systems of linear equations. POCS methods have found many applications ranging from computer tomography to digital signal and image processing. The Kaczmarz method is one of the most popular solvers for overdetermined systems of linear equations due to its speed and simplicity. Here we introduce and analyze an extension of the Kaczmarz method which iteratively projects the estimate onto a solution space given from two randomly selected rows. We show that this projection algorithm provides exponential convergence to the solution in expectation. The convergence rate significantly improves upon that of the standard randomized Kaczmarz method when the system has coherent rows. We also show that the method is robust to noise, and converges exponentially in expectation to the noise floor. Experimental results are provided which confirm that in the coherent case our method significantly outperforms the randomized Kaczmarz method.

1. INTRODUCTION

We consider a consistent system of linear equations of the form

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b},$$

where $\boldsymbol{b} \in \mathbb{C}^m$ and $\boldsymbol{A} \in \mathbb{C}^{m \times n}$ is a full-rank $m \times n$ matrix that is overdetermined, having more rows than columns ($m \geq n$). When the number of rows of $\boldsymbol{A}$ is large, it is far too costly to invert the matrix to solve for $\boldsymbol{x}$, so one may utilize an iterative solver such as the Projection onto Convex Sets (POCS) method, used in many applications of signal and image processing [1, 18]. The Kaczmarz method is often preferred, iteratively cycling through the rows of $\boldsymbol{A}$ and orthogonally projecting the estimate onto the solution space given by each row [10]. Precisely, let us denote by $\boldsymbol{a_1}, \boldsymbol{a_2}, \ldots, \boldsymbol{a_m}$ the rows of $\boldsymbol{A}$ and $b_1, b_2, \ldots, b_m$ the coordinates of $\boldsymbol{b}$. For simplicity, we will assume throughout that the matrix $\boldsymbol{A}$ is *standardized*, meaning that each of its rows has unit Euclidean norm; generalizations from this case will be straightforward. Given some trivial initial estimate $\boldsymbol{x_0}$, the Kaczmarz method cycles through the rows of $\boldsymbol{A}$ and in the kth iteration projects the previous estimate $\boldsymbol{x_k}$ onto the solution hyperplane of $\langle \boldsymbol{a_i}, \boldsymbol{x} \rangle = b_i$ where $i = k \bmod m$,

$$\boldsymbol{x_{k+1}} = \boldsymbol{x_k} + (b_i - \langle \boldsymbol{a_i}, \boldsymbol{x_k} \rangle)\boldsymbol{a_i}.$$

Theoretical results about the rate of convergence of the Kaczmarz method have been difficult to obtain, and most are based on quantities which are themselves hard to compute [3, 7]. Even more importantly, the method as we have just described depends heavily on the ordering of the rows of $\boldsymbol{A}$. A malicious or unlucky ordering may therefore lead to extremely slow convergence. To overcome this, one can select the rows of $\boldsymbol{A}$ in a *random* fashion rather than cyclically [9, 12]. Strohmer and Vershynin analyzed a randomized version of the Kaczmarz method that in each iteration selects a row of $\boldsymbol{A}$ with probability proportional to its Euclidean norm [20, 19]. Thus in the standardized case we consider here, a row of $\boldsymbol{A}$ is chosen uniformly at random. This randomized Kaczmarz method is described by the following pseudocode.

ALGORITHM 1.1: Randomized Kaczmarz

Input: Standardized matrix $\boldsymbol{A}$, vector $\boldsymbol{b}$
Output: An estimation $\boldsymbol{x_k}$ of the unique solution $\boldsymbol{x}$ to $\boldsymbol{Ax} = \boldsymbol{b}$

Set $\boldsymbol{x_0}$.   { Trivial initial approximation }
$k \leftarrow 0$

repeat
   $k \leftarrow k + 1$
   Select $r \in \{1, 2, \ldots, n\}$ { Randomly select a row of $\boldsymbol{A}$ }
   Set $\boldsymbol{x_k} \leftarrow \boldsymbol{x_{k-1}} + (b_r - \langle \boldsymbol{a_r}, \boldsymbol{x_{k-1}} \rangle)\boldsymbol{a_r}$ { Perform projection }

Note that this method as stated selects each row *with replacement*, see [17] for a discussion on the differences in performance when selecting with and without replacement. Strohmer and Vershynin show that this method exhibits exponential convergence in expectation [20, 19],

$$(1.1) \qquad \mathbb{E}\|\boldsymbol{x_k} - \boldsymbol{x}\|_2^2 \leq \left(1 - \frac{1}{R}\right)^k \|\boldsymbol{x_0} - \boldsymbol{x}\|_2^2, \quad \text{where} \quad R \stackrel{\text{def}}{=} \|\boldsymbol{A}\|_F^2 \|\boldsymbol{A}^{-1}\|^2.$$

Here and throughout, $\| \cdot \|_2$ denotes the vector Euclidean norm, $\| \cdot \|$ denotes the matrix spectral norm, $\| \cdot \|_F$ denotes the matrix Frobenius norm, and the inverse $\|\boldsymbol{A}^{-1}\| = \inf\{M : M\|\boldsymbol{Ax}\|_2 \geq \|\boldsymbol{x}\|_2 \text{ for all } \boldsymbol{x}\}$ is well-defined since $\boldsymbol{A}$ is full-rank. This bound shows that when $\boldsymbol{A}$ is well conditioned, the randomized Kaczmarz method will converge exponentially to the solution in just $\mathrm{O}(n)$ iterations (see Section 2.1 of [20] for details). The cost of each iteration is the cost of a single projection and takes $\mathrm{O}(n)$ time, so the total runtime is just $\mathrm{O}(n^2)$. This is superior to Gaussian elimination which takes $\mathrm{O}(mn^2)$ time, especially for very large systems. The randomized Kaczmarz method even substantially outperforms the well-known conjugate gradient method in many cases [20].

Leventhal and Lewis show that for certain probability distributions, the expected rate of convergence can be bounded in terms of other natural linear-algebraic quantities. They propose generalizations to other convex systems [11]. Recently, Chen and

Powell proved that for certain classes of random matrices $\boldsymbol{A}$, the randomized Kaczmarz method convergences exponentially to the solution not only in expectation but also almost surely [16].

In the presence of noise, one considers the possibly inconsistent system $\boldsymbol{Ax} + \boldsymbol{w} \approx \boldsymbol{b}$ for some error vector $\boldsymbol{w}$. In this case the randomized Kaczmarz method converges exponentially fast to the solution within an error threshold [13],

$$(1.2) \qquad \mathbb{E}\|\boldsymbol{x_k} - x\|_2 \leq \left(1 - \frac{1}{R}\right)^{k/2} \|\boldsymbol{x_0} - x\|_2 + \sqrt{R}\|\boldsymbol{w}\|_\infty,$$

where R the the scaled condition number as in (1.1) and $\|\cdot\|_\infty$ denotes the largest entry in magnitude of its argument. This error is sharp in general [13]. Modified Kaczmarz algorithms can also be used to solve the least squares version of this problem, see for example [4, 5, 8, 2] and the references therein.

1.1. **Coherent systems.** Although the convergence results for the randomized Kaczmarz method hold for any consistent system, the factor $\frac{1}{R}$ in the convergence rate may be quite small for matrices with many correlated rows. Consider for example the reconstruction of a bandlimited function from nonuniformly spaced samples, as often arises in geophysics as it can be physically challenging to take uniform samples. Expressed as a system of linear equations, the sampling points form the rows of a matrix $\boldsymbol{A}$; for points that are close together, the corresponding rows will be highly correlated.

To be precise, we examine the *coherence* of a standardized matrix $\boldsymbol{A}$ by defining the quantities

$$(1.3) \qquad \Delta = \Delta(\boldsymbol{A}) = \max_{j \neq k} |\langle \boldsymbol{a_j}, \boldsymbol{a_k} \rangle| \quad and \quad \delta = \delta(\boldsymbol{A}) = \min_{j \neq k} |\langle \boldsymbol{a_j}, \boldsymbol{a_k} \rangle|.$$

Note that because $\boldsymbol{A}$ is standardized, $0 \leq \delta \leq \Delta \leq 1$. It is clear that when $\boldsymbol{A}$ has high coherence parameters, $\|\boldsymbol{A}^{-1}\|$ is very small and thus the factor R in (1.1) is also small, leading to a weak bound on the convergence. Indeed, when the matrix has highly correlated rows, the angles between successive orthogonal projections are small and convergence is stunted. We can explore a wider range of orthogonal directions by looking towards solution hyperplanes spanned by *pairs* of rows of $\boldsymbol{A}$. We thus propose a modification to the randomized Kaczmarz method where each iteration performs an orthogonal projection onto a two-dimensional subspace spanned by a randomly-selected pair of rows. We point out that the idea of projecting in each iteration onto a subspace obtained from multiple rows rather than a single row has been previously investigated numerically, see e.g. [6, 1].

