2022

Check Yourself Before You WREK Yourself: Unpacking and Generalizing Randomized Extended Kaczmarz

William Gilroy

Check Yourself Before You WREK Yourself: Unpacking and Generalizing Randomized Extended Kaczmarz

Will Gilroy

Jamie Haddock, Advisor

Heather Zinn-Brooks, Reader

HARVEY MUDD COLLEGE

Department of Mathematics

May, 2022

Abstract

Linear systems are fundamental in many areas of science and engineering. With the advent of computers there now exist extremely *large* linear systems that we are interested in. Such linear systems lend themselves to iterative methods. One such method is the family of algorithms called Randomized Kaczmarz methods.

Among this family, there exists a Randomized Kaczmarz variant called Randomized Extended Kaczmarz which solves for least squares solutions in inconsistent linear systems. Among Kaczmarz variants, Randomized Extended Kaczmarz is unique in that it modifies input system in a special way to solve for the least squares solution.

In this work we unpack the geometry underlying Randomized Extended Kaczmarz (REK) by uniting proofs by Zouzias and Freris (2013) and Du (2018), leading to more insight about *why* REK works. We also provide novel proofs showing: that REK will converge with an alternative sequence of $\mathbf{z}$ updates, and giving a closed form for REK's original $\mathbf{z}$ updates. Lastly we have done some work generalizing the ideas behind REK and QuantileRK (Haddock et al., 2020) to lay foundations for a new Randomized Kaczmarz variant called *Weighted Randomized Extended Kaczmarz* (WREK) which aim to solve weighted least squares problems with dynamic reweightings.

Contents

Abstract iii

Acknowledgments xi

Preface xiii

1 Background on Linear Systems 1
- 1.1 Matrix Jargon . . . 1
- 1.2 Consistent Linear Systems . . . 4
- 1.3 Inconsistent Systems . . . 7
- 1.4 "Solutions" to Inconsistent Systems . . . 10
- 1.5 Notation . . . 13

2 Randomized Kaczmarz 15
- 2.1 Convergence Properties . . . 17
- 2.2 REK and WREK . . . 21

3 Randomized Extended Kaczmarz 23
- 3.1 Definition and Convergence Properties . . . 23
- 3.2 What is Up With the **z** Vectors? . . . 26
- 3.3 Theoretical Foundations of REK . . . 28
- 3.4 REK Convergence for an Arbitrary Sequence of **b** Vectors . . . 39

4 Weighted Randomized Extended Kaczmarz 43
- 4.1 Extending Randomized Extended Kaczmarz . . . 43
- 4.2 Dynamic Reweightings . . . 45
- 4.3 Behaviour of the WLS solutions . . . 47
- 4.4 Proposed Definition for WREK . . . 52
- 4.5 Future Directions & Conclusion . . . 54

Bibliography 55

List of Figures

1.1 Visualization of the geometry of the given linear system. 6
1.2 Geometric depiction of a rank one linear system as given in example (1.2.1). 7
1.3 The geometry of an inconsistent linear system. There is no solution to this linear system as seen since there is no point in $\mathbb{R}^2$ where all three hyperplanes intersect simultaneously. 8
1.4 The geometry of a noisy linear system. The original linear system is represented by the dotted lines. The hyperplanes have been modified by a small amount of noise. 9
1.5 The geometry of a corrupted linear system. In this case nearly all the hyperplanes intersect at a single point but with a few highly perturbed hyperplanes. One solution we might want is the point where most of the hyperplanes intersect, in this case on the left. However, the least squares solution for such a system is far away from that point, instead it is somewhere between the corrupted hyperplanes and the point we would actually want. 10

2.1 Iterations from Randomized Kaczmarz. At each step $\mathbf{x}^{(k)}$ is orthogonally projected onto a randomly selected affine hyperplane. Each hyperplane represents a single equation from our linear system. 16
2.2 Recall that in corrupted linear systems the solution we would want is the one where the corrupted hyperplanes are ignored. Here, the desired solution is denoted by $\mathbf{x}^*$. However, Randomized Kaczmarz will never converge to $\mathbf{x}^*$. Instead, Randomized Kaczmarz may get close, but every time a corrupted hyperplane is projected onto the $\mathbf{x}$ iterates will move far away from $\mathbf{x}^*$. . . . 21

3.1 Steps of REK iterations demonstrating the system updates and the solution vector estimate updates $\mathbf{x}^{(k)}$. The hyperplanes of the original inconsistent system is represented with the black lines and the modified system is represented by the red dotted lines. 25

3.2 A geometric depiction of Du's Fundamental Triangle. The black lines indicate the inconsistent system $A\mathbf{x} = \mathbf{b}$, indicate the convergent system $A\mathbf{x} = \mathbf{b}_{col(A)}$. The magenta triangle depicts Du's Fundamental Triangle. Here both $\mathbf{x}^{(k)}$ and $\hat{\mathbf{x}}^{(k)}$ are orthogonal projection from $\mathbf{x}^{(k-1)}$. Note that $\mathbf{x}^{(k-1)}$ and $\mathbf{x}^{(k)}$ is not necessarily on either system's hyperplanes because REK's iterates are some modified system between the original system and the convergent system. . 34

4.1 Here we have an inconsistent linear system visualized by the black lines. Our single reweighting only affects the thick hyperplane. The magenta line indicates the line along which the pseudosolutions move. 50

List of Algorithms

1 Randomized Kaczmarz 16
2 Randomized Extended Kaczmarz 24
3 QuantileRK 46
4 Weighted Randomized Extended Kaczmarz 53

Acknowledgments

Thank you to Prof Haddock for guiding and mentoring me over these months. What a wonderful journey. I will always appreciate your support and encouragement.

Thank you also to the Large Linear Systems UCLA Applied Mathematics Group, we created some delightful stories together.

Thank you to the many Mudd Profs (Maths and otherwise!) who have given me important support over the years. Mathematics used to be one of my least favourite subjects until I met you all and rekindled a new relationship with the subject. Thank you for all the help, through every the stumble and celebration. Thank you Patty for helping me understand that it's not bad to see things a bit differently, and for introducing me to LW.

Thank you to all my friends and their families who invited me in during the years we could not be on campus and I could not go home, I really do not know what I would have done without you all. Thank you to all my friends, I'm so happy to have spent these years with you. If any of my Physics frosh are here, thank you for spending your evenings with us, meeting you all was such a joy. You're all going to do wonderful things.

Thank you for taking the time to read This Work.

Preface

To my dear reader.

In my experience, the preface is a space for me as an author to speak to you as a reader. Readers, like humans, are non-homogenous. You're bringing particular experiences, and intentions to the table. I will acknowledge you as best I can. That is, I will try my best to write for a human being. At this moment, I will acknowledge certain groups of people that I imagine might find their way to a text like this.

To the undergraduate reader. Welcome! Perhaps you are interested in seeing what Harvey Mudd Maths students get up to. Perhaps you are thinking of doing Maths Thesis yourself and you want to see what kind of work Prof Haddock gets up to. Maybe you're here for an entirely different reason. Whatever the reason, I'm glad you're here. Compared to other senior theses you might read I think the background to read this text will be relatively managable. I would say that you only need one or two undergraduate classes on linear algebra to follow this text in as much detail as you want. I've tried to make clear the ideas I will be leaning on, and so some parts of this document can provide as a reference point or a very quick review. The numerical analysis in this text can be a hive of symbols but nothing super arcane is being pulled out. If you prefer, there is a good amount of geometric intuition that I emphasize which you are free to lean on. I would recommend reading from the start, skipping around, or reading at whatever pace feels fun and engaging.

To the academic. Welcome! Thanks for taking an interest in my thesis. I imagine you are experienced in navigating Mathematical texts. Nevertheless, if you are interested in learning about Randomized Kaczmarz then I would recommend starting at Chapter two. If you are familiar with Randomized Kaczmarz then my main contributions can be found in chapters three and four. My main contributions are outlined in the Abstract of this document.

Lastly, in my experience, an important part a part of Mathematics is the doing. I believe "doing maths" is a multifaceted human experience. I think it should not be as difficult as it is for someone with the right curiousity and desire to "do mathematics". To this end, I want to emphasize all the non-rational parts of mathematics; all the time spend investigating, conjecturing, failing, proving, confusing, realizing, hoping, forgetting, enjoying. I will do my best to acknowledge these non-rational parts of "doing maths" in the following ways.

Usually texts like this are for me to tell you the adventures I've been on, to tell you what hill I've climbed and how exactly I climbed it. That sort of writing is mostly fun for me only. However, I've already had my fun climbing this hill and so, apart from articulating myself to clarify my ideas, that kind of writing is not useful at all. That said, I want to graduate and we exist in a certain subculture which expects a certain kind of writing, so I will meet you halfway. I will write this document about my adventures, but I will point out places which *might* be interesting to explore for yourself (of course, what you find fun depends on who you are, but I will do my best to guess). At various points in this text I will leave a small ☕ which indicates that I think the following idea might be fun to think about, or that there's a deep idea lurking here which is worth experiencing, or that this is an affluent place look.

I will also try to write in such a way that emphasizes where all the pieces fit into the bigger story of this hill that we're climbing together. Of course, there are parts which cannot be told, only experienced. I think Mathematics is its least fun if I am a tour guide and you are a tourist. So instead, let's go as friends exploring together.

One last thing. Life is bigger outside of Mathematics. For you, I, and everyone else. I want to take a moment to acknowledge wherever we are, in whatever sense is most meaningful.

Chapter 1

Background on Linear Systems

In this document we are motivated by linear systems and how to solve them numerically. From a very abstract level, a system is described by an operator which maps input functions to output functions. Furthermore, a system is linear if and only if it satisfies the property that the net response caused by two or more inputs is equivalent to the sum of the responses that each of these inputs cause individually. This informal definition outlines why linear systems appear in places like engineering — where our input and output functions might be special kinds of signals, mechanical systems, or other kinds of systems[1]. This definition might also show why linear systems arise in quantum mechanics where the state of a particle can be represented by linear combinations of states which remain constant under the time operator. In many of these applied systems we are interested in the quantities which are output by a system. In the case of a quantum system we may want to compare our output quantity to experimental measurements. As a result, there is much interest in solving linear systems, exactly and analytically where we can, and numerically where we cannot. Throughout this document we are interested in the cases where analytical methods fail, where we instead focus on numerical methods.

At its heart the ideas explored in this thesis are all to the end of solving linear systems. However, before we can focus our attention fully on linear systems it will be important to state explicitly some of the matrix concepts which are pertinent to our discussion of linear systems. In this chapter we present the foundational theory to grapple with the numerical work in the rest of this document. To that end, let $A \in \mathbb{R}^{m \times n}$.

1.1 Matrix Jargon

This Section in particular will be quite terse as we provide rapid summaries of the foundational linear algebra ideas which will be important for the points of research within

[1]https://en.wikipedia.org/wiki/Linear_time-invariant_system

this document. We do not recommend this Section as a way to study the ideas presented here. Instead, we intend for this Section to serve as a high-level overview of some of the relevant matrix theory.

1.1.1 Important Matrix Subspaces

There are some important vector subspaces that we will be addressing consistently throughout this document. Here we will present some informal definitions. The *column space of A*, denote $\text{col}(A)$, is the vector space spanned by the columns of A. Similarly, the *row space of A*, denoted $\text{row}(A)$ or $\text{col}(A^T)$, is the vector space spanned by the rows of A. We will also be interested in the *orthogonal complement to the column space of A*, denoted $\text{col}(A)^\perp$ which is the vector space of vectors which are orthogonal to the column space. The orthogonal complement to the column space of A is also given by the nullspace of A^T. Recall that the null space of a matrix A is the set of vectors $\mathbf{x} \in \mathbb{R}^n$ such that $A\mathbf{x} = \mathbf{0}$.

1.1.2 Rank and Nullity

The *rank* of A is the dimension of $\text{col}(A)$. If the rank of A is $\min(m, n)$ then we say A is full rank. Intuitively the rank of A tells you how large of a space A can map vectors onto. As we will discuss later, in the case of a consistent linear system the system $A\mathbf{x} = \mathbf{b}$ will have a unique solution when $m \geq n$ and A is full rank. On the other hand if A is less than full rank then the system $A\mathbf{x} = \mathbf{b}$ will have many solutions.

The nullity of A is the dimension of the null space of A. By the rank-nullity theorem we have that the rank of A plus the nullity of A equals m.

1.1.3 Vector space decomposition

Given an ambient space and a vector subspace we can always decompose the ambient subspace into the direct sum of that vector space and its orthogonal complement space. In particular if we consider our ambient space to be $\mathbb{R}^m$ then we can decompose $\mathbb{R}^m = \text{col}(A) \oplus \text{col}(A)^\perp$. This means for any $\mathbf{b} \in \mathbb{R}^m$ we can decompose $\mathbf{b} = \mathbf{b}_{col(A)} + \mathbf{b}_{col(A)^\perp}$ where $\mathbf{b}_{col(A)}$ denotes the orthogonal projection of $\mathbf{b}$ onto the column space of A and $\mathbf{b}_{col(A)^\perp}$ denotes the orthogonal projection of $\mathbf{b}$ onto the orthogonal complement of the column space of A. Note that for any vector $\mathbf{b} \in \mathbb{R}^m$ the system $A\mathbf{x} = \mathbf{b}_{col(A)}$ is always consistent.

1.1.4 Singular Values

The *singular values* of A are the eigenvalues of the symmetric matrix $A^T A$. Note that since $A^T A$ is symmetric, all of its eigenvalues are real ☕ . We denote the m singular values $\sigma_{min}(A) \leq \sigma_1(A) \leq \cdots \leq \sigma_{max}(A)$. We also have $\|A\mathbf{x}\|_2^2 \leq \sigma_{max}^2(A)\,\|\mathbf{x}\|_2^2$ and the rank of A is the number of nonzero singular values.

The singular values of A will be of interest to us because the speed of convergence of various numerical methods will depend the singular values. In this sense, the singular values govern the "difficulty" of solving a linear system numerically.

Now we define some quantities, the relevence of which will become clear later. Set $\kappa_F^2(A) = \|A\|_F^2 \left\|A^\dagger\right\|_2^2$ and define the *condition number of A* is $\kappa^2(A) = \sigma_{max}^2(A)/\sigma_{min}^2(A)$. Also note $\kappa^2(A) \leq \kappa_F^2(A)$.

1.1.5 Pseudoinverses

If A is square and of full rank then it will have an inverse A^{-1}. If you can easily compute the matrix inverse then a linear system $A\mathbf{x} = \mathbf{b}$ can be solved as $\mathbf{x} = A^{-1}\mathbf{b}$. Not every matrix has an inverse, but every matrix has a *Moore-Penrose* pseudoinverse (MP inverse), denoted $A^\dagger$. The MP pseudoinverse is a generalized inverse. The MP pseudoinverse has a number of rich properties but must important for us is that $A^\dagger\mathbf{b}$ gives the minimum-norm solution to the least-squares problem. This problem is discussed in detail in Section (1.4.1).

The Moore-Penrose pseudoinverse can be calculated using singular value decomposition[2]. However, for us our linear systems will be so large that we cannot compute the MP pseudoinverse.

1.1.6 Various Matrix Norms

For vectors $\mathbf{x} \in \mathbb{R}^n$ we will usually only be interested in the *Euclidean norm* or *2-norm* of that vector,

$$\|\mathbf{x}\|_2^2 = \sum_{i=0}^{n-1} x_i^2 = \mathbf{x} \cdot \mathbf{x} = \mathbf{x}^T\mathbf{x}.$$

Note that we generally consider the square of the given norms because it will be useful for later analysis. On the other hand, when it comes to matrix norms we will be interested in more than one. We can have multiple matrix norms because a "norm" is just a mapping from a matrix A to a non-negative real number satisfying a predefined set of properties. For our purposes we will simply list the matrix norms which will appear in this document. Firstly the *spectral norm of A* is given by

$$\|A\|_2^2 = \sigma_{max}^2(A).$$

And the *Frobenius norm* of A is given by

$$\|A\|_F^2 = \sum_{i=0}^{m-1}\sum_{j=0}^{n-1} |a_{ij}|^2.$$

[2]For those who are interested, https://mathformachines.com/posts/least-squares-with-the-mp-inverse/

Note that $\|A\|_2^2 \leq \|A\|_F^2$. A pertinent related fact is the *Cauchy-Schwarz inequality* for vectors. For any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ we have $|\langle \mathbf{u}, \mathbf{v} \rangle|^2 \leq \|\mathbf{u}\|_2^2 \|\mathbf{v}\|_2^2$. In particular $\|A\mathbf{x}\|_2^2 \leq \|A\|_2^2 \|\mathbf{x}\|_2^2 = \sigma_{max}^2 \|\mathbf{x}\|_2^2$. Finally, note that $\left\|A^\dagger\right\|_2^2 = 1/\sigma_{min}^2(A)$.

1.2 Consistent Linear Systems

Above we discussed the spirit and applications of linear systems. Now we will define linear systems formally

Definition 1.2.1. A *linear system* is a set of m linear equations in n variables, each equation takes the form

$$a_{i0}x_0 + a_{i1}x_1 + \cdots + a_{in}x_n = b_i,$$

where, for each i we do not have $a_{ij} = 0$ for all j. That is, we assume there are no equations of the form $0 = b_i$. Often we will leverage matrix and vector algebra to represent such a system as

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

where $A \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$ are given.

We say that a linear system is *consistent* if there exists a vector $\mathbf{x}^* \in \mathbb{R}^n$ which makes the above equation true. If we know that a system is consistent we will often be interested in finding a solution $\mathbf{x}^*$. One example of how to do this, which the reader may be familiar with, is Gaussian elimination. From a high level, Gaussian elimination involves performing one of three elementary row operations to manipulate the system of equations until we are left with the solutions to the variables $x_0, x_1, \cdots, x_n$. One issue with this method is that it requires the entire matrix to be represented at each step of the process. There exist cases where this cannot be done. For example, if the matrix is extremely large, then it is often infeasible to consider the entire matrix, even if we were to perform Gaussian elimination on a computer. In this document we are motivated by linear systems which are large, in particular those linear systems which are tall where $m \gg n$. An example of such a system is in *computer tomography* where, for example, a number of X-ray samples are reconstructed computationally to create a high quality image for medical uses. Some discussion on the setup and the matrices which are used in computer tomography problems be found at the end of Section two by Gustafsson (1996).

For the rest of this chapter we will discuss some foundational theory behind different kinds of linear systems.

1.2.1 Geometry of Consistent Linear Systems

If we have a linear system $A\mathbf{x} = \mathbf{b}$ then notice each linear equation defines an affine[3] hyperplane and then the solution to our linear equation is given by the intersection of each of these hyperplanes

Definition 1.2.2 (Geometric Interpretation of a Linear System)**.** A linear system $A\mathbf{x} = \mathbf{b}$ with $A \in \mathbb{R}^{m\times n}$ induces a geometry of m affine hyperplanes in $\mathbb{R}^n$. In particular, each hyperplane is given by "each row" of A,

$$H_i = \{\mathbf{x} \in \mathbb{R}^n : A^{(i)}\mathbf{x} = b_i\}.$$

And then the solution to our linear system is given by the intersection over all such hyperplanes

$$L = \bigcap_{i=0}^{m} H_i.$$

If A is full-rank then L is a point. As the rank of A decreases the dimension of L increases. For example if A is one less than full rank then L will be one dimensional.

Recall that we can uniquely define any hyperplane in $\mathbb{R}^n$ in terms of a *normal vector* and a distance from the origin. Choose a vector $\mathbf{n} \in \mathbb{R}^n$ and a distance $d \in \mathbb{R}$, then we can uniquely define any hyperplane as the subspace of $\mathbb{R}^n$ which is orthogonal to $\mathbf{n}$ and a distance d from the origin.

