KACZMARZ AND RANDOMIZED KACZMARZ METHOD

by

Nurideen Abubakari

AN EXPOSITORY PAPER

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented June 13, 2017
Commencement June 2017

ACKNOWLEDGEMENTS

I would like to thank Professor David V. Finch for his immense guidance and support in this entire paper. He also helped me to get teaching position at Linn-Benton Community College. I would also like to thank all other committee members, Professor Adel Faridani, Professor Mary De Raeve Beisiegel for their insights, suggestions and their time for my final exam. I would also like to thank all the math department staff for their support and help. Thanks to all my colleagues and friends for their support and help. A special thanks to my family for their unwavering love and support.

TABLE OF CONTENTS

LIST OF FIGURES

1 INTRODUCTION

Stefan Kaczmarz, (see [6, 7]) in 1937 introduced an iterative algorithm for solving a system of linear algebraic equations of the form $Rf = g$ in Euclidean spaces. The method is now called the Kaczmarz method and since then has been rediscovered under different names. In 1970, the method was rediscovered by Richard Gorden, Robert Bender, and Gabor Herman under the name Algebraic Reconstruction Technique (ART) [4] and is used in the field of image reconstruction. It is also known by cyclic projection or successive projection and is closely related to an earlier result by Von Neumann which appeared in some lecture notes in 1933 but not published until 1950. The applications of this method have advanced since then from computer tomography to digital signal processing.

The Kaczmarz method iteratively solves system of linear algebraic equations of the form

$$R^i f = g^i \quad i = 1, ..., m \quad or \quad Rf = g \tag{1.1}$$

where $R \in \mathbb{R}^{m\times n}$ and $g \in \mathbb{R}^m$ in Euclidean spaces. The iteration is defined by

$$f_{j+1} = f_j + \frac{(g^i - R^i f_j)}{\|R^i\|_2^2}(R^i)^*, \tag{1.2}$$

where f_j is the $j-th$ iterate, $i = (\text{j mod m}) + 1$, R^i is the $i-th$ row of the matrix R, $(R^i)^*$ is the transpose of R^i, g^i is the $i-th$ component of the right-hand side vector, and $\|R^i\|_2$ denotes the Euclidean norm of the vector R^i. This notation will be used throughout this paper. Let $f_0 \in \mathbb{R}^n$ be the starting vector. Kaczmarz originally considered systems with square matrix and showed that for a nonsingular matrix R, the sequence (f_j) converges to the solution regardless of the initial approximation f_0. The algorithm (1.2) above operates by first taking an arbitrary initial approximation f_0 and at each iteration j, the current iterate is successively projected orthogonally onto the solution hyperplane $R^i f = g^i$ (see [14], p.248). To explain the computational steps involved in this method, we will use a very simple example of 2×2 matrix R to demonstrate. Let us consider a system of two equations and two unknowns: $x + 2y = 5$ and $x - 2y = 1$. Each of these equations represent a hyperplane and when there exist a unique solution to these equations, then the hyperplanes will intersect at that solution. Basic algebra approach gives us the solution $x = 3$ and $y = 1$ as the unique solution to the equations. We can represent these equations in matrix form as $Rf = g$ i.e.,

$$R = \begin{bmatrix} 1 & 2 \\ 1 & -2 \end{bmatrix}, f = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}, g = \begin{bmatrix} 5 \\ 1 \end{bmatrix}$$

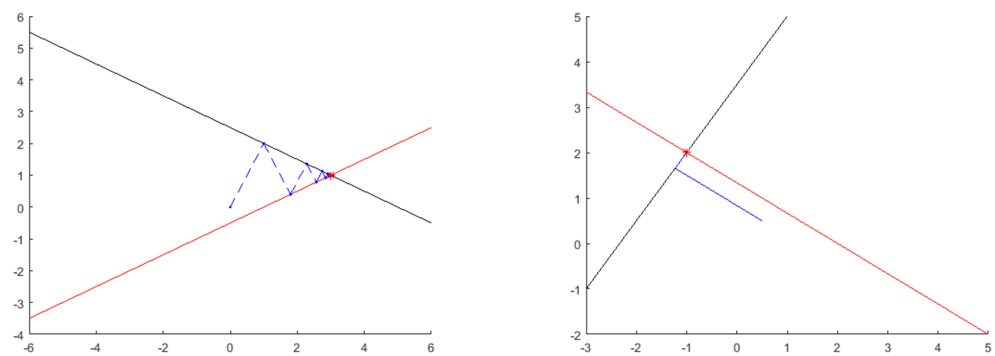
$R^1 = \begin{bmatrix} 1 & 2 \end{bmatrix}$, $g^1 = 5$. Starting with our initial approximation ("guess") $f_0 = \begin{bmatrix} 0 & 0 \end{bmatrix}^*$, projecting this initial guess onto the first hyperplane, reprojecting the resulting point onto the second hyperplane, and then projecting back onto the first equation, and so on. When there is a unique solution, the iterations will always converge to that solution ("point") as in (1.2). So, applying the algorithm in (1.2) to the system of equations above we get that

$$f_1 = f_0 + \frac{(g^1 - R^1 f_0)}{R^1 (R^1)^*} (R^1)^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix} + \frac{\left(\begin{bmatrix} 5 \end{bmatrix} - \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right)}{\begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix}} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

which satisfies the first equation $x + 2y = 5$. This shows that f_0 is projected onto the hyperplane represented by the first equation to yield $\begin{bmatrix} 1 \\ 2 \end{bmatrix} = f_1$. Next, we project f_1 onto the hyperplane represented by the second equation to get the iterate f_2. i.e.,

$$f_2 = f_1 + \frac{(g^2 - R^2 f_1)}{R^2 (R^2)^*}(R^2)^* = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \frac{\left(\begin{bmatrix} 1 \end{bmatrix} - \begin{bmatrix} 1 & -2 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right)}{\begin{bmatrix} 1 & -2 \end{bmatrix}\begin{bmatrix} 1 \\ -2 \end{bmatrix}} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} \frac{9}{5} \\ \frac{2}{5} \end{bmatrix}$$

which satisfies the second equation. This illustrates that the sequence of the iterates, in this case the lines intersect, so the iteration converges as $j \to \infty$ to the solution of the system of equations. That is, the process continues projecting from one hyperplane to the other until the approximation f_j converges to $f = \begin{bmatrix} 3 & 1 \end{bmatrix}^*$.On the other hand, when the two lines are orthogonal, the sequence will converge in one iteration. We use the figure below to illustrate this example.