With this as our goal, a single iteration of the modified algorithm will consist of the following steps. Let $\boldsymbol{x_k}$ denote the current estimation in the kth iteration.

- Select two distinct rows $\boldsymbol{a_r}$ and $\boldsymbol{a_s}$ of the matrix $\boldsymbol{A}$ at random
- Compute the translation parameter ε
- Perform an intermediate projection: $\boldsymbol{y} \leftarrow \boldsymbol{x_k} + \varepsilon(b_r - \langle \boldsymbol{x_k}, \boldsymbol{a_r} \rangle)\boldsymbol{a_r}$

- Perform the final projection to update the estimation: $\boldsymbol{x_{k+1}} \leftarrow \boldsymbol{y} + (b_s - \langle \boldsymbol{y}, \boldsymbol{a_s} \rangle)\boldsymbol{a_s}$

In general, the optimal choice of ε at each iteration of the two-step procedure corresponds to subtracting from $\boldsymbol{x_k}$ its orthogonal projection onto the solution space $\{\boldsymbol{x} : \langle \boldsymbol{a_r}, \boldsymbol{x} \rangle = b_r \text{ and } \langle \boldsymbol{a_s}, \boldsymbol{x} \rangle = b_s\}$, which motivates the name two-subspace Kaczmarz method. By *optimal choice* of ε, we mean the value ε_{opt} minimizing the residual $\|\boldsymbol{x} - \boldsymbol{x_{k+1}}\|_2^2$. Expanded, this reads

$$\|\boldsymbol{x} - \boldsymbol{x_{k+1}}\|_2^2 = \|\varepsilon(b_r - \langle \boldsymbol{x_k}, \boldsymbol{a_r} \rangle)(\boldsymbol{a_r} - \langle \boldsymbol{a_s}, \boldsymbol{a_r} \rangle \boldsymbol{a_s}) + \boldsymbol{x_k} - \boldsymbol{x} + (b_s - \langle \boldsymbol{x_k}, \boldsymbol{a_s} \rangle)\boldsymbol{a_s}\|_2^2.$$

Using that the minimizer of $\|\gamma\boldsymbol{w} + \boldsymbol{z}\|_2^2$ is $\gamma = -\frac{\langle \boldsymbol{w}, \boldsymbol{z} \rangle}{\|\boldsymbol{w}\|_2^2}$, we see that

$$\varepsilon_{opt} = \frac{-\langle \boldsymbol{a_r} - \langle \boldsymbol{a_s}, \boldsymbol{a_r} \rangle \boldsymbol{a_s}, \boldsymbol{x_k} - \boldsymbol{x} + (b_s - \langle \boldsymbol{x_k}, \boldsymbol{a_s} \rangle)\boldsymbol{a_s} \rangle}{(b_r - \langle \boldsymbol{x_k}, \boldsymbol{a_r} \rangle)\|\boldsymbol{a_r} - \langle \boldsymbol{a_s}, \boldsymbol{a_r} \rangle \boldsymbol{a_s}\|_2^2}.$$

Note that the unknown vector $\boldsymbol{x}$ appears in this expression only through its observable inner products, and so ε_{opt} is computable. After some algebra, one finds that the two-step procedure with this choice of ε_{opt} can be re-written as follows [15].

ALGORITHM 1.2: Two-subspace Kaczmarz

Input: Matrix $\boldsymbol{A}$, vector $\boldsymbol{b}$
Output: An estimation $\boldsymbol{x_k}$ of the unique solution $\boldsymbol{x}$ to $\boldsymbol{Ax} = \boldsymbol{b}$

Set $\boldsymbol{x_0}$.   { Trivial initial approximation }
$k \leftarrow 0$

repeat
   $k \leftarrow k + 1$
   Select $r, s \in \{1, 2, \ldots, n\}$ { Select two distinct rows of $\boldsymbol{A}$ uniformly at random }
   Set $\mu_k \leftarrow \langle \boldsymbol{a_r}, \boldsymbol{a_s} \rangle$ { Compute correlation }
   Set $\boldsymbol{y_k} \leftarrow \boldsymbol{x_{k-1}} + (b_s - \langle \boldsymbol{x_{k-1}}, \boldsymbol{a_s} \rangle)\boldsymbol{a_s}$ { Perform intermediate projection }
   Set $\boldsymbol{v_k} \leftarrow \frac{\boldsymbol{a_r} - \mu_k \boldsymbol{a_s}}{\sqrt{1 - |\mu_k|^2}}$ { Compute vector orthogonal to $\boldsymbol{a_s}$ in direction of $\boldsymbol{a_r}$ }
   Set $\beta_k \leftarrow \frac{b_r - b_s \mu_k}{\sqrt{1 - |\mu_k|^2}}$ { Compute corresponding measurement }
   $\boldsymbol{x_k} \leftarrow \boldsymbol{y_k} + (\beta_k - \langle \boldsymbol{y_k}, \boldsymbol{v_k} \rangle)\boldsymbol{v_k}$ { Perform projection }

Our main result shows that the two-subspace Kaczmarz algorithm provides the same exponential convergence rate as the standard method in general, and substantially improved convergence when the rows of $\boldsymbol{A}$ are coherent [15]. Figure 1 plots two iterations of the one-subspace random Kaczmarz and compares this to a single iteration of the two-subspace Kaczmarz algorithm.

Theorem 1.1. *Let $\boldsymbol{A}$ be a full-rank standardized matrix with n columns and $m > n$ rows and suppose $\boldsymbol{Ax} = \boldsymbol{b}$. Let $\boldsymbol{x_k}$ denote the estimation to the solution $\boldsymbol{x}$ in the*
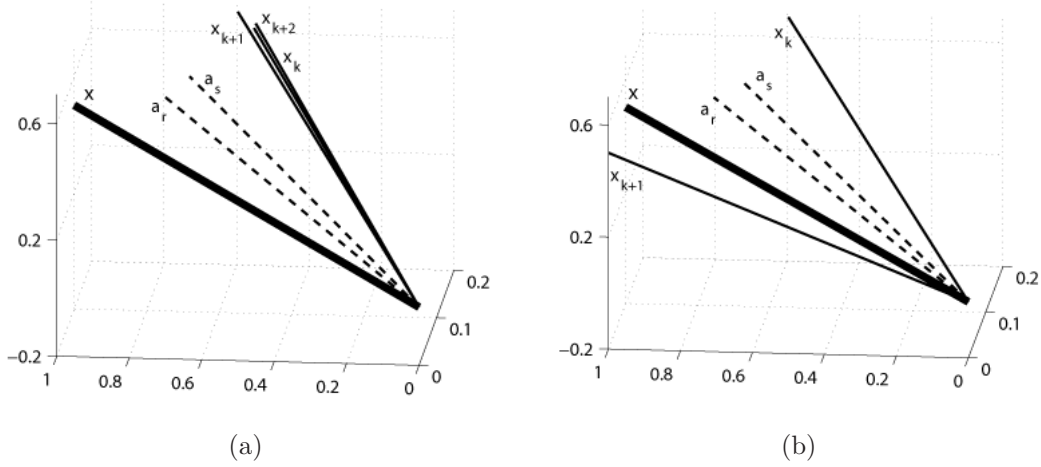
FIGURE 1. For coherent systems, the one-subspace randomized Kaczmarz algorithm (a) converges more slowly than the two-subspace Kaczmarz algorithm (b).

kth iteration of the two-subspace Kaczmarz method. Then

$$\mathbb{E}\|\boldsymbol{x} - \boldsymbol{x_k}\|_2^2 \leq \left(\left(1 - \frac{1}{R} \right)^2 - \frac{D}{R} \right)^k \|\boldsymbol{x} - \boldsymbol{x}_0\|_2^2,$$

where $D = \min \left\{ \frac{\delta^2(1-\delta)}{1+\delta}, \frac{\Delta^2(1-\Delta)}{1+\Delta} \right\}$, Δ and δ are the coherence parameters (1.3), *and $R = \|\boldsymbol{A}\|_F^2 \|\boldsymbol{A}^{-1}\|^2$ denotes the scaled condition number.*

Remarks. **1.** When $\Delta = 1$ or $\delta = 0$ we recover the same convergence rate as provided for the standard Kaczmarz method (1.1) since the two-subspace method utilizes two projections per iteration.

2. The bound presented in Theorem 1.1 is a pessimistic bound. Even when $\Delta = 1$ or $\delta = 0$, the two-subspace method improves on the standard method if any rows of $\boldsymbol{A}$ are highly correlated (but not equal). This is evident in the proof of Theorem 1.1 in Section 2 but we present this bound for simplicity. See also Section 4 for more details on improved convergence bounds.

Figure 2 shows the value of D of Theorem 1.1 for various values of Δ and δ. This demonstrates that in the best case (when $\delta \approx \Delta \approx 0.62$), the convergence rate is improved by at least a factor of 0.1.