Example 1.2.1. Suppose we have the linear system

$$A = \begin{bmatrix} 5 & 6 \\ 10 & 3 \\ 1 & -1 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Notice $A \in \mathbb{R}^{3\times 2}$ is full rank; this can be seen since the columns of A are linearly independent. Since $n = 2$ this system defines affine hyperplanes in the plane $\mathbb{R}^2$. Hyperplanes are one dimension less than the ambient space (also referred to as *codimension* 1) and so in this case our hyperplanes are of dimension $2 - 1 = 1$ which corresponds to lines in the plane. Since this document happens to be locally embedded within $\mathbb{R}^2$ we can straightforwardly visualize the geometry of this linear system, as shown in Figure (1.1).

[3]What does affine mean here? It's a technical thing used to distinguish the hyperplane geometry we care about from *projective* geometry and hyperplane geometry that we are familiar with from linear algebra. It is important to note that the hyperplanes we are familiar with from linear algebra, have the condition that they need to include the 0 vector. This means that these hyperplanes must pass through the origin. However, we want to include hyperplanes which do not necessarily pass through the origin.

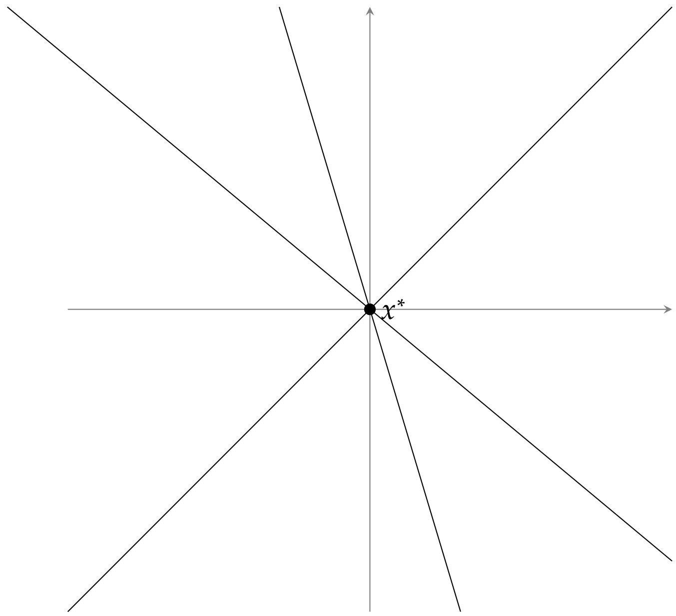


Figure 1.1 Visualization of the geometry of the given linear system.

Notice also that the solution is given by $\mathbf{x}^* = [0, 0]^T$ where all of the hyperplanes intersect. In this case L was a zero-dimensional point, however, if A was rank one then we would have a one-dimensional solution space. For example if we had

$$A = \begin{bmatrix} 1 & -1 \\ 2 & -2 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

then A is rank one, since the column space of A has dimension 1. Then the solution set are all the vectors $\mathbf{x}^* = [xy]^T$ which satisfy $y = x$. A geometric depiction is given in Figure (1.2).

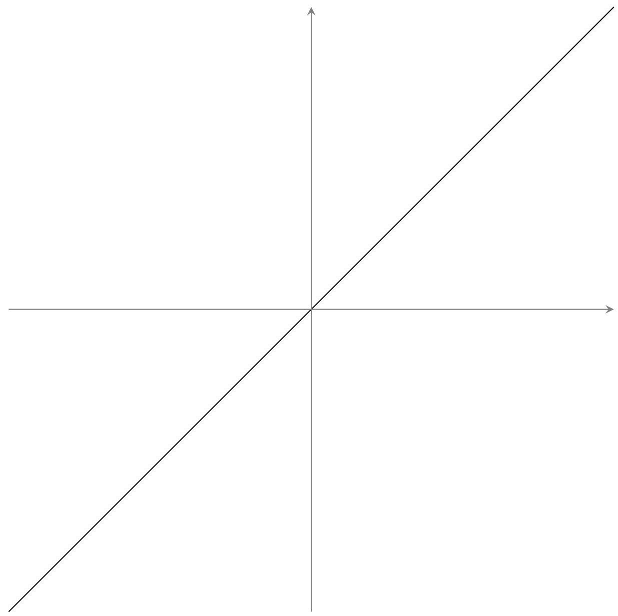


Figure 1.2 Geometric depiction of a rank one linear system as given in example (1.2.1).

1.3 Inconsistent Systems

Often linear systems will not be consistent. In terms of applications we might expect perfectly consistent linear systems but then empirically measure *inconsistent linear systems*. Informally, these are systems where there is no exact vector $\mathbf{x}^*$ which solves the system defined by A and $\mathbf{b}$. We might measure such systems because of noise in measuring signals or realities of empirical measurement. In these cases, we usually want a solution which is “the next best thing” to an exact solution. In this Section we will discuss what an inconsistent system is, what different kinds of inconsistent systems we might come across, and what kinds of “next best” solutions that we might hope to find.

Definition 1.3.1. (Inconsistent Linear System) A linear system $A\mathbf{x} = \mathbf{b}$ is *inconsistent* if there is no $\mathbf{x}$ which satisfies $A\mathbf{x} = \mathbf{b}$. To avoid confusion with consistent linear systems we will often write $A\mathbf{x} = \mathbf{b}$ when we expect there to be an exact solution, otherwise we will refer to the problem defined by A and $\mathbf{b}$.

Remark 1.3.1. (Geometry of Inconsistent Linear Systems) Recall that we can understand linear systems in terms of hyperplanes H_i defined by each of the equations in the systems. When the linear system is inconsistent then we find

$$\bigcap_{i=0}^{n} H_i = \emptyset.$$

In other words, there is no point in $\mathbb{R}^n$ where all of the hyperplanes intersect.

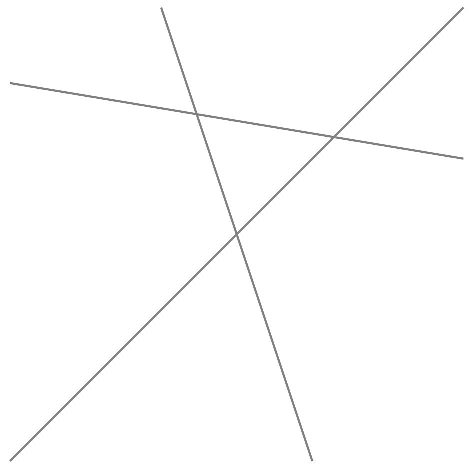


Figure 1.3 The geometry of an inconsistent linear system. There is no solution to this linear system as seen since there is no point in $\mathbb{R}^2$ where all three hyperplanes intersect simultaneously.

Remark 1.3.2 (When are Linear Systems Inconsistent?)**.** The linear problem is defined by both A and $\mathbf{b}$. A linear system is always consistent when $\mathbf{b} \in \text{col}(A)$ (♨ consider what it means to perform $A\mathbf{x}$). And so, since a system can either be consistent or inconsistent and since we can always decompose $\mathbb{R}^n = \text{col}(A) \oplus \text{col}(A)^\perp$, it follows that a system is inconsistent when $\mathbf{b}_{col(A)^\perp} \neq \mathbf{0}$. In other words when $\mathbf{b}$ has any "part" which lies outside the column space of A then our system will be inconsistent — i.e., no linear combination of the columns of A will ever "reach" that "part" of $\mathbf{b}$ which lives outside A's column space. However, this means that inconsistency is due to both A and $\mathbf{b}$.

1.3.1 Types of Inconsistent Linear Systems

There are two kinds of inconsistent linear systems that we are interested in. Informally we might call one kind "those in which the errors are small but ubiquitous" and the other "those in which the errors are large in a few places." If we have a system where the error is "large and ubiquitous" then we may need heavier hammers than simply Mathematics.

Definition 1.3.2 (Noisy Linear Systems)**.** Suppose $A\mathbf{x} = \mathbf{b}$ is a consistent system. Then a *noisy linear system* is one given by A, and $\mathbf{b} + \mathbf{r}$, where $\mathbf{r}$ has a "small magnitude." For our purposes we will not assume exactly what it means for $\mathbf{r}$ to be "small". However, as an example, you might sample $\mathbf{r}$ from multivariate Gaussian distribution with expectation $\mathbf{0}$ and "small" diagonal covariances.

Noisy linear systems are often of interest because they model noise which might come from sensor measurements. A geometric depiction is given in Figure (1.4). Intuitively, we might want to deal with noisy linear systems by finding the point "in the average" of all the hyperplanes. In this case we can "average out" the error and have a reasonable approximation of the underlying consistent system.

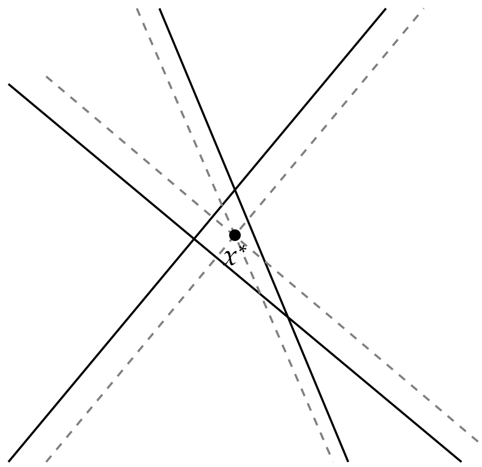


Figure 1.4 The geometry of a noisy linear system. The original linear system is represented by the dotted lines. The hyperplanes have been modified by a small amount of noise.

Another case of inconsistent systems are *corrupted linear systems*.

Definition 1.3.3 (Corrupted Linear Systems)**.** If $A\mathbf{x} = \mathbf{b}$ is a consistent linear system then a *corrupted linear system* is one given by A and $\mathbf{b} + \mathbf{b}^C$ where $\mathbf{b}^C$ is sparse but where the non-zero entries are large.

This definition captures the linear system with a few adversarial components. A geometric depiction is given in Figure (1.5). Since a very small portion of the system is modified in a corrupted linear system, it would be ideal if we could detect the corrupted equations and ignore them completely. If we could strip away the few corrupted equations then we could simply treat the remaining system as consistent. This is in contrast to the noisy linear system where we might not worry about removing the error entirely but rather "average it out".

It is worth noting that there are different kinds of corrupted systems. For example, instead of corruptions being found in $\mathbf{b}$ you might instead corrupt the entries of A. For this document we will only be considering corruptions to $\mathbf{b}$. Note that when $\mathbf{b}$ is modified the new hyperplanes shift paralell along their original normal vector as can be seen in Figure (1.4).

We will now turn our attention to different ways of solving these kinds of inconsistent systems.

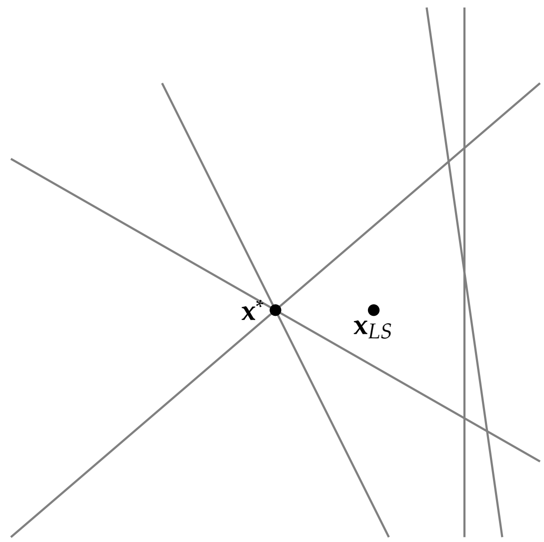


Figure 1.5 The geometry of a corrupted linear system. In this case nearly all the hyperplanes intersect at a single point but with a few highly perturbed hyperplanes. One solution we might want is the point where most of the hyperplanes intersect, in this case on the left. However, the least squares solution for such a system is far away from that point, instead it is somewhere between the corrupted hyperplanes and the point we would actually want.

1.4 "Solutions" to Inconsistent Systems

By definition, we cannot find exact solutions to inconsistent linear systems, however, we can find various approximate solutions by optimizing different objectives.

1.4.1 Least Squares

One approximation is called the *least squares solution* which is found by solving the *least squares problem*.

Definition 1.4.1 (The Least Squares Problem)**.** Given a linear system defined by A and $\mathbf{b}$ the *least squares solution* $\mathbf{x}^* \in \mathbb{R}^n$ is a vector $\mathbf{x}^*$ such that

$$||\mathbf{b} - A\mathbf{x}^*||_2^2 \leq ||\mathbf{b} - A\mathbf{x}||_2^2$$

for all other vectors $\mathbf{x} \in \mathbb{R}^n$. ☕ Equivalently the *least squares solution* is the vector $\mathbf{x}^* \in \mathbb{R}^m$ which optimizes

$$\min_{\mathbf{x}\in\mathbb{R}^n} ||A\mathbf{x} - \mathbf{b}||_2^2 .$$

In the context of the least squares problem, we will use the term *pseudosolution* to refer to the least squares solution of a linear system.

When A is not full rank then we expect there to be multiple vectors which satisfy the least squares problem. Having multiple solutions to the least squares problem is not a concern for us and in this case we will let $\mathbf{x}^*$ denote the least squares solution of minimum 2-norm.

To distinguish the Least Squares Problem from the *Weighted Least Squares Problem* we may also call the least squares problem the *Ordinary Least Squares Problem*.

When we expect the error to be small, e.g., in noisy linear systems, then the least squares solution is usually a good approximation for the solution of the system. For example, in Figure (1.4) the vector $\mathbf{x}^*$ also happens to be the solution to the least squares problem for that system.

We can compute the least squares solution in a few ways. We can compute the least squares solution via the normal equations

$$A^T A\mathbf{x} = A^T\mathbf{b}.$$

If A is full column rank then the normal equations have solution

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

On the other hand, we can use the *Moore-Penrose* pseudo inverse

$$\mathbf{x}^* = A^\dagger\mathbf{b}.$$

Both of these methods require loading the entire matrix into memory. As a result these methods cannot be used when A is sufficiently large. For this reason we will later turn to iterative methods for solving the least squares problem.

When the expected error is small and affects all of our equations equally then the least squares solution is in some sense the optimal solution[4]. However, least squares may not be optimal when different equations are affected by different amounts of error. One example of this is the corrupted system where most equations are unaffected by the error and a few are heavily affected by the error. As an example, in Figure (1.5) the least squares solution, denoted by $\mathbf{x}^*$ is far away from the place where most of the hyperplanes intersect to the left of $\mathbf{x}^*$. In this case we would prefer a different solution to $\mathbf{x}^*$. In this case we present another kind of optimization problem in the following Section.

1.4.2 Weighted Least Squares Problem

For an inconsistent system the weighted least squares problem solves for a vector $\mathbf{x}^*$ where a weight is assigned to each of the equations. Intuitively the weight of an equation is

[4]There is a rich statistics theory. For those interested, one might say that the least squares solution is the Best Linear Unbiased Estimator (BLUE) when the conditions for linear regression are met.

how "reliable" you expect a particular equation to be. As an example, if you are using the outcomes of games to assign ratings to teams then you may "trust" some outcomes more than others and thus assign a higher weight to them. For example, if a team's key player was injured for one game then you may want to down-weigh that game because it might not be as good an indication of that team's performance compared to when they do have their key player. Formally the weighted least squares problem is given as follows

Definition 1.4.2 (Weighted Least Squares Problem)**.** Let A and $\mathbf{b}$ define an inconsistent system, and let $W \in \mathbb{R}^{m \times m}$ be a diagonal matrix of non-negative weights where W_{ii} is the weight assigned to equation i. The *Weighted Least Squares Problem* (WLS) is to find a vector $\mathbf{x} \in \mathbb{R}^n$ which minimizes

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|W\mathbf{b} - WA\mathbf{x}\|_2^2 .$$

In the context of the context of WLS we call $\mathbf{x}^*$ the *pseudosolution.*

Note that the Ordinary Least Squares problem is a special case of the Weighted Least Squares problem where $W_{ii} = 1$ for all i. Like the OLS problem there may not always be a unique vector which solve the WLS problem. In this case we usually let $\mathbf{x}^*$ denote the pseudosolution of minimum 2-norm.

Remark 1.4.1 (Reweightings do not change the geometry of a linear system ☕)**.** If $(A^{(i)})^T \mathbf{x} = b_i$ defines a hyperplane then $W_{ii}(A^{(i)})^T \mathbf{x} = W_{ii} b_i$ defines the same hyperplane, so long as $W_{ii} \neq 0$. Instead, the weighting changes the position of the pseudosolution compared to OLS.

Remark 1.4.2 (WLS Problems not uniquely defined by W)**.** Let us fix a matrix A and a vector $\mathbf{b}$. Suppose we are also given a matrix W to define a WLS problem. Notice that aW for some nonzero $a \in \mathbb{R}$ defines the same WLS problem as W. Consider the following example. Let W_1 be defined by $(W_1)_{11} = a$ and $(W_1)_{ii} = 1$ for $i \neq 1$. And let W_2 be defined by $(W_2)_{11} = 1$ and $(W_2)_{ii} = 0$ for $i \neq 1$. ☕ As $a \to \infty$, both W_1 and W_2 define the same WLS problem.

Later we will be considering sequences of WLS problems, to make such statements well-defined let us constrain $\|W\|_F^2 = 1$ unless otherwise stated.

For those who have a background in statistics, if the variances of each equation σ_i^2 are known[5] then it can be shown that the weighted least squares solution is a Best Linear Unbiased Estimator (BLUE) if we choose $W_{ii} = 1/\sigma_i^2$. In this sense, the weighted least squares solution is optimal if we have heteroskedasicity (i.e., unequal but diagonal variances).

If our inconsistent linear system is corrupted then we may want to solve the WLS problem. As an extreme example, suppose we knew which equations had been corrupted then we could set $W_{c_i} = 0$ for all corrupted equations indexed by c_i. Intuitively, one may

[5]This is usually difficult to do, see https://www.itl.nist.gov/div898/handbook/pmd/section1/pmd143.htm

want to use WLS when we do not trust all the data equally, but we do not want to prune the equations we trust less. The difficulty with Weighted Least Squares is assigning the weights to each of the equations, as estimating the variances of each of the equations can be difficult.

Similar to the Ordinary Least Squares Problem there are normal equations for finding the weighted least squares solution of a problem.

$$A^T W A \mathbf{x}^* = A^T W \mathbf{b}.$$

Similar to the normal equations for least squares problems, the normal equations methods require loading the entire matrix A into memory which is not feasible for large linear systems.

This is the required linear system background to understand the numerical methods contributions of this text. In the following chapter we will introduce the Randomized Kaczmarz method as a way of solving large linear systems.

1.5 Notation

Let $[k] = \{0, 1, 2, \cdots, k-1\}$. Unless otherwise stated we let $A \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$. Let $I_m \in \mathbb{R}^{m \times m}$ denote the identity matrix in m dimensions. We let $A^{(i)}$ denote the ith row of A and $A_{(i)}$ denote the i column of A, both taken as column vectors. Often α will denote the quantity $1 - 1/\kappa_F^2(A)$ where $\kappa_F^2(A) = \|A\|_F^2 \left\|A^\dagger\right\|_2^2$. For a linear system $A\mathbf{x} = \mathbf{b}$ we let $\mathbf{x}^*$ denote the vector which exactly solves the linear system. In a least squares problem $\mathbf{x}^*$ will denote the minimum norm solution. And in a weighted least squares problem we let $\mathbf{x}^*$ denote the minimum norm solution.

Chapter 2

Randomized Kaczmarz

Methods such as computing the inverse or Gauss elimination involve loading the entire matrix into memory. Often, the linear systems that we are interested in are so large that loading the entire matrix into memory is not feasible. In these cases we turn to *iterative methods* which numerically solve the linear system by only using a portion of the input matrix A at a time. Within this chapter we will focus our attention on a particular numerical method called Randomized Kaczmarz[1].