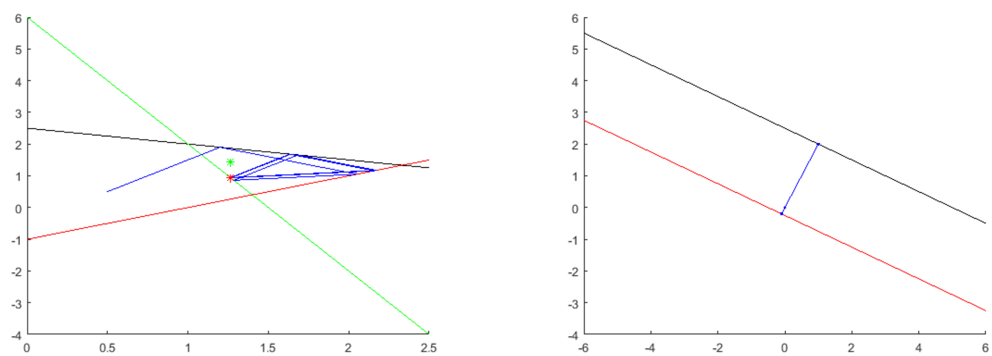
(a) converges in infinitely many steps (b) converges in one iteration

FIGURE 1.1: examples where the projections converge

This shows that the Kaczmarz algorithm converges to the solution of the system of equations if there is a unique solution. However, in general this is not always the case. For instance, the algorithm does not converge for two parallel equations (see figure 2 below), which also corresponds to an inconsistent system which has no solution. Now, if we introduce a third equation $4x + y = 6$, then the system of equations become overdetermined inconsistent. The projection of f_2 onto this third equation is:

$$f_3 = f_2 + \frac{(g^3 - R^3 f_2)}{R^3 (R^3)^*}(R^3)^* = \begin{bmatrix} \frac{9}{5} \\ \frac{2}{5} \end{bmatrix} + \frac{\left(\begin{bmatrix} 6 \end{bmatrix} - \begin{bmatrix} 4 & 1 \end{bmatrix} \begin{bmatrix} \frac{9}{5} \\ \frac{2}{5} \end{bmatrix} \right)}{\begin{bmatrix} 4 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix}} \begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{121}{85} \\ \frac{26}{85} \end{bmatrix}$$

the projections from one hyperplane to another between the three equations is trapped in a triangle pattern but do not converge as $j \to \infty$. We use the figures below to illustrate the cases where the Kaczmarz algorithm does not converge.



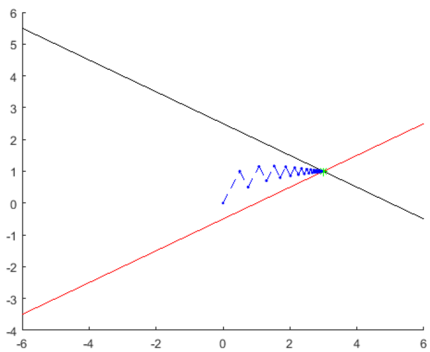
(a) overdetermined systems (b) two parallel lines

FIGURE 1.2: examples where the projections do not converge

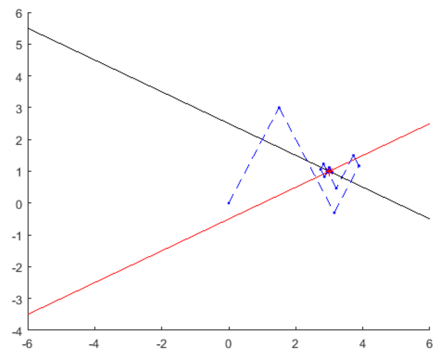
In this case we turn to other modified versions of the method to improve convergence. For example, to reduce the effect of the "chaotic" behavior and to accelerate the convergence, some kind of smoothing, also known as relaxation parameter, is frequently introduced into the Kaczmarz algorithm, extending it to

$$f_{j+1} = f_j + \omega \frac{(g^i - R^i f_j)}{\|R^i\|_2^2}(R^i)^* \tag{1.3}$$

where ω is a relaxation parameter that extends the projections either in front of the hyperplane ($\omega < 1$), exactly on the hyperplane ($\omega = 1$), or beyond the hyperplane ($\omega > 1$), and we assume henceforth that $0 < \omega < 2$ (see, [1] p. 217). Below are figures illustrating the result in for $\omega < 1$ and $\omega > 1$ conditions on ω in the consistent case.



(a) $\omega < 1$



(b) $\omega > 1$

FIGURE 1.3: conditions on ω in the consistent case

In this paper we will first look at Kaczmarz method [6] in the Hilbert space settings using the concepts of projections and relaxation parameter. We will then introduce Whitney and Meany's version of Kaczmarz method [16] and give an example. We will then look at the convergence properties of randomized extended Kaczmarz method in the finite dimensional setting based on the recent paper of Anna Ma, Deanna Needell, and Aaditya Ramdas (see [8]).

1.1 Organization of the paper

In section 2, we will look at Kaczmarz method in the Hilbert space, introduce Whitney and Meany's version of Kaczmarz method [16] and give an example. Section 3, will look at the convergence properties of randomized extended Kaczmarz method in the finite dimensional setting based on the paper of Anna Ma, Deanna Needell, and Aaditya Ramdas (see [8]).

2 KACZMARZ METHOD IN HILBERT SPACE

In the early 1930's Von Neumann [15] and later in 1962 Halperin [5] proved results for iterative projection methods in Hilbert space and then in 1977 McCormick [9] investigated the Kaczmarz method in Hilbert spaces. We will first consider the Kaczmarz method for projections following the presentations (in [5, 2, 12]) and then, we will also look at the relaxed form of the Kaczmarz method. In a later section, we refer to Whitney and Meany relaxed version of the Kaczmarz method ([16]) which makes use of the relaxation parameter.

Let P_j be the orthogonal projection operator onto a closed subspace M_j for $j = 1, ..., r$ of a Hilbert space H and let P_M be the projection operator onto the intersection $M = \cap_{j=1}^{r} M_j$ In this paper we will use the theorem in [12] and also our proof is along the same general lines as in [12], in that the steps of this proof are each well digested.

Theorem 2.1. *Let P_j $(j = 1, ..., r)$ be the orthogonal projection onto a closed subspace M_j of a Hilbert space H and let P_M be the orthogonal projection onto the intersection M $= \cap_{j=1}^{r} M_j$. If $Q = P_r ... P_1$ then $Q^k \to P_M$ strongly as $k \to \infty$. That is, for each $x \in H$*

$$\lim_{k \to \infty} Q^k x = P_M(x)$$

The following lemma is essential for the proof we give of Theorem 2.1.

Lemma 2.1. *For each x in the Hilbert space H, $\|Q^k x - Q^{k+1} x\| \to 0$ as $k \to \infty$*

Proof. Let x be in H. Since Q is a product of projections, we have $\|Q\| \leq 1$ which guarantees that the sequence $\left\{\|Q^k x\|\right\}$ is decreasing. It is bounded below and so converges. In particular,

$$\|Q^k x\| - \|Q^{k+1} x\| \to 0 \tag{2.1}$$

as $k \to \infty$. By the Pythagorean theorem, we have that for any orthogonal projection P and for any x in H

$$\|x - Px\|^2 = \|x\|^2 - \|Px\|^2. \tag{2.2}$$

In order to use the Pythagorean theorem, we take operators that are related by projection and defined as $Q_j = P_j Q_{j-1}$. Let $Q_0 = I$ and $Q_r = Q$. Then, the triangle inequality and (2.2) with $Q_j Q^k x$ in place of x give

$$\begin{aligned}
\|Q^k x - Q^{k+1}x\|^2 &= \|\sum_{j=0}^{r-1}(Q_j Q^k x - Q_{j+1}Q^k x)\|^2 \le [\sum_{j=0}^{r-1}\|Q_j Q^k x - Q_{j+1}Q^k x\|]^2 \\
&\le r\sum_{j=0}^{r-1}\|Q_j Q^k x - P_{j+1}Q_j Q^k x\|^2 = r\sum_{j=0}^{r-1}(\|Q_j Q^k x\|^2 - \|P_{j+1}Q_j Q^k x\|^2) \\
&= r(\sum_{j=0}^{r-1}\|Q_j Q^k x\|^2 - \sum_{n=1}^{r}\|Q_n Q^k x\|^2) \qquad \text{let } n = j+1 \\
&= r(\|Q_0 Q^k x\|^2 - \|Q_r Q^k x\|^2) = r(\|Q^k x\|^2 - \|Q^{k+1}x\|^2)
\end{aligned}$$

From (2.1) we have that $r(\|Q^k x\|^2 - \|Q^{k+1}x\|^2) \to 0$ since $\|Q^k x\| - \|Q^{k+1}x\| \to 0$. Hence, the claim of lemma 2.1 follows.

