ANALYZING SIMULTANEOUS ITERATIONS

by

Benwei Shi

A Senior Thesis submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Bachelor of Computer Science

School of Computing
The University of Utah
May 2018

Approved:

Jeff Phillips
Supervisor

April 10, 2018
Date

H. James de St. Germain
Director of Undergraduate Studies
School of Computing

Date

Ross Whitaker
Director
School of Computing

Date

ABSTRACT

Simultaneous Iteration (or Block Power Iteration) is a fast and simple method to approximate low-rank singular value decomposition for any matrix. This paper introduces a new error bound of the approximation matrix, and analyzes it in the two dimensional case. We show the error independent on the gaps between singular values, by providing a formula for the required number of iterations.

CONTENTS

CHAPTER 1

INTRODUCTION

In recent years, the demands on data number and dimension have grown rapidly. To use or analyze data in such ultra big matrices, several new algorithms have been developed to sketch approximations of these matrices. Usually an approximation is provided as a matrix factorization containing the underlying principle structure of the original matrix.

But in the big-data regime, both dimension and data number can be too big even for sketching the approximation by using standard algorithms, due to the time and/or space complexity. Fortunately, most of the ultra high dimensional data are also ultra sparse, so we can only represent the non-zero entries. And streaming is a solution for huge number of data points.

Recent research has developed several algorithms to deal with this situation. One of them is Sparse Frequent Directions described in [2] by Dr. Mina Ghashami, Dr. Edo Liberty, and Dr. Jeff M. Phillips. They then found a simplified version called Fast Sparse Frequent Directions (FSFD), which is composed by two algorithms. First is called Simultaneous Iteration (SimItr), a space efficient implementation of the popular randomized block power method, designed to deal with high dimensional ultra sparse data and the low rank approximation. Second is a efficient streaming-friendly Frequent Directions algorithm, designed to merge the summaries by SimItr, makes the algorithm friendly to streaming and parallelizing.

The algorithm FSFD is very successful in practice, but the error bound is still incomplete. Since the SimItr is a part of FSFD, we should analyze the error of SimItr first. There is an unfinished error bound analysis by the authors of [2] and Dr. Christopher Musco. That is basically only missing a proof of an appropriate error bound of SimItr. There are several error bounds of Simultaneous Iteration studied, like Frobenius Norm Error, Spectral Norm Error, and Per Vector Error, see [3]. However, for analyzing the error bound of FSFD, we need to bound the error of Simultaneous Iteration in another way:

$$\text{Vector Norm Error:} \qquad \|Aw\|^2 \geq (1-\epsilon)\|Pw\|^2 \tag{1.1}$$

for any unitary vector w, where P is the low rank approximation of A by Simultaneous Iteration. Let's call this as the *vector norm error*, which is the subject of this paper. This vector norm error can also answer a natural question about the accuracy of P, when $\|Pw\|^2$ is big, how we can guarantee that $\|Aw\|^2$ is also big. In particular, the goal will be to bound the number of iterations required for SimItr in terms of ϵ and only the dimensions of A, so that this property holds.

CHAPTER 2

BACKGROUND AND PRIOR WORK

We provide the background about matrix approximation and Simultaneous Iteration in this chapter.

2.1 Matrix Approximation

Singular value decomposition (SVD) is one of the best ways to understand the underlying principle structure of a matrix and many linear algebra concept. Using SVD, any data matrix $A \in \mathbb{R}^{n \times d}$ can be written as $A = U\Sigma V^T$. U and V have orthonormal columns called A's left and right singular vectors. Σ is a nonnegative diagonal matrix contain A's singular values in nondecreasing order.

The A's best rank k approximation is $A_k = U_k \Sigma_k V_k^T$, where U_k, Σ_k, and V_k contain only top k columns of U, Σ, and V respectively. We can also project A to it's top k singular vectors V_k to get $A_k = AV_kV_k^T$. Finding out the top principle subspace is the most common way to approximate a matrix. This implies that, to estimate the approximation, we only need to estimate the top principle subspace.