Given a consistent system, Randomized Kaczmarz is an iterative method which finds the solution to linear systems by exploiting the geometry of the hyperplanes. In particular, from a high-level geometry standpoint, it performs a sequence of orthogonal projections onto a randomly selected hyperplane, eventually converging to the solution $\mathbf{x}^*$.

Definition 2.0.1 (Randomized Kaczmarz)**.** Let $A\mathbf{x} = \mathbf{b}$ be a linear system. Define any initial iterate $\mathbf{x}^{(0)} \in \mathrm{col}(A^T)$. The iterates of the *Randomized Kaczmarz* algorithm are given by

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \frac{b_{i_k} - \left\langle \mathbf{x}^{(k)}, \mathbf{A}^{(i_k)} \right\rangle}{||\mathbf{A}^{(i_k)}||_2^2} \mathbf{A}^{(i_k)},$$

where index i_k is chosen with probability $q_i = \frac{||A^{(i)}||_2^2}{||A||_F^2}$.

This algebraic expression can be intense at first sight, however, we will break it down together in the following remarks. We also define Randomized Kaczmarz in algorithm form in Algorithm (1).

[1]Kaczmarz was a polish fellow, knowing this might aid with pronouncing his name and the name of the algorithms. From what I understand, Google Translate has a decent pronunciation for this name.

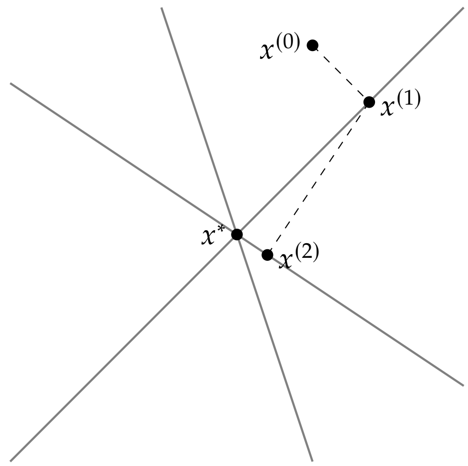


Figure 2.1 Iterations from Randomized Kaczmarz. At each step $\mathbf{x}^{(k)}$ is orthogonally projected onto a randomly selected affine hyperplane. Each hyperplane represents a single equation from our linear system.

Algorithm 1 Randomized Kaczmarz

function RandomizedKaczmarz($A, \mathbf{b}, T$) ▷ where $A \in \mathbb{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$, and $T \in \mathbb{N}$
  Initialize $\mathbf{x}^{(0)} \in \mathbb{R}^n$ arbitrarily
  for $k = 1, 2, 3, \ldots, T$ **do**
    Pick i_k with probability $q_i := \frac{\|A^{(i)}\|_2^2}{\|A\|_F^2}, i \in [m]$
    Set $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \frac{b_{i_k} - \langle \mathbf{x}^{(k)}, A^{(i_k)}\rangle}{\|A^{(i_k)}\|_2^2} A^{(i_k)}$
  end for
  return $\mathbf{x}^{(T)}$
end function

Remark 2.0.1 (Kaczmarz Updates are Projections onto Hyperplanes)**.** Whilst the RK updates may look intimidating, we can break it down and extract a nice geometry. First consider

$$\ell = \mathbf{b}_{i_k} - \left\langle \mathbf{x}^{(k)}, \mathbf{A}^{(i_k)} \right\rangle .$$

The inner product $\left\langle \mathbf{x}^{(k)}, \mathbf{A}^{(i_k)} \right\rangle$ can be thought intuitively like the matrix vector multiplication $A\mathbf{x}$ executed on a single row $A^{(i_k)}$. Then ℓ is the difference between $A\mathbf{x}$ and $\mathbf{b}$ for a single equation of the system. And so, intuitively, the modification term $\frac{b_{i_k} - \left\langle \mathbf{x}^{(k)}, A^{(i_k)} \right\rangle}{||A^{(i_k)}||_2^2} A^{(i_k)}$ will move the current $\mathbf{x}$ iterate in the direction of $A^{(i_k)}$ as to solve equation i_k exactly. Geometrically this corresponds to projecting $\mathbf{x}^{(k)}$ onto the hyperplane given by a single hyperplane H_{i_k}. Later we will show that these projections are orthogonal, which will be important for guarenteeing RK's convergence.

Remark 2.0.2 (Row Distribution is Superficial)**.** The probability of picking each row q_i is usually given in a form which may make one ask which rows are more likely to be chosen than others. It turns out that this probability distribution of picking the rows is superficial. Consider, we can normalize[2] the rows of A by setting $A^{(i)} = A^{(i)}/\left\|A^{(i)}\right\|_2^2$ and $b_i = b_i/\left\|A^{(i)}\right\|$. Here $\left\|A^{(i)}\right\|_2^2 = 1$ for all i. Moreover, the solution to the whole system does not change by such a transformation. Suppose $B \in \mathbb{R}^{m\times m}$ is diagonal such that $B_{ii} = 1/\left\|A^{(i)}\right\|_2^2$. And so if $\mathbf{x}^*$ solves $A\mathbf{x} = \mathbf{b}$ then $\mathbf{x}^*$ will also solve $BA\mathbf{x} = B\mathbf{b}$. That is the geometry of the underlying system is unchanged. Since $\left\|A^{(i)}\right\|_2^2 = 1$ for all i it follows that q_i defines a uniform distribution. That is, the original distribution for choosing the rows is uniform.

2.1 Convergence Properties

Randomized Kaczmarz projects onto hyperplanes in such a way that the algorithm is guaranteed to converge in a very fast time. Orthogonal projections are key to the convergence of RK. So, to understand how this convergence works we must first show that these projections are actually orthogonal.

Proposition 1 (RK Updates are Orthogonal Projections)**.** *Let $A\mathbf{x} = \mathbf{b}$ be some linear system. On a given iterate of Randomized Kaczmarz, the error $\mathbf{x}^* - \mathbf{x}^{(k+1)}$ and the update vector $\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}$ are orthogonal.*

Proof. We will show this by definition. Consider the following

$$\begin{aligned}
(\mathbf{x}^* - \mathbf{x}^{(k+1)})^T(\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}) &= (\mathbf{x}^* - \mathbf{x}^{(k+1)})^T \frac{b_{i_k} - \left\langle x^{(k)}, A^{(i_k)}\right\rangle}{\left\|A^{(i_k)}\right\|_2^2} A^{(i_k)} \\
&= (\mathbf{x}^* - \mathbf{x}^{(k+1)})^T \beta A^{(i_k)} \\
&= \beta(\mathbf{x}^*)^T A^{(i_k)} - \beta(\mathbf{x}^{(k+1)})^T A^{(i_k)} \\
&= \beta b_{i_k} - \beta(\mathbf{x}^{(k+1)})^T A^{(i_k)} && \text{Since } A\mathbf{x}^* = \mathbf{b} \\
&= \beta b_{i_k} - \beta b_{i_k} \\
&= 0,
\end{aligned}$$

[2]Here, we do not have to worry about $\left\|A^{(i)}\right\|_2^2 = 0$. Recalling the properties of norm, this would only happen if $A^{(i)} = \mathbf{0}$, however, these are discounted from discussion.

Where the second to last line follows because $A^{(i_k)}\mathbf{x}^{(k)} = b_{(i_k)}$ by definition of Kaczmarz updates. Informally, a single RK update makes it so that $\mathbf{x}^{(k+1)}$ satisfies the hyperplane that it projects onto. Thus Randomized Kaczmarz updates are orthogonal projections. □

This orthogonality means that we can show the following.

Proposition 2 (RK Error is Monotonically Non-Increasing)**.** *The norm of the error* $||\mathbf{x}^* - \mathbf{x}^{(k)}||_2^2$ *is monotonically non-increasing.*

Proof. Consider the following

$$\begin{aligned} \left\|\mathbf{x}^* - \mathbf{x}^{(k+1)}\right\|_2^2 &= \left\|\mathbf{x}^* - \mathbf{x}^{(k)} + \mathbf{x}^{(k)} - \mathbf{x}^{(k+1)}\right\|_2^2 \\ &\leq \left\|\mathbf{x}^* - \mathbf{x}^{(k)}\right\|_2^2 + \left\|\mathbf{x}^{(k)} - \mathbf{x}^{(k+1)}\right\|_2^2. && \text{By the triangle inequality,} \end{aligned}$$

And so since, $(\mathbf{x}^* - \mathbf{x}^{(k+1)})$ and $(\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)})$ are orthogonal, we have

$$\begin{aligned} \left\|\mathbf{x}^* - \mathbf{x}^{(k+1)}\right\|_2^2 &\leq \left\|\mathbf{x}^* - \mathbf{x}^{(k)}\right\|_2^2 + \left(\left\|\mathbf{x}^* - \mathbf{x}^{(k)}\right\| - \left\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\right\|_2^2\right) \\ 2\left\|\mathbf{x}^* - \mathbf{x}^{(k+1)}\right\|_2^2 &\leq 2\left\|\mathbf{x}^* - \mathbf{x}^{(k)}\right\|_2^2 \\ \left\|\mathbf{x}^* - \mathbf{x}^{(k+1)}\right\|_2^2 &\leq \left\|\mathbf{x}^* - \mathbf{x}^{(k)}\right\|_2^2. \end{aligned}$$

Thus the error vectors are monotonically increasing. (☕ Can you see this proposition geometrically in Figure 2.1?) □

At the very least we have shown that the iterations of Randomized Kaczmarz will never take us farther from the pseudo solution. However, we can do one better and show that RK, in expectation, converges exponentially fast (also called *linearly in expectation*).

Theorem 2.1.1 (Randomized Kaczmarz Converges Linearly in Expectation)**.** *Suppose A is full column rank. After k iterations of Randomized Kaczmarz we have the following*

$$\mathbb{E}||\mathbf{x}^* - \mathbf{x}^{(k+1)}||_2^2 \leq \left(1 - \frac{1}{\kappa_F^2(A)}\right)^k ||\mathbf{x}^* - \mathbf{x}^{(0)}||_2^2,$$

where A and **b** *define a consistent linear system and* $\mathbf{x}^{(0)} \in col(A^T)$.

This result was first proved in Strohmer and Vershynin (2009), and we will closely follow their proof with some modifications presented by Zouzias and Freris (2013).

Proof. Let Z be a random variable over $[m]$ with probability mass function $P(Z = i) = \left\|A^{(i)}\right\|_2^2 / \|A\|_F^2$. We will first show that

$$\mathbb{E}||\mathbf{x}^* - \mathbf{x}^{(k+1)}||_2^2 \leq \left(1 - \frac{1}{\kappa_F^2(A)}\right) ||\mathbf{x}^* - \mathbf{x}^{(k)}||_2^2,$$

and then repeating the same argument $k-1$ more times will give us the desired result. By Proposition (1) it suffices to show that $\mathbb{E}_Z \left\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\right\|_2^2 \geq 1/\kappa_F^2(A) \left\|\mathbf{x}^{(k)} - \mathbf{x}^*\right\|_2^2$. By definition of $\mathbf{x}^{(k+1)}$, and by linearity of inner product, we have

$$\begin{aligned}
\mathbb{E}_Z \left\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\right\|_2^2 &= \mathbb{E}_Z \left[\left(\frac{b_z - \left\langle \mathbf{x}^{(k)}, A^{(Z)}\right\rangle}{\left\|A^{(Z)}\right\|_2^2}\right)^2 \left\|A^{(Z)}\right\|_2^2\right] \\
&= \mathbb{E}_Z \left[\frac{\left\langle \mathbf{x}^* - \mathbf{x}^{(k)}, A^{(Z)}\right\rangle^2}{\left\|A^{(Z)}\right\|_2^2}\right] && \text{Since } A\mathbf{x}^* = \mathbf{b}. \\
&= \sum_{i=0}^{m-1} \frac{\left\langle \mathbf{x}^* - \mathbf{x}^{(k)}, A^{(Z)}\right\rangle^2}{\left\|A^{(Z)}\right\|_2^2} \cdot \frac{\left\|A^{(Z)}\right\|_2^2}{\|A\|_F^2} \\
&= \sum_{i=0}^{m-1} \frac{\left\langle \mathbf{x}^* - \mathbf{x}^{(k)}, A^{(Z)}\right\rangle^2}{\|A\|_F^2} \\
&= \frac{\left\|A(\mathbf{x}^* - \mathbf{x}^{(k)})\right\|_2^2}{\|A\|_F^2} \\
&\leq \frac{\sigma_{min}^2(A) \left\|(\mathbf{x}^* - \mathbf{x}^{(k)}\right\|_2^2}{\|A\|_F^2} \\
&= \frac{1}{\kappa_F^2(A)} \left\|(\mathbf{x}^* - \mathbf{x}^{(k)}\right\|_2^2.
\end{aligned}$$

The second to last line follows by the Cauchy-Schwarz inequality. We can apply this reasoning because $\mathbf{x}^{(k)} \in \mathrm{col}(A^T)$ which follows from and definition of RK updates $\mathbf{x}^{(0)} \in \mathrm{col}(A^T)$. (☕ Why did choosing the hyperplanes randomly help with our convergence speed? What happens if we choose the hyperplanes deterministically in a cycle?) □

In the case where A is not full column rank the final position of $\mathbf{x}$ will vary with the initial iterate $\mathbf{x}^{(0)}$. This exponential convergence rate dependent only on the input matrix is the gold standard for Randomized Kaczmarz and its variations. It is the gold standard in the sense that when designing different variants of RK we aim to have a convergence rate as fast as linear in expectation.

We have seen wonderful results from Randomized Kaczmarz. However, it all comes with one important caveat — the input system must have an exact solution. However, not every linear system is consistent. So next we turn our attention to the behaviour of RK on inconsistent systems.

2.1.1 Behavior on Inconsistent Systems

Recall that the behavior of RK is to orthogonally project onto each hyperplane in our linear system until it converges to the solution. When our linear system is inconsistent then in the limit $T \to \infty$, RK will not converge to any particular point in $\mathbb{R}^n$, because its convergence will be bounded by the "horizon" defined by the geometry of the hyperplanes in the inconsistent system.

This intuition is captured by the following theorem, first proved by Needell (2010).

Theorem 2.1.2 (Horizon of Convergence)**.** *Assume that the system* $A\mathbf{x} = \mathbf{y}$ *has some solution* $\mathbf{x}^*$ *for* $\mathbf{y} \in \mathbb{R}^m$*. Let* $\mathbf{x}^{(k)}$ *denote the k-th iterate of Randomized Kaczmarz applied on the system* $A\mathbf{x} = \mathbf{b}$ *where* $\mathbf{b} = \mathbf{y} + \mathbf{w}$ *for some fixed* $\mathbf{w} \in \mathbb{R}^m$*. It follows that*

$$\mathbb{E}\left\|\mathbf{x}^{(k)} - \mathbf{x}^*\right\|_2^2 \leq \left(1 - \frac{1}{\kappa_F^2(A)}\right)^k \left\|\mathbf{x}^{(0)} - \mathbf{x}^*\right\|_2^2 + \frac{||\mathbf{w}||_2^2}{||A||_F^2}.$$

This theorem says that on an inconsistent linear system RK will converge to the solution exponentially in expectation, except its convergence will be bounded proportional to $||\mathbf{w}||_2^2$, where $||\mathbf{w}||_2^2$ is in some sense the "displacement" of our system from a consistent system. Notice that a linear system is always convergent if $\mathbf{w} = \mathbf{0}$, i.e., when $\mathbf{b} \in col(A)$. An example of this behavior is depicted in Figure (2.2).

Theorem (2.1.2) says that Randomized Kaczmarz will converge to some bounded horizon exponentially quickly. As discussed earlier, there is no exact solution to inconsistent systems and so we instead prefer a least squares solution or weighted least squares solution. However, RK is bound to projecting onto hyperplanes of the input system. This behaviour is a superpower of RK since it grants us orthogonal projections and exponential convergence in the consistent case. However, it is RK's downfall in the case where our hyperplanes do not intersect at a meaningful point. In response to this, there have been modifications to RK to grant better convergence in the case of inconsistent systems. (Try to draw a figure similar to Figure (2.2) for a noisy linear system. Can you see the horzion $\mathbf{w}$ on each of these pictures?)

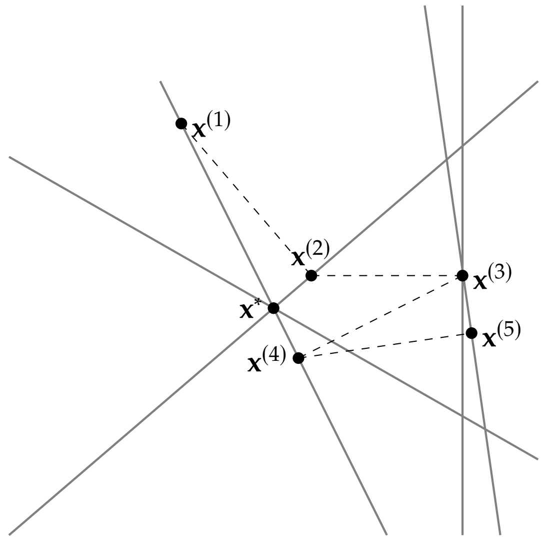


Figure 2.2 Recall that in corrupted linear systems the solution we would want is the one where the corrupted hyperplanes are ignored. Here, the desired solution is denoted by $\mathbf{x}^*$. However, Randomized Kaczmarz will never converge to $\mathbf{x}^*$. Instead, Randomized Kaczmarz may get close, but every time a corrupted hyperplane is projected onto the $\mathbf{x}$ iterates will move far away from $\mathbf{x}^*$.

2.2 REK and WREK

In the case where we would prefer the least squares solution of a linear system, we have a Randomized Kaczmarz variant called *Randomized Extended Kaczmarz* (REK). The main contributions of this document come from expositing and generalizing the theory behind REK. The focus of our next chapter will be entirely on REK.

In the summer of 2021 Haddock's UCLA summer research group proposed a new variation of RK called Weighted Randomized Extended Kaczmarz (WREK). This variation is quite new but aims to solve weighted least squares problems in a similar way to REK. We will discuss this variant more in the final chapter, Chapter 4.

Chapter 3

Randomized Extended Kaczmarz

We now turn our attention to Randomized Extended Kaczmarz (REK). REK was first proposed by Zouzias and Freris (2013) as a modification to RK to converge to the least squares solution of an inconsistent system. Unlike other variants of RK, REK iteratively modifies the input system and then solves the new system. One may ask what resemblance a different linear system has to our original linear system. We will unpack the theory behind REK to hopefully shed understanding on how REK carefully modifies the given linear system to result in a meaningful convergent solution. Understanding REK and its convergence properties will set the foundation for our investigation into WREK.

If you are familiar with REK then my personal contributions are as follows. I have done work expositing Zouzias and Freris (2013) original proof of convergence. Additionally, one might argue that the original proof of REK's convergence does not reveal the same geometry that we see in proofs of RK's convergence. To me, it was important work to reveal more of the underlying geometry behind REK. To this end I also exposit the ideas in Du (2018) in a more similar langauge to Zouzias and Freris (2013) to provide a unified way of thinking about the two proofs. In particular, I do some work to emphasize the underlying geometry of REK. In addition to this, I provide some novel propositions including an explicit closed form of the $\mathbf{z}$ updates in REK.

3.1 Definition and Convergence Properties

Throughout this section, unless stated otherwise, we let $A \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$.

Randomized Extended Kaczmarz iterates the same way as Randomized Kaczmarz, but also iteratively modifies the system as it iterates. The consequence of the modifications is to project $\mathbf{b}$ onto the column space of A. By doing so the system $A\mathbf{x} = \mathbf{b}_{col(A)}$ is consistent and it turns out the solution to this new system is the least squares solution to the old system. We describe Randomized Extended Kaczmarz in algorithm form.