Now, for the proof of Theorem 2.1. We will use lemma 2.1 and continuity to show that $Q^k y \to 0$ strongly when y lies in $[\mathcal{R}(I-Q)]$ (the closure of the range of $(I-Q)$).

So, for $y \in \mathcal{R}(I-Q)$ lemma 2.1 shows the sequence (Q^k) of Q converges to zero on the range of $(I-Q)$. That is $\|Q^k y\| \to 0$ strongly for $y \in \mathcal{R}(I-Q)$. We will show that the same holds for $y \in [\mathcal{R}(I-Q)]$ using continuity. Take $y \in [\mathcal{R}(I-Q)]$. Given $\varepsilon > 0$, we need to find N such that for $k \ge N$, $\|Q^k y\| < \varepsilon$. Now, choose $y_1 \in \mathcal{R}(I-Q)$ so that $\|y - y_1\| < \frac{\varepsilon}{2}$ and take N such that for $k \ge N$, $\|Q^k y_1\| < \frac{\varepsilon}{2}$.

$$\begin{aligned}
\|Q^k y\| &= \|Q^k(y - y_1) + Q^k(y_1)\| \\
&\le \|Q^k(y - y_1)\| + \|Q^k y_1\|
\end{aligned}$$

$$\begin{aligned} &\le \|(y-y_1)\| + \|Q^k y_1\| \quad \text{since} \quad \|Q^k\| \le 1 \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \\ &= \varepsilon. \end{aligned}$$

□

To prove the convergence, we need the following lemma.

Lemma 2.2. *[10] Let Q be a linear map in H with $\|Q\| \le 1$, then*

$$\mathcal{N}(I-Q) = \mathcal{N}(I-Q^{\star}).$$

Proof. Suppose $Qf = f$ then $\|f\|^2 = \langle Qf, f\rangle = \langle f, Q^* f\rangle \le \|f\|^2$ with equality in Cauchy-Schwarz only when $Q^* f = f$ as $\|Q^* f\| \le \|f\|$.

Conversely, if $Q^* f = f$ then the preceding argument gives $Q^{**} f = f$ or $Qf = f$ Therefore, $\mathcal{N}(I-Q) = \mathcal{N}(I-Q^{\star})$. □

Lemma 2.3. $\mathcal{N}(I-Q) = M$. *The proof is trivial.*

Since we know that the Hilbert space H has the decomposition $H = [\mathcal{R}(I-Q)] \oplus \mathcal{N}(I-Q^*)$, then by lemma 2.2 we have $H = [\mathcal{R}(I-Q)] \oplus \mathcal{N}(I-Q)$. For any x in H we have the orthogonal decomposition $x = y + z$ with y in $[\mathcal{R}(I-Q)]$ and z in M.

Then multiplying both sides of the orthogonal decomposition of x by Q^k, we get $Q^k x = Q^k y + Q^k z$, and since $Q^k y \to 0$ and $Q^k z = z$ for all k, we have that $Q^k x = Q^k y + Q^k z = Q^k y + z \to z$ strongly as $k \to \infty$. Thus, $Q^k x \to z$ strongly as $k \to \infty$.

We state and prove a theorem parallel to theorem 2.1 by replacing P_j with $I - \omega I + \omega P_j$ for $0 < \omega < 2$. That is;

Theorem 2.2. *Let $T_j = (1-\omega)I + \omega P_j$ $(j = 1, ..., r)$ with P_j an orthogonal projection onto the closed subspace M_j of a Hilbert space H and let P_M be the orthogonal projection* *onto the intersection* $M = \cap_{j=1}^{r} M_j$. *If* $T = T_r \ldots T_1$ *then* $T^k \to P_M$ *strongly as* $k \to \infty$. *That is, for each* $x \in H$

$$\lim_{k\to\infty} T^k x = P_M(x)$$

Proof. For each x in H, we will show that $\|T^k x - T^{k+1}x\| \to 0$ as $k \to \infty$. Since T is a contraction, we have $\|T\| \leq 1$ which guarantees that the sequence $\{\|T^k x\|\}$ is decreasing. It is bounded below and so converges. In particular,

$$\|T^k x\| - \|T^{k+1}x\| \to 0 \tag{2.3}$$

as $k \to \infty$. In the previous theorem, we used the Pythagorean theorem to state the fact that for any orthogonal projection P and for any x in H

$$\|x - Px\|^2 = \|x\|^2 - \|Px\|^2. \tag{2.4}$$

But here, we have that

$$\|x - Tx\|^2 = \|(\omega I - \omega P)x\|^2 \quad \text{since } x - (I - \omega I + \omega p)x = (\omega I - \omega p)x \tag{2.5}$$

$$= \omega^2 \|(I - P)x\|^2 \tag{2.6}$$

$$= \omega^2 (\|x\|^2 - \|Px\|^2). \tag{2.7}$$

Next, we have that

$$\begin{aligned}
\|Tx\|^2 &= \langle (1-\omega)x + \omega Px, (1-\omega)x + \omega Px \rangle \\
&= (1-\omega)^2\|x\|^2 + 2\langle (1-\omega)x, \omega Px \rangle + \omega^2\|Px\|^2 \\
&= (1-\omega)^2\|x\|^2 + 2(1-\omega)\omega \langle Px, Px \rangle + \omega^2\|Px\|^2 \\
&= (1-\omega)^2\|x\|^2 + 2(1-\omega)\omega\|Px\|^2 + \omega^2\|Px\|^2
\end{aligned}$$

$$= (1-\omega)^2\|x\|^2 + (2\omega-\omega^2)\|Px\|^2$$

Now, subtracting $\|Tx\|^2$ from $\|x\|^2$ we get

$$\begin{aligned}
\|x\|^2 - \|Tx\|^2 &= (1-(1-\omega)^2)\|x\|^2 - (2\omega-\omega^2)\|Px\|^2 \\
&= (2\omega-\omega^2)\|x\|^2 - (2\omega-\omega^2)\|Px\|^2 \\
&= (2\omega-\omega^2)(\|x\|^2 - \|Px\|^2) \\
&= (2\omega-\omega^2)\frac{1}{\omega^2}\|x-Tx\|^2 \quad \text{from (2.5)} \\
&= \frac{\omega(2-\omega)}{\omega^2}\|x-Tx\|^2 \\
&= \frac{2-\omega}{\omega}\|x-Tx\|^2
\end{aligned}$$

Hence,

$$\begin{aligned}
\|x-Tx\|^2 &= \frac{\omega}{2-\omega}(\|x\|^2 - \|Tx\|^2) \\
&= K(\|x\|^2 - \|Tx\|^2) \quad \text{where } K = \frac{\omega}{2-\omega} \text{ and since } 0<\omega<2, \text{ K is positive.}
\end{aligned}$$