To find out the dominant principle subspace, we can work on the A's covariance matrix A^TA (or AA^T when $n < d$).

$$A^TA = V\Sigma U^T U \Sigma V^T = V\Sigma^2 V^T$$

It is well know that the singular values of A are the square roots of the eigenvalues of A^TA or AA^T, and the right or left singular vectors of A are the eigenvectors of A^TA or AA^T.

2.2 Simultaneous Iteration

Simultaneous iteration is also called as block power iteration. The idea of SimItr is to apply the power iteration to several vectors simultaneously. It takes an input matrix $A \in \mathbb{R}^{n \times d}$ ($d < n$), and three parameters $k \ll d$, $\varepsilon \in (0, 1)$, and $\delta \in (0, 1)$, generate a rank k

approximation P of A, within the error bound by ϵ, with probability $1-\delta$.

Algorithm 1 SIMULTANEOUS ITERATION

Input: $A \in \mathbb{R}^{n\times d}$, an integer $k \ll d$, and $\varepsilon \in (0,1)$, $\delta \in (0,1)$
$t = f(d,\epsilon,\delta)$
$Z = \text{qr}((A^T A)^t G)$
$[U,\Lambda] = \text{SVD}(Z^T A^T A Z)$ #s.t. $U\Lambda U^T = Z^T A^T A Z$
Return $P = \sqrt{\Lambda} U^T Z^T$

It first calculates the required number of iterations $t = f(d,\epsilon,\delta)$ using only d, ϵ, and δ, initializes k random vectors by a random Gaussian matrix G with 0 mean and 1 standard deviation. After t times, the subspace $(A^T A)^t G$ should converge to the subspace V_k with suitable assumptions:

1. $V_k^T G$ is nonsingular.

2. There is enough gap between σ_k and σ_{k+1}.

We use δ to represent the probability of getting G not good enough for the first assumption.

The second assumption seems like the result should depend on the properties of A. The algorithm returns high quality principal values, but the quality of the principal components still depends on the gaps between the singular values. But we don't have to worry about it. We may not get a good subspace close to the top k principle subspace, just because the gaps between the singular values are not big enough. In other words, if there are some singulars are similar, then the importances of these subspace are also similar, failing to find the right subspace in this case should not effect the error that much. Therefore, we can say SimItr returns the dominant subspace if there are any. This makes it possible to find the t for required ϵ independently of the properties of A. If we measure the error by residue spectral norm or Frobenius norm, the speed and accuracy are independent of singular value gaps, the required number of iterations is $t = O(\log(d)/\epsilon)$ with probability at least 99/100. This property has been well studied by Musco [3], Woodruff [4], and Boutsidis, Drineas, Magdon-Ismail [1]. However, none of these works address the vector norm error bound we seek.

CHAPTER 3

RESULT

The result of this paper is summarized in Theorem 3.0.1. This theorem gives exact required t for given ϵ and δ.

Theorem 3.0.1. *Let $A \in \mathbb{R}^{n\times 2}$ be any given data matrix, $g \in \mathbb{R}^2$ be a Gaussian random vector, z be the unit vector of $(A^T A)^t g$, $P = Azz^T$. Then for a given $\epsilon \in (0,1)$, it holds that $\|Aw\|^2 \geq (1-\epsilon)\|Pw\|^2$ for any unit vector $w \in \mathbb{R}^2$ with probability $1-\delta$, as long as*

$$t \geq \frac{\log(\epsilon^{-1}) + 2\log(\delta^{-1})}{2\log\left(\frac{1+\sqrt{\epsilon}}{1-\sqrt{\epsilon}}\right)} + \frac{1}{2}$$

Sometimes, we see ϵ as a constant, and only want the big O notation. Foe instance, to analyze FSFD, we only need $\epsilon = 1/2$.

Corollary 3.0.1. *In Theorem 3.0.1, when $\epsilon = 1/2$, we require*

$$t = O\left(\log(\delta^{-1})\right)$$

Figure 3.1 on this page is the chart about $\alpha = 1-\epsilon$ and t with $\delta = 0.01$.