Algorithm 2 Randomized Extended Kaczmarz

function RandomizedExtendedKaczmarz($A, \mathbf{b}, T$) ▷ where $A \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, and $T \in \mathbb{N}$

  Initialize $\mathbf{x}^{(0)} = \mathbf{0}$ and $z^{(0)} = \mathbf{b}$

  for $k = 0, 1, 2, \ldots, T-1$ **do**

    Pick $i_k \in [m]$ with probability $q_i := \frac{\|A^{(i)}\|_2^2}{\|A\|_F^2}, i \in [m]$

    Pick $j_k \in [n]$ with probability $p_j := \frac{\|A_{(j)}\|_2^2}{\|A\|_F^2}, j \in [n]$

    Set $\mathbf{z}^{(k+1)} = \mathbf{z}^{(k)} - \frac{\left\langle A_{(j_k)}, \mathbf{z}^{(k)} \right\rangle}{\left\| A_{(j_k)} \right\|} A_{(j_k)}$

    Set $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \frac{b_{i_k} - z_{i_k}^{(k+1)} - \langle \mathbf{x}^{(k)}, A^{(i_k)} \rangle}{\|A^{(i_k)}\|_2^2} A^{(i_k)}$

  end for

  return $\mathbf{x}^{(T)}$

end function

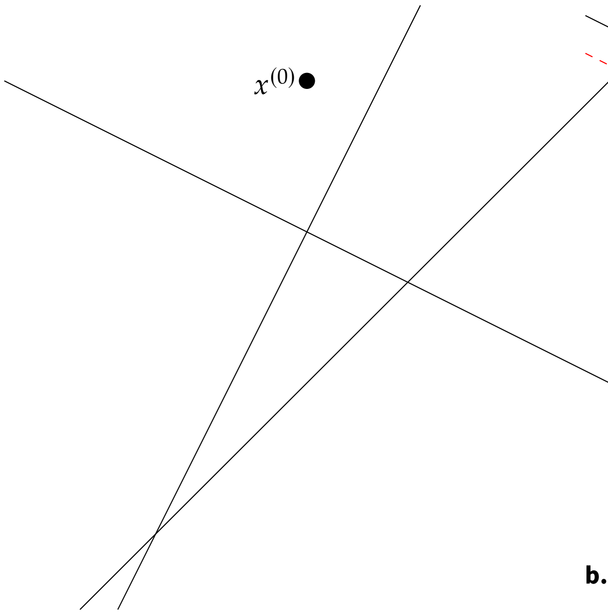
Note Zouzias and Freris (2013) present a stopping criterion that the interested reader may investigate. We will discuss some of the spirit and ideas behind REK.

Remark 3.1.1 (Reducing the Horizon)**.** Recall Theorem (2.1.2) states that RK on an inconsistent system converges up to a horizon. In particular if we have a measurement vector $\mathbf{b} = \mathbf{y} + \mathbf{w}$ such that $\mathbf{y} \in \mathrm{col}(A)$ and $\mathbf{w} \in \mathrm{col}(A)^\perp$ then RK will converge more tightly if $\|\mathbf{w}\|_2^2$ is smaller. The idea with REK is to iteratively reduce the horizon to zero by projecting $\mathbf{b}$ onto a randomly chosen column of A. As we will show later, such a process will eventually orthogonally project $\mathbf{b}$ onto the column space of A. This works because we can decompose any vector space into a subspace and its orthogonal complement, in particular $\mathbb{R}^m = col(A) \oplus col(A)^\perp$.

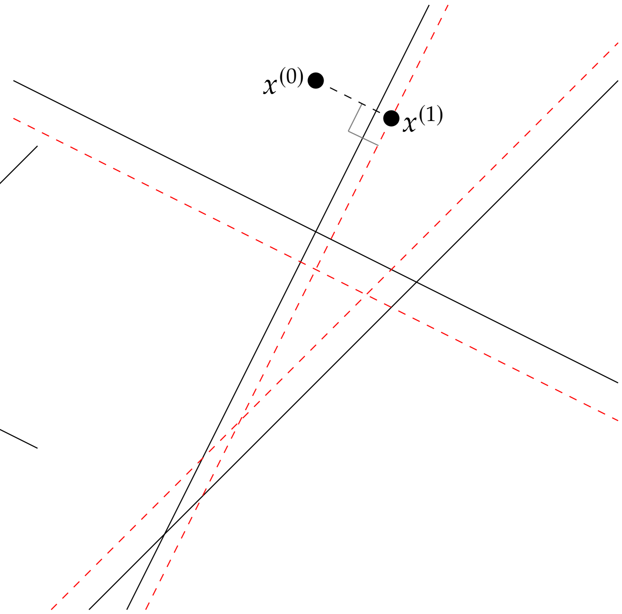
Remark 3.1.2 (REK is just double RK)**.** The updates to $\mathbf{x}$ and $\mathbf{z}$ look very similar to each other in Algorithm (2) and this is not a coincidence. ☕ The updates to the $\mathbf{z}$ vector are the result of running Randomized Kaczmarz on the system $A^T\mathbf{z} = \mathbf{0}$.

3.1.1 Geometric Interpretation of REK

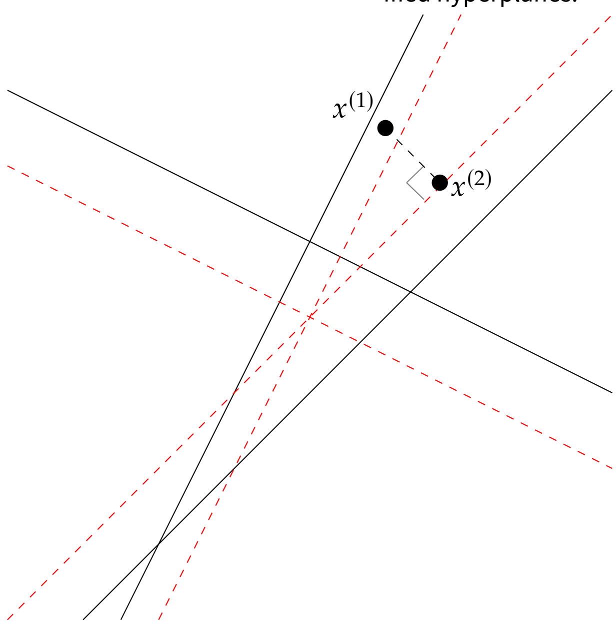
Before we move onto the theoretical foundations of REK, we present some geometric interpretations of the action of REK based on the underlying hyperplanes. A geometric interpretation is displayed and discussed in Figure (3.1). (☕ The actual change in geometry of the hyperplanes is a bit more subtle than the following figure. Sketch/Graph an $m \times 2$ linear system. Project $\mathbf{b}$ onto a column of A and see how the hyperplanes change.)



a. Original unmodified system with initial **x** iterate



b. System after one step of modification along with first **x** update. Notice that the new **x** iterate projects onto the new modified hyperplanes.



c. One more step of iteration. Modified system is now consistent, seen since all the hyperplanes intersect at a point. Next **x** iterate projects on the now consistent system.

Figure 3.1 Steps of REK iterations demonstrating the system updates and the solution vector estimate updates $\mathbf{x}^{(k)}$. The hyperplanes of the original inconsistent system is represented with the black lines and the modified system is represented by the red dotted lines.

3.2 What is Up With the z Vectors?

By understanding what the $\mathbf{z}$ vectors are doing one may understand how REK works. In Zouzias and Freris (2013) we know that the $\mathbf{z}$ vectors converge to $\mathbf{b}_{col(A)^{\perp}}$. However, what do the $\mathbf{z}$ iterates look like? What are they doing? Qualitatively, REK is doing something like taking projections of $\mathbf{b}$ onto a random column of A and then using some sum of these projections in the $\mathbf{x}$ updates. But, what does this mean? Can we get anymore insight into the $\mathbf{z}$ updates? In this Section we will be unpacking the iterative definition of the $\mathbf{z}$ updates into a closed form.

Definition 3.2.1 ($\mathbf{b}$ Iterates)**.** Let

$$\mathbf{b}^{(k)} = \mathbf{b} - \mathbf{z}^{(k)}.$$

At each step of REK these $\mathbf{b}^{(k)}$ iterates are the actual vectors input to the $\mathbf{x}$ updates.

Rather than discussing what the $\mathbf{z}$ vectors look like in each step, for understanding it will be easier to talk about the $\mathbf{b}^{(k)}$ updates. The final form for $\mathbf{b}^{(k)}$ will be quite slick. First we will illustrate with an example.

Example 3.2.1 (First $\mathbf{z}$ Iterates)**.** We will unpack the first couple iterations of the $\mathbf{z}$ updates for REK. Recall that we have

$$\mathbf{z}^{(k+1)} = \mathbf{z}^{(k)} - \frac{\left\langle A_{(j_{k+1})}, \mathbf{z}^{(k)} \right\rangle}{\left\| A_{(j_{k+1})} \right\|} A_{(j_{k+1})}$$

with $\mathbf{z}^{(0)} = \mathbf{b}$. Let $\alpha_k = \frac{1}{\left\| A_{(j_k)} \right\|}$. Also let

$$\mathbf{b}_{(j_k)} = \alpha_k \left\langle A_{(j_k)}, \mathbf{b} \right\rangle A_{(j_k)}$$

be a projection term. Finally, let us define the correction terms by

$$\mathbf{c}_k = \alpha_k \left\langle A_{(j_k)}, \mathbf{b}^{(k-1)} \right\rangle A_{(j_k)}$$

for $k > 1$ and let

$$\mathbf{c}_1 = \mathbf{0}.$$

Now consider the following first few iterations of the $\mathbf{z}$ vectors. We have $\mathbf{z}^{(0)} = \mathbf{b}$ and so $\mathbf{b}^{(0)} = \mathbf{0}$. Then following the updating rule for $\mathbf{z}$ we have

$$\mathbf{z}^{(1)} = \mathbf{b} - \mathbf{b}_{(j_1)}.$$

Then, it follows

$$\mathbf{b}^{(1)} = \mathbf{b}_{(j_1)} = \mathbf{b}_{(j_1)} - \mathbf{c}_1.$$

So, the first $\mathbf{b}^{(k)}$ iterate is just the projection of $\mathbf{b}$ onto a random column of A. Next,

$$\begin{aligned}\mathbf{z}^{(2)} &= \mathbf{z}^{(1)} - \alpha_2 \left\langle A_{(j_2)}, \mathbf{z}^{(1)} \right\rangle A_{(j_2)} \\ &= \mathbf{b} - \mathbf{b}_{(j_1)} - \alpha_2 \left\langle A_{(j_2)}, \mathbf{b} - \mathbf{b}_{(j_1)} \right\rangle A_{(j_2)}\end{aligned}$$

Using linearity and symmetry of inner product on real numbers we have

$$\begin{aligned}\mathbf{z}^{(2)} &= \mathbf{b} - \mathbf{b}_{(j_1)} - \alpha_2(\left\langle A_{(j_2)}, \mathbf{b} \right\rangle - \left\langle A_{(j_2)}, \mathbf{b}_{(j_1)} \right\rangle) A_{(j_2)} \\ &= \mathbf{b} - \mathbf{b}_{(j_1)} - \mathbf{b}_{(j_2)} + \mathbf{c}_2 + \mathbf{c}_1.\end{aligned}$$

Then, it follows

$$\mathbf{b}^{(2)} = \mathbf{b}_{(j_1)} + \mathbf{b}_{(j_2)} - \mathbf{c}_2 - \mathbf{c}_1.$$

By a similar line of reasoning we have

$$\mathbf{b}^{(3)} = \mathbf{b}_{(j_1)} + \mathbf{b}_{(j_2)} + \mathbf{b}_{(j_3)} - \mathbf{c}_1 - \mathbf{c}_2 - \mathbf{c}_3.$$

Now we generalize this as a Proposition. (☕ If you followed the preceding example then you have the tools to prove the following proposition yourself!)

Proposition 3 (Closed Form of **b** Iterates). *Let* $\mathbf{b}^{(k)}$ *be the iterates of* $\mathbf{b}$. *Let* $\mathbf{b}_{(j_k)}$ *and* $\mathbf{c}_k$ *be defined as in Example* (3.2.1). *Then the* **b** *iterates are sums of projection terms and correction terms*

$$\mathbf{b}^{(k)} = \sum_{i=1}^{k} \mathbf{b}_{(j_i)} - \mathbf{c}_i.$$

Proof. We prove this by induction on k. Above we have shown the base case for $k = 1$. By our inductive hypothesis we can write

$$\mathbf{b}^{(k)} = \sum_{i=1}^{k} \mathbf{b}_{(j_i)} - \mathbf{c}_i.$$

Recalling our definition for $\mathbf{b}^{(k)}$ this is equivalent to

$$\mathbf{z}^{(k)} = \mathbf{b} - \sum_{i=1}^{k} \mathbf{b}_{(j_i)} - \mathbf{c}_i.$$

Now consider $\mathbf{b}^{(k+1)}$. By definition, we have $\mathbf{b}^{(k+1)} = \mathbf{b} - \mathbf{z}^{(k+1)}$ where

$$\mathbf{z}^{(k+1)} = \mathbf{z}^{(k)} - \alpha_{k+1} \left\langle A_{(j_{k+1})}, \mathbf{z}^{(k)} \right\rangle A_{(j_{k+1})}.$$

By our inductive hypothesis we have

$$\mathbf{z}^{(k+1)} = \mathbf{b} - \sum_{i=1}^{k} (\mathbf{b}_{(j_i)} - \mathbf{c}_i) - \alpha_{k+1} \left\langle A_{(j_{k+1})}, \mathbf{b} - \mathbf{b}^{(k)} \right\rangle A_{(j_{k+1})}.$$

Using linearity of real inner product we have

$$\begin{aligned}
\mathbf{z}^{(k+1)} &= -\alpha_{k+1} \left(\left\langle A_{(j_{k+1})}, \mathbf{b} \right\rangle - \left\langle A_{(j_{k+1})}, \mathbf{b}^{(k)} \right\rangle \right) A_{(j_{k+1})} + \mathbf{b} - \sum_{i=1}^{k} \mathbf{b}_{(j_i)} - \mathbf{c}_i \\
&= -b_{j_k} + c_k + \mathbf{b} - \sum_{i=1}^{k} \mathbf{b}_{(j_i)} - \mathbf{c}_i \\
&= \mathbf{b} - \sum_{i=1}^{k+1} \mathbf{b}_{(j_i)} - \mathbf{c}_i .
\end{aligned}$$

Then, by definition, it follows

$$\mathbf{b}^{(k+1)} = \sum_{i=1}^{k+1} \mathbf{b}_{(j_i)} - \mathbf{c}_i .$$

□

We have a closed form for the **b** iterates. These vectors are significant because they are input into the **x** iterate updates. We can interpret the above result as follows. Each **b** iterate "adds in" a projection term $\mathbf{b}_{(j_k)}$. However, when we "add in" the j_kth column we need to subtract the portion of $\mathbf{b}^{(k-1)}$ which lies in the j_kth column, so that it is not "double counted" in $\mathbf{b}^{(k)}$. This rings a similar flavor as inclusion-exclusion.

3.3 Theoretical Foundations of REK

Similar to other Kaczmarz methods, REK is guaranteed to converge exponentially in expectation. The first proof of convergence was presented by Zouzias and Freris (2013) and improved on by Du (2018). Du claims to provide a tighter bound on the convergence of REK and uses a different strategy. We discuss both proofs here because understanding each perspective may give us more diverse tools for laying the foundations of WREK.

Whilst most ideas presented here can be found in Zouzias and Freris (2013) and in Du (2018) we present these ideas with significant exposition towards understanding of the author's methods and their significance. We also present the author's ideas in allusion to extension for WREK.

3.3.1 Zouzias et al.'s Proof of Convergence

REK modifies its underlying system and one may ask what resemblance the modified system holds to the original system. The following proposition means that the modified system holds a special significance to the original system.

Proposition 4 (Fundamental Proposition of REK). *Let A be any non-zero real $m \times n$ matrix and let $\mathbf{b} \in \mathbb{R}^m$. Let $\mathbf{x}^*$ be the minimum norm solution to the least squares problem defined by A and $\mathbf{b}$. Then $\mathbf{x}^*$ is an exact solution to $A\mathbf{x} = \mathbf{b}_{col(A)}$.*

Proof. By definition $\mathbf{x}^*$ is the vector which optimizes $\min_{\mathbf{x}} ||A\mathbf{x} - \mathbf{b}||_2^2$. We can rewrite this problem as $\min_{\mathbf{z}} ||\mathbf{z} - \mathbf{b}||_2^2$ such that $\mathbf{z} = A\mathbf{x}$. Notice $\mathbf{z}$ is restricted to lie in the column space of A. So, by definition, the solution to this new optimization problem is the vector which is in the column space of A and is closest to $\mathbf{b}$. By definition this solution is $\mathbf{z} = \mathbf{b}_{col(A)}$ the projection of $\mathbf{b}$ onto the column space of A. However, we also have $\mathbf{z} = A\mathbf{x}$. So $\mathbf{x}_{LS}$ which solves the least squares problem also solves $A\mathbf{x} = \mathbf{b}_{col(A)}$. Since $A\mathbf{x} = \mathbf{b}_{col(A)}$ has an exact solution the least squares solution to this system is the same as its exact solution. Thus if $\mathbf{x}^*$ is the least squares solution to $A\mathbf{x} = \mathbf{b}$ then $\mathbf{x}_{LS}$ is also the solution to $A\mathbf{x} = \mathbf{b}_{col(A)}$. □

This significance of this proposition is that REK can modify the given linear system so that $\mathbf{b} \to \mathbf{b}_{col(A)}$ and RK on this new system will result in the least squares solution of the original system.

Theorem 3.3.1 (Convergence of $\mathbf{z}$). *Let $\mathbf{z}^{(k)}$ denote the kth iterate of RK applied to the system $A^T\mathbf{z} = \mathbf{0}$. In exact arithmetic, it holds*

$$\mathbb{E}\left\|\mathbf{z}^{(k)} - \mathbf{b}_{col(A)^\perp}\right\|_2^2 \leq \left(1 - 1/\kappa_F^2(A)\right)^k \left\|\mathbf{b}_{col(A)}\right\|_2^2.$$

Now, we present the rate of convergence of REK as presented by Zouzias et al..

Theorem 3.3.2 (Zouzias et al.. REK Convergence). *After $T > 1$ iterations, in exact arithmetic, REK with input matrix A (possibly rank deficient) and $\mathbf{b}$ computes a vector $\mathbf{x}^{(T)}$ such that*

$$\mathbb{E}\left\|\mathbf{x}^{(T)} - \mathbf{x}^*\right\|_2^2 \leq \left(1 - \frac{1}{\kappa_F^2(A)}\right)^{\lfloor T/2 \rfloor} (1 + 2\kappa^2(A)) \left\|\mathbf{x}^*\right\|_2^2.$$

We will now go through Zouzias et al.'s proof. We will take the steps of the proof slower than Zouzias et al. and we will add additional explanation. At the very least this reiteration of REK's convergence will provide some expository value, and we will indicate where we might find that this proof may be useful to show theoretical foundations for WREK.

Proof. Let $\alpha = 1 - 1/\kappa_F^2(A)$ and let $\mathbb{E}_k$ be the expected value conditioned on the first k iterations of REK. The idea of this proof is to bound the convergence by considering each half of the T steps separately.