1.2. **Organization.** The remainder of the report is organized as follows. In Section 2 we state and prove the main lemmas which serve as the proof of Theorem 1.1. Section 3 discusses the two-subspace Kaczmarz method in the presence of noise and shows that in this case the method exhibits exponential convergence to an error
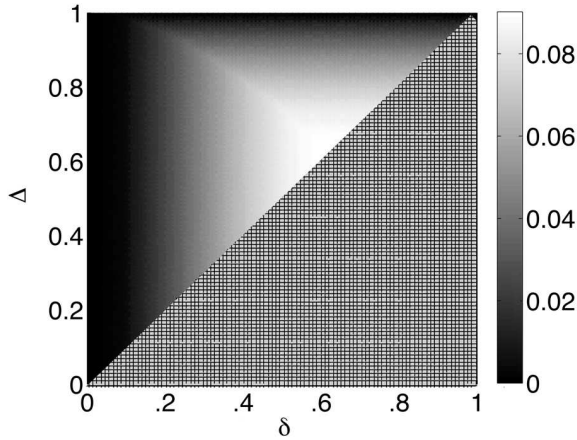
FIGURE 2. A plot of the improved convergence factor D as a function of the coherence parameters δ and $\Delta \geq \delta$.

threshold. Section 4 presents further modifications of the two-subspace Kaczmarz method which provide even more improvements on the provable convergence bounds. A discussion of these methods is provided in Section 6. We conclude with numerical experiments demonstrating the improvements from our method in Section 5.

2. MAIN RESULTS

We now present the proof of Theorem 1.1. We first derive a bound for the expected progress made in a single iteration. Since the two row indices are chosen independently at each iteration, we will be able to apply the bound recursively to obtain the desired overall expected convergence rate.

Our first lemma shows that the expected estimation error in a single iteration of the two-subspace Kaczmarz method is decreased by a factor strictly less than that of the standard randomized method.

Lemma 2.1. *Let $\boldsymbol{x_k}$ denote the estimation to the solution of $\boldsymbol{Ax} = \boldsymbol{b}$ in the kth iteration of the two-subspace Kaczmarz method. Denote the rows of $\boldsymbol{A}$ by $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots \boldsymbol{a}_m$. Then we have the following bound,*

$$\mathbb{E}\|\boldsymbol{x} - \boldsymbol{x_k}\|_2^2 \leq \left(1 - \frac{1}{R}\right)^2 \|\boldsymbol{x} - \boldsymbol{x_{k-1}}\|_2^2 - \frac{1}{m^2 - m}\sum_{r<s} C_{r,s}^2 \left(\langle \boldsymbol{x} - \boldsymbol{x_{k-1}}, \boldsymbol{a_r}\rangle^2 + \langle \boldsymbol{x} - \boldsymbol{x_{k-1}}, \boldsymbol{a_s}\rangle^2\right),$$

where $C_{r,s} = \frac{|\mu_{r,s}| - \mu_{r,s}^2}{\sqrt{1 - \mu_{r,s}^2}}$, $\mu_{r,s} = \langle \boldsymbol{a_r}, \boldsymbol{a_s}\rangle$, and $R = \|\boldsymbol{A}^{-1}\|^2 \|\boldsymbol{A}\|_F^2$ denotes the scaled condition number.

Proof. We fix an iteration k and for convenience refer to $\boldsymbol{v}_k$, μ_k, and $\boldsymbol{y}_k$ as $\boldsymbol{v}$, μ, and $\boldsymbol{y}$, respectively. We will also denote $\gamma = \langle \boldsymbol{a_r}, \boldsymbol{v}\rangle$.

First, observe that by the definitions of $\boldsymbol{v}$ and $\boldsymbol{x_k}$ we have

$$\boldsymbol{x_k} = \boldsymbol{x_{k-1}} + \langle \boldsymbol{x} - \boldsymbol{x_{k-1}}, \boldsymbol{a_s}\rangle \boldsymbol{a_s} + \langle \boldsymbol{x} - \boldsymbol{x_{k-1}}, \boldsymbol{v}\rangle \boldsymbol{v}.$$

Since $\boldsymbol{a_s}$ and $\boldsymbol{v}$ are orthonormal, this gives the estimate

$$(2.1) \qquad \|\boldsymbol{x} - \boldsymbol{x_k}\|_2^2 = \|\boldsymbol{x} - \boldsymbol{x_{k-1}}\|_2^2 - |\langle \boldsymbol{x} - \boldsymbol{x_{k-1}}, \boldsymbol{a_s}\rangle|^2 - |\langle \boldsymbol{x} - \boldsymbol{x_{k-1}}, \boldsymbol{v}\rangle|^2$$

We wish to compare this error with the error from the standard randomized Kaczmarz method. Since we utilize two rows per iteration in the two-subspace Kaczmarz method, we compare its error with the error from two iterations of the standard method. Let $\boldsymbol{z}$ and $\boldsymbol{z'}$ be two subsequent estimates in the standard method following the estimate $\boldsymbol{x_{k-1}}$, and assume $\boldsymbol{z} \neq \boldsymbol{z'}$. That is,

$$(2.2) \qquad \boldsymbol{z} = \boldsymbol{x_{k-1}} + (b_r - \langle \boldsymbol{x_{k-1}}, \boldsymbol{a_r}\rangle)\boldsymbol{a_r} \quad \text{and} \quad \boldsymbol{z'} = \boldsymbol{z} + (b_s - \langle \boldsymbol{z}, \boldsymbol{a_s}\rangle)\boldsymbol{a_s}.$$

Recalling the definitions of $\boldsymbol{v}$, μ and γ, we have

$$(2.3) \qquad \boldsymbol{a_r} = \mu \boldsymbol{a_s} + \gamma \boldsymbol{v} \quad \text{with} \quad \mu^2 + \gamma^2 = 1.$$

Substituting this into (2.2) yields

$$\boldsymbol{z} = \boldsymbol{x_{k-1}} + \mu\langle \boldsymbol{x} - \boldsymbol{x_{k-1}}, \boldsymbol{a_r}\rangle \boldsymbol{a_s} + \gamma\langle \boldsymbol{x} - \boldsymbol{x_{k-1}}, \boldsymbol{a_r}\rangle \boldsymbol{v}.$$

Now substituting this into (2.2) and taking the orthogonality of $\boldsymbol{a_s}$ and $\boldsymbol{v}$ into account,

$$\boldsymbol{z'} = \boldsymbol{x_{k-1}} + \langle \boldsymbol{x} - \boldsymbol{x_{k-1}}, \boldsymbol{a_s}\rangle \boldsymbol{a_s} + \gamma\langle \boldsymbol{x} - \boldsymbol{x_{k-1}}, \boldsymbol{a_r}\rangle \boldsymbol{v}.$$

For convenience, let $\boldsymbol{e_{k-1}} = \boldsymbol{x} - \boldsymbol{x_{k-1}}$ denote the error in the $(k-1)$st iteration of two-subspace Kaczmarz. Then we have

$$
\begin{aligned}
\|\boldsymbol{x} - \boldsymbol{z'}\|_2^2 &= \|\boldsymbol{e_{k-1}} - \langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s}\rangle \boldsymbol{a_s} - \gamma\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_r}\rangle \boldsymbol{v}\|_2^2 \\
&= \|\boldsymbol{e_{k-1}} - \langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s}\rangle \boldsymbol{a_s} - \langle \boldsymbol{e_{k-1}}, \boldsymbol{v}\rangle \boldsymbol{v} - (\gamma\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_r}\rangle - \langle \boldsymbol{e_{k-1}}, \boldsymbol{v}\rangle)\boldsymbol{v}\|_2^2 \\
&= \|\boldsymbol{e_{k-1}}\|_2^2 - |\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s}\rangle|^2 - |\langle \boldsymbol{e_{k-1}}, \boldsymbol{v}\rangle|^2 + |\gamma\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_r}\rangle - \langle \boldsymbol{e_{k-1}}, \boldsymbol{v}\rangle|^2.
\end{aligned}
$$

The third equality follows from the orthonormality of $\boldsymbol{a_s}$ and $\boldsymbol{v}$. We now expand the last term,

$$
\begin{aligned}
|\gamma\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_r}\rangle - \langle \boldsymbol{e_{k-1}}, \boldsymbol{v}\rangle|^2 &= |\gamma\langle \boldsymbol{e_{k-1}}, \mu\boldsymbol{a_s} + \gamma\boldsymbol{v}\rangle - \langle \boldsymbol{e_{k-1}}, \boldsymbol{v}\rangle|^2 \\
&= |\gamma^2\langle \boldsymbol{e_{k-1}}, \boldsymbol{v}\rangle + \gamma\mu\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s}\rangle - \langle \boldsymbol{e_{k-1}}, \boldsymbol{v}\rangle|^2 \\
&= |\mu^2\langle \boldsymbol{e_{k-1}}, \boldsymbol{v}\rangle - \gamma\mu\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s}\rangle|^2.
\end{aligned}
$$