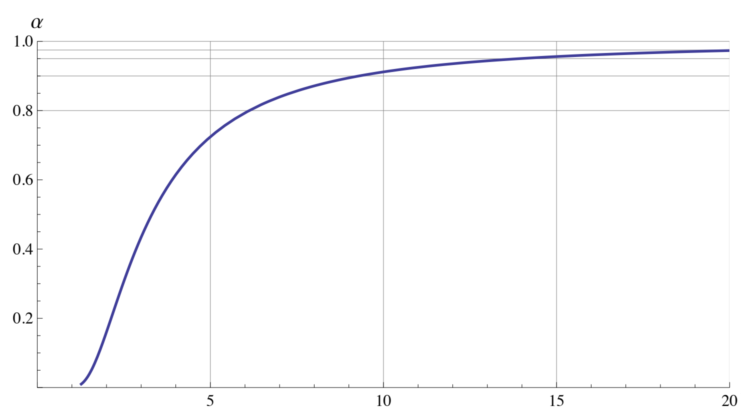

Figure 3.1. Plot of function $t = \frac{-\log(1-\alpha)-2\log(0.01)}{2\log\left(\frac{1+\sqrt{1-\alpha}}{1-\sqrt{1-\alpha}}\right)} + \frac{1}{2}$.

CHAPTER 4

TWO DIMENSION ANALYSIS

In this section, we analyze the Simultaneous Iteration by proving the main result in two dimension (the simplest) case. In other words, we want to show that, given an $\epsilon \in (0, 1)$ close to 0, find out the required t as small as possible, so that, $\|Aw\|^2 \geq (1-\epsilon)\|Pw\|^2$ with a high probability.

In 2D, we have $d = 2$, $k = 1$. The input matrix $A \in \mathbb{R}^{n\times 2}$ has singular vectors $V = [v_1, v_2]$, $U = [u_1, u_2]$, and singular values $\Sigma = \mathsf{diag}(\sigma_1, \sigma_2)$, $\sigma_1 \geq \sigma_2 \geq 0$. Since $k = 1$, there is only 1 column vector in matrix Z, we let $z_1 = Z = SimItr(A, 1, \epsilon, \delta)$, let z_2 be the null space of z_1, or a unit vector orthogonal to z_1 in other words. Let θ be the angle between v_1 and z_1, γ be the angle between Az_1 and Az_2. To simplify equations, we use $s = \frac{\sigma_2}{\sigma_1}$. Let $w_1 = \langle w, z_1\rangle z_1 = z_1 z_1^T w$ be the project vector of w onto z_1, so that $Pw = Az_1z_1^T w = Aw_1$. Let $w_2 = w - w_1$ be the project vector of w onto z_2. **Figure 4.1** on the current page shows most of the notations in a typical case.

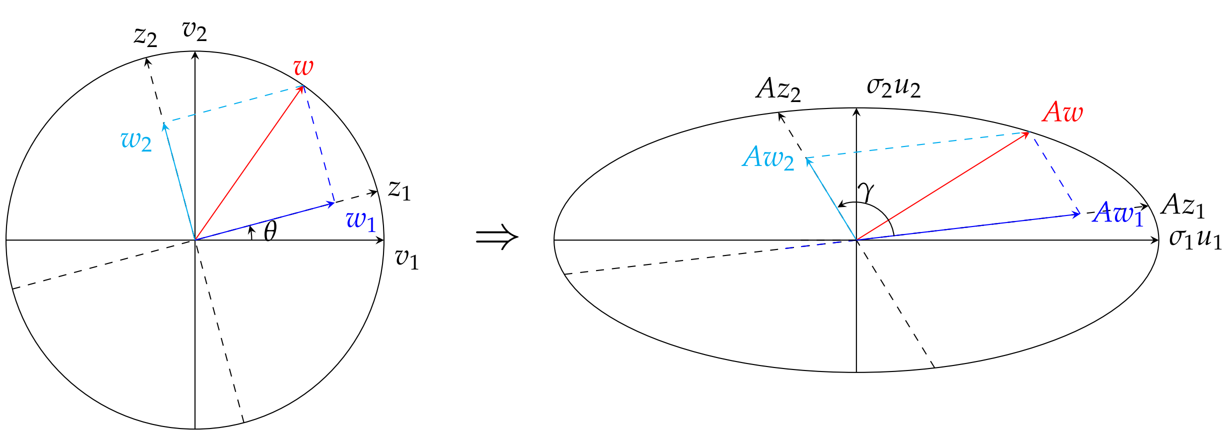

Figure 4.1. w in V space and Aw in U space.