Note that the steps of modifying $\mathbf{z}$ are independent of the steps modifying $\mathbf{x}$ and so Theorem (3.3.1) implies that for every $l \geq 0$

$$\mathbb{E}\left\|\mathbf{z}^{(l)} - \mathbf{b}_{Col(A)^\perp}\right\|_2^2 \leq \alpha^l \left\|\mathbf{b}_{col(A)}\right\|_2^2 \leq \left\|\mathbf{b}_{col(A)}\right\|_2^2 \tag{3.1}$$

This inequality will be used to bound the horizon of RK with respect to $\mathbf{b}_{col(A)}$. Fix a parameter $k^* = \left\lfloor \frac{T}{2} \right\rfloor$. After the k^*th iteration of REK it follows from Theorem (2.1.2) that

$$\mathbb{E}_{(k^*-1)}\left\|\mathbf{x}^{(k^*)} - \mathbf{x}^*\right\|_2^2 \leq \alpha \left\|\mathbf{x}^{(k^*-1)} - \mathbf{x}^*\right\| + \frac{\left\|\mathbf{b}_{col(A)^\perp} - \mathbf{z}^{(k^*-1)}\right\|_2^2}{\|A\|_F^2}.$$

This can be seen because we can decompose $\mathbf{b} = \mathbf{b}_{col(A)} + \mathbf{b}_{col(A)^\perp}$ and note that $A\mathbf{x} = \mathbf{b}_{Col(A)}$ has an exact solution since $\mathbf{b}_{col(A)} \in col(A)$.

Averaging the inequality over the random variables $i_1, j_1, \cdots, i_{k^*-1}, j_{k^*-1}$ and using the linearity of expectation it follows that

$$\begin{aligned}
\mathbb{E}\left\|\mathbf{x}^{(k^*)} - \mathbf{x}^*\right\|_2^2 &\leq \alpha\mathbb{E}\left\|\mathbf{x}^{(k^*-1)} - \mathbf{x}^*\right\|_2^2 + \frac{\mathbb{E}\left\|\mathbf{b}_{col(A)^\perp} - \mathbf{z}^{(k^*-1)}\right\|_2^2}{\|A\|_F^2} \\
&\leq \alpha\mathbb{E}\left\|\mathbf{x}^{(k^*-1)} - \mathbf{x}^*\right\|_2^2 + \frac{\left\|\mathbf{b}_{col(A)}\right\|_2^2}{\|A\|_F^2} && \text{Using inequality (3.1)} \\
&\leq \alpha\left(\alpha\mathbb{E}\left\|\mathbf{x}^{(k^*-2)} - \mathbf{x}^*\right\|_2^2 + \frac{\left\|\mathbf{b}_{col(A)}\right\|_2^2}{\|A\|_2^2}\right) + \frac{\left\|\mathbf{b}_{col(A)}\right\|_2^2}{\|A\|_F^2} && \text{Using the same reasoning as above} \\
&\leq \cdots \leq \alpha^{k^*}\left\|\mathbf{x}^{(0)} - \mathbf{x}^*\right\|_2^2 + \sum_{l=0}^{k^*-1} \alpha^l \frac{\left\|\mathbf{b}_{col(A)}\right\|_2^2}{\|A\|_F^2} && \text{Iterating these steps a total of } k^* \text{ times.}
\end{aligned} \tag{3.2}$$

$$\begin{aligned}
&\leq \alpha^{k^*}\left\|\mathbf{x}^{(0)} - \mathbf{x}^*\right\|_2^2 + \sum_{l=0}^{\infty} \alpha^l \frac{\left\|\mathbf{b}_{col(A)}\right\|_2^2}{\|A\|_F^2} && \text{Since } \alpha \leq 1, \\
&\leq \alpha^{k^*}\left\|\mathbf{x}^{(0)} - \mathbf{x}^*\right\|_2^2 + \frac{1}{1-\alpha}\frac{\left\|\mathbf{b}_{col(A)}\right\|_2^2}{\|A\|_F^2} && \text{By the infinite geometric series.} \\
&\leq \alpha^{k^*}\left\|\mathbf{x}^{(0)} - \mathbf{x}^*\right\|_2^2 + \kappa(A)_F^2\frac{\left\|\mathbf{b}_{col(A)}\right\|_2^2}{\|A\|_F^2} && \text{By definition of } \alpha.
\end{aligned}$$

$$\leq \alpha^{k^*} \left\| \mathbf{x}^{(0)} - \mathbf{x}^* \right\|_2^2 + \|A\|_F^2 \left\|A^\dagger\right\|_2^2 \frac{\left\|\mathbf{b}_{col(A)}\right\|_2^2}{\|A\|_F^2} \qquad \text{By definition of } \kappa(A)_F^2$$

$$\leq \alpha^{k^*} \left\| \mathbf{x}^{(0)} - \mathbf{x}^* \right\|_2^2 + \frac{\left\|\mathbf{b}_{col(A)}\right\|_2^2}{\left\|A^\dagger\right\|_2^2}$$

$$\leq \alpha^{k^*} \left\| \mathbf{x}^{(0)} - \mathbf{x}^* \right\|_2^2 + \frac{\left\|\mathbf{b}_{col(A)}\right\|_2^2}{\sigma_{min}^2(A)}$$

$$\mathbb{E}\left\| \mathbf{x}^{(k^*)} - \mathbf{x}^* \right\|_2^2 \leq \alpha^{k^*} \|\mathbf{x}^*\|_2^2 + \frac{\left\|\mathbf{b}_{col(A)}\right\|_2^2}{\sigma_{min}^2(A)} \qquad \text{Since } \mathbf{x}^{(0)} = \mathbf{0}. \tag{3.3}$$

Inequality (3.3) gives us a bound on the first half of the iteration. We will use a similar line of reasoning to find the desired bound on the total number of iteration steps. Similar to inequality (3.1), notice that for any $l \geq 0$ we have

$$\mathbb{E}\left\| \mathbf{b}_{col(A)^\perp} - \mathbf{z}^{(l+k^*)} \right\|_2^2 \leq \alpha^{l+k^*} \left\|\mathbf{b}_{col(A)}\right\|_2^2 \leq \alpha^{k^*} \left\|\mathbf{b}_{col(A)}\right\|_2^2 \tag{3.4}$$

Let $0 < k < \lfloor T/2 \rfloor$, using similar reasoning to above we have

$$\mathbb{E}\left\| \mathbf{x}^{(k+k^*)} - \mathbf{x}^* \right\|_2^2 \leq \alpha \mathbb{E}\left\| \mathbf{x}^{(k+k^*-1)} - \mathbf{x}^* \right\|_2^2 + \frac{\mathbb{E}\left\|\mathbf{b}_{col(A)^\perp} - \mathbf{z}^{k+k^*-1}\right\|_2^2}{\|A\|_F^2}$$

$$\leq \alpha \mathbb{E}\left\| \mathbf{x}^{(k+k^*-1)} - \mathbf{x}^* \right\|_2^2 + \frac{\alpha^{k^*}\left\|\mathbf{b}_{col(A)}\right\|_2^2}{\|A\|_F^2} \qquad \text{By inequality (3.4)}$$

$$\leq \alpha \left(\mathbb{E}\left\| \mathbf{x}^{(k+k^*-2)} - \mathbf{x}^* \right\|_2^2 + \frac{\alpha^{k^*}\left\|\mathbf{b}_{col(A)}\right\|_2^2}{\|A\|_F^2} \right) + \frac{\alpha^{k^*}\left\|\mathbf{b}_{col(A)}\right\|_2^2}{\|A\|_F^2}$$

$$\leq \cdots \leq \alpha^k \mathbb{E}\left\| \mathbf{x}^{(k^*)} - \mathbf{x}^* \right\|_2^2 + \sum_{l=0}^{k-1} \alpha^l \frac{\alpha^{k^*}\left\|\mathbf{b}_{col(A)}\right\|_2^2}{\|A\|_F^2} \tag{3.5}$$

$$\leq \alpha^k \mathbb{E}\left\| \mathbf{x}^{(k^*)} - \mathbf{x}^* \right\|_2^2 + \sum_{l=0}^{\infty} \alpha^l \frac{\alpha^{k^*}\left\|\mathbf{b}_{col(A)}\right\|_2^2}{\|A\|_F^2} \qquad \text{Since } \alpha < 1$$

$$\leq \alpha^k \mathbb{E}\left\| \mathbf{x}^{(k^*)} - \mathbf{x}^* \right\|_2^2 + \sum_{l=0}^{\infty} \alpha^l \frac{\alpha^{k^*}\left\|\mathbf{b}_{col(A)}\right\|_2^2}{\|A\|_F^2}$$

$$= \alpha^k \mathbb{E}\left\| \mathbf{x}^{(k^*)} - \mathbf{x}^* \right\|_2^2 + \frac{1}{1-\alpha} \frac{\alpha^{k^*}\left\|\mathbf{b}_{col(A)}\right\|_2^2}{\|A\|_F^2} \qquad \text{By the infinite geometric series} \tag{3.6}$$

$$
\begin{aligned}
&= \alpha^{k}\mathbb{E}\left\|\mathbf{x}^{(k^*)} - \mathbf{x}^*\right\|_2^2 + \alpha^{k^*}\frac{\left\|\mathbf{b}_{col(A)}\right\|_2^2}{\sigma_{min}^2(A)} \\
&\leq \alpha^{k}\left(\alpha^{k^*}\left\|\mathbf{x}^*\right\|_2^2 + \frac{\left\|\mathbf{b}_{col(A)}\right\|_2^2}{\sigma_{min(A)}^2}\right) + \alpha^{k^*}\frac{\left\|\mathbf{b}_{col(A)}\right\|_2^2}{\sigma_{min}^2(A)} && \text{By inequality (3.3)} \\
&= \alpha^{k+k^*}\left\|\mathbf{x}^*\right\|_2^2 + (\alpha^{k^*} + \alpha^{k})\frac{\left\|\mathbf{b}_{col(A)}\right\|_2^2}{\sigma_{min}^2(A)} \\
&\leq \alpha^{k+k^*}\left\|\mathbf{x}^*\right\|_2^2 + (\alpha^{k^*} + \alpha^{k})\kappa^2(A)\left\|\mathbf{x}^*\right\|_2^2 && \text{Since } \left\|\mathbf{b}_{col(A)}\right\|_2^2 \leq \sigma_{max}^2(A)\left\|\mathbf{x}^*\right\|_2^2 \\
&\leq \alpha^{k+k^*}\left\|\mathbf{x}^*\right\|_2^2 + (2\alpha^{k^*})\kappa^2(A)\left\|\mathbf{x}^*\right\|_2^2 \\
&\leq \alpha^{k^*}\left\|\mathbf{x}^*\right\|_2^2 + (2\alpha^{k^*})\kappa^2(A)\left\|\mathbf{x}^*\right\|_2^2 .
\end{aligned}
\tag{3.7}
$$

So, in summary, we have

$$
\mathbb{E}\left\|\mathbf{x}^{(k+k^*)} - \mathbf{x}^*\right\|_2^2 \leq \alpha^{k^*}(1 + 2\kappa^2(A))\left\|\mathbf{x}^*\right\|_2^2 .
$$

□

There's a lot happening in this proof. One may ask why we performed a series of manipulations on the first k^* iterations and then perform almost the same iterations on the second k iterations. First, notice that the action of REK is concurrent RK iterates with iterative system modifications to reduce the horizon of convergence. And so intuitively, we have this strange two part convergence proof because the horizon of convergence does not converge as fast as the Randomized Kaczmarz iterates. In particular the Randomized Kaczmarz steps in REK converge exponentially, seen in the $\alpha^{k^*}\left\|\mathbf{x}^{(k^*-1)} - \mathbf{x}^*\right\|_2^2$ terms, whereas horizon terms end up decaying geometrically, as seen in the $\sum_{l=0}^{k^*-1}\alpha^{l}\frac{\|b_{col(A)}\|_2^2}{\|A\|_F^2}$ terms. As a result, we split up the convergence proof into two parts because intuitively "by the time we reach the second k iterates, the horizon will have converged sufficiently, for the overall convergence rate we are trying to show".

This proof by Zouzias and Freris (2013) and seems difficult to generalize. Since the introduction of REK there has emerged a new proof of convergence by Du (2018). As mentioned previously, Du claims to improve on the convergence presented by Zouzias et al.. and takes a new strategy to do so. One question we want to answer in this project is how different are these proofs? If they are fundamentally different then this may give us more tools and insight to generalize the theoretical results of REK to WREK. We will now unpack Du's proof of convergence.

3.3.2 Du's Proof of Convergence

First we give Du's rate of convergence with the notation modified to match the notation found in this document.

Theorem 3.3.3 (Du REK Convergence)**.** *After k iterations with $\mathbf{x}^{(0)} \in col(A^T)$ and $\mathbf{z}^{(0)} \in \mathbf{b} + col(A)$, in exact arithmetic it holds*

$$\mathbb{E}\left\|\mathbf{x}^{(k)} - \mathbf{x}^*\right\|_2^2 \leq \alpha^k \left\|\mathbf{x}^{(0)} - \mathbf{x}^*\right\|_2^2 + \frac{\alpha^k(1-\alpha^k)}{\sigma_{min}^2(A)} \left\|\mathbf{b}_{col(A)^\perp} - \mathbf{z}^{(0)}\right\|_2^2.$$

Du claims to improve Zouzias et al.'s bound. Since there is a form difference between Du's bound and Zouzias et al.'s bound it is worth exploring the difference in the bounds so far. Recall Zouzias et al.'s bound, after k iterations we have

$$\mathbb{E}\left\|\mathbf{x}^{(k)} - \mathbf{x}^*\right\|_2^2 \leq \alpha^{\lfloor k/2 \rfloor}(1 + 2\kappa^2(A)) \left\|\mathbf{x}^*\right\|_2^2.$$

In terms of the number of iterations we see that Zouzias et al.'s bound decays like $\alpha^{\lfloor k/2 \rfloor}$. And so it looks like Du's speed upgrade converges much faster than Zouzias et. al.'s time (α^k vs $\alpha^{k/2}$). One may argue that this speed upgrade is an artifact of Zouzias et al..'s approach of splitting up the first $k/2$ iterations from the second $k/2$ iterations, indeed we explore this more in section (3.3.4).

3.3.3 Du's Proof of REK Convergence

Now we will unpack the proof of Du's REK convergence rate. This proof is notable because Du's approach is fundamentally different from Zouzias' and so unpacking this proof may give us more tools for laying the theoretical foundations for WREK.

Before going into the details of Du's proof, we will provide some geometric intuition for the strategy of the proof. Du puts a bound on the convergence of REK by decomposing the error as follows.

Proposition 5 (Du's Fundamental Triangle)**.** *Given the $x^{(k)}$th iterate of REK we have*

$$\left\|\mathbf{x}^{(k)} - \mathbf{x}^*\right\|_2^2 = \left\|\mathbf{x}^{(k)} - \hat{\mathbf{x}}^{(k)}\right\|_2^2 + \left\|\hat{\mathbf{x}}^{(k)} - \mathbf{x}^*\right\|_2^2.$$

Where $\hat{\mathbf{x}}^{(k)}$ is a one step RK update on the system $A\mathbf{x} = \mathbf{b}_{col(A)}$ starting from $\mathbf{x}^{(k-1)}$. This geometry is displayed in Figure (3.2)*.*

The approach of Du is to first prove that this triangle holds on each step of REK. Then he bounds the error $\left\|\mathbf{x}^{(k)} - \mathbf{x}^*\right\|_2^2$ by bounding each of the terms of the right hand side.

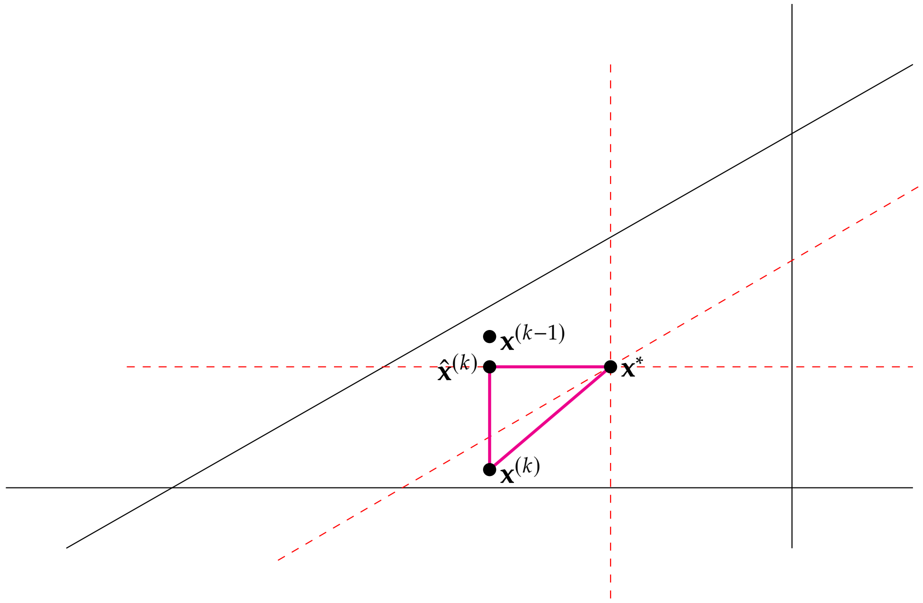


Figure 3.2 A geometric depiction of Du's Fundamental Triangle. The black lines indicate the inconsistent system $A\mathbf{x} = \mathbf{b}$, indicate the convergent system $A\mathbf{x} = \mathbf{b}_{col(A)}$. The magenta triangle depicts Du's Fundamental Triangle. Here both $\mathbf{x}^{(k)}$ and $\hat{\mathbf{x}}^{(k)}$ are orthogonal projection from $\mathbf{x}^{(k-1)}$. Note that $\mathbf{x}^{(k-1)}$ and $\mathbf{x}^{(k)}$ is not necessarily on either system's hyperplanes because REK's iterates are some modified system between the original system and the convergent system.

Before we begin the formal proof, we introduce a couple of lemmas which will help us with the proof.

Lemma 3.3.1. *Let A be any nonzero real matrix. For every $\mathbf{u} \in col(A)$, it holds*

$$\mathbf{u}^T \left(I - \frac{AA^T}{\|A\|_F} \right) \mathbf{u} \le \alpha \|\mathbf{u}\|_2^2 .$$

Lemma 3.3.2 (Double Projections)**.** *Let $\mathbf{a}$ be any nonzero vector. Then*

$$\left(\frac{\mathbf{a}\mathbf{a}^T}{\|\mathbf{a}\|_2^2} \right)^2 = \frac{\mathbf{a}\mathbf{a}^T}{\|\mathbf{a}\|_2^2} \text{ and } \left(I - \frac{\mathbf{a}\mathbf{a}^T}{\|\mathbf{a}\|_2^2} \right)^2 = I - \frac{\mathbf{a}\mathbf{a}^T}{\|\mathbf{a}\|_2^2}.$$

Following Du (2018) we present this new bound on REK's convergence.