This gives

$$\|\boldsymbol{x} - \boldsymbol{z'}\|_2^2 = \|\boldsymbol{e_{k-1}}\|_2^2 - |\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s}\rangle|^2 - |\langle \boldsymbol{e_{k-1}}, \boldsymbol{v}\rangle|^2 + |\mu^2\langle \boldsymbol{e_{k-1}}, \boldsymbol{v}\rangle - \gamma\mu\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s}\rangle|^2.$$

Combining this identity with (2.1), we now relate the expected error in the two-subspace Kaczmarz algorithm, $\mathbb{E}\|\boldsymbol{x} - \boldsymbol{x_k}\|_2^2$ to the expected error of the standard method, $\mathbb{E}\|\boldsymbol{x} - \boldsymbol{z'}\|_2^2$ as follows:

$$(2.4) \qquad \mathbb{E}\|\boldsymbol{x} - \boldsymbol{x_k}\|_2^2 = \mathbb{E}\|\boldsymbol{x} - \boldsymbol{z'}\|_2^2 - \mathbb{E}|\mu^2\langle \boldsymbol{e_{k-1}}, \boldsymbol{v}\rangle - \gamma\mu\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s}\rangle|^2.$$

It thus remains to analyze the last term. Since we select the two rows r and s independently from the uniform distribution over pairs of distinct rows, the expected error is just the average of the error over all $m^2 - m$ ordered choices r, s. To this end we introduce the notation $\mu_{r,s} = \langle \boldsymbol{a_r}, \boldsymbol{a_s} \rangle$. Then by definitions of $\boldsymbol{v}$, μ and γ,

$$\mathbb{E}|\mu^2 \langle \boldsymbol{e_{k-1}}, \boldsymbol{v} \rangle - \gamma \mu \langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s} \rangle|^2$$

$$= \frac{1}{m^2 - m} \sum_{r \neq s} \left| \frac{\mu_{r,s}^2}{\sqrt{1 - \mu_{r,s}^2}} (\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_r} \rangle - \mu_{r,s} \langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s} \rangle) - \mu_{r,s} \sqrt{1 - \mu_{r,s}^2} \langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s} \rangle \right|^2$$

$$= \frac{1}{m^2 - m} \sum_{r \neq s} \left| \frac{\mu_{r,s}^2}{\sqrt{1 - \mu_{r,s}^2}} \langle \boldsymbol{e_{k-1}}, \boldsymbol{a_r} \rangle - \left(\frac{\mu_{r,s}^3}{\sqrt{1 - \mu_{r,s}^2}} + \mu_{r,s} \sqrt{1 - \mu_{r,s}^2} \right) \langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s} \rangle \right|^2$$

$$= \frac{1}{m^2 - m} \sum_{r \neq s} \left| \frac{\mu_{r,s}^2}{\sqrt{1 - \mu_{r,s}^2}} \langle \boldsymbol{e_{k-1}}, \boldsymbol{a_r} \rangle - \left(\frac{\mu_{r,s}}{\sqrt{1 - \mu_{r,s}^2}} \right) \langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s} \rangle \right|^2.$$

We now recall that for any $\theta, \pi, u,$ and v,

$$(\theta u - \pi v)^2 + (\theta v - \pi u)^2 \geq (|\pi| - |\theta|)^2 (u^2 + v^2).$$

Setting $\theta_{r,s} = \frac{\mu_{r,s}^2}{\sqrt{1 - \mu_{r,s}^2}}$ and $\pi_{r,s} = \frac{\mu_{r,s}}{\sqrt{1 - \mu_{r,s}^2}}$, we have by rearranging terms in the symmetric sum,

$$\mathbb{E}|\mu^2 \langle \boldsymbol{e_{k-1}}, \boldsymbol{\theta} \rangle - \gamma \mu \langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s} \rangle|^2$$

$$= \frac{1}{m^2 - m} \sum_{r \neq s} |\theta_{r,s} \langle \boldsymbol{e_{k-1}}, \boldsymbol{a_r} \rangle - \pi_{r,s} \langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s} \rangle|^2$$

$$= \frac{1}{m^2 - m} \sum_{r < s} |\theta_{r,s} \langle \boldsymbol{e_{k-1}}, \boldsymbol{a_r} \rangle - \pi_{r,s} \langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s} \rangle|^2 + |\theta_{r,s} \langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s} \rangle - \pi_{r,s} \langle \boldsymbol{e_{k-1}}, \boldsymbol{a_r} \rangle|^2$$

$$\geq \frac{1}{m^2 - m} \sum_{r < s} (|\pi_{r,s}| - |\theta_{r,s}|)^2 \left((\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_r} \rangle)^2 + (\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s} \rangle)^2 \right)$$

(2.5)

$$= \frac{1}{m^2 - m} \sum_{r < s} \left(\frac{|\mu_{r,s}| - \mu_{r,s}^2}{\sqrt{1 - \mu_{r,s}^2}} \right)^2 \left((\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_r} \rangle)^2 + (\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s} \rangle)^2 \right).$$

Since selecting two rows without replacement (i.e. guaranteeing not to select the same row back to back) can only speed the convergence, we have from (1.1) that the error from the standard randomized Kaczmarz method satisfies

$$\mathbb{E}\|\boldsymbol{x} - \boldsymbol{z}'\|_2^2 \leq (1 - 1/R)^2 \|\boldsymbol{x} - \boldsymbol{x_{k-1}}\|_2^2.$$

Combining this with (2.4) and (2.5) yields the desired result.

□

Although the result of Lemma 2.1 is tighter, using the coherence parameters δ and Δ of (1.3) allows us to present the following looser but simpler result.

Lemma 2.2. *Let $\boldsymbol{x_k}$ denote the estimation to $\boldsymbol{Ax} = \boldsymbol{b}$ in the kth iteration of the two-subspace Kaczmarz method. Denote the rows of $\boldsymbol{A}$ by $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots \boldsymbol{a}_m$. Then*

$$\mathbb{E}\|\boldsymbol{x} - \boldsymbol{x_k}\|_2^2 \leq \left(\left(1 - \frac{1}{R}\right)^2 - \frac{D}{R}\right)\|\boldsymbol{x} - \boldsymbol{x_{k-1}}\|_2^2,$$

where $D = \min\left\{\frac{\delta^2(1-\delta)}{1+\delta}, \frac{\Delta^2(1-\Delta)}{1+\Delta}\right\}$, δ and Δ are the coherence parameters as in (1.3), and $R = \|\boldsymbol{A}^{-1}\|^2\|\boldsymbol{A}\|_F^2$ denotes the scaled condition number.

Proof. By Lemma 2.1 we have
(2.6)
$$\mathbb{E}\|\boldsymbol{x} - \boldsymbol{x_k}\|_2^2 \leq \left(1 - \frac{1}{R}\right)^2 \|\boldsymbol{x} - \boldsymbol{x_{k-1}}\|_2^2 - \frac{1}{m^2 - m}\sum_{r<s} C_{r,s}^2 \left(\langle \boldsymbol{x} - \boldsymbol{x_{k-1}}, \boldsymbol{a}_r\rangle^2 + \langle \boldsymbol{x} - \boldsymbol{x_{k-1}}, \boldsymbol{a}_s\rangle^2\right),$$

where
$$C_{r,s} = \frac{|\langle \boldsymbol{a}_r, \boldsymbol{a}_s\rangle| - \langle \boldsymbol{a}_r, \boldsymbol{a}_s\rangle^2}{\sqrt{1 - \langle \boldsymbol{a}_r, \boldsymbol{a}_s\rangle^2}}.$$

By the assumption that $\delta \leq |\langle \boldsymbol{a}_r, \boldsymbol{a}_s\rangle| \leq \Delta$, we have

$$C_{r,s}^2 \geq \min\left\{\frac{\delta^2(1-\delta)}{1+\delta}, \frac{\Delta^2(1-\Delta)}{1+\Delta}\right\} = D.$$

Thus we have that
$$\frac{1}{m^2 - m}\sum_{r<s} C_{r,s}^2 \left(\langle \boldsymbol{x} - \boldsymbol{x_{k-1}}, \boldsymbol{a}_r\rangle^2 + \langle \boldsymbol{x} - \boldsymbol{x_{k-1}}, \boldsymbol{a}_s\rangle^2\right)$$

$$\geq \frac{D}{m^2 - m}\sum_{r<s} \left(\langle \boldsymbol{x} - \boldsymbol{x_{k-1}}, \boldsymbol{a}_r\rangle^2 + \langle \boldsymbol{x} - \boldsymbol{x_{k-1}}, \boldsymbol{a}_s\rangle^2\right)$$

$$= \frac{D(m-1)}{m^2 - m}\sum_{r=1}^{m} \langle \boldsymbol{x} - \boldsymbol{x_{k-1}}, \boldsymbol{a}_r\rangle^2$$

(2.7)
$$\geq \frac{D}{m} \cdot \frac{\|\boldsymbol{x} - \boldsymbol{x_{k-1}}\|_2^2}{\|\boldsymbol{A}^{-1}\|_2^2}.$$

In the last inequality we have employed the fact that for any $\boldsymbol{z}$,

$$\sum_{r=1}^{m} \langle \boldsymbol{z}, \boldsymbol{a}_r\rangle^2 \geq \frac{\|\boldsymbol{z}\|_2^2}{\|\boldsymbol{A}^{-1}\|_2^2}.$$

Combining (2.7) and (2.6) along with the definition of R yields the claim.