4.1 Bound error by angle

Lemma 4.1.1. *Let $A \in \mathbb{R}^{n\times 2}$, $Z \in \mathbb{R}^{2\times 2} = [z_1, z_2]$ be an arbitrary basis of $\mathbb{R}^2$ space. Let $P = Az_1z_1^T$, γ be the angle between Az_1 and Az_2. For any unit vector $w \in \mathbb{R}^2$*

$$\|Aw\|^2/\|Pw\|^2 \geq 1 - \cos^2\gamma$$

Proof. Let $w_1 = \langle w, z_1\rangle z_1 = z_1z_1^Tw$ be the project vector of w onto z_1, so that $Pw = Az_1z_1^Tw = Aw_1$. Let $w_2 = w - w_1$ be the project vector of w onto z_2.

We first expand $\|Aw\|^2$ and $\|Pw\|^2$ separately,

$$\begin{aligned}
\|Aw\|^2 =&\|A(w_1 + w_2)\|^2 = \|Aw_1 + Aw_2\|^2 \\
=&\|Aw_1\|^2 + \|Aw_2\|^2 + 2\langle Aw_1, Aw_2\rangle \\
=&\|Aw_1\|^2 + \|Aw_2\|^2 + 2\|Aw_1\|\|Aw_2\|\cos\gamma
\end{aligned}$$

and

$$\|Pw\|^2 =\|Az_1z_1^Tw\|^2 = \|Aw_1\|^2$$

Combining above two, we have

$$\begin{aligned}
\frac{\|Aw\|^2}{\|Pw\|^2} =&\frac{\|Aw_1\|^2 + \|Aw_2\|^2 + 2\|Aw_1\|\|Aw_2\|\cos\gamma}{\|Aw_1\|^2} \\
=&1 + \frac{\|Aw_2\|^2}{\|Aw_1\|^2} + 2\cos\gamma\frac{\|Aw_2\|}{\|Aw_1\|} \\
\geq&1 - \cos^2\gamma
\end{aligned}$$

The last inequality follows $ax^2 + bx + c \geq c - \frac{b^2}{4a}$ for $a \geq 0$, where $x = \frac{\|Aw_N\|}{\|Aw_Z\|}$, $a = 1$, $b = 2\cos\gamma$, and $c = 1$. So $c - \frac{b^2}{4a} = 1 - \frac{(2\cos\gamma)^2}{4} = 1 - \cos^2\gamma$. □

This result is a great starting point to analyze our problem in any dimension, not only in two dimension. It replaces ϵ with a well defined $\cos^2\gamma$ without introducing any extra error. Furthermore, this implies the next step, to find out the minimum angle between all possible pairs of vectors in subspace AZ and AN, which is known as the minimum principal angle between them. In two dimension case, there are only one principal angle, which is also the angle between Az_1 and Az_2.

4.2 Thinking in two dimension and proof idea

By Lemma 4.1.1 we have $\frac{\|Aw\|^2}{\|Pw\|^2} \geq 1 - \cos^2\gamma$. This gives us a way to think this problem geometrically especially in two dimension case. To show that ϵ can be close to 0, now we can show $\cos^2\gamma$ can be close to 0 instead, that is γ can be close to $\pi/2$. Actually, we will consider how the right angle between z_1 and z_2 changes by matrix A. For a vector x, $Ax = U\Sigma V^T x$, we can think the change between x and Ax as a linear transformation by matrix A, which is a map from V space to U space with squeeze and stretch. For examples, $Av_1 = \sigma_1 u_1$, and $Av_2 = \sigma_2 u_2$, v_1, v_2 map to u_1, u_2 with scale σ_1, σ_2 respectively.