Thus,

$$\|x-Tx\|^2 = K(\|x\|^2 - \|Tx\|^2) \tag{2.8}$$

Let $Q_0 = I$ and for $j = 1, ..., r$ recursively define $Q_j = ((1-\omega)I + \omega P_j)Q_{j-1}$ i.e. $Q_j = T_j Q_{j-1}$, so that $Q_r = T$. The triangle inequality and (8) with $Q_j T^k x$ in place of x lead to

$$\begin{aligned}
\|T^k x - T^{k+1}x\|^2 &= \|\sum_{j=0}^{r-1}(Q_j T^k x - Q_{j+1}T^k x)\|^2 \leq [\sum_{j=0}^{r-1}\|Q_j T^k x - Q_{j+1}T^k x\|]^2 \\
&\leq r\sum_{j=0}^{r-1}\|Q_j T^k x - Q_{j+1}T^k x\|^2
\end{aligned}$$

$$= r\sum_{j=0}^{r-1}(\|T_jQ_{j-1}T^kx - T_{j+1}Q_jT^kx\|^2) \qquad \text{since } Q_{j+1} = T_{j+1}Q_j$$

$$= r\left(\frac{\omega}{2-\omega}\right)\left(\sum_{j=1}^{r-1}\|Q_{j-1}T^kx\|^2 - \sum_{n=1}^{r}\|Q_nT^kx\|^2\right) \qquad \text{from above}$$

$$= rK(\|Q_0T^kx\|^2 - \|Q_rT^kx\|^2)$$

$$= rK(\|T^kx\|^2 - \|T^{k+1}x\|^2) \qquad \text{since } Q_0 = I \text{ and } Q_r = T.$$

From (2.3) we have that $rK(\|T^kx\|^2 - \|T^{k+1}x\|^2) \to 0$ since $\|T^kx\| - \|T^{k+1}x\| \to 0$. Hence, the claim of lemma 2.1 follows. This shows the sequence (T^k) of T converges to zero on the range of $(I-T)$. So $T^ky \to 0$ strongly that is $\|T^ky\| \to 0$ for $y \in \mathcal{R}(I-T)$. Therefore, from the proof of theorem 2.1 we know this remains true for y in $[\mathcal{R}(I-T)]$ and orthogonal decomposition of any $x \in H$ according to $[\mathcal{R}(I-Q)] \oplus \mathcal{N}(I-Q)$ by lemma 2.2. We can conclude that $T^kx \to P_Mx$ strongly as $k \to \infty$ for any $x \in H$.

□

The theorem above only shows that the sequence of iterates $Q^k(x)$ and $T^k(x)$ converge to $P_M(x)$ for every x. We will now consider the relaxed version of Kaczmarz method which was first presented by Whitney and Meany who called it the single pattern LMS (least mean squares) algorithm [16]. Here and throughout this paper, S will denote the pseudoinverse of R. The pseudoinverse (S) satisfies the following properties.

1. $RSR = R$ $\quad$ S is an $n \times m$ matrix
2. $SRS = S$
3. $(RS)^* = RS$
4. $(SR)^* = SR$

We will let $R^{\{i\}}$ where $i = 1, 2, 3, 4$ denote the conditions of the pseudoinverse S. We will make some conclusions from the conditions of the pseudoinverse;

$(RS)^2 = RSRS = R(SRS) = RS$. So RS is a projection and since $(RS)^* = RS$, it is an orthogonal projection. Similarly, SR is an orthogonal projection. So $\mathcal{R}(RS) \subseteq \mathcal{R}(R)$. To see that $\mathcal{R}(RS) = \mathcal{R}(R)$, we note from $RSR = R$ that $\mathcal{R}(R) = \mathcal{R}(RSR) \subseteq \mathcal{R}(RS)$. We claim that $\mathcal{R}(SR) = \mathcal{R}(R^*)$. $SR = (SR)^* = R^*(S)^*$, so $\mathcal{R}(SR) = \mathcal{R}(R^*(S)^*) \subseteq \mathcal{R}(R^*)$. But from condition (1) above we have that $R^* = R^*(S)^*R^*$. So $\mathcal{R}(R^*) = \mathcal{R}(R^*(S)^*R^*) \subseteq \mathcal{R}(R^*(S)^*)$.

Here we will look at the method applied to the linear algebraic equation $Rf = g$ where $R \in \mathbb{R}^{m\times n}$ and $g \in \mathbb{R}^m$. Whitney and Meany relaxation algorithm has the form (see [3], p.140-141)

$$
\begin{aligned}
f_{j+1} &= f_j - \omega e_i^T(g^i - R^i f_j)R^* e_i \\
&= (I - \omega R^* e_i e_i^* R) f_j + \omega g^i R^* e_i
\end{aligned}
$$

where I is the $n \times n$ identity matrix, ω is a scalar, $1 \le i \le n$ and $i \equiv (\text{j mod m}) + 1$. We will use the diagonal matrix $D = \text{diag } (1/\|R^*e_i\|_2^2)$ to normalize the linear algebraic equation $Rf = g$ to obtain the normalized problem $D^{1/2}Rf = D^{1/2}g$. Then the relaxation algorithm above takes the form

$$
\begin{aligned}
f_{j+1} &= f_j - \omega \frac{e_i^*(g^i - R^i f_j)}{\|R^*e_i\|_2^2} R^* e_i && (2.9)\\
&= (I - \omega P_i) f_j + \omega \frac{g^i}{\alpha_i} R^* e_i && (2.10)\\
&= [(1-\omega)I + (\omega I - \omega P_i)] f_j + \omega \frac{g^i}{\alpha_i} R^* e_i && (2.11)\\
&= [(1-\omega)I + \omega(I - P_i)] f_j + \omega \frac{g^i}{\alpha_i} R^* e_i && (2.12)\\
&= [(1-\omega)I + \omega P_i^{\perp}] f_j + \omega \frac{g^i}{\alpha_i} R^* e_i && (2.13)
\end{aligned}
$$

where

$$
P_i = \frac{R^* e_i e_i^* R}{\|R^*e_i\|_2^2}, \quad \alpha_i = \|R^* e_i\|_2^2 \tag{2.14}
$$

This algorithm (2.10) is clearly a relaxed version of the Kaczmarz method applied to the linear algebraic system $Rf = g$. Let $Q_{k,j}(\omega) = (I - \omega P_k) \cdots (I - \omega P_j) \quad (k \geq j)$ and $Q_{k,j}(\omega) = I \quad (k < j)$. So $Q_{k,j}(\omega)$ is a product of contractions and hence a contraction. Applying $I - \omega P_1, \cdots, I - \omega P_m$ in succession, (2.10) gives

$$f_{km} = Q_{m,1}(\omega) f_{(k-1)m} + \omega \sum_{i=1}^{m} Q_{m,i+1}(\omega) \frac{g^i}{\alpha_i} R^* e_i \quad (k = 1, 2, \cdots). \tag{2.15}$$

Define the $n \times m$ matrix $A(\omega)$ whose $i-th$ column vector is $Ae_i = \omega Q_{m,i+1}(\omega) \frac{R^* e_i}{\alpha_i}$

$$A(\omega) = \omega \left[Q_{m,2}(\omega) \frac{R^* e_1}{\alpha_1}, \cdots, Q_{m,m+1}(\omega) \frac{R^* e_m}{\alpha_m} \right]. \tag{2.16}$$

In terms of $A(\omega)$ (2.15) becomes

$$f_{km} = Q_{m,1}(\omega) f_{(k-1)m} + A(\omega) g. \tag{2.17}$$

We will need later that the right hand side of (2.16) is a linear combination of the columns of R^*.