$\square$

Applying Lemma 2.2 recursively and using the fact that the selection of rows in each iteration is independent yields our main result Theorem 1.1.

3. Noisy Systems

Next we consider systems which have been perturbed by noise. The inconsistent system $\boldsymbol{b} = \boldsymbol{Ax}$ now becomes (the possibly inconsistent system) $\boldsymbol{b} = \boldsymbol{Ax} + \boldsymbol{w}$ for some error vector $\boldsymbol{w}$. As evident from (1.2), the standard method with noise exhibits exponential convergence down to an error threshold, which is proportional to $\|\boldsymbol{w}\|_\infty$. Our main result in the noisy case is that the two-subspace version again exhibits even faster exponential convergence, down to a threshold also proportional to $\|\boldsymbol{w}\|_\infty$.

Theorem 3.1. *Let $\boldsymbol{A}$ be a full rank matrix with m rows and suppose $\boldsymbol{b} = \boldsymbol{Ax} + \boldsymbol{w}$ is a noisy system of equations. Let $\boldsymbol{x_k}$ denote the estimation to the solution $\boldsymbol{x}$ in the kth iteration of the two-subspace Kaczmarz method. Then*

$$\mathbb{E}\|\boldsymbol{x} - \boldsymbol{x_k}\|_2 \le \eta^{k/2}\|\boldsymbol{x} - \boldsymbol{x_0}\|_2 + \frac{3}{1 - \sqrt{\eta}} \cdot \frac{\|\boldsymbol{w}\|_\infty}{\sqrt{1 - \Delta^2}},$$

where $\eta = \left(1 - \frac{1}{R}\right)^2 - \frac{D}{R}$, $D = \min\left\{\frac{\delta^2(1-\delta)}{1+\delta}, \frac{\Delta^2(1-\Delta)}{1+\Delta}\right\}$, Δ and δ are the coherence parameters (1.3), and $R = \|\boldsymbol{A}^{-1}\|^2\|\boldsymbol{A}\|_F^2$ denotes the scaled condition number.

As in the case of our main result Theorem 1.1, this bound is not tight. The same improvements mentioned in the remarks about Theorem 1.1 can also be applied here. In particular, the dependence on Δ seems to be only an artifact of the proof (see Section 5. Nonetheless, this result still shows that the two-subspace Kaczmarz method provides expected exponential convergence down to an error threshold which is analagous to that of the standard method. The convergence factors are again substantially better than the standard method for coherent systems.

Proof of Theorem 3.1. Fix an iteration k and denote by $\boldsymbol{y}$, $\boldsymbol{v}$, μ and β the values of $\boldsymbol{y_k}$, $\boldsymbol{v_k}$, μ_k and β_k for convenience. Let $\boldsymbol{y}'$ and β' be the values of $\boldsymbol{y_k}$, and β_k as if there were noise (i.e. $\boldsymbol{w} = 0$). In other words, we have

$$\boldsymbol{y} = \boldsymbol{y}' + w_r \boldsymbol{a_r} \quad \text{and} \quad \beta = \beta' + \frac{w_r + \mu w_s}{\sqrt{1 - \mu^2}}.$$

Then by the definition of $\boldsymbol{x_k}$, we have

$$\boldsymbol{x_k} = \boldsymbol{x_k^*} + w_r \boldsymbol{a_r} + \frac{w_r + \mu w_s}{\sqrt{1 - \mu^2}}\boldsymbol{v},$$

where $\boldsymbol{x_k^*} = \boldsymbol{y'} + (\beta' - \langle \boldsymbol{x_{k-1}}, \boldsymbol{v} \rangle)\boldsymbol{v}$ denotes the next estimation from $\boldsymbol{x_{k-1}}$ if there were no noise. Therefore, we have that

$$\|\boldsymbol{x} - \boldsymbol{x_k}\|_2 \leq \|\boldsymbol{x} - \boldsymbol{x_k^*}\|_2 + \|w_r \boldsymbol{a_r} + \frac{w_r + \mu w_s}{\sqrt{1 - \mu^2}} \boldsymbol{v}\|_2$$

$$\leq \|\boldsymbol{x} - \boldsymbol{x_k^*}\|_2 + |w_r|\|\boldsymbol{a_r}\|_2 + \left| \frac{w_r + \mu w_s}{\sqrt{1 - \mu^2}} \right| \|\boldsymbol{v}\|_2$$

$$= \|\boldsymbol{x} - \boldsymbol{x_k^*}\|_2 + |w_r| + \left| \frac{w_r + \mu w_s}{\sqrt{1 - \mu^2}} \right|$$

$$\leq \|\boldsymbol{x} - \boldsymbol{x_k^*}\|_2 + \frac{3\|\boldsymbol{w}\|_\infty}{\sqrt{1 - \Delta^2}}.$$

By Jensen's inequality and Lemma 2.1,

$$\mathbb{E}\|\boldsymbol{x} - \boldsymbol{x_k^*}\|_2 \leq \sqrt{\eta}\mathbb{E}\|\boldsymbol{x} - \boldsymbol{x_{k-1}}\|_2.$$

Combining the above recursively yields

$$\mathbb{E}\|\boldsymbol{x} - \boldsymbol{x_k}\|_2 \leq \eta^{k/2}\|\boldsymbol{x} - \boldsymbol{x_0}\|_2 + \frac{3\|\boldsymbol{w}\|_\infty}{\sqrt{1 - \Delta^2}} \sum_{j=1}^{k-1} \eta^{j/2}$$

$$\leq \eta^{k/2}\|\boldsymbol{x} - \boldsymbol{x_0}\|_2 + \frac{3}{1 - \sqrt{\eta}} \cdot \frac{\|\boldsymbol{w}\|_\infty}{\sqrt{1 - \Delta^2}},$$

which proves the claim. $\qquad\square$

4. Further improvements

Next we state and prove a lemma which demonstrates even more improvements on the convergence rate from the standard method in the case where the correlations between the rows are non-negative. If this is not the case, we may alter one step of the two-subspace method to generalize the result to matrices with arbitrary correlations. This modification will decrease the factor yet again in the exponential convergence rate of the two-subspace method. We consider the noiseless case here, although results analagous to those in Section 3 can easily be obtained using the same methods.

We first define an $m^2 \times n$ matrix $\boldsymbol{\Omega}$ whose rows ω are differnces of the rows of $\boldsymbol{A}$:

(4.1) $$\boldsymbol{\omega}_{m(j-1)+i} = \begin{cases} \frac{\boldsymbol{a_j} - \boldsymbol{a_i}}{\|\boldsymbol{a_j} - \boldsymbol{a_i}\|}, & j, i = 1, ..., m, \quad j \neq i, \\ \boldsymbol{0}, & j = i \end{cases}$$

We may now state our main lemma.

Lemma 4.1. *Let $\boldsymbol{x_k}$ denote the estimation to $\boldsymbol{Ax} = \boldsymbol{b}$ in the kth iteration of the two-subspace Kaczmarz method. For indices r and s, set $\mu_{r,s} = \langle \boldsymbol{a_r}, \boldsymbol{a_s} \rangle$. We have the following bound,*

$$\mathbb{E}\|\boldsymbol{x} - \boldsymbol{x_k}\|_2^2 \leq \left(1 - \frac{1}{R}\right)^2 \|\boldsymbol{x} - \boldsymbol{x_{k-1}}\|_2^2 - \frac{1}{m^2 - m} \sum_{r<s} C_{r,s}\left(\langle \boldsymbol{x} - \boldsymbol{x_{k-1}}, \boldsymbol{a_r} \rangle^2 + \langle \boldsymbol{x} - \boldsymbol{x_{k-1}}, \boldsymbol{a_s} \rangle^2\right)$$

(4.2)
$$- \frac{1}{m^2 - m} \sum_{j,i=1}^{m} E_{i,j} \langle \boldsymbol{x} - \boldsymbol{x_{k-1}}, \boldsymbol{\omega_{m(j-1)+i}} \rangle^2$$

where $C_{r,s} = \frac{\mu_{r,s}^2(1-\mu_{r,s})}{1+\mu_{r,s}}$ and $E_{i,j} = 4\mu_{i,j}^3$.