Now we can see there are two special cases, in both of them, γ is guarantee to be exactly $\pi/2$. First, when $s = 1$ ($\sigma_1 = \sigma_2$), the squeeze or stretch is uniform, so the angle remains $\pi/2$. Second, when $s = 0$ ($\sigma_2 = 0$), we should get $z_1 = \pm v_1$ as a result of Simultaneous Iteration algorithm, and $z_2 = \pm v_2$. So $Az_1 = \pm\sigma_1 u_1$, $Az_2 = \pm\sigma_2 u_2$. Since u_1, u_2 are orthogonal to each other, therefore $\gamma = \pi/2$.

Note that the Simultaneous Iteration algorithm with $k = 1$ is just a Power Method, it can generate z_1 very close to v_1 with very high probability, as long as there is some gap between σ_1 and σ_2, we analysis this in Lemma 4.3.3 for two dimensional case. Now we can expand the above special cases to the more general case. First, when s is close to 1, γ is close to $\pi/2$, no matter where z_1 lives, this is shown in Lemma 4.3.2. Second, when s is small, z_1 should close to v_1, so we have γ close to $\pi/2$ again, this is shown in Lemma 4.3.4. See **Figure 4.1** on page 6 for a better understanding.

To analysis γ in this way, we introduce θ as the angle between v_1 and z_1, so we can measure the closeness between them by $\tan^2\theta$. Lemma 4.3.1 express $\cos^2\gamma$ as a function of s and $\tan^2\theta$. With all the mentioned lemmas, we can start to prove the theorem.

4.3 Lemmas needed in two dimension

In this section, we prove all the lemmas needed in the proof of the theorem in two dimension. The ideas behind the lemmas have been stated in the last paragraph of the last section.

Lemma 4.3.1. *Let $A \in \mathbb{R}^{n\times 2}$, let $U, \Sigma, V^T = \mathsf{SVD}(A)$, where $V = [v_1, v_2]$, $\Sigma = [\sigma_1, \sigma_2]$, let $s = \frac{\sigma_2}{\sigma_1}$. Let $Z \in \mathbb{R}^{2\times 2} = [z_1, z_2]$ be an arbitrary basis of $\mathbb{R}^2$ space. Let θ is the angle between v_1 and z_1, γ be the angle between Az_1 and Az_2.*

$$\cos^2\gamma = \frac{\left(s^2-1\right)^2}{1+s^4+s^2(\tan^2\theta+\cot^2\theta)}$$

Proof.

$$\begin{aligned}
\cos^2\gamma = &\left(\frac{\langle Az_1, Az_2\rangle}{\|Az_1\|\|Az_2\|}\right)^2 \\
=&\frac{(z_1^T A^T A z_2)^2}{(z_1^T A^T A z_1)(z_2^T A^T A z_2)} \\
=&\frac{(z_1^T V\Sigma^2 V^T z_2)^2}{(z_1^T V\Sigma^2 V^T z_1)(z_2^T V\Sigma^2 V^T z_2)} && \text{since } A^TA = V\Sigma U^T U\Sigma V^T = V\Sigma^2 V^T \\
=&\frac{\left((V^Tz_1)^T\Sigma^2(V^Tz_2)\right)^2}{\left((V^Tz_1)^T\Sigma^2(V^Tz_1)\right)\left((V^Tz_2)^T\Sigma^2(V^Tz_2)\right)} \\
=&\frac{\left(\mp\sigma_1^2\sin\theta\cos\theta \pm \sigma_2^2\sin\theta\cos\theta\right)^2}{\left(\sigma_1^2\cos^2\theta+\sigma_2^2\sin^2\theta\right)\left(\sigma_1^2\sin^2\theta+\sigma_2^2\cos^2\theta\right)} && V^Tz_1 = (\cos\theta, \sin\theta), V^Tz_2 = (\cos(\theta\pm\pi/2), \sin(\theta\pm\pi \\
=&\frac{\left(\sigma_1^2-\sigma_2^2\right)^2}{\left(\sigma_1^2+\sigma_2^2\tan^2\theta\right)\left(\sigma_1^2+\sigma_2^2\cot^2\theta\right)} && \text{dividing by } \sin^2\theta\cos^2\theta \\
=&\frac{\left(1-s^2\right)^2}{\left(1+s^2\tan^2\theta\right)\left(1+s^2\cot^2\theta\right)} && \text{dividing by } \sigma_1^4 \\
=&\frac{\left(1-s^2\right)^2}{1+s^4+s^2\left(\tan^2\theta+\cot^2\theta\right)}
\end{aligned}$$