Lemma 2.4. *$\mathcal{R}(R^*)$ is invariant under $Q_{m,j}$ for all j.*

Proof. To show that $Q_{m,j} R^* z$ belongs to $\mathcal{R}(R^*)$, it suffices to show $Q_{m,j} R^* z$ lies in $\mathcal{N}(R)^\perp$. Take $v \in \mathcal{N}(R)$ then $P_k v = 0$ for all k, so $(I - \omega P_k) v = v$ and thus

$$\begin{aligned}
\langle Q_{m,j} R^* z, v \rangle &= \left\langle R^* z, Q_{m,j}^* v \right\rangle \\
&= \langle R^* z, (I - \omega P_j) \cdots (I - \omega P_m) v \rangle \\
&= \langle R^* z, v \rangle \\
&= 0 \quad \text{since } R^* z \in \mathcal{R}(R^*) \text{ and } v \text{ is orthogonal to } \mathcal{R}(R^*).
\end{aligned}$$

□

Next, we evaluate $A(\omega)R$. Let $Ae_i = Q_{m,i+1}(\omega)\frac{R^*e_i}{\alpha_i}$. Multiply both sides by e_i^*R, we get $Ae_ie_i^*R = \omega Q_{m,i+1}(\omega)\frac{R^*e_ie_i^*R}{\alpha_i}$.

So

$$\begin{aligned} A(\omega)R &= \sum_{i=1}^{m} Ae_ie_i^*R = \sum_{i=1}^{m} Q_{m,i+1}(\omega)\omega P_i \quad \text{since } P_i = \frac{R^*e_ie_i^*R}{\alpha_i} \\ &= \sum_{i=1}^{m}(Q_{m,i+1}(\omega) - Q_{m,i}(\omega)) \quad \text{since } Q_{m,i} = Q_{m,i+1}(I - \omega P_i) \\ &= Q_{m,m+1}(\omega) - Q_{m,1}(\omega) = I - Q_{m,1}(\omega) \end{aligned}$$

We will rewrite (2.15) by considering the recurrence relation

$$W_k = Q_{m,1}(\omega)W_{k-1} + A(\omega), \tag{2.18}$$

where W_0 is an arbitrary $n \times m$ matrix. If (2.18) is multiplied by g, we obtain (2.15) with $f_{km} = W_k g \quad (k \geq 0)$.

Theorem 2.3. *Let R be a $m \times n$ real matrix and let the matrices $W_k(k \geq 1)$ be defined by (2.18), where $Q_{m,1}(\omega)$ and $A(\omega)$ are defined as above, W_0 is an arbitrary matrix, and $0 < \omega < 2$.*

Then $\sum_{i=1}^{\infty} Q_{m,1}^i A(\omega)$ converges and with $W(\omega)$ denoting the sum

(i) $W_k \to (I - SR)W_0 + W(\omega)$, S is the pseudoinverse of R

(ii) $W(\omega) \in R^{\{1,2,4\}}$, but $W(\omega) \notin R^{\{3\}}$

in general $R^{\{i\}}$ is the conditions of the pseudoinverse (S)

(iii) If $rank(R) = n$, then $W_k \to (R^*DR)^{-1}R^*D$ when $\omega \to 0$

Proof. First, we show $\|Q_{m,1}|_{\mathcal{R}(R^*)}\| < 1$. $\|(I - \omega P_i)y\|_2 < \|y\|_2$ unless $(I - \omega P_i)y = y$. Thus $\|Q_{m,1}(y)\|_2 < \|y\|_2$ unless $(I - \omega P_i)y = y$ for all i, or $y \in \mathcal{N}(R)$. Since $\mathcal{R}(R^*)$ is finite dimensional and by lemma 2.4 is invariant under $Q_{m,1}$, there exists z in the unit sphere in $\mathcal{R}(R^*)$ such that

$$\|Q_{m,1}|_{\mathcal{R}(R^*)}\| = \|Q_{m,1}z\|_2.$$

Then, by the preceding $\|Q_{m,1}(z)\|_2 < \|z\|_2 = 1$. Let $\alpha = \|Q_{m,1}|_{\mathcal{R}(R^*)}\|$. Since $\|Q_{m,1}|_{\mathcal{R}(R^*)}\| = \alpha < 1$, if $z \in \mathcal{R}(R^*)$, then by induction

$$\|Q^k_{m,1}(\omega)z\|_2 \leq \alpha^k \|z\|_2 \quad (k \geq 1). \tag{2.19}$$

Now the recurrence relations (2.18) leads to the formula

$$W_k = Q^k_{m,1}(\omega)W_0 + \left(\sum_{i=0}^{k-1} Q^i_{m,1}(\omega)\right) A(\omega). \tag{2.20}$$

Since $I - \omega P_i = (1-\omega)I + \omega P_i^{\perp}$, by theorem 2.3 $Q^k_{m,1}(\omega)$ converges to orthogonal projection on

$$\cap\mathcal{R}(P_i^{\perp}) = \cap\mathcal{R}(I - P_i) = \cap\mathcal{N}(P_i) = \mathcal{N}(R).$$

Thus $Q^k_{m,1} \to P_{\mathcal{N}(R)} = (I - SR)$. Since $\|Q_{m,1}|_{\mathcal{R}(R^*)}\| = \alpha < 1$ the series $\sum_{i=0}^{\infty} Q^i_{m,1}(\omega)A(\omega)$ converges. Setting

$$W(\omega) = \sum_{i=0}^{\infty} Q^i_{m,1}(\omega)A(\omega).$$

Then

$$\begin{aligned} W(\omega)R &= \lim_{k\to\infty}\left(\sum_{i=0}^{k-1} Q^i_{m,1}(\omega)\right)A(\omega)R \\ &= \lim_{k\to\infty}\left(\sum_{i=0}^{k-1} Q^i_{m,1}(\omega)\right)(I - Q_{m,1}(\omega)) \\ &= \lim_{k\to\infty}(I - Q^k_{m,1}(\omega)) \\ &= I - P_{\mathcal{N}(R)} \\ &= I - (I - SR) = SR \quad \textit{since} \quad P_{\mathcal{N}(R)} = I - SR \end{aligned} \quad . \tag{2.21}$$

Multiplying (2.21) on the right by SR, one has

$$W(\omega)R = W(\omega)RSR = \lim_{k\to\infty}(SR - Q^k_{m,1}(\omega)SR) = SR. \tag{2.22}$$

Since the columns of $Q^k_{m,1}(\omega)A(\omega)$ belong to $\mathcal{R}(R^*)$ for $k = 1, 2, \cdots$, the columns of $W(\omega)$ belong to $\mathcal{R}(R^*)$. Hence from (2.19), $Q^k_{m,1}(\omega)W(\omega) \to 0$, when $k \to \infty$. Multiplying (2.21) on the right by $W(\omega)$, one has

$$W(\omega)RW(\omega) = \lim_{k\to\infty}(W(\omega) - Q^k_{m,1}(\omega)W(\omega)) = W(\omega).$$