Remark. If the correlations $\mu_{r,s} = \langle \boldsymbol{a_r}, \boldsymbol{a_s} \rangle$ between the rows of $\boldsymbol{A}$ are non-negative, then the constants $E_{i,j}$ are all non-negative and thus this lemma offers a strict improvement over Lemma 2.1. However, if this is not the case, this bound may actually be worse than that of Lemma 2.1. To overcome this, we may simply modify the two-subspace Kaczmarz algorithm so that in each iteration $\mu_{r,s}$ is non-negative (by possibly using $-\boldsymbol{a_r}$ instead of $\boldsymbol{a_r}$ when needed for example). This modification seems necessary only for the proof, and empirical results for the modified and unmodified methods remain the same. From this point on, we will assume this modification is in place.

Proof of Lemma 4.1. We again fix an iteration k and for convenience refer to $\boldsymbol{v_k}$, μ_k, and $\boldsymbol{y_k}$ as $\boldsymbol{v}$, μ, and $\boldsymbol{y}$, respectively. We will also let $\boldsymbol{e_{k-1}} = \boldsymbol{x} - \boldsymbol{x_{k-1}}$ be the error in the $(k-1)$st iteration.

By (2.4) and (1.1), we have that

(4.3) $\mathbb{E}\|\boldsymbol{x} - \boldsymbol{x_k}\|_2^2 = (1 - 1/R)\|\boldsymbol{x} - \boldsymbol{x_{k-1}}\|_2^2 + \mathbb{E}|\mu^2 \langle \boldsymbol{e_{k-1}}, \boldsymbol{v} \rangle - \gamma\mu\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s} \rangle|^2.$

We now analyze the last term carefully. To take expectation we must look over all combinations of choices r, s, so to that end denote $\langle \boldsymbol{a_r}, \boldsymbol{a_s} \rangle$ by $\mu_{r,s}$. Since we select two rows uniformly at random (with replacement), using the definitions of $\boldsymbol{c}$, μ and γ, we have

$$\mathbb{E}|\mu^2\langle \boldsymbol{e_{k-1}}, \boldsymbol{c}\rangle - \gamma\mu\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s}\rangle|^2$$

$$= \frac{1}{m^2 - m}\sum_{r\neq s}\left|\frac{\mu_{r,s}^2}{\sqrt{1-\mu_{r,s}^2}}(\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_r}\rangle - \mu_{r,s}\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s}\rangle) - \mu_{r,s}\sqrt{1-\mu_{r,s}^2}\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s}\rangle\right|^2$$

$$= \frac{1}{m^2 - m}\sum_{r\neq s}\left|\frac{\mu_{r,s}^2}{\sqrt{1-\mu_{r,s}^2}}\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_r}\rangle - \left(\frac{\mu_{r,s}^3}{\sqrt{1-\mu_{r,s}^2}} + \mu_{r,s}\sqrt{1-\mu_{r,s}^2}\right)\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s}\rangle\right|^2$$

$$= \frac{1}{m^2 - m}\sum_{r\neq s}\frac{\mu_{r,s}^2}{1-\mu_{r,s}^2}\left(\mu_{r,s}\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_r}\rangle - \langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s}\rangle\right)^2$$

Now observe that

$$\left(\mu_{r,s}\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_r}\rangle - \langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s}\rangle\right)^2 + \left(\mu_{r,s}\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s}\rangle - \langle \boldsymbol{e_{k-1}}, \boldsymbol{a_r}\rangle\right)^2$$

$$= (1-\mu_{r,s})^2\left(\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_r}\rangle^2 + \langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s}\rangle^2\right) + 2\mu_{r,s}\left(\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_r} - \boldsymbol{a_s}\rangle\right)^2$$

$$= (1-\mu_{r,s})^2\left(\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_r}\rangle^2 + \langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s}\rangle^2\right) + 4\mu_{r,s}(1-\mu_{r,s}^2)\left(\langle \boldsymbol{e_{k-1}}, \frac{\boldsymbol{a_r} - \boldsymbol{a_s}}{\|\boldsymbol{a_r} - \boldsymbol{a_s}\|}\rangle\right)^2$$

Thus taking advantage of the symmetry in the sum we have,

$$\mathbb{E}|\mu^2\langle \boldsymbol{e_{k-1}}, \boldsymbol{c}\rangle - \gamma\mu\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s}\rangle|^2$$

$$= \frac{1}{m^2 - m}\sum_{r<s}\frac{\mu_{r,s}^2(1-\mu_{r,s})}{1+\mu_{r,s}}\left(\langle \boldsymbol{e_{k-1}}, \boldsymbol{a_r}\rangle^2 + \langle \boldsymbol{e_{k-1}}, \boldsymbol{a_s}\rangle^2\right) + 4\mu_{r,s}^3\left(\langle \boldsymbol{e_{k-1}}, \frac{\boldsymbol{a_r} - \boldsymbol{a_s}}{\|\boldsymbol{a_r} - \boldsymbol{a_s}\|}\rangle\right)^2$$

(4.4)

Combining with (4.3) yields the claim.

$\square$

We may now use the coherence parameters δ and Δ from (1.3) to obtain the following simplified result.

Lemma 4.2. *Let $\boldsymbol{x_k}$ denote the estimation to $\boldsymbol{Ax} = \boldsymbol{b}$ in the kth iteration of the two-subspace Kaczmarz method. Then,*

$$\mathbb{E}\|\boldsymbol{x} - \boldsymbol{x_k}\|_2^2 \leq \left(\left(1 - \frac{1}{R}\right)^2 - \frac{D}{R} - \frac{E}{Q}\right)\|\boldsymbol{x} - \boldsymbol{x_{k-1}}\|_2^2,$$

where $D = \min\left\{\frac{\delta^2(1-\delta)}{1+\delta}, \frac{\Delta^2(1-\Delta)}{1+\Delta}\right\}$, $E = 4\delta^3$ and $R = \|\boldsymbol{A}^{-1}\|^2\|\boldsymbol{A}\|_F^2$ and $Q = \|\boldsymbol{\Omega}^{-1}\|^2\|\boldsymbol{\Omega}\|_F^2$ denote the scaled condition numbers of $\boldsymbol{A}$ and $\boldsymbol{\Omega}$ (from (4.1), respectively.

Proof. In light of Lemma 4.1 and the proof of Lemma 2.2, it suffices to show that

$$\frac{1}{m^2-m}\sum_{j,i=1}^m E_{i,j}\langle\boldsymbol{x}-\boldsymbol{x}_{k-1},\boldsymbol{\omega}_{\boldsymbol{m(j-1)+i}}\rangle^2 \geq \frac{E}{Q}\|\boldsymbol{x}-\boldsymbol{x}_{k-1}\|_2^2.$$

By the definition (1.3) of δ and $\|\Omega^{-1}\|$, we have

$$\frac{1}{m^2-m}\sum_{i\neq j}^m E_{i,j}\langle\boldsymbol{x}-\boldsymbol{x}_{k-1},\boldsymbol{\omega}_{\boldsymbol{m(j-1)+i}}\rangle^2 \geq \frac{4\delta^3}{m^2-m}\sum_{i\neq j}^m\langle\boldsymbol{x}-\boldsymbol{x}_{k-1},\boldsymbol{\omega}_{\boldsymbol{m(j-1)+i}}\rangle^2$$

$$\geq \frac{4\delta^3}{m^2-m}\cdot\frac{\|\boldsymbol{x}-\boldsymbol{x}_{k-1}\|_2^2}{\|\Omega^{-1}\|}$$

$$\geq \frac{E}{Q}.$$

The last equality follows since the rows of Ω are unit norm and equal to zero for $i = j$.

$\square$

Applying Lemma 4.2 recursively yields our main theorem.

Theorem 4.3. *Let $\boldsymbol{x_k}$ denote the estimation to $\boldsymbol{Ax} = \boldsymbol{b}$ in the kth iteration of the two-subspace Kaczmarz method. Then,*

$$\mathbb{E}\|\boldsymbol{x}-\boldsymbol{x_k}\|_2^2 \leq \left(\left(1-\frac{1}{R}\right)^2 - \frac{D}{R} - \frac{E}{Q}\right)^k \|\boldsymbol{x}-\boldsymbol{x_0}\|_2^2,$$

where $D = \min\left\{\frac{\delta^2(1-\delta)}{1+\delta}, \frac{\Delta^2(1-\Delta)}{1+\Delta}\right\}$, $E = 4\delta^3$ and $R = \|\boldsymbol{A}^{-1}\|^2\|\boldsymbol{A}\|_F^2$ and $Q = \|\boldsymbol{\Omega}^{-1}\|^2\|\boldsymbol{\Omega}\|_F^2$ denote the scaled condition numbers of $\boldsymbol{A}$ and $\boldsymbol{\Omega}$ (from (4.1), respectively.