□

Lemma 4.3.2. *Let $s \in [0,1]$, given a fixed $\epsilon \in (0,1)$, when $s^2 \geq \frac{1-\sqrt{\epsilon}}{1+\sqrt{\epsilon}}$, $\frac{\left(s^2-1\right)^2}{1+s^4+s^2(\tan^2\theta+\cot^2\theta)} \leq \epsilon$.*

Proof.

$$\begin{aligned}
&\frac{\left(1-s^2\right)^2}{1+s^4+s^2(\tan^2\theta+\cot^2\theta)} \\
=&\frac{\left(1-s^2\right)^2}{1+s^4+2s^2+s^2(\tan^2\theta-2+\cot^2\theta)} \\
=&\frac{\left(1-s^2\right)^2}{1+s^4+2s^2+s^2(\tan\theta-\cot\theta)^2} \\
\leq&\frac{\left(1-s^2\right)^2}{1+s^4+2s^2} \\
=&\left(\frac{1-s^2}{1+s^2}\right)^2
\end{aligned}$$

Now we want

$$
\begin{aligned}
\left(\frac{1-s^2}{1+s^2}\right)^2 \leq& \epsilon \\
\frac{1-s^2}{1+s^2} \leq& \sqrt{\epsilon} \\
1-s^2 \leq& s^2\sqrt{\epsilon}+\sqrt{\epsilon} \\
1-\sqrt{\epsilon} \leq& s^2\left(1+\sqrt{\epsilon}\right) \\
s^2 \geq& \frac{1-\sqrt{\epsilon}}{1+\sqrt{\epsilon}}
\end{aligned}
$$

□

Lemma 4.3.3. *Let $A \in \mathbb{R}^{n\times 2}$ with singular vectors $V = [v_1, v_2]$ and singular values $\sigma_1 \geq \sigma_2 \geq 0$, let $s = \sigma_2/\sigma_1$. Let $g \in \mathbb{R}^2$ be a i.i.d. Gaussian random vector, let z be the unit vector of $(A^TA)^t g$, let θ be the angle between v_1 and z. With probability $1-\delta$, after t iterations, $\tan^2\theta < s^{4t}\delta^{-2}$.*

Proof. Let $g^{(0)} = g$, $g^{(t)} = (A^TA)^t g$. Let $\theta^{(0)}$ be the angle between v_1 and $g^{(0)}$, $\theta^{(t)} = \theta$ be the angle between v_1 and $g^{(t)}$(or z).

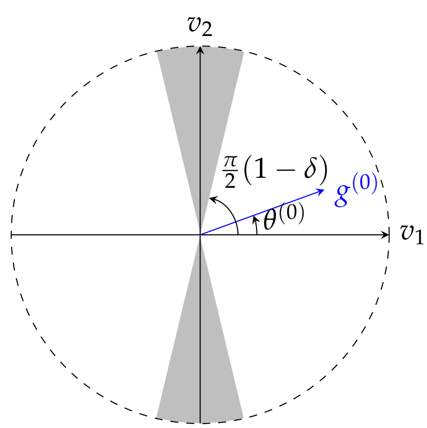

Figure 4.2. Gaussian random vector appears in white direction with probability $1-\delta$.