Proof. (Proposition (5) and Theorem (3.3.3)) Let

$$\hat{\mathbf{x}}^{(k)} = \mathbf{x}^{(k-1)} - \frac{(A^{(i)})^T \mathbf{x}^{(k-1)} - (\mathbf{b}_{col(A)})_i}{\|A^{(i)}\|_2^2} A^{(i)}.$$

Then it follows

$$
\begin{aligned}
\hat{\mathbf{x}}^{(k)} - \mathbf{x}^* &= \mathbf{x}^{(k-1)} - \mathbf{x}^* - \frac{(A^{(i)})^T \mathbf{x}^{(k-1)} - (\mathbf{b}_{col(A)})_i}{\left\|A^{(i)}\right\|_2^2} A^{(i)} \\
&= \mathbf{x}^{(k-1)} - \mathbf{x}^* - \frac{(A^{(i)})^T \mathbf{x}^{(k-1)} - (A^{(i)})^T \mathbf{x}^*}{\left\|A^{(i)}\right\|_2^2} A^{(i)} \\
&= \left(\mathbf{I} - \frac{A^{(i)}(A^{(i)})^T}{\left\|A^{(i)}\right\|_2^2}\right)\left(\mathbf{x}^{(k-1)} - \mathbf{x}^*\right).
\end{aligned} \tag{3.8}
$$

Moreover,

$$
\mathbf{x}^{(k)} - \hat{\mathbf{x}}^{(k)} = \frac{(\mathbf{b}_{col(A)^\perp})_i - \mathbf{z}_i^k}{\left\|A^{(i)}\right\|_2^2} A^{(i)}.
$$

We show that $\hat{\mathbf{x}}^{(k)} - \mathbf{x}^*$ and $\mathbf{x}^{(k)} - \hat{\mathbf{x}}^{(k)}$ are orthogonal. Consider,

$$
(\mathbf{x}^{(k)} - \hat{\mathbf{x}}^{(k)})^T (\hat{\mathbf{x}}^k - \mathbf{x}^*) = \left(\frac{(\mathbf{b}_{col(A)^\perp})_i - \mathbf{z}_i^{(k)}}{\left\|A^{(i)}\right\|_2^2} A^{(i)}\right)^T \left(I - \frac{A^{(i)}(A^{(i)})^T}{\left\|A^{(i)}\right\|_2^2}\right)(\mathbf{x}^{(k-1)} - \mathbf{x}^*)
$$

Let $\beta = \frac{(\mathbf{b}_{col(A)^\perp})_i - \mathbf{z}_i^{(k)}}{\left\|A^{(i)}\right\|_2^2}$, then

$$
\begin{aligned}
(\mathbf{x}^{(k)} - \hat{\mathbf{x}}^{(k)})^T (\hat{\mathbf{x}}^k - \mathbf{x}^*) &= \beta (A^{(i)})^T \left(I - \frac{A^{(i)}(A^{(i)})^T}{\left\|A^{(i)}\right\|_2^2}\right)(\mathbf{x}^{(k-1)} - \mathbf{x}^*) \\
&= \beta (A^{(i)})^T (\mathbf{x}^{(k-1)} - \mathbf{x}^*) - \beta \frac{(A^{(i)})^T A^{(i)} (A^{(i)})^T}{\left\|A^{(i)}\right\|_2^2}(\mathbf{x}^{(k-1)} - \mathbf{x}^*) \\
&= \beta (A^{(i)})^T (\mathbf{x}^{(k-1)} - \mathbf{x}^*) - \beta \frac{\left\|A^{(i)}\right\|_2^2 (A^{(i)})^T}{\left\|A^{(i)}\right\|_2^2}(\mathbf{x}^{(k-1)} - \mathbf{x}^*) \\
&= \beta (A^{(i)})^T (\mathbf{x}^{(k-1)} - \mathbf{x}^*) - \beta (A^{(i)})^T (\mathbf{x}^{(k-1)} - \mathbf{x}^*) \\
&= 0.
\end{aligned}
$$

Hence, we have

$$
\left\|\mathbf{x}^{(k)} - \mathbf{x}^*\right\|_2^2 = \left\|\mathbf{x}^{(k)} - \hat{\mathbf{x}}^{(k)}\right\|_2^2 + \left\|\hat{\mathbf{x}}^{(k)} - \mathbf{x}^*\right\|_2^2.
$$

Now, let $\mathbb{E}_{k-1}$ denote the expectation conditioned on the first $k-1$ iterations of REK. That is, conditioned on i_l for $0 \le l \le k-1$ and j_l for $0 \le l \le k-1$, the first $k-1$ rows chosen and the first $k-1$ columns chosen respectively. And let $\mathbb{E}^i_{k-1}$ denote the expectation conditioned on the first $k-1$ iterations of REK and the kth column chosen. Lastly, let $\mathbb{E}^j_{k-1}$ denote the expectation conditioned on the first $k-1$ iterations of REK and the kth row chosen. Then, by the law of total expectation we have

$$\mathbb{E}_{k-1}[\cdot] = \mathbb{E}^j_{k-1}[\mathbb{E}^i_{k-1}[\cdot]].$$

To find our final result we will show the expected norms on each term in Du's Fundamental Triangle. First consider

$$\begin{aligned}
\mathbb{E}_{k-1}\left[\left\|\mathbf{x}^{(k)} - \hat{\mathbf{x}}^{(k)}\right\|_2^2\right] &= \mathbb{E}_{k-1}\left[\frac{((\mathbf{b}_{col(A)^\perp})_i - \mathbf{z}_i^k)^2}{\|A(i)\|_2^2}\right] \\
&= \mathbb{E}^j_{k-1}\left[\mathbb{E}^i_{k-1}\left[\frac{((\mathbf{b}_{col(A)^\perp})_i - \mathbf{z}_i^k)^2}{\left\|A^{(i)}\right\|_2^2}\right]\right] && \text{By law of total expectation} \\
&= \mathbb{E}^j_{k-1}\left[\sum_{i=0}^{n}\left[\frac{((\mathbf{b}_{col(A)^\perp})_i - \mathbf{z}_i^k)^2}{\left\|A_(i)\right\|_2^2}\cdot\frac{\left\|A^{(i)}\right\|_2^2}{\|A\|_F^2}\right]\right] && \text{By def. of expectation.} \\
&= \mathbb{E}^j_{k-1}\left[\frac{\left\|(\mathbf{b}_{col(A)^\perp}) - \mathbf{z}^{(k)}\right\|_2^2}{\|A\|_F^2}\right] && \text{By def. of 2-norm.} \\
&= \frac{1}{\|A\|_F^2}\mathbb{E}_{k-1}\left[\left\|(\mathbf{b}_{col(A)^\perp}) - \mathbf{z}^{(k)}\right\|_2^2\right] \\
&\leq \frac{\alpha^k}{\|A\|_F^2}\mathbb{E}\left[\left\|\mathbf{z}^{(0)} - (\mathbf{b}_{col(A)^\perp})\right\|_2^2\right] && \text{By Theorem (3.3.1).}
\end{aligned}$$

Now consider

$$\begin{aligned}
\mathbb{E}_{k-1}\left[\left\|\hat{\mathbf{x}}^{(k)} - \mathbf{x}^*\right\|_2^2\right] &= \mathbb{E}_{k-1}\left[(\hat{\mathbf{x}}^{(k)} - \mathbf{x}^*)^T(\hat{\mathbf{x}}^{(k)} - \mathbf{x}^*)\right] && \text{By def. 2-norm.} \\
&= \mathbb{E}_{k-1}\left[(\hat{\mathbf{x}}^{(k-1)} - \mathbf{x}^*)^T\left(I - \frac{A^{(i)}(A^{(i)})^T}{\left\|A^{(i)}\right\|_2^2}\right)^2(\hat{\mathbf{x}}^{(k-1)} - \mathbf{x}^*)\right] && \text{By Equality (3.8)} \\
&= \mathbb{E}_{k-1}\left[(\hat{\mathbf{x}}^{(k-1)} - \mathbf{x}^*)^T\left(I - \frac{A^{(i)}(A^{(i)})^T}{\left\|A^{(i)}\right\|_2^2}\right)(\hat{\mathbf{x}}^{(k-1)} - \mathbf{x}^*)\right] && \text{By Lemma (3.3.2)} \\
&= \sum_{i=0}^{m}(\hat{\mathbf{x}}^{(k-1)} - \mathbf{x}^*)^T\left(I - \frac{A^{(i)}(A^{(i)})^T}{\left\|A^{(i)}\right\|_2^2}\right)(\hat{\mathbf{x}}^{(k-1)} - \mathbf{x}^*)\cdot\frac{\left\|A^{(i)}\right\|_2^2}{\|A\|_F^2}.
\end{aligned}$$

Note that we are conditioning on the first $k-1$ iterations and so the only random variable We are considering is the i chosen at the kth iteration. So by linearity of sum and matrix multiplication we have

$$\mathbb{E}_{k-1}\left[\left\|\hat{\mathbf{x}}^{(k)}-\mathbf{x}^*\right\|_2^2\right]=(\hat{\mathbf{x}}^{(k-1)}-\mathbf{x}^*)^T\sum_{i=0}^{m}\left[\left(I-\frac{A^{(i)}(A^{(i)})^T}{\left\|A^{(i)}\right\|_2^2}\right)\cdot\frac{\left\|A^{(i)}\right\|_2^2}{\|A\|_F^2}\right](\hat{\mathbf{x}}^{(k-1)}-\mathbf{x}^*),$$

I is a constant and recall that by our pmf we have $\sum_i \left\|A^{(i)}\right\|_2^2/\|A\|_F^2=1$. Thus, we have

$$\begin{aligned}
\mathbb{E}_{k-1}\left[\left\|\hat{\mathbf{x}}^{(k)}-\mathbf{x}^*\right\|_2^2\right]&=(\hat{\mathbf{x}}^{(k-1)}-\mathbf{x}^*)^T\left(I-\sum_{i=0}^{m}\left[\frac{A^{(i)}(A^{(i)})^T}{\left\|A^{(i)}\right\|_2^2}\cdot\frac{\left\|A^{(i)}\right\|_2^2}{\|A\|_F^2}\right]\right)(\hat{\mathbf{x}}^{(k-1)}-\mathbf{x}^*)\\
&=(\hat{\mathbf{x}}^{(k-1)}-\mathbf{x}^*)^T\left(I-\frac{1}{\|A\|_F^2}\sum_{i=0}^{m}\left[A^{(i)}(A^{(i)})^T\right]\right)(\hat{\mathbf{x}}^{(k-1)}-\mathbf{x}^*)\\
&=(\hat{\mathbf{x}}^{(k-1)}-\mathbf{x}^*)^T\left(I-\frac{A^TA}{\|A\|_F^2}\right)(\hat{\mathbf{x}}^{(k-1)}-\mathbf{x}^*) && \text{By def. of } A^TA\\
&\leq \alpha\left\|\mathbf{x}^{(k-1)}-\mathbf{x}^*\right\|_2^2. && \text{By lemma 3.3.1}
\end{aligned}$$

We can use Lemma (3.3.1) because $\mathbf{x}^{(k)}-\mathbf{x}^*\in\text{col}(A^T)$ for all k. By assumption, we have $\mathbf{x}^{(0)}\in\text{col}(A^T)$ and by definition $\mathbf{x}^*\in\text{col}(A^T)$. And so since vector spaces are closed under addition and scalar multiplication, we have $\mathbf{x}^{(0)}-\mathbf{x}^*\in\text{col}(A^T)$. Then $\mathbf{x}^{(k)}-\mathbf{x}^*\in\text{col}(A^T)$ follows by induction on k. Plugging both of these inequalities into Du's fundamental triangle (Proposition (5)) gives us

$$\begin{aligned}
\mathbb{E}\left[\left\|\mathbf{x}^{(k)}-\mathbf{x}^*\right\|_2^2\right]&=\mathbb{E}\left[\left\|\mathbf{x}^{(k)}-\hat{\mathbf{x}}^{(k-1)}\right\|_2^2\right]+\mathbb{E}\left[\left\|\hat{\mathbf{x}}^{(k)}-\mathbf{x}^*\right\|_2^2\right]\\
&\leq\frac{\alpha^k}{\|A\|_F^2}\left\|\mathbf{z}^{(k)}-(b_{col(A)^\perp})\right\|_2^2+\alpha\mathbb{E}\left[\left\|\mathbf{x}^{(k-1)}-\mathbf{x}^*\right\|_2^2\right]\\
&\leq\cdots\leq\frac{\alpha^k}{\|A\|_F^2}\left\|\mathbf{z}^{(k)}-(b_{col(A)^\perp})\right\|_2^2\sum_{l=0}^{k-1}\alpha^l+\alpha^k\left\|\mathbf{x}^{(k-1)}-\mathbf{x}^*\right\|_2^2.
\end{aligned}$$

The last statement is by repeating the above arguments $k-1$ more times. We then have

$$\mathbb{E}\left[\left\|\mathbf{x}^{(k)}-\mathbf{x}^*\right\|_2^2\right]\leq\frac{\alpha^k}{\|A\|_F^2}\left\|\mathbf{z}^{(k)}-(b_{col(A)^\perp})\right\|_2^2\left(\frac{1-\alpha^k}{1-\alpha}\right)+\alpha^k\left\|\mathbf{x}^{(k-1)}-\mathbf{x}^*\right\|_2^2.\quad\text{By finite geometric series.}$$

The finally, by definition of α,

$$\mathbb{E}\left[\left\|\mathbf{x}^{(k)} - \mathbf{x}^*\right\|_2^2\right] \leq \frac{\alpha^k(1-\alpha^k)}{\sigma_{min}^2(A)}\left\|\mathbf{z}^{(k)} - (b_{col(A)^\perp})\right\|_2^2 + \alpha^k\left\|\mathbf{x}^{(k-1)} - \mathbf{x}^*\right\|_2^2.$$

□

Some remarks before we move on. We note that the strategy of Du's approach has a clear geometric motivation in the form of Du's Fundamental Triangle. As we think about the foundations of WREK this will be significant because if we can just show that Du's Fundamental Triangle holds for Weighted Least Sqaures problems then we can leverage most of the machinery from this proof out of the box.

3.3.4 Finding Du's Result in Zousias' Approach

One of Du's fundamental claims is that he improves on Zousias' bound significantly. This improvement seems doubly significant when considering that Du seems to take a fundamentally different approach to Zouzias et al.. However, we will show that Du's convergence rate can be nearly obtained from Zouzias et al..'s approach with only a few modifications. Starting from inequality (3.2) consider

$$\mathbb{E}||\mathbf{x}^{(k^*)} - \mathbf{x}^*||_2^2 \leq \alpha^{k^*}||\mathbf{x}^{(0)} - \mathbf{x}^*||_2^2 + \sum_{l=0}^{k^*-1}\alpha^l\frac{||\mathbf{b}_{col(A)}||}{||A||_F^2}.$$

At this point the authors expand the upper bound on this sum by using the fact that $\alpha < 1$ and then they use the infinite geometric series. However, let us instead use the finite geometric series. We then obtain

$$\begin{aligned}
\mathbb{E}||\mathbf{x}^{(k^*)} - \mathbf{x}^*||_2^2 &\leq \alpha^{k^*}||\mathbf{x}^{(0)} - \mathbf{x}^*||_2^2 + \sum_{l=0}^{k^*-1}\alpha^l\frac{||\mathbf{b}_{col(A)}||_2^2}{||A||_F^2} \\
&\leq \alpha^{k^*}||\mathbf{x}^{(0)} - \mathbf{x}^*||_2^2 + \frac{1-\alpha^{k^*}}{1-\alpha}\cdot\frac{||\mathbf{b}_{col(A)}||}{||A||_F^2} \\
&\leq \alpha^{k^*}||\mathbf{x}^{(0)} - \mathbf{x}^*||_2^2 + \frac{1-\alpha^{k^*}}{\sigma_{min}^2(A)}||\mathbf{b}_{col(A)}||_2^2. && \text{By definition of } \alpha
\end{aligned}$$

.

Then folowing all of Zouzias et al.'s steps until inequality (3.5) we have

$$\begin{aligned}
\mathbb{E}\left\|\mathbf{x}^T - \mathbf{x}^*\right\|_2^2 &\leq \alpha^k \left\|\mathbf{x}^{(k^*)} - \mathbf{x}^*\right\|_2^2 + \frac{\alpha^{k^*}\left\|\mathbf{b}_{col(A)}\right\|_2^2}{\|A\|_F^2}\sum_{l=0}^{k-1}\alpha^l \\
&\leq \alpha^k \left\|\mathbf{x}^{(k^*)} - \mathbf{x}^*\right\|_2^2 + \frac{\alpha^{k^*}\left\|\mathbf{b}_{col(A)}\right\|_2^2}{\|A\|_F^2}\cdot\frac{1-\alpha^k}{1-\alpha} && \text{By finite geometric series.} \\
&\leq \alpha^k\left(\alpha^{k^*}\left\|\mathbf{x}^{(0)} - \mathbf{x}^*\right\|_2^2 + \frac{1-\alpha^{k^*}}{\sigma_{min}^2(A)}\left\|\mathbf{b}_{col(A)}\right\|_2^2\right) + \frac{\alpha^{k^*}(1-\alpha^k)}{\sigma_{min}^2(A)}\left\|\mathbf{b}_{col(A)}\right\|_2^2 \\
&\leq \alpha^T\left\|\mathbf{x}^{(0)} - \mathbf{x}^*\right\|_2^2 + \frac{\alpha^{k^*}(1-\alpha^k) + \alpha^k(1-\alpha^{k^*})}{\sigma_{min}^2(A)}\left\|\mathbf{b}_{col(A)}\right\|_2^2 \\
&\leq \alpha^T\left\|\mathbf{x}^{(0)} - \mathbf{x}^*\right\|_2^2 + \frac{\alpha^{k^*} + \alpha^k - 2\alpha^T}{\sigma_{min}^2(A)}\left\|\mathbf{b}_{col(A)}\right\|_2^2 \\
&\leq \alpha^T\left\|\mathbf{x}^{(0)} - \mathbf{x}^*\right\|_2^2 + \frac{2\alpha^{k^*}(1-\alpha^k)}{\sigma_{min}^2(A)}\left\|\mathbf{b}_{col(A)}\right\|_2^2.
\end{aligned}$$

Although there are some differences between this rate and Du's exact rate the above demonstrates that we can find something like Du's convergence rate in Zousias et al.'s proof. This then shows that the significant contribution of Du's approach is the new approach with a clear geometric motivation, rather than his convergence upgrade.

3.4 REK Convergence for an Arbitrary Sequence of b Vectors

Zouzias et al.. proved that REK converges when the $\mathbf{z}^{(k)}$ updates converge to $\mathbf{b}_{col(A)}$. Here we show a generalization of Zouzias et al..'s proof to demonstrate that *REK* will converge for a sequence of vectors which converge to a vector in the column space of A. As we think to theoretical foundations for WREK this proof serves as an indication that some of the REK convergence properties may carry over to more general problems.

