Guarantees of Fast Band Restricted Thresholding Algorithm for Low-Rank Matrix Recovery Problem

Fujun Zhao,1 Jigen Peng,2 Kai Sun,1 and Angang Cui1

1School of Mathematics and Statistics, Xi’an Jiaotong University, Xi’an, 710049, China
2School of Mathematics and Information Science, Guangzhou University, Guangzhou 510006, China

Correspondence should be addressed to Jigen Peng; jgpengxjtu@126.com

Received 1 March 2020; Accepted 7 April 2020; Published 30 April 2020

Guest Editor: Junpeng Shi

Copyright ©2020 Fujun Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Affine matrix rank minimization problem is a famous problem with a wide range of application backgrounds. This problem is a combinatorial problem and deemed to be NP-hard. In this paper, we propose a family of fast band restricted thresholding (FBRT) algorithms for low rank matrix recovery from a small number of linear measurements. Characterized via restricted isometry constant, we elaborate the theoretical guarantees in both noise-free and noisy cases. Two thresholding operators are discussed and numerical demonstrations show that FBRT algorithms have better performances than some state-of-the-art methods. Particularly, the running time of FBRT algorithms is much faster than the commonly singular value thresholding algorithms.

1. Introduction

The affine matrix rank minimization (AMRM) problem, which is to recover a low-rank matrix from only a small number of linear measurements, can be described as the following optimization problem:

\[
\min_{X \in \mathbb{C}^{n_1 \times n_2}} \text{rank}(X) \\
\text{s.t.} \quad \|\mathcal{A}(X) - b\|_2 \leq \eta,
\]

(1)

where \(\mathcal{A} \in \mathbb{L}(\mathbb{C}^{n_1 \times n_2}, \mathbb{C}^d)\) is a given linear map, \(b \in \mathbb{C}^d\) is a given vector, and \(\eta \geq 0\) is error tolerance. This problem has got much attention in recent years and many applications arising in various areas can be captured by solving model (1), for example, matrix completion [1, 2], background modeling [3], subspace clustering [4], phase retrieval [5, 6], image inpainting [7], and other applications [8, 9]. The lines and the symbols in all figures are too small to see. It is better to enlarge them. Unfortunately, ARMR problem is known to be NP-hard. Therefore, without further assumptions on \(\mathcal{A}\) and \(X\), solving this problem would be computationally intractable. To overcome this shortcoming, many researchers have focused on replacing \(\text{rank}(X)\) with other penalties \(G(X)\), such as Schatten \(p\)-norm \((0 < p < 1)\) [10], Minimax Concave Plus (MCP) [11], Smoothly Clipped Absolute Deviation (SCAD) [12], Logarithm [13], Geman [14], and Laplace [15], and considered the following optimization problem:

\[
\min_{X \in \mathbb{C}^{n_1 \times n_2}} G(X) \\
\text{s.t.} \quad \|\mathcal{A}(X) - b\|_2 \leq \eta,
\]

(2)
or the corresponding unconstrained optimization problem:

\[
\min_{X \in \mathbb{C}^{n_1 \times n_2}} \frac{1}{2}\|\mathcal{A}(X) - b\|_2^2 + \lambda \cdot G(X),
\]

(3)

where \(\lambda > 0\) is a regularization parameter. Model (3) can be transformed into a fixed point problem described as

\[
X = \mathcal{H}(X - s \cdot \mathcal{A}^*(AX - b)),
\]

(4)

where \(\mathcal{A}^*\) is the dual operator of \(\mathcal{A}\), \(\mathcal{H}\) is the thresholding operator, and \(s\) is a step size parameter. Naturally, we have corresponding iterative thresholding algorithm. Given a fixed penalty \(G(\cdot)\), there have been many theoretical guarantees. However, there are still some challenges as follows:

(I) Most of convergence results were developed for model (3) with fixed \(\lambda\), and it is difficult to choose an appropriate parameter \(\lambda\).
We display the numerical simulations, and then conclude this paper. Theorem 1 (see [23]). Given a linear operator $\mathcal{A}: C^{n_1 \times n_2} \rightarrow C^d$ and $1 \leq r \leq \min(m, n)$, the restricted isometry constant $\delta_r$ is the smallest nonnegative number, such that, for any matrix $X \in C^{n_1 \times n_2}$, with rank$(X) \leq r$, we have the following inequality:

$$\left(1 - \delta_r\right)\|X\|_F^2 \leq \|\mathcal{A}(X)\|_2^2 \leq \left(1 + \delta_r\right)\|X\|_F^2,$$

where $\|\cdot\|_F$ is the Frobenius norm.