This is a simple 2D Power Method process. $g^{(0)}$ is the initial Gaussian random vector, the distribution of its direction is uniform, with probability $1-\delta$, we have $g^{(0)}$ in the white direction shown in the **Figure 4.2** on the current page. That is

$$
\tan^2\theta^{(0)} \leq \tan^2\left(\frac{\pi}{2}(1-\delta)\right)
$$

Now let look at each iteration, or from $(t-1)$th iteration to tth iteration in other words,

$$\tan\theta^{(t)} = \frac{\langle V_2, A^TAg^{(t-1)}\rangle}{\langle V_1, A^TAg^{(t-1)}\rangle} = \frac{V_2^TA^TAg^{(t-1)}}{V_1^TA^TAg^{(t-1)}} = \frac{V_2^TVS^2V^Tg^{(t-1)}}{V_1^TVS^2V^Tg^{(t-1)}} = \frac{\sigma_2^2V_2^Tg^{(t-1)}}{\sigma_1^2V_1^Tg^{(t-1)}} = s^2\frac{\langle V_2, g^{(t-1)}\rangle}{\langle V_1, g^{(t-1)}\rangle} = s^2\tan\theta^{(t-1)}$$

Combining above two, we have

$$\tan^2\theta = \tan^2\theta^{(t)} \le s^{4t}\tan^2\left(\frac{\pi}{2}(1-\delta)\right)$$

Since $0 \le \delta \le 1$, we know $0 \le \frac{\pi}{2}(1-\delta) \le \frac{\pi}{2}$, so

$$\begin{aligned} &\cot\left(\frac{\pi}{2}(1-\delta)\right) = \tan\left(\frac{\pi}{2} - \frac{\pi}{2}(1-\delta)\right) \ge \frac{\pi}{2}\delta > \delta \\ \Leftrightarrow &\tan\left(\frac{\pi}{2}(1-\delta)\right) < \delta^{-1} \end{aligned}$$

Finally, we have

$$\tan^2\theta < s^{4t}\delta^{-2}$$

□

Lemma 4.3.4. *Let $s \in [0,1]$, $\delta \in [0,1]$. Given a fixed $\epsilon \in (0,1)$, when $s^2 < \frac{1-\sqrt{\epsilon}}{1+\sqrt{\epsilon}}$, $\frac{\left(s^2-1\right)^2}{1+s^4+s^2(\tan^2\theta+\cot^2\theta)} < \epsilon$, if*

$$t \ge \frac{\ln(\epsilon^{-1}) + 2\ln(\delta^{-1})}{2\ln\left(\frac{1+\sqrt{\epsilon}}{1-\sqrt{\epsilon}}\right)} + \frac{1}{2}$$

Proof.

$$\begin{aligned} &\frac{\left(1-s^2\right)^2}{1+s^4+s^2(\tan^2\theta+\cot^2\theta)} \\ <&\frac{1}{s^2(\tan^2\theta+\cot^2\theta)} && \text{since } 0 \le s \le 1 \\ \le&\frac{1}{s^2(\cot^2\theta)} = s^{-2}\tan^2\theta \\ <&s^{4t-2}\delta^{-2} && \text{by lemma 4.3.3} \\ <&\left(\frac{1-\sqrt{\epsilon}}{1+\sqrt{\epsilon}}\right)^{2t-1}\delta^{-2} \end{aligned}$$

Now we want

$$\begin{aligned}
& \left(\frac{1-\sqrt{\epsilon}}{1+\sqrt{\epsilon}}\right)^{2t-1}\delta^{-2} \le \epsilon \\
\iff \quad & \left(\frac{1-\sqrt{\epsilon}}{1+\sqrt{\epsilon}}\right)^{2t-1} \le (\epsilon)\delta^2 \\
\iff \quad & (2t-1)\log\left(\frac{1-\sqrt{\epsilon}}{1+\sqrt{\epsilon}}\right) \le \log(\epsilon) + 2\log\delta \\
\iff \quad & t \ge \frac{\log(\epsilon)+2\log(\delta)}{2\log\left(\frac{1-\sqrt{\epsilon}}{1+\sqrt{\epsilon}}\right)} + \frac{1}{2} = \frac{\log(\epsilon^{-1})+2\log(\delta^{-1})}{2\log\left(\frac{1+\sqrt{\epsilon}}{1-\sqrt{\epsilon}}\right)} + \frac{1}{2}
\end{aligned}$$

□

4.4 Proof the Theorem in two dimension

We prove Theorem 3.0.1 in this section.