Proposition 6 (Convergence of REK for alternative **b** projections)**.** *Let* $\mathbf{b}' \in col(A)$ *be any vector in the column space of* A*. Then let* $\mathbf{b}^{(k)} \in \mathbb{R}^m$ *be a sequence of vectors with* $\mathbf{b}^{(0)} = \mathbf{0}$ *which converge to* $\mathbf{b}'$ *at least as fast as*

$$\left\|\mathbf{b}^{(k)} - \mathbf{b}'\right\|_2^2 \leq \alpha^k \|\mathbf{b}\|_2^2.$$

after k iterations. Denote $\mathbf{x}^*$ *as the solution to* $A\mathbf{x} = \mathbf{b}'$*. We apply Randomized Kaczmarz with input* A *and* $\mathbf{b}'$*. After* $T > 1$ *iterations this process computes a vector* $\mathbf{x}^{(T)}$ *such that*

$$\mathbb{E}\left\|\mathbf{x}^{(T)} - \mathbf{x}^*\right\|_2^2 \leq \alpha^{\lfloor T/2\rfloor}(1 + 2\kappa^2(A))\|\mathbf{x}^*\|.$$

Proof. Let $\alpha = 1 - 1/\kappa_F^2(A)$. Also note that, by assumption, for every $l \geq 0$ the following inequalities hold

$$\mathbb{E}||\mathbf{b}^{(l)} - \mathbf{b}'||_2^2 \leq \alpha^l ||\mathbf{b}'||_2^2 \leq ||\mathbf{b}||_2^2 \tag{3.9}$$

Fix a parameter $k^* = \lfloor T/2 \rfloor$. After k^* iterations it follows from Theorem (2.1.2) that

$$\begin{aligned}
\mathbb{E}||\mathbf{x}^{(k^*)} - \mathbf{x}^*||_2^2 &\leq \alpha ||\mathbf{x}^{(k^*-1)} - x^*||_2^2 + \frac{\mathbb{E}||\mathbf{b}^{(k^*-1)} - \mathbf{b}'||_2^2}{||A||_F^2} \\
&\leq \alpha \mathbb{E}||\mathbf{x}^{(k^*-1)} - \mathbf{x}^*||_2^2 + \frac{||\mathbf{b}'||_2^2}{||A||_F^2}. && \text{By Inequality (3.9)}
\end{aligned}$$

Then by another application of Theorem 7 and Inequality (3.9) we have

$$\begin{aligned}
\mathbb{E}||\mathbf{x}^{(k^*)} - \mathbf{x}^*||_2^2 &\leq \alpha \left(\alpha \mathbb{E}||\mathbf{x}^{(k^*-2)} - \mathbf{x}^*||_2^2 + \frac{||\mathbf{b}'||_2^2}{||A||_F^2} \right) + \frac{||\mathbf{b}'||_2^2}{||A||_F^2} \\
&\leq \alpha^2 \mathbb{E}||\mathbf{x}^{(k^*-2)} - \mathbf{x}^*||_2^2 + (\alpha + \alpha^0) \frac{||\mathbf{b}'||_2^2}{||A||_F^2} \\
&\leq \alpha^2 \left(\alpha \mathbb{E}||\mathbf{x}^{(k^*-3)} - \mathbf{x}^*|| + \frac{||\mathbf{b}'||_2^2}{||A||_F^2} \right) + (\alpha + \alpha^0) \frac{||\mathbf{b}'||_2^2}{||A||_F^2} \\
&\leq \cdots \leq \alpha^k \mathbb{E}||\mathbf{x}^{(0)} - \mathbf{x}^*||_2^2 + \sum_{l=0}^{k^*-1} \alpha^l \frac{||\mathbf{b}'||_2^2}{||A||_F^2} \\
&\leq \alpha^{k^*} \mathbb{E}||\mathbf{x}^{(0)} - \mathbf{x}^*||_2^2 + \sum_{l=0}^{\infty} \alpha^l \frac{||\mathbf{b}'||_2^2}{||A||_F^2} && \text{Since norms non-negative and } \alpha > 0 \\
&\leq \alpha^{k^*} \mathbb{E}||\mathbf{x}^*||_2^2 + \sum_{l=0}^{\infty} \alpha^l \frac{||\mathbf{b}'||_2^2}{||A||_F^2} && \text{Since } \mathbf{x}^{(0)} = 0 \\
&\leq \mathbb{E}||\mathbf{x}^*||_2^2 + \sum_{l=0}^{\infty} \alpha^l \frac{||\mathbf{b}'||_2^2}{||A||_F^2}. && \text{Since } \alpha = 1/(1 - \kappa_F^2(A)) < 1.
\end{aligned}$$

Recall that by the geometric series we have $\sum_{l=0}^{\infty} \alpha^l = 1/(1-\alpha) = \kappa_F^2(A)$. Also recall $\kappa_F^2(A) = ||A||_F^2 ||A^\dagger||_2^2$. We thus have

$$\mathbb{E}||\mathbf{x}^{(k^*)} - \mathbf{x}^*||_2^2 \leq ||\mathbf{x}^*||_2^2 + \frac{||\mathbf{b}'||_2^2}{\sigma_{min}^2}, \tag{3.10}$$

since $||A^\dagger||_F^2 = 1/\sigma_{min}^2$.

Now we reapply a similar line of reasoning to find the convergence rate stated. Similar to Inequality (3.9), for every $l \geq 0$ we have the following inequality,

$$\mathbb{E}||\mathbf{b}^{(k)} - \mathbf{b}'||_2^2 \leq \alpha^{k^*+l}||\mathbf{b}'||_2^2 \leq \alpha^{k^*}||\mathbf{b}'||_2^2. \tag{3.11}$$

Then, for any $0 < k < \lfloor T/2 \rfloor$, we have the following

$$\begin{aligned}
\mathbb{E}||\mathbf{x}^{(k+k^*)} - \mathbf{x}^*||_2^2 &\leq \alpha||\mathbf{x}^{(k+k^*-1)} - \mathbf{x}^*||_2^2 + \frac{||\mathbf{b}^{(k+k^*-1)} - \mathbf{b}'||_2^2}{||A||_F^2} \\
&\leq \alpha\mathbb{E}||\mathbf{x}^{(k+k^*-1)} - \mathbf{x}^*||_2^2 + \alpha^{(k^*)}\frac{||\mathbf{b}'||_2^2}{||A||_F^2} && \text{By Inequality (3.11)} \\
&\leq \alpha\left(\alpha\mathbb{E}||\mathbf{x}^{(k+k^*-2)} - \mathbf{x}^*||_2^2 + \alpha^{(k^*)}\frac{||\mathbf{b}'||_2^2}{||A||_F^2}\right) + \frac{||\mathbf{b}'||_2^2}{||A||_F^2} \\
&\leq \cdots \leq \alpha^k\mathbb{E}||\mathbf{x}^{(k^*)} - \mathbf{x}^*||_2^2 + \sum_{l=0}^{k-1}\alpha^{(k^*+l)}\frac{||\mathbf{b}'||_2^2}{||A||_F^2} \\
&= \alpha^k\mathbb{E}||\mathbf{x}^{(k^*)} - \mathbf{x}^*||_2^2 + \frac{\alpha^{k^*}||\mathbf{b}'||_2^2}{||A||_F^2}\sum_{l=0}^{k-1}\alpha^l \\
&\leq \alpha^k\mathbb{E}||\mathbf{x}^{(k^*)} - \mathbf{x}^*||_2^2 + \frac{\alpha^{k^*}||\mathbf{b}'||_2^2}{||A||_F^2}\sum_{l=0}^{\infty}\alpha^l && \text{Since } \alpha > 0 \\
&\leq \alpha^k\mathbb{E}||\mathbf{x}^{(k^*)} - \mathbf{x}^*||_2^2 + \frac{\alpha^{k^*}||\mathbf{b}'||_2^2}{\sigma_{min}^2} && \text{By infinite geometric series} \\
&\leq \alpha^k(||\mathbf{x}^*||_2^2 + ||\mathbf{b}'||_2^2/\sigma_{min}^2) + \frac{\alpha^{k^*}||\mathbf{b}'||_2^2}{\sigma_{min}^2} && \text{By Inequality (3.10)} \\
&= \alpha^k||\mathbf{x}^*||_2^2 + \frac{a^k + a^{k^*}}{\sigma_{min}^2}||\mathbf{b}'||_2^2
\end{aligned}$$

Since $\mathbf{b}' = A\mathbf{x}^*$. We have $||\mathbf{b}'||_2^2 = ||A\mathbf{x}^*||_2^2 \leq ||A||_2^2||\mathbf{x}^*||_2^2 = \sigma_{max}||\mathbf{x}^*||_2^2$, by the Cauchy-Schwarz inequality, hence

$$
\begin{aligned}
\mathbb{E}||\mathbf{x}^{(k+k^*)-\mathbf{x}^*}||_2^2 &\leq \alpha^k||\mathbf{x}^*||_2^2 + \frac{(a^k + a^{k^*})\sigma_{max}^2}{\sigma_{min}^2}||\mathbf{x}^*||_2^2 \\
&\leq \alpha^k||\mathbf{x}^*||_2^2 + (\alpha^k + \alpha^{k^*})\kappa^2(A)||\mathbf{x}^*||_2^2 && \text{By definition of } \kappa^2(A) \\
&\leq \alpha^k||\mathbf{x}^*||_2^2 + 2\alpha^{k^*}\kappa^2(A)||\mathbf{x}||_2^2 && \text{By definition of } k \text{ and } k^* \\
&\leq \alpha^{k^*}||\mathbf{x}^*||_2^2 + 2\alpha^{k^*}\kappa^2(A)||\mathbf{x}||_2^2 \\
&= \alpha^{k^*}(1 + 2\kappa^2(A))||\mathbf{x}^*||_2^2.
\end{aligned}
$$

And by definition of α, we are done. □

Note that there are no obvious ways of generating sequences of **b** vectors which converge to some place in the column space of A, apart from the one presented by Zouzias and Freris (2013). As a result this is more of a proof of concept that a more general version of REK should converge to some solution.

In this chapter, we have studied how and why REK works. Recall that REK solves for the least squares solution of a given linear system. Such solutions are desired for noisy linear systems, however, as discussed in Section (1.4), least squares solutions are not desired for corrupted linear systems. That is to say REK will not work well for corrupted linear systems. Nevertheless, our discussion in this chapter, and specifically Proposition (6), demonstrates that the there is potential to take REK's idea of modifying the given linear system to find other meaningful solutions. Generalizing REK to solve corrupted linear systems is our motivation going into the following chapter about WREK.

Chapter 4

Weighted Randomized Extended Kaczmarz

4.1 Extending Randomized Extended Kaczmarz

REK is a fantastic method for solving solving large noisy linear systems quickly. However, as discussed at the end of Chapter 3, the least squares solution is not desirable for corrupted linerar systems, and so REK is not suited for solving corrupted linear systems. Generalizing REK for corrupted linear systems is our motivation going into this chapter.

Apart from its ability to solve noisy linear systems quickly, Randomized Extended Kaczmarz is interesting because, unlike other variants of RK, REK modifies the underlying system as it iterates and the iterates on this new system give a meaningful answer to our original system. In particular, we can "orthogonally project our given system down into its consistent subspace" and solve that new system to solve the least squares problem for our original system. As we begin to think about corrupted linear systems, a motivating question moving forward is

are there other ways to modify a given linear system so that the solution to the modified system tells us something meaningful about the original linear system?

Let us ask this question more precisely. As we saw in Section (1.4.1), given an inconsistent linear system there are many ways to find an approximate solution. Can we find other RK variants similar to REK which solve for these other approximate solutions?

4.1.1 REK and Weighted Least Squares Problems

As we saw in Section (1.4.2), given an inconsistent linear system and matrix of weights W we can find the weighted least squares solution. Following our motivating question at the top of this Section, we ask "is there a way to modify A and $\mathbf{b}$ to solve the weighted least squares problem?"

Lemma 4.1.1. (*REK solves WLS problems*) *Let $A \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$ define a linear system. Then let A, $\mathbf{b}$, and $W \in \mathbb{R}^{m\times m}$ define a Weighted Least Squares problem. Then REK with the inputs WA and $W\mathbf{b}$ solves the given Weighted Least Squares problem.* [1]

Proof. The OLS solution of the system defined by WA and $W\mathbf{b}$ is equivalent to the WLS solution to the problem defined by A, $\mathbf{b}$, and W. Recall that in the WLS problem we are searching for a vector $\mathbf{x}^* \in \mathbb{R}^n$ defined by

$$\mathbf{x}^* = \arg\min_{\mathbf{x}\in\mathbb{R}^n} \|W\mathbf{b} - WA\mathbf{x}\|_2^2 .$$

Moreover, recall that the least squares problem for a system $A' \in \mathbb{R}^{m\times n}$ and $\mathbf{b}' \in \mathbb{R}^m$ solves for a vector $(\mathbf{x}^*)' \in \mathbb{R}^n$ defined by

$$(\mathbf{x}^*)' = \arg\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{b}' - A'\mathbf{x}\|_2^2 .$$

Comparing definitions gives us that the OLS solution on WA and $W\mathbf{b}$ is the WLS solution defined by A, $\mathbf{b}$, and W.

Recall Theorem (3.3.2) gives us that REK with input WA and $W\mathbf{b}$ will converge to the least squares solution of that input system. Therefore, REK will converge to the solution of the Weighted Least Squares problem defined by A, $\mathbf{b}$, and W. □

Remark 4.1.1. Note that by the above lemma and Proposition (4) the WLS solution defined by A, $\mathbf{b}$, W is given by the exact solution to

$$WA\mathbf{x} = (W\mathbf{b})_{col(WA)}.$$

The above lemma means that, certainly, there is a meaningful way to modify a given linear system to find its weighted least solution. Namely, you map $A \mapsto WA$ and $\mathbf{b} \mapsto W\mathbf{b}$ and then iteratively modify WA and $W\mathbf{b}$ as REK does. In the end we will end up with the weighted least squares solution to the original system.

Notice that the weighted least squares problem requires a weighting matrix prior to solving the problem. However, recall that we are motivated by large linear systems where $m \gg n$. In this case it might be unreasonable to expect a known, meaningful weighting on all m equations. In particular, it can be shown that the WLS solution is a Best Linear Unbiased Estimator if we select $W_{ii} = 1/\sigma_i^2$ where σ_i^2 is the variance of equation i. Generally, estimating the variances of a number of equations is difficult (See Section (1.4.2)). Thus, for large linear systems, where we have so many equations that we cannot load the whole

[1] You could prove this yourself using ideas we have already seen earlier in the text!

matrix into working memory, estimating the variances to solve the WLS problem seems intractable. Currently, we can still solve the OLS problem with REK. However, we saw in Section (1.4.2) that the OLS solution is far from the desired solution in corrupted linear systems. Instead, the WLS solution is desirable for such systems. So we have a method which solves for the WLS solution, a solution desirable for corrupted linear systems, but the current setup of the WLS problem is not amenable to our case of large linear systems. All hope is not lost. There is another RK variant which deals with corrupted linear systems in a different way. To find and motivate new ideas, we will take a brief detour to explore *QuantileRK*.

4.2 Dynamic Reweightings

At the end of the last section we saw that in WLS problems it is difficult to assign a weighting for large linear systems. However, WLS problems seem to be good for solving corrupted linear systems. Instead of assigning a weight to each equation in the system a priori, we might instead learn the weightings as we iterate. Let us call this idea *dynamic reweightings*. Our goal is to investigate whether we can construct an RK variant which incorporates dynamic reweightings to solve corrupted linear systems.

Part of dynamic reweightings is to use *local system information* to generate RK variant iterates. We do not define a precise notion of local system information, but the idea is to use quantities which can be easily measured at each iteration. We are concerned with large linear systems and so local system information is desirable to keep the RK variant fast. If one understands the idea of *residuals* this is an example of something we might count as local information. If the reader has not heard of residuals, we will discuss this idea in the following subsection.

4.2.1 QuantileRK

In Haddock et al. (2020) the authors discuss a variation of Randomized Kaczmarz called *QuantileRK*. QuantileRK is a variant of RK for corrupted linear systems. Recall that corrupted systems are those where most of the linear equations intersect at a point, but a small number of equations have a large error.

Speaking generally, QuantileRK converges in corrupted systems in the following sense. Given a corrupted linear system QuantileRK will converge to the point where most of the equations intersect. Recall that in Section (2.1.1) we saw that RK would not converge to any point. Before we proceed we briefly define the *residuals* in linear systems. Let A and $\mathbf{b}$ defined a linear system. Let $\mathbf{x} \in \mathbb{R}^n$. Then the *residual of equation i with respect to* $\mathbf{x}$ is the quantity $\left\|\left\langle A^{(i)}, \mathbf{x}\right\rangle - b_i\right\|_2^2$. In a sense, the residual of equation i with respect to $\mathbf{x}$ is how far away $\mathbf{x}$ is from solving equation i, equivalently how far away $\mathbf{x}$ is from hyperplane i. We will give the algorithm for QuantileRK, although, for our purposes it is not as important to understand in as much detail as RK and REK.

Algorithm 3 QuantileRK

function QUANTILERK(A, $\mathbf{b}$, T, q) ▷ where $A \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, and $T \in \mathbb{N}$
  Initialize $\mathbf{x}^{(0)} \in \mathbb{R}^n$ arbitrarily
  for $k = 1, 2, 3, \ldots, T$ **do**
    Pick i_k with probability $q_i := \frac{\|A^{(i)}\|_2^2}{\|A\|_F^2}, i \in [m]$
    if $\left\|\left\langle A_{(j_k)}, \mathbf{x}^{(k)} \right\rangle \mathbf{b}_k\right\|_2^2 \leq Q_q(k)$ **then**
      Set $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \frac{b_{i_k} - \langle \mathbf{x}^{(k)}, A^{(i_k)} \rangle}{\|A^{(i_k)}\|_2^2} A^{(i_k)}$
    else
      Set $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)}$
    end if
  end for
  return $\mathbf{x}^{(T)}$
end function

Here $Q_q(k)$ is the q-th quantile of all the residuals of all equations with respect to $\mathbf{x}^{(k)}$. The main idea[2] behind QuantileRK is to use the residuals to change the probability that a hyperplane is chosen. In Chapter (2) we saw that, if we normalize our linear system, the distribution for choosing which hyperplane to project onto hyperplane is uniform. In QuantileRK the probability distribution for picking each hyperplane is uniform, however, some hyperplanes have 0 probability of being chosen if they are "too far away" in the residual sense. That is, QuantileRK uses strictly local information about the system based on the current $\mathbf{x}$ iterate to change the way that it will iterate next. This is significant because it is not necessary to know which hyperplanes are corrupted for QuantileRK to converge.

4.2.2 Using Local Information to Generate New Solutions

Here, we want to generalize this idea of using local information to solve for other approximate solutions to inconsistent systems. QuantileRK uses local information to adjust the probabilities of selecting each hyperplane at each step. In particular, if a hyperplane is "too far away" then its probability of selection is set to zero. One way that we can change the probability of selecting a row to project onto is as follows. We can use information from A, $\mathbf{b}$, and $\mathbf{x}^{(k)}$ to generate a probability distribution for row selection at each step. Alternatively, recalling the definition of the row probabilities, we can iteratively apply a reweighting to

[2]Please note this is more of a summary than our other discussions of RK and its variants. There is much more richness and technical differences to QuantileRK which are not discussed for the purposes here.

the system. The probability of choosing hyperplane i to project the RK iterate onto is given by

$$q_i := \frac{\left\|A^{(i)}\right\|_2^2}{\|A\|_F^2}.$$

And so a reweighting WA changes the norm of each $A^{(i)}$ by W_{ii} and thus changes the probabilities by a corresponding factor. So to change the probablities of row selection we can use local system information at each step to generate a reweighting matrix W_i. However, more things happen when we apply a reweighting to the system.

When we apply a reweighting to our linear system, the position of the pseudosolution changes. Recall that under the mapping $A \mapsto W_i A$ and $\mathbf{b} \mapsto W_i\mathbf{b}$ REK solves for the WLS solution defined by A, $\mathbf{b}$, and W instead of the least squares solution defined by A and $\mathbf{b}$. That is, everytime a new reweighting is applied $W_i A$ and $W_i\mathbf{b}$, REK will converge to a different vector.

To summarize, the idea is as follows. We seek an algorithm which uses local system information to generate a reweighting W_i which converges to some reweighting W such that REK applied to WA and $W\mathbf{b}$ will converge to a solution that we care about. In particular, we might want this new algorithm to converge to the desired solution in a corrupted linear system. This is an idea which looks like it has some promise, it builds off previous variants of RK of which we have a good understanding. However, there is some foundational theory to establish. In particular, how does the pseudosolution change for a sequence of reweightings W_i? Another question to ask is whether any part of REK might break if we have apply a sequence of reweightings. We have seen that REK already modifies the underlying system as we iterate. How would REK's $\mathbf{z}$ updates interact with a varying WA and $W\mathbf{b}$? We will investigate these questions in the following Sections.

4.3 Behaviour of the WLS solutions

We are lead here motivated by an RK variant with dynamic reweightings. For such an RK variant to be convergent to any solution it must be that the dynamic reweightings are convergent. Consider the following example

Example 4.3.1 (Non-convergent reweightings)**.** Let $A \in \mathbb{R}^{m\times n}$ and $\mathbf{b} \in \mathbb{R}^m$ define a linear system. Let W_n be a sequence of reweightings such that

$$(W_n)_{ii} = \begin{cases} 1 & \text{if } i = n \bmod m \\ 0 & \text{otherwise} \end{cases}.$$

In this reweighting scheme we "turn off" every row except a single row and we cycle which row is "turned on". Recall that in WLS problems we constrain $\|W_n\|_F^2 = 1$ unless stated otherwise.

This sequence of rewightings induces a sequence of pseudosolutions $(\mathbf{x}^*)_n$ which each solve the WLS problem[3] defined by A, $\mathbf{b}$, and W_n. Since W_n does not converge it follows that $(\mathbf{x}^*)_n$ also does not converge. Since our new RK variant is to solve some WLS problem defined by our sequence of reweightings, it follows that no such algorithm would converge. (☕ Can you visualize the geometry of the pseudosolutions in this reweighting scheme?)