5. NUMERICAL RESULTS

Next we perform several experiments to compare the convergence rate of the two-subspace randomized Kaczmarz with that of the standard randomized Kaczmarz method. As discussed, both methods exhibit exponential convergence in expectation, but in many regimes the constant factor in the exponential bound of the two-subspace method is much smaller, yielding much faster convergence.

To test these methods, we construct various types of 500×50 matrices $\boldsymbol{A}$. To get a range of δ and Δ, we set the entries of $\boldsymbol{A}$ to be independent indentically distributed uniform random variables on some interval $[c, 1]$. Changing the value of c will appropriately change the values of δ and Δ. Note that there is nothing special about this interval, other intervals (both negative and positive or both) of varying widths yield the same results. For each matrix construction, both the randomized Kaczmarz and two-subspace randomized methods are run with the same initial (randomly selected) estimate. The estimation errors are computed at each iteration. Since each iteration of the two-subspace method utilizes two rows of the matrix $\boldsymbol{A}$, we call a single iteration of the standard method two iterations in the Algorithm 1.1 for fair comparison.

Figure 3 demonstrates the regime where the two-subspace method offers the most improvement over the standard method. Here the matrix A has highly coherent rows, with $\delta \approx \Delta \approx 1$.
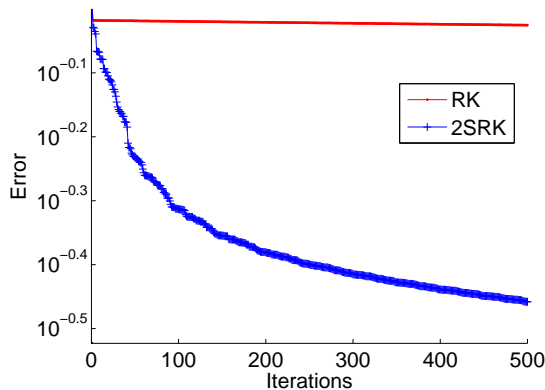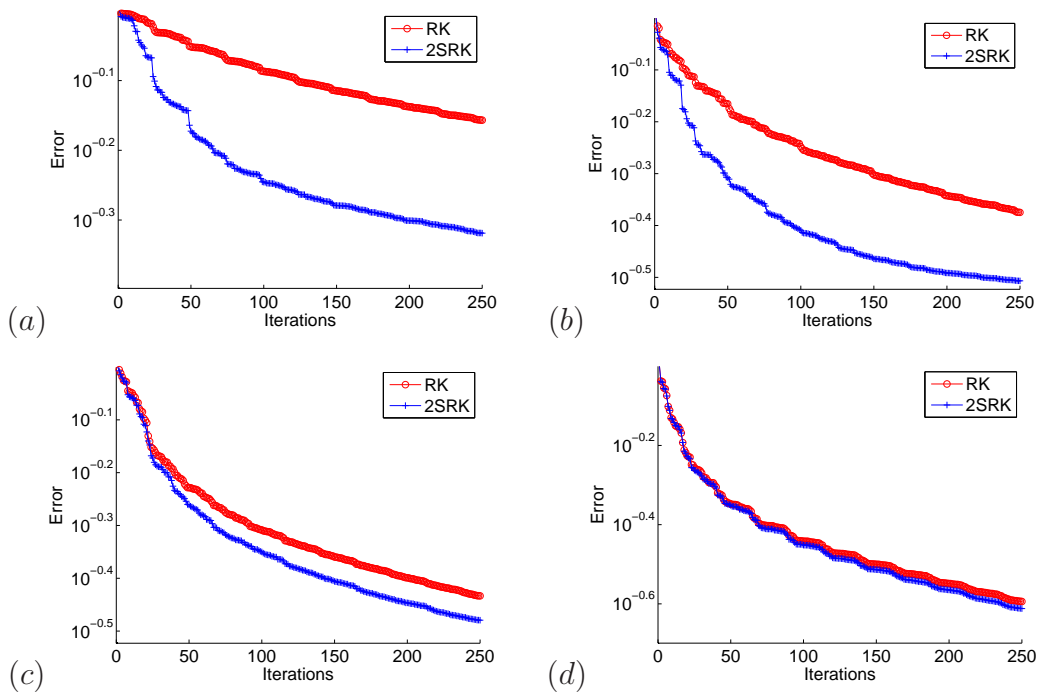

FIGURE 3. A log-linear plot of the error per iteration for the randomized Kaczmarz (RK) and two-subspace RK (2SRK). Matrix $\boldsymbol{A}$ has uniformly distributed highly coherent rows with $\delta = 0.992$ and $\Delta = 0.998$.

Our result Theorem 1.1 suggests that as δ becomes smaller the two-subspace method should offer less and less improvements over the standard method. When $\delta = 0$ the convergence rate bound of Theorem 1.1 is precisely the same as that of the standard method (1.1). Indeed, we see this precise behavior as is depicted in Figure 4.

Next we performed experiments on noisy systems. We used the same dimensions and construction of the matrix $\boldsymbol{A}$ as well as the signal type. Then we added i.i.d. Gaussian noise with norm 0.1 to the measurements $\boldsymbol{b}$. Figure 5 demonstrates the exponential convergence of the methods in the presence of noise for various values of δ and Δ.

FIGURE 4. A log-linear plot of the error per iteration for the randomized Kaczmarz (RK) and two-subspace RK (2SRK). Matrix $\boldsymbol{A}$ has uniformly distributed highly coherent rows with (a) $\delta = 0.837$ and $\Delta = 0.967$, (b) $\delta = 0.534$ and $\Delta = 0.904$, (c) $\delta = 0.018$ and $\Delta = 0.819$, and (d) $\delta = 0$ and $\Delta = 0.610$.

Plots (a), (b), and (c) demonstrate exponential convergence to the error threshold (or below) with improvements over the standard method. Plot (d) shows the semi-convergence effect of the two-subspace Kaczmarz method. It is an open problem to determine at which point to terminate the method for optimal error without knowledge of the solution $\boldsymbol{x}$. One option is to simply to terminate after $\mathrm{O}(n^2)$ iterations, as this is the amount of iterations needed for convergence (see the discussion in [20]). This is of course not optimal, and as Figure 6 shows, using this halting criterion may cause the two-subspace method to perform worse than the standard Kaczmarz method. We leave overcoming this challenge for future work.

6. DISCUSSION

As is evident from Theorems 1.1 and 4.3, the two-subspace Kaczmarz method provides exponential convergence in expectation to the solution of $\boldsymbol{Ax} = \boldsymbol{b}$. The constant in the rate of convergence for the two-subspace Kaczmarz method is at most equal to that of the best known results for the randomized Kaczmarz method (1.1). When the matrix $\boldsymbol{A}$ has many correlated rows, the constant is significantly lower than that of the standard method, yielding substantially faster convergence. This
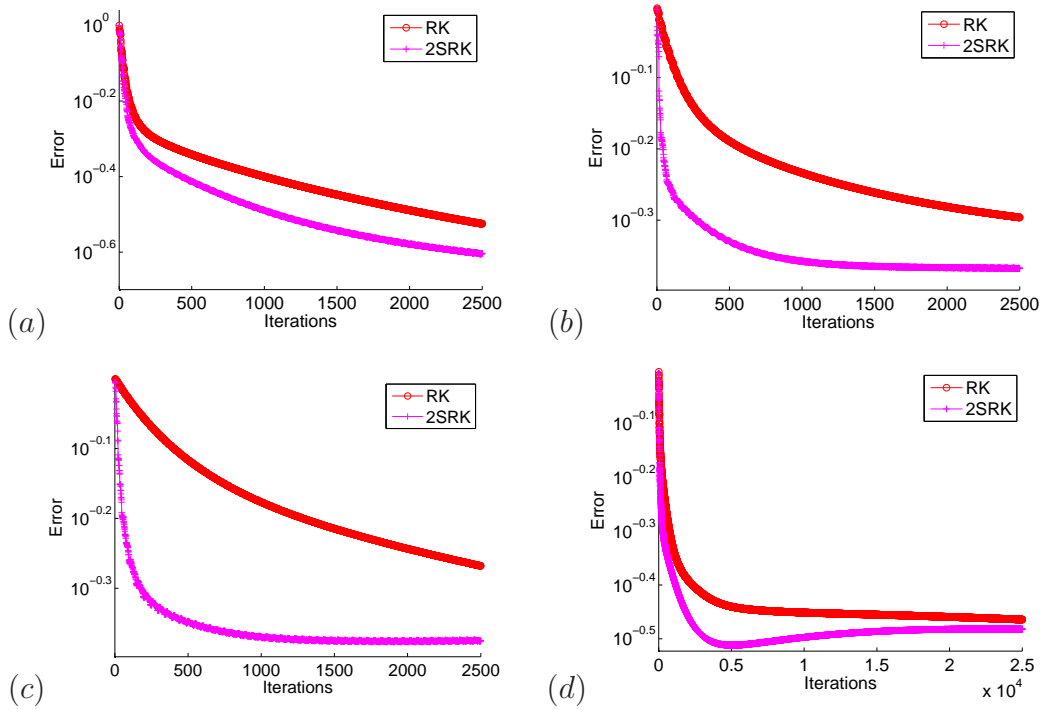
FIGURE 5. A log-linear plot of the error per iteration for the randomized Kaczmarz (RK) and two-subspace RK (2SRK). Matrix $\boldsymbol{A}$ has uniformly distributed highly coherent rows with (a) $\delta = 0.651$ and $\Delta = 0.932$, (b) $\delta = 0.933$ and $\Delta = 0.985$, (c) $\delta = 0.964$ and $\Delta = 0.992$, and (d) $\delta = 0.965$ and $\Delta = 0.991$.
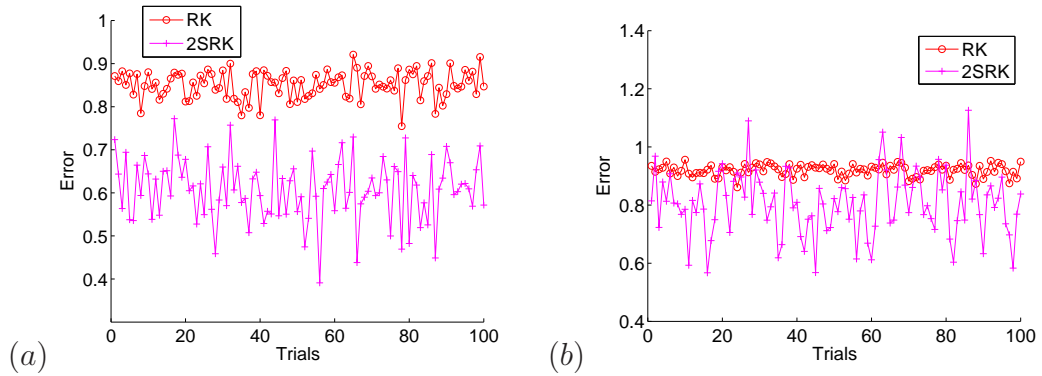