Also, (2.21) shows that $W(\omega)R$ is symmetric and that $RW(\omega)R = RSR = R$. Thus $W(\omega)$ satisfies the first, second and fourth conditions of pseudoinverse. We may deduce from (2.20), (2.21), and (2.22) that

$$W_k \to (I - SR)W_0 + W(\omega) \tag{2.23}$$

when $k \to \infty$. If rank(R) = n, then $\mathcal{R}(R^*)$ has dimension n. The infinite series $\sum_{k=0}^{\infty} Q^k_{m,1}(\omega)$ therefore converges to $(I - Q_{m,1}(\omega))^{-1}$, and hence, $W(\omega) = (I - Q_{m,1}(\omega))^{-1}A(\omega)$. Now when $\omega \to 0$,

$$\frac{1}{\omega}(I - Q_{m,1}(\omega)) \to \sum_{i=0}^{m} P_i = R^*DR$$

and

$$\frac{1}{\omega}A(\omega) \to R^*D.$$

It follows that when $\omega \to 0$,

$$W(\omega) = \left[\frac{I - Q_{m,1}(\omega)}{\omega}\right]^{-1} \frac{A(\omega)}{\omega} \to (R^*DR)^{-1}R^*D.$$

We can easily check that $(R^*DR)^{-1}R^*D \in R^{\{1,2,4\}}$. $\square$

Example 2.1. *Let* $R = \begin{bmatrix} 1 & 0 \\ \cos\theta & \sin\theta \\ 0 & 1 \end{bmatrix}$, $R^* = \begin{bmatrix} 1 & \cos\theta & 0 \\ 0 & \sin\theta & 1 \end{bmatrix}$ *and set* $\omega = 1$. *Then*

$$Q_{m,i}(\omega) = (I - \omega P_m)\cdots(I - \omega P_i)$$
$$Q_{3,1}(\omega) = (I - \omega P_3)(I - \omega P_2)(I - \omega P_1)$$

so

$$Q_{3,1}(1) = (I - P_3)(I - P_2)(I - P_1) \qquad \textit{since } \omega = 1$$

P_1 *is projection on* $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$. $I - P_1$ *is projection on* $\begin{bmatrix} 1 \\ 0 \end{bmatrix}^{\perp} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, *so has matrix representation* $\begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$.

P_2 is projection on $\begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$. *$I - P_2$ is projection on* $\begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}^{\perp} = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$, *so has matrix representation* $\begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix} \begin{bmatrix} -\sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} \sin^2\theta & -\sin\theta\cos\theta \\ -\sin\theta\cos\theta & \cos^2\theta \end{bmatrix}$.

P_3 is projection on $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$. *$I - P_3$ is projection on* $\begin{bmatrix} 0 \\ 1 \end{bmatrix}^{\perp} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$, *so has matrix representation* $\begin{bmatrix} -1 \\ 0 \end{bmatrix} \begin{bmatrix} -1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$.

Then

$$
\begin{aligned}
Q_{3,1}(1) &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \sin^2\theta & -\sin\theta\cos\theta \\ -\sin\theta\cos\theta & \cos^2\theta \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \\
&= \begin{bmatrix} 0 & -\sin\theta\cos\theta \\ 0 & 0 \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
Q_{3,2}(1)R^*e_1 &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \sin^2\theta & -\sin\theta\cos\theta \\ -\sin\theta\cos\theta & \cos^2\theta \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \sin^2\theta \\ 0 \end{bmatrix} \\
Q_{3,3}(1)R^*e_2 &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} = \begin{bmatrix} \cos\theta \\ 0 \end{bmatrix} \\
Q_{3,4}(1)R^*e_3 &= I \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}
\end{aligned}
$$

but

$$
A(1) = [Q_{3,2}(1)R^*e_1, \quad Q_{3,3}(1)R^*e_2, \quad Q_{3,4}(1)R^*e_3]
$$

$$A(1) = \begin{bmatrix} \sin^2\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

So

$$\begin{aligned} W(1) &= (\sum_{i=0}^{\infty} Q_{3,1}^i(1))A(1) \\ &= \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & -\sin\theta\cos\theta \\ 0 & 0 \end{bmatrix} \right) A(1) \\ &= \begin{bmatrix} 1 & -\sin\theta\cos\theta \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sin^2\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \sin^2\theta & \cos\theta & -\sin\theta\cos\theta \\ 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

Hence

$$\begin{aligned} RW(1) &= \begin{bmatrix} 1 & 0 \\ \cos\theta & \sin\theta \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sin^2\theta & \cos\theta & -\sin\theta\cos\theta \\ 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \sin^2\theta & \cos\theta & -\sin\theta\cos\theta \\ \sin^2\theta\cos\theta & \cos^2\theta & -\sin\theta\cos^2\theta + \sin\theta \\ 0 & 0 & 1 \end{bmatrix}, \end{aligned}$$

which is not symmetric. Therefore, $W(\omega)$ need not satisfy the third condition of the pseudoinverse.

Next, we give the actual pseudoinverse (S) of the matrix R in this case. That is,

$$S = (R^*R)^{-1}R^*$$

But

$$R^*R = \begin{bmatrix} 1 & \cos\theta & 0 \\ 0 & \sin\theta & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \cos\theta & \sin\theta \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1+\cos^2\theta & \sin\theta\cos\theta \\ \sin\theta\cos\theta & 1+\sin^2\theta \end{bmatrix}$$

and let

$$\begin{aligned} B &= (1+\cos^2\theta)(1+\sin^2\theta) - (\sin\theta\cos\theta)(\sin\theta\cos\theta) \\ &= 1+\sin^2\theta+\cos^2\theta+\sin^2\theta\cos^2\theta-\sin^2\theta\cos^2\theta \\ &= 2 \quad \text{since } \sin^2\theta+\cos^2\theta = 1. \end{aligned}$$

Then

$$\begin{aligned} (R^*R)^{-1} &= \begin{bmatrix} 1+\cos^2\theta & \sin\theta\cos\theta \\ \sin\theta\cos\theta & 1+\sin^2\theta \end{bmatrix}^{-1} \\ &= \frac{1}{B}\begin{bmatrix} 1+\sin^2\theta & -\sin\theta\cos\theta \\ -\sin\theta\cos\theta & 1+\cos^2\theta \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} 1+\sin^2\theta & -\sin\theta\cos\theta \\ -\sin\theta\cos\theta & 1+\cos^2\theta \end{bmatrix} \end{aligned}$$

So

$$
\begin{aligned}
S &= (R^*R)^{-1}R^* \\
&= \frac{1}{2}\begin{bmatrix} 1+\sin^2\theta & -\sin\theta\cos\theta \\ -\sin\theta\cos\theta & 1+\cos^2\theta \end{bmatrix}\begin{bmatrix} 1 & \cos\theta & 0 \\ 0 & \sin\theta & 1 \end{bmatrix} \\
&= \frac{1}{2}\begin{bmatrix} 1+\sin^2\theta & \cos\theta + -\sin^2\theta\cos\theta + \sin^2\theta\cos\theta & -\sin\theta\cos\theta \\ -\sin\theta\cos\theta & -\sin\theta\cos^2\theta + \sin\theta + \sin\theta\cos^2\theta & 1+\cos^2\theta \end{bmatrix} \\
&= \frac{1}{2}\begin{bmatrix} 1+\sin^2\theta & \cos\theta & -\sin\theta\cos\theta \\ -\sin\theta\cos\theta & \sin\theta & 1+\cos^2\theta \end{bmatrix}.
\end{aligned}
$$