Given the above example we see that we at least want our sequence of dynamic reweightings to converge. We want to develop an algorithm which solves some WLS problem which is derived from a sequence of reweightings. In this sense, we want our algorithm to converge to some WLS solution. To this end it would be useful to understand how the pseudosolution changes as the reweightings change. In the rest of this section we develop foundations to show how the pseudosolution moves with respect to a sequence of reweightings.

4.3.1 Single Reweightings

To begin we explore how the pseudosolution moves when the weighting on a single equation is changed. To begin we will consider what happens to the pseudosolution between two reweightings.

Example 4.3.2 (Single Reweightings)**.** Without loss of generality suppose on the first reweighting we reweight the first equation

$$(W_1)_{ii} = \begin{cases} (W_0)_{ii} & \text{if } i \neq 1 \\ (W_0)_{ii} + \delta & \text{if } i = 1 \end{cases}$$

where $(W_0)_{ii} = 1$ for all i and where $\delta \geq -(W_0)_{11}$. Also suppose that we normalize W_1 so that $\|W_1\|_F^2 = 1$. We ask, how does $(\mathbf{x}^*)_1$ change to $(\mathbf{x}^*)_2$?

First, consider a couple of edge cases. Suppose $\delta = -(W_0)_{11}$ or equivalently $(W_1)_{11} = 0$. In this case, as long as WA is still overdetermined, then the pseudosolution will be the least squares solution on the system with every equation in $A, \mathbf{b}$ without the first equation, because the remaining solutions are equally weighted. Since our systems are large we can expect that this new system will still be overdetermined and so the pseudosolution will be unique. Let $(\mathbf{x}^*)_{-\infty}$ denote the pseudosolution associated with the weighting where $(W_1)_{11} = 0$.

Suppose δ is very large. We can consider what happens as $\delta \to \infty$ which results in $(W_1)_{11} \to \infty$. This is equivalent to $(W_1)_1 = 1$ and $(W_1)_1 = 0$ for all $i \neq 1$. In this case WA is no longer over determined and so the WLS problem does not have a unique solution. That said,

[3]It is worth noting that with this set of reweightings there is not a single vector $(\mathbf{x}^*)_n$ which solves the WLS problem. As stated in Section (1.4.2) we let $(\mathbf{x}^*)_n$ denote the pseudosolution of minimum 2-norm.

the pseudosolution $(\mathbf{x}^*)_1$ will be somewhere on the hyperplane defined by the first equation. Let $(\mathbf{x}^*)_\infty$ denote a pseudosolution associated to the weighting where $(W_1)_{11} \to \infty$.

Now consider our example where we reweight the first equation to some other positive number. Let $(\mathbf{x}^*)_0$ denote the pseudosolution defined by $W_0 A, W_0 \mathbf{b}$. Since $W_0 = I_m$ then it follows that $(\mathbf{x}^*)$ is the least squares solution to the system defined by A and $\mathbf{b}$. In the next step we only modify the weight associated with equation 1. There are two meaningful cases. Either the weight for equation $\delta > 0$ or $\delta < 0$. If $\delta < 0$ then, since the weighting for a given equation is continuous, the new pseudosolution moves towards $(\mathbf{x}^*)_{-\infty}$. Similarly, if $\delta > 0$ then the new pseudosolution moves towards some $(\mathbf{x}^*)_\infty$. We are interested in the changes between pseudosolutions, $\Delta\mathbf{x}^* = (\mathbf{x}^*)_1 - (\mathbf{x}^*)_0$. Since $\Delta\mathbf{x}^*$ is a vector then the change in pseudosolutions moves along a line. We have argued above that the new pseudosolution moves towards $(\mathbf{x}^*)_{\pm\infty}$ and so a reasonable guess would be that the pseudosolutions move along a line towards $(\mathbf{x}^*)_{\pm\infty}$. If we continue with additional reweightings to only the first equation then $(\mathbf{x}^*)_{\pm\infty}$ will be the same point for each reweighting. And so for the case of reweighting a single equation we can characterize the movement of the pseudosolution for all reweightings. We will generalize this arugment as a conjecture. But first, some notation.

Definition 4.3.1 (WLS Solutions Without Some Equations)**.** Let A, $\mathbf{b}$ define an inconsistent linear system and let W be a weighting matrix. Recall that $A \in \mathbb{R}^{m\times n}$ and so we have m equations. We index our m equations with $[m]$. Let a collection of equation indices $n_1, n_2, \cdots, n_k$ be distinct integers between 1 and m and $k < m$. Then let

$$\mathbf{x}^\circ_{n_1,n_2,\cdots,n_k|WA}$$

denote the pseudosolution to the weighted least squares problem defined by A, $\mathbf{b}$, and W without equations $n_1, n_2, \cdots, n_k$.

Remark 4.3.1 (Relationship between $\mathbf{x}^\circ_{1|WA}$ and $(\mathbf{x}^*)_{-\infty}$)**.** In Example (4.3.2) we defined $(\mathbf{x}^*)_{\pm\infty}$. With the above definition notice that $(\mathbf{x}^*)_{-\infty} = \mathbf{x}^\circ_{1|WA}$. As discussed in Example (4.3.2), generally we expect $(\mathbf{x}^*)_{-\infty}$ to be a single vector we expect there to be many $(\mathbf{x}^*)_\infty$. As a result we focus on $(\mathbf{x}^*)_{-\infty}$ which is generalized by the above definition.

Conjecture 4.3.1 (Single Reweighting Pseudosolution Movement)**.** Let A, $\mathbf{b}$ define an inconsistent linear system. Define a sequence of reweightings by $W_0 = I_m$ and

$$(W_n)_{ii} = \begin{cases} (W_{n-1})_i + \delta_n & \text{if } i = k \\ (W_{n-1})_i & \text{if } i \neq k \end{cases}$$

where k is a fixed row index, and $-(W_{n-1})_k \leq \delta_n \in \mathbb{R}$.

Then the induced sequence of pseudosolutions $(\mathbf{x}^*)_n$ is such that $(\mathbf{x}^*)_n - (\mathbf{x}^*)_{n-1}$ lies on a line defined by

$$(\mathbf{x}^*)_{n-1} \quad \text{and} \quad \mathbf{x}^\circ_{k|W_{n-1}A}$$

This is visualized in Figure (4.1).

Remark 4.3.2. I have written the above proposition in such a way that will generalize when we will investigate how the pseudosolutions move with multiple reweightings. However, it follows that the above proposition has some redundancies. In particular, in the single reweighting case it can be shown that all the pseudosolutions $(\mathbf{x}^*)_n$ lie on the same line defined by $\mathbf{x}^\circ_{k|A}$ and $\mathbf{x}^*$.

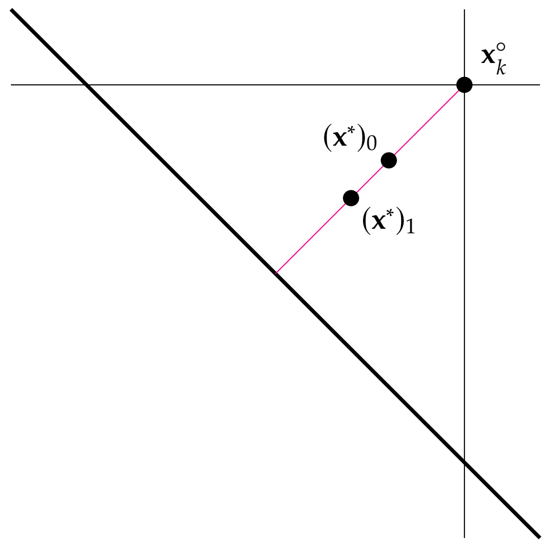


Figure 4.1 Here we have an inconsistent linear system visualized by the black lines. Our single reweighting only affects the thick hyperplane. The magenta line indicates the line along which the pseudosolutions move.

The above is given as a conjecture because we do not currently have a proof. That said, we have strong experimental evidence to suggest that the above conjecture holds. A high-level description of our experiment is as follows. We define a sequence of reweightings as in Conjecture (4.3.1) and then compute the pseudosolutions at each step using $\mathbf{x}^* = (A^TWA)^{-1}A^TW\mathbf{b}$. We then computed $(\Delta\mathbf{x}^*)_n = (\mathbf{x}^*)_n - (\mathbf{x}^*)_{n-1}$. If we let $\mathbf{d} = (\mathbf{x}^\circ_{k|A}) - (\mathbf{x}^*)_0$ then we verified that the direction of the updates $(\Delta\mathbf{x}^*)_n$ followed $\mathbf{d}$ by verifying the dot product $|\mathbf{d} \cdot (\Delta\mathbf{x}^*)_n| = 1$ for each n.

A future work is to prove this conjecture. A first line of attack might be to inspect $(\Delta\mathbf{x}^*)_n$ by using the definition of the WLS solution using the normal equations.

We now discuss the movement of multiple reweightings. However, the following discussion depends on the preceding discussion. As such much of the next ideas are given as conjectures, although they have been tested with similar numerical experiments.

4.3.2 Multiple Reweighting

Now that we have how the pseudosolution moves according to a single reweighting we want to understand how the pseudosolution moves according to more than one reweighting. Without loss of generality, suppose we reweigh equations one and two,

$$(W_1)_{ii} = \begin{cases} (W_0)_{ii} + \delta & \text{if } i = 1 \\ (W_0)_{ii} + \varepsilon & \text{if } i = 2 \\ (W_0)_{ii} & \text{Otherwise} \end{cases}$$

We ask how does $(\mathbf{x}^*)_1$ change to $(\mathbf{x}^*)_2$. In the previous part we found that a single reweighting followed a line defined by $(\mathbf{x}^*)_0$ and $\mathbf{x}^\circ_{1|W_0A}$. Note that, the final pseudosolution $(\mathbf{x}^*)_2$ is the same as the final pseudosolution given by the following reweightings

$$(W_1')_{ii} = \begin{cases} (W_0)_{ii} + \delta & \text{if } i = 1 \\ (W_0)_{ii} & \text{Otherwise} \end{cases} \qquad (W_2')_{ii} = \begin{cases} (W_1)_{ii} + \varepsilon & \text{if } i = 2 \\ (W_1)_{ii} & \text{Otherwise} \end{cases}$$

This is because $W_2' = W_1$ and so, as long as $W_2'A = W_1'A$ are both overdetermined, they must have the same pseudosolution defined by $(A^TWA)^{-1}A^TW\mathbf{b}$. We generalize this as a lemma. The following lemma is true regardless of Conjecture (4.3.1).

Lemma 4.3.1 (Decomposing Multiple Reweightings). *If a reweighting is given by*

$$\begin{aligned} (W_n)_{11} &= (W_{n-1})_{11} + \delta_1 \\ (W_n)_{22} &= (W_{n-1})_{22} + \delta_2 \\ &\cdots \\ (W_n)_{mm} &= (W_{n-1})_{mm} + \delta_m \end{aligned}$$

Then this reweighting is equivalent to the sequence of reweightings

$$\begin{aligned} (W_n')_{ii} &= \begin{cases} (W_{n-1})_{11} + \delta_1 & \textit{if } i = 1 \\ (W_{n-1}) & \textit{otherwise} \end{cases} \\ (W_{n+1}')_{ii} &= \begin{cases} (W_n')_{22} + \delta_2 & \textit{if } i = 2 \\ (W_n') & \textit{otherwise} \end{cases} \\ &\cdots \\ (W_{n+m}')_{ii} &= \begin{cases} (W_{n+m-1}')_{mm} + \delta_m & \textit{if } i = m \\ (W_{n+m-1}') & \textit{otherwise} \end{cases} \end{aligned}$$

In the sense that the final pseudosolution for both sets of reweightings will be the same vector.

Proof. This follows since W, A, $\mathbf{b}$ will be the same at the end of both sequences of reweightings and a pseudosolution is determined by exactly these objects. □

By decomposing the multiple reweighting as two single reweightings we can unpack the movement of the pseudosolution using our understanding of how a single reweighting moves the pseudosolution. The following conjecture depends on Conjecture (4.3.1). If Conjecture (4.3.1) is true then the following conjecture is also true by Lemma (4.3.1).

Conjecture 4.3.2 (Pseudosolution for Two Reweightings Follows a Sequence of Lines)**.** Let A, $\mathbf{b}$ define an inconsistent linear system. Let k_1, k_2 be equation indices. Then consider a reweighting

$$(W_1)_{ii} = \begin{cases} (W_0)_{ii} + \delta & \text{if } i = k_1 \\ (W_0)_{ii} + \varepsilon & \text{if } i = k_2 \\ (W_0)_{ii} & \text{Otherwise} \end{cases}$$

The pseudosolution for this reweighting a line defined by the lines at each of the single reweightings.

Proof. By lemma (4.3.1) we can decompose this reweighting as

$$(W_1')_{ii} = \begin{cases} (W_0)_{ii} + \delta & \text{if } i = k_1 \\ (W_0)_{ii} & \text{Otherwise} \end{cases} \qquad (W_2')_{ii} = \begin{cases} (W_1)_{ii} + \varepsilon & \text{if } i = k_2 \\ (W_1)_{ii} & \text{Otherwise} \end{cases}$$

with $((\mathbf{x}')^*)_3 = (\mathbf{x}^*)_2$. Assuming Conjecture (4.3.1) we know that $((\mathbf{x}')^*)_2 - ((\mathbf{x}')^*)_1$ follows a line $(\Delta(\mathbf{x}')^*)_1$ defined by $\mathbf{x}^\circ_{k_1|(W_1')A}$ and $((\mathbf{x}')^*)_1$. Similarly, $((\mathbf{x}')^*)_3 - ((\mathbf{x}')^*)_2$ follows a line $(\Delta(\mathbf{x}')^*)_2$ defined by $\mathbf{x}^\circ_{k_2|(W_2')A}$ and $((\mathbf{x}')^*)_2$. It follows that the final pseudosolution is a linear combination $(\mathbf{x}^*)_2 = ((\mathbf{x}')^*)_3 = c_1(\Delta(\mathbf{x}')^*)_1 + c_2(\Delta(\mathbf{x}')^*)_2$ which is also a line. (☕ Can you draw a similar figure to Figure (4.1) for two reweightings? How about three or more reweightings?) □

Note that it seems plausible to determine how far a single reweighting moves a pseudosolution. If we had that information then we would be able to compute c_1, c_2 without needing to compute $((\mathbf{x}')^*)_2, ((\mathbf{x}')^*)_3$.

The previous couple of subsections is enough to completely characterize the movement of the pseudosolution for any general reweighting. One can generalize the above arguments into an inductive form, or use a similar argument for more than two reweightings in a single step.

4.4 Proposed Definition for WREK

In the above Sections we have laid some foundations for how dynamic reweightings behave in terms of the induced sequence of pseudosolutions. There is more theoretical foundations which can be laid. We could ask which sequences of reweightings give meaningful pseudosolutions? In particular, given our inspiration from QuantileRK, one might attempt to construct a sequence of reweightings using only local residual information to generate a sequence of pseudosolutions which converge to the desired solution for corrupted linear systems. As well as understanding dynamic reweightings we want an RK variant which will converge to the pseudosolution of interest.

Now we will turn to *Weighted Randomized Extended Kaczmarz* (WREK), a proposed Randomized Kaczmarz variant which seeks to use dynamic reweightings to converge to a pseudosolution of interest. The algorithm for WREK is as follows.

Algorithm 4 Weighted Randomized Extended Kaczmarz

function WREK($\mathbf{A}$, $\mathbf{b}$, T, W, reweight($\cdot$)) ▷ where $A \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, and $T \in \mathbb{N}$
    $\mathbf{x}_0 = \mathbf{0}$
    $\mathbf{z}_0 = b$
    for $j = 1, .., T$ **do**
        Pick $i_k \in m$ with probability $q_i := \frac{\|\mathbf{A}^{(i)}\|_2^2}{\|\mathbf{A}\|_F^2}, i \in [m]$
        Pick $j_k \in n$ with probability $q_i := \frac{\|\mathbf{A}_{(j)}\|_2^2}{\|\mathbf{A}\|_F^2}, i \in [n]$
        $\mathbf{W}$ = reweight(A, b, i_k, j_k, W)
        Set $z^{(k+1)} = z^{(k)} - \frac{\langle (WA)_{(j_k)}, z^{(k)} \rangle}{\|(WA)_{(j_k)}\|_2^2}(WA)_{(j_k)}$
        Set $x^{(k+1)} = x^{(k)} + \frac{(Wb) - z^{(k)} - \langle (WA)^{(i_k)}, x^k \rangle}{\|(WA)^{(i_k)}\|_2^2}(WA)^{(i_k)}$
    end for
    return $\mathbf{x}_N$
end function

We remark that the action of WREK is the same as the action of REK except that on each pass you reweight the A and $\mathbf{b}$. The particular reweighting scheme is left as an input and will depend on further investigations on useful sequences of reweightings which use local system information. One may also remark that, by reading off the definition, if $W_n = I_m$ for all n then the definition of WREK matches the definition of REK exactly. And so in REK should be a special case of WREK.

In this way, our proposed definition of WREK looks like "the minimal change to REK to incorporate dynamic reweightings." However, there is no guarantee that this is the correct way to extend REK. The following are some important directions for further investigations.

4.5 Future Directions & Conclusion

Before we conclude this work we discuss some future directions to carry the investigation.

- One could set up numerical experiments to investigate whether the proposed WREK behaves like REK in the case where $W_n = I_m$ for all n.
- Another sanity check for this proposed definition of WREK is to investigate whether WREK behaves like RK in the case where $A, \mathbf{b}$ define a consistent system and $W_n = I_m$ for all n.
- After these sanity checks one may want to investigate reweighting schemes for corrupted linear systems. Given QuantileRK, one may investigate a reweighitng scheme which uses system residuals at each iteration.
- There are also theoretical investigations to be made. For example, in Section (3.2) we found a closed form for REK's $\mathbf{z}$ updates. We can perform a similar investigation into WREK's $\mathbf{z}$ vectors. We could then ask whether the corresponding $\mathbf{b}^{(k)}$ updates make sense.
- The current definition of WREK modifies the linear system. However, another avenue of investigation is to instead modifying the probability distributions to pick each row and column. One difference between these two methods is that applying a reweighting to A changes the column space of A which in turn changes the space of consistent measure vectors $\mathbf{b}'$. That said, there could be other dangers to modifying the column space of A which have not yet been investigated.
- One could attempt to prove Conjecture (4.3.1) and similar propositions for many reweightings in a single step.

Throughout this work we have considered large linear systems and numerical methods to solve them quickly. We have been particularly interested in Randomized Extended Kaczmarz's action of system modifications to solve noisy linear systems. To that end we unpacked how and why Randomized Extended Kaczmarz works so that we could lay foundations for a generalization to solve corrupted linear systems.

Thank you for joining me on this journey; climbing this hill with me. I hope you enjoyed traversing this space as much as I did and I hope you were able to develop your own relationship with some of the ideas here. I wish you the best on the most important step one can take, the next one.

Bibliography

Du, Kui. 2018. Refined upper bounds for the convergence of the randomized extended kaczmarz and gauss-seidel algorithms. *arXiv preprint arXiv:180103250* .

Gustafsson, Bertil. 1996. Mathematics for computer tomography. *Physica Scripta* 1996(T61):38.

Haddock, Jamie, Deanna Needell, Elizaveta Rebrova, and William Swartworth. 2020. Quantile-based iterative methods for corrupted systems of linear equations. *arXiv preprint arXiv:200908089* .

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

Strohmer, Thomas, and Roman Vershynin. 2009. A randomized kaczmarz algorithm with exponential convergence. *Journal of Fourier Analysis and Applications* 15(2):262–278.

Zouzias, Anastasios, and Nikolaos M. Freris. 2013. Randomized extended Kaczmarz for solving least squares. *SIAM J Matrix Anal A* 34(2):773–793.