FIGURE 6. A plot of the error after 500 iterations for the randomized Kaczmarz (RK) and two-subspace RK (2SRK). Matrix $\boldsymbol{A}$ has uniformly distributed highly coherent rows with (a) $\delta = 0.981$ and $\Delta = 0.995$, and (b) $\delta = 0.993$ and $\Delta = 0.998$.

has positive implications for many applications such as nonuniform sampling in Fourier analysis, as discussed in Section 1.

We emphasize that the bounds presented in our main theorems are weaker than what we actually prove, and that even when δ is small, if the rows of $\boldsymbol{A}$ still have many correlations, Lemmas 2.1 and 4.1 still guarantee improved convergence. For example, if the matrix $\boldsymbol{A}$ has correlated rows but contains a pair of identical rows and a pair of orthogonal rows, it will of course be that $\delta = 0$ and $\Delta = 1$. However, we see from the proofs of our main theorems that the two-subspace method still guarantees substantial improvement over the standard method. Numerical experiments in cases like this produce results identical to those in Section 5.

It is clear both from the numerical experiments and Theorem 1.1 that the two-subspace Kaczmarz performs best when the correlations $\langle \boldsymbol{a_r}, \boldsymbol{a_s} \rangle$ are bounded away from zero. In particular, the larger δ is the faster the convergence of the two-subspace method. The dependence on Δ, however, is not as straightforward. Theorems 1.1 and 4.3 suggest that when Δ is very close to 1 the two-subspace method should provide similar convergence to the standard method. However, in the experiments of Section 5 we see this is not the case. This dependence on Δ appears to be only an artifact of the proof.

6.1. **Noisy systems.** As is the case for many iterative algorithms, the presence of noise introduces complications both theoretically and empirically. Theorem 3.1 guarantees expected exponential convergence to the noise threshold. For pessimistic values of D, the noise threshold provided by Theorem 3.1 is greater than that of the standard method, (1.2), by a factor of $\sqrt{R}$. In addition, large values of Δ produce large error thresholds in this bound. As in the noiseless case, we believe this dependence on Δ to be an artifact of the proof.

A further and important complication that noise introduces is a semi-convergence effect, a well-known effect in Algebraic Reconstruction Technique (ART) methods (see e.g. [5]). For example, in Figure 5 (d), the estimation error for the two-subspace method decreases to a point and then begins to increase. It remains an open problem to determine an optimal stopping condition without knowledge of the solution $\boldsymbol{x}$.

6.2. **Future Work.** The issue of detecting semiconvergence is a very deep problem. The simple solution would be to terminate the algorithm once the residual $\|\boldsymbol{Ax_k} - \boldsymbol{b}\|_2$ decreases below some threshold. However, the residual decreases in each iteration even when the estimation error begins to increase. Determining the residual threshold beyond which one should terminate is not an easy problem and work in this area continues to be done.

We also hope to improve the error threshold bound of Theorem 3.1 for the two-subspace method. We conjecture that the $(1 - \Delta)$ term can be removed or improved, and that the dependence on R can be reduced to $\sqrt{R}$ in the error term of Theorem 3.1.

Finally, a natural extension to our method would be to use more than two rows in each iteration. Indeed, extensions of the two-subspace algorithm to arbitrary subspaces can be analyzed [14].

REFERENCES

[1] C. Cenker, H.G. Feichtinger, M. Mayer, H. Steier, and T. Strohmer. New variants of the POCS method using affine subspaces of finite codimension, with applications to irregular sampling. In *Conf. SPIE 92 Boston*, pages 299–310, 1992.

[2] Y. Censor, P.P.B. Eggermont, and D. Gordon. Strong underrelaxation in Kaczmarz's method for inconsistent systems. *Numer. Math.*, 41(1):83–92, 1983.

[3] F. Deutsch and H. Hundal. The rate of convergence for the method of alternating projections. *J. Math. Anal. Appl.*, 205(2):381–405, 1997.

[4] P. Drineas, M.W. Mahoney, S. Muthukrishnan, and T. Sarlós. Faster least squares approximation. *Numerische Mathematik*, 117(2):217–249.

[5] T. Elfving, T. Nikazad, and P. C. Hansen. Semi-convergence and relaxation parameters for a class of SIRT algorithms. *Electron. T. Numer. Ana.*, 37:321–336, 2010.

[6] H. G. Feichtinger and T .Strohmer. A Kaczmarz-based approach to nonperiodic sampling on unions of rectangular lattices. In *Proc. Conf. SampTA-95*, pages 32–37, 1995.

[7] A. Galàntai. On the rate of convergence of the alternating projection method in finite dimensional spaces. *J. Math. Anal. Appl.*, 310(1):30–44, 2005.

[8] M. Hanke and W. Niethammer. On the acceleration of Kaczmarz's method for inconsistent linear systems. *Linear Alg. Appl.*, 130:83–98, 1990.

[9] G.T. Herman and L.B. Meyer. Algebraic reconstruction techniques can be made computationally efficient. *IEEE T. Med. Imaging*, 12(3):600–609, 1993.

[10] S. Kaczmarz. Angenäherte auflösung von systemen linearer gleichungen. *Bull. Internat. Acad. Polon.Sci. Lettres A*, pages 335–357, 1937.

[11] D. Leventhal and AS Lewis. Randomized methods for linear constraints: Convergence rates and conditioning. *Math. Oper. Res.*, 35(3):641–654, 2010.

[12] F. Natterer. *The Mathematics of Computerized Tomography*. Wiley, New York, 1986.

[13] D. Needell. Randomized Kaczmarz solver for noisy linear systems. *BIT Num. Math.*, 50(2):395–403, 2010.

[14] D. Needell and J. A. Tropp. Convergence of randomized Kaczmarz method. In preparation, 2012.

[15] D. Needell and R. Ward. Two-subspace projection method for coherent overdetermined systems. Submitted for publication, 2012.

[16] A. Powell and X. Chen. Almost sure convergence for the Kaczmarz algorithm with random measurements. Submitted, 2012.

[17] B. Recht and C. Re. Beneath the valley of the noncommutative arithmetic-geometric mean inequality: Conjectures, case studies, and consequences. Submitted for publication, 2012.

[18] K. M. Sezan and H. Stark. Applications of convex projection theory to image recovery in tomography and related areas. In H. Stark, editor, *Image Recovery: Theory and application*, pages 415–462. Acad. Press, 1987.

[19] T. Strohmer and R. Vershynin. A randomized solver for linear systems with exponential convergence. In *RANDOM 2006 (10th International Workshop on Randomization and Computation)*, number 4110 in Lecture Notes in Computer Science, pages 499–507. Springer, 2006.

[20] T. Strohmer and R. Vershynin. A randomized Kaczmarz algorithm with exponential convergence. *J. Fourier Anal. Appl.*, 15:262–278, 2009.