Thus, it follows that S is a left inverse of R:

$$
\begin{aligned}
SR = (R^*R)^{-1}R^*R = I \quad &\text{where is identity matrix} \\
&= \frac{1}{2}\begin{bmatrix} 1+\sin^2\theta & \cos\theta & -\sin\theta\cos\theta \\ -\sin\theta\cos\theta & \sin\theta & 1+\cos^2\theta \end{bmatrix}\begin{bmatrix} 1 & 0 \\ \cos\theta & \sin\theta \\ 0 & 1 \end{bmatrix} \\
&= \frac{1}{2}\begin{bmatrix} 1+\sin^2\theta+\cos^2\theta & \sin\theta\cos\theta - \sin\theta\cos\theta \\ -\sin\theta\cos\theta + \sin\theta\cos\theta & \sin^2\theta + 1 + \cos^2\theta \end{bmatrix} \\
&= \frac{1}{2}\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \quad \text{since } \sin^2\theta + \cos^2\theta = 1 \\
&= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I
\end{aligned}
$$

Theorem 2.4. *For any matrix $R \in \mathbb{R}^{m\times n}$ and any $g \in \mathbb{R}^m$. Let the vectors $f_k (k \geq 1)$ be* *defined by* (2.10) *with* $0 < \omega < 2$ *and* f_0 *arbitrary. Then*

$$\lim_{k \to \infty} f_k = f(\omega) = P_{\mathcal{N}(R)} f_0 + W(\omega) g$$

and the speed of convergence is linear. If $Rf = g$ *has a solution then* $f(\omega)$ *is a solution for any* f_0*, and* $W(\omega)g$ *is the minimum norm solution. If rank(R)* $= n$ *then* $f(\omega) \to P_{\mathcal{N}(R)} f_0 + (R^* D R)^{-1} R^* D g$

Proof. Suppose $g \neq 0$ and let an $n \times m$ matrix W_0 be determined so that $W_0 g = f_0$, and let W_k be defined as in the previous theorem. Then $f_{km} = W_k g \to (I - SR) W_0 g + W(\omega) g = (I - SR) f_0 + W(\omega) g = f(\omega)$ □

3 CONVERGENCE PROPERTIES OF THE RANDOMIZED EXTENDED KACZMARZ METHOD

The Kaczmarz method being one of the most popularly used iterative algorithm for solving linear algebraic systems of the form $Rf = g$, has undergone a lot of modifications and extensions in recent years to become faster and more efficient. One of the important modifications of the original Kaczmarz algorithm is the so-called randomized Kaczmarz algorithm which was introduced by Strohmer and Vershynin [13] who proved that it converges with expected exponential rate for consistent, overdetermined linear algebraic systems. An interesting follow up was immediately done by Deanna Needell to analyze the behavior of the randomized method for the case where the linear system is corrupted by noise [11].

3.1 Randomized Kaczmarz method

Here, we will describe the randomized Kaczmarz method proposed by Strohmer and Vershynin [13]. Consider the linear algebraic system $Rf = g$ as in (1.1), taking R, g as inputs and starting from an arbitrary initial approximation for f (for example, $f_0 = 0$), the Randomized Kaczmarz method repeats the following in each iteration. First, a row $i \in \{1, \cdots, m\}$ is chosen at random with probability proportional to its Euclidean norm, i.e.,

$$P_r(row = i) = \frac{\|R^i\|_2^2}{\|R\|_F^2},$$

where $\|R\|_F$ is the Frobenius norm of matrix R, i.e., $\|R\|_F^2 = \sum_{i=1}^{m} \|R^i\|^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} R_{ij}^2$. Then, project the current iterate onto that row, i.e.,

$$f_{j+1} = f_j + \frac{(g^i - R^i f_j)}{\|R^i\|_2^2}(R^i)^* \tag{3.1}$$

where R^* again denotes the (conjugate) transpose of R. One can easily see that $Rf = g$ can be updated iteratively as

$$R^i f_{j+1} = g^i \tag{3.2}$$

The main result of (3.1) states that f_j converges exponentially fast to the solution of (1.1), and the rate of convergence depends on the scaled condition number $\kappa(R) = \|R\|_F \|R^{-1}\|_2$

Theorem 3.1. *Let f^* be the solution of $Rf = g$ and assume that R has full column rank. Then (3.1) converges to f^* in expectation, with the average error*

$$\mathbb{E}\|f_j - f^*\|_2^2 \leq (1 - \kappa(R)^{-2})^j \cdot \|f_0 - f^*\|_2^2. \tag{3.3}$$

To prove this theorem, we will first show that

$$\mathbb{E}\|f_{j+1} - f^*\|^2 \leq (1 - \kappa(R)^{-2}) \cdot \|f_j - f^*\|^2, \text{ where } \mathbb{E} \text{ is expectation conditioned on } f_j.$$

Proof. Lets assume that $Rf = g$ has a solution, then by orthogonality $f_{j+1} - f^*$ is orthogonal to $f_{j+1} - f_j$. So, it holds that $\|f_{j+1} - f^*\|^2 = \|f_j - f^*\|^2 - \|f_{j+1} - f_j\|^2$. Taking expectation of both sides by we obtain

$$\begin{aligned}
\mathbb{E}\|f_{j+1} - f^*\|^2 &= \|f_j - f^*\|^2 - \mathbb{E}\|f_{j+1} - f_j\|^2 \\
&= \|f_j - f^*\|^2 - \mathbb{E}\left[\left\|\frac{g^i - R^i f_j}{\|R^i\|^2}(R^i)^*\right\|^2\right] && \text{definition of iteration} \\
&= \|f_j - f^*\|^2 - \mathbb{E}\left[\frac{(g^i - R^i f_j)^2}{\|R^i\|^4}\|(R^i)^*\|^2\right] && \text{take scalars outside norm} \\
&= \|f_j - f^*\|^2 - \mathbb{E}\left[\frac{(R^i f^* - R^i f_j)^2}{\|R^i\|^2}\right] && \text{cancellation and use of } R^i f^* = g^i \\
&= \|f_j - f^*\|^2 - \sum_{i=1}^{m} \frac{\|R^i\|^2}{\|R\|_F^2} \frac{(R^i(f_j - f^*))^2}{\|R^i\|^2} && \text{definition of expectation} \\
&= \|f_j - f^*\|^2 - \frac{1}{\|R\|_F^2} \sum_{i=1}^{m} (R^i(f_j - f^*))^2 && \|R\|_F^2 \text{ does not depend on } i
\end{aligned}$$

$$
\begin{aligned}
&= \|f_j - f^*\|^2 - \frac{1}{\|R\|_F^2}\|R(f_j - f^*)\|^2 && \|Rf\|^2 = \sum_{i=1}^{m}(R^i f)^2 \\
&= \|f_j - f^*\|^2\left(1 - \frac{\|R(f_j - f^*)\|^2}{\|R\|_F^2\|f_j - f^*\|^2}\right) && \text{common factor of } \|f_j - f^*\|^2 \\
&\leq \left(1 - \frac{\|R\|^2\|f_j - f^*\|^2}{\|R\|_F^2\|f_j - f^*\|^2}\right)\|f_j - f^*\|^2 \\
&= \left(1 - \frac{\|R\|^2}{\|R\|_F^2}\right)\|f_j - f^*\|^2 && \text{cancel common term } \|f_j - f^*\|^2 \\
&= \left(1 - \frac{1}{\|R\|_F^2\|R^{-1}\|^2}\right)\|f_j - f^*\|^2 \\
&= \left(1 - \frac{1}{\kappa(R)^2}\right)\|f_j - f^*\|^2 && \|R\|_F^2\|R^{-1}\|^2 = \kappa(R)^2 \\
&= \left(1 - \kappa(R)^{-2}\right)\|f_j - f^*\|^2
\end{aligned}
$$