Proof. When $\sigma_1 = 0$, since $0 \le \sigma_2 \le \sigma_1$, we know that $\sigma_2 = 0$, therefore $\|Pw\| = \|Azz^T w\| = 0$. It is easy to see that, when $\|Pw\| = 0$, α can be any value. We assume $\|Pw\| \neq 0$ and $\sigma_1 \neq 0$ from now on. So we can do division by $\|Pw\|$ and σ_1.

By Lemma 4.1.1 we have:

$$\frac{\|Aw\|^2}{\|Pw\|^2} \ge 1 - \cos^2\gamma$$

Where γ is the angle between vectors Az_1 and Az_2.

Let $\theta \in [0, \pi/2]$, the angle between v_1 and z_1. By Lemma 4.3.1 we have:

$$\cos^2\gamma = \frac{\left(s^2-1\right)^2}{1+s^4+s^2(\tan^2\theta+\cot^2\theta)}$$

The rest of the proof is divided into two cases. Both of them show that $\frac{\left(s^2-1\right)^2}{1+s^4+s^2(\tan^2\theta+\cot^2\theta)} \le \epsilon$ under some conditions.

First, for a given ϵ, when s^2 is big enough, $\cos^2\gamma$ can be small enough.

$$\text{When } s^2 \ge \frac{1-\sqrt{\epsilon}}{1+\sqrt{\epsilon}}, \frac{\left(s^2-1\right)^2}{1+s^4+s^2(\tan^2\theta+\cot^2\theta)} \le \epsilon \qquad \text{Lemma 4.3.2}$$

Second, when s^2 is not big enough, we need t big enough to generate a good z_1 with small θ, see Lemma 4.3.4, so that, $\cos^2\gamma$ can be small enough with high probability.

$$\text{When } s^2 < \frac{1-\sqrt{\epsilon}}{1+\sqrt{\epsilon}}, \frac{\left(s^2-1\right)^2}{1+s^4+s^2(\tan^2\theta+\cot^2\theta)} < 1-\alpha$$

$$\text{with probability } 1-\delta, \text{ and require } t \ge \frac{\log(\epsilon^{-1})+2\log(\delta^{-1})}{2\log\left(\frac{1+\sqrt{\epsilon}}{1-\sqrt{\epsilon}}\right)} + \frac{1}{2} \qquad \text{Lemma 4.3.4}$$

Now we can conclude that, given an $\epsilon \in (0,1)$, with probability $1-\delta$, $\|Aw\|^2 \geq (1-\epsilon)\|Pw\|^2$, as long as

$$t \geq \frac{\log(\epsilon^{-1}) + 2\log(\delta^{-1})}{2\log\left(\frac{1+\sqrt{\epsilon}}{1-\sqrt{\epsilon}}\right)} + \frac{1}{2}.$$

$\square$

CHAPTER 5

CONCLUSIONS

We analyzed the behavior of algorithm Simultaneous Iteration in two dimension in detail. We provided the intuition and proof idea, all the proof detail, and the complete proof of the main theorem in two dimension. This provide many details of the algorithm in two dimension, and prepares for understanding it in high dimensions. Based on pastoal progress, we are confident that this solution is a critical building block towards a full analysis in the high-dimensional case where d and k are unbounded.

REFERENCES

[1] C. Boutsidis, P. Drineas, and M. Magdon-Ismail, *Near-Optimal Column-Based Matrix Reconstruction*, Proceedings - Annual IEEE Symposium on Foundations of Computer Science, FOCS, (2011), pp. 305–314.

[2] M. Ghashami, E. Liberty, and J. M. Phillips, *Efficient Frequent Directions Algorithm for Sparse Matrices*, in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16, New York, New York, USA, 2016, ACM Press, pp. 845–854.

[3] C. C. Musco and C. C. Musco, *Randomized Block Krylov Methods for Stronger and Faster Approximate Singular Value Decomposition*, in Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1, MIT Press, apr 2015, pp. 1396–1404.

[4] D. P. Woodruff, *Sketching as a Tool for Numerical Linear Algebra*, Foundations and Trends® in Theoretical Computer Science, 10 (2014), pp. 1–157.