Now, we will apply the law of total expectation and induction law on j to obtain (3.3), that is

$$
\begin{aligned}
\mathbb{E}\|f_j - f^*\|^2 &= \mathbb{E}_{j-1}\left[\mathbb{E}\|f_j - f^*\|^2\right] \\
&\leq \mathbb{E}_{j-1}\left(1 - \kappa(R)^{-2}\right)\|f_{j-1} - f^*\|^2 && \text{from (3.1) the base case j = 1 is proved} \\
&= \left(1 - \kappa(R)^{-2}\right)\mathbb{E}_{j-2}\left[\mathbb{E}_{j-1}\|f_{j-1} - f^*\|^2\right] \\
&= \left(1 - \kappa(R)^{-2}\right)\mathbb{E}_{j-2}\left(1 - \kappa(R)^{-2}\right)\|f_{j-2} - f^*\|^2 \\
&= \left(1 - \kappa(R)^{-2}\right)^2\mathbb{E}_{j-3}\left[\mathbb{E}_{j-2}\|f_{j-2} - f^*\|^2\right] && \text{take scalars outside the expectation} \\
&= \left(1 - \kappa(R)^{-2}\right)^2\mathbb{E}_{j-3}\left(1 - \kappa(R)^{-2}\right)\|f_{j-3} - f^*\|^2 \\
&= \left(1 - \kappa(R)^{-2}\right)^3\mathbb{E}_{j-4}\left[\mathbb{E}_{j-3}\|f_{j-3} - f^*\|^2\right]
\end{aligned}
$$

Now, let k be a natural number with $k \leq 3$ be given and suppose (3.3) is true for $k - 1$

and show it holds for k. Then

$$\begin{aligned}\mathbb{E}\|f_k - f^*\|^2 &= \left(1-\kappa(R)^{-2}\right)^{k-1}\mathbb{E}_{k-k}\left[\mathbb{E}_{k-k+1}\|f_{k-k+1} - f^*\|^2\right] \\ &= \left(1-\kappa(R)^{-2}\right)^{k-1}\mathbb{E}_0\left[\mathbb{E}_1\|f_1 - f^*\|^2\right] \\ &= \left(1-\kappa(R)^{-2}\right)^{k-1}\left[\mathbb{E}_1\|f_1 - f^*\|^2\right] \\ &= \left(1-\kappa(R)^{-2}\right)^{k-1}\left(1-\kappa(R)^{-2}\right)\|f_0 - f^*\|^2 \\ &= \left(1-\kappa(R)^{-2}\right)^{k}\|f_0 - f^*\|^2\end{aligned}$$

□

Thus, (3.3) holds for $j = k$ which completes the proof.

3.2 Randomized Extended Kaczmarz method

. The randomized Kaczmarz method in the case of inconsistent systems fails to converge to the least square solution as we will expect, since the method at each iteration projects completely onto a selected solution space without being able to reduce the error term. To overcome this, Zouzias and Freris [17] suggested a version of the randomized Kaczmarz method which iteratively reduces the compoonent of g orthogonal to the range of R by random projection. This method, named the randomized extended Kaczmarz method has the following iteration formulation, which can be initialized by setting $f_0 = 0$ and $z_0 = g$:

$$f_{j+1} := f_j + \frac{(g^i - z_j^i - R^i f_j)}{\|R^i\|_2^2}(R^i)^*, \quad z_{j+1} = z_j - \frac{\langle R_{(j)}, z_j\rangle}{\|R_{(j)}\|_2^2}R_{(j)}. \tag{3.4}$$

Here, a random column $j \in \{1, ..., n\}$ is chosen with a probability proportional to its Euclidean norm, i.e.,

$$P_r(column = j) = \frac{\|R_{(j)}\|_2^2}{\|R\|_F^2}, \tag{3.5}$$

where $R_{(j)}$ denotes the jth column of R. Here, z_j estimates the component of g which is orthogonal to the range of R, allowing for the iterates f_i to converge to the true least-squares solution of the system. Zouzias and Freris [17] proved that the randomized extended Kaczmarz method converges linearly in expectation to this solution f_{LS}.

BIBLIOGRAPHY

1. Joseph M Elble, Nikolaos V Sahinidis, and Panagiotis Vouzis. Gpu computing with Kaczmarz's and other iterative algorithms for linear systems. *Parallel computing*, 36(5):215–231, 2010.

2. A Galántai. On the rate of convergence of the alternating projection method in finite dimensional spaces. *Journal of mathematical analysis and applications*, 310(1):30–44, 2005.

3. Aurél Galántai. *Projectors and projection methods*, volume 6. Springer Science & Business Media, 2013.

4. Richard Gordon, Robert Bender, and Gabor T Herman. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *Journal of theoretical Biology*, 29(3):471IN1477–476IN2481, 1970.

5. Israel Halperin. The product of projection operators. *Acta Sci. Math.(Szeged)*, 23(1):96–99, 1962.

6. Stefan Kaczmarz. Angenähert auflösung von systemen linearer gleichungen. *International Bulletin of the Polish Academy of Sciences and Letters*, 35:355–357, 1937.

7. Stefan Kaczmarz. Approximate solution of systems of linear equations. *International Journal of Control*, 57(6):1269–1271, 1993.

8. Anna Ma, Deanna Needell, and Aaditya Ramdas. Convergence properties of the randomized extended Gauss–Seidel and Kaczmarz methods. *SIAM Journal on Matrix Analysis and Applications*, 36(4):1590–1604, 2015.

9. SF McCormick. The methods of Kaczmarz and row orthogonalization for solving linear equations and least squares problems in Hilbert space. *Indiana Univ. Math. J*, 26(6):1137–1150, 1977.

10. Frank Natterer. *The mathematics of computerized tomography.* SIAM, 2001.

11. Deanna Needell. Randomized Kaczmarz solver for noisy linear systems. *BIT Numerical Mathematics*, 50(2):395–403, 2010.

12. Anupan Netyanun and Donald C Solmon. Iterated products of projections in Hilbert space. *The American Mathematical Monthly*, 113(7):644–648, 2006.

13. Thomas Strohmer and Roman Vershynin. A randomized Kaczmarz algorithm with exponential convergence. *Journal of Fourier Analysis and Applications*, 15(2):262, 2009.

14. Roman Sznajder et al. Kaczmarz algorithm revisited. *Czasopismo Techniczne*, 2015(Nauki Podstawowe Zeszyt 2-NP (20) 2015):247–254, 2016.

15. John Von Neumann. On rings of operators. Reduction theory. *Annals of Mathematics*, pages 401–485, 1949.

16. TM Whitney and RK Meany. Two algorithms related to the method of steepest descent. *SIAM journal on Numerical Analysis*, 4(1):109–118, 1967.

17. Anastasios Zouzias and Nikolaos M Freris. Randomized extended Kaczmarz for solving least squares. *SIAM Journal on Matrix Analysis and Applications*, 34(2):773–793, 2013.