For a given linear map $\mathcal{A}$, it is difficult to calculate the restricted isometry constant $\delta_r$. However, for Gaussian random linear map, there has been a result that the number of measurements $d \geq c_r r (n_1 + n_2) \log (n_1 n_2)$ ($c_r$ only depends on $r$) is sufficient to yield an RIC of $\delta$ with high probability [23].

In the noisy case, the following theorem claims that the solution of model (1) cannot be far from the original low-rank matrix, if $\mathcal{A}$ satisfies a certain RIC. The related theorem about compressive sensing is mentioned in [24], see equation (5.24).

**Theorem 1.** Let $Z, Z''$ be two solution of (1) within error tolerance $\eta$ (i.e., $\|\mathcal{A}(Z) - b\|_2^2, \|\mathcal{A}(Z'') - b\|_2^2 \leq \eta$). If rank$(Z) = \text{rank}(Z'') = r$ and $\delta_{2r} < 1$, we obtain a stability claim of the form

$$\|Z - Z''\|_2^2 < \frac{4\eta^2}{1 - \delta_{2r}}. \quad (7)$$

If we take $\eta = 0$ in Theorem 1, we obtain the following uniqueness result, in a noise-free case.

**Corollary 1.** Let $Z$ be a $r$-rank solution of $\mathcal{A}(X) = b$. If $\mathcal{A}$ satisfies

$$\delta_{2r} < 1, \quad (8)$$

then $Z$ is the unique lowest rank solution.

Taking into account the abovementioned results, it implies that the original low-rank matrix $X$ is the solution of model (1), if $\mathcal{A}$ satisfies a certain RIC.

**3. Fast Band Restricted Thresholding Algorithm**

In this section, we will design a new fast band restricted thresholding (FBRT) algorithm. For this purpose, we need to develop some thresholding algorithms for (1) in previous work and compare FBRT algorithm with other algorithms.

Similar to compressive sensing problem, both iterative hard thresholding and iterative soft thresholding, also known as SVP [16] and SVT [2], are simple and efficient algorithms for low-rank matrix recovery. In pursuit of better results, some alternative algorithms, such as HFFA [22], thresholding function for Schatten 2/3-norm [18], SCAD [12], and firm thresholding [25], have been proposed. We present these commonly used thresholding functions in Figure 1, and it is obvious that all of them satisfy the following definitions.
A function \( h_t \) of \( \mathbb{R}^7 \) with parameter \( \tau > 0 \) is called as band restricted thresholding (BRT) function, if it satisfies

(I) \( h_t(u) = 0 \), with \( u \leq \tau \)

(II) \( h_t(u) \leq h_t(v) \), with \( \tau < u \leq v \)

(III) There exists a constant \( c \in [0, 1] \) (band parameter), such that \( u - c\tau \leq h_t(u) \leq u \), with \( u \geq \tau \).

According to Definition 2, the following Table 1 shows the corresponding band parameter \( c \) of these BRT functions.

**Table 1: Band parameter \( c \) for different thresholding functions.**

<table>
<thead>
<tr>
<th>( c )</th>
<th>( h_{1,s} )</th>
<th>( h_{1,b} )</th>
<th>( h_{1,3} )</th>
<th>( h_{1,1/2} )</th>
<th>( h_{1,1/3} )</th>
<th>( h_{1,SCAD} )</th>
<th>( h_{1,f\text{irm}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \epsilon )</td>
<td>0</td>
<td>1</td>
<td>1/3</td>
<td>3/4</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

\[
\mathcal{P}_{S_k}(Z) = U_k[U_k^*Z + ZV_kV_k^* - U_kU_k^*ZV_kV_k^*] = (I - U_kU_k^*)Z(I - V_kV_k^*).
\]

Meanwhile, we derive the fast band restricted thresholding (FBRT) algorithm described as in Algorithm 1.

**Remark 1.** The threshold parameter \( \tau \) plays an important role which affects the performance of FBRT algorithm. Here, we let \( \tau = \sigma_{k+1} \), and we will show the theoretical performance of FBRT algorithm in the following discussion.

**Remark 2.** The stopping criterion is \( \| \mathcal{A}(X^{(i)}) - b \|_2 \leq \epsilon \). Taking into account Theorem 1, \( X^{(i)} \) cannot be far from \( X \), if \( \mathcal{A} \) satisfies a certain RIC. On the contrary, it is an important situation of classified discussion that \( X^{(i)} \) is close to \( X \), and the detail will be shown in the following section.

It is worth noting that \( X^{(k)} + \mathcal{A}^*(b - \mathcal{A}(X^{(k)})) \) is always a full rank matrix and computing the SVD will use \( O(N^3) \) \( (N = (m + n)) \) floating point operations. In the mean time, \( W^{(k)} \) in every iteration of FBRT algorithm is a \((2r)\)-rank matrix, and according to the QR factorization, we can obtain that

\[
W^{(k)} = QR = QR^*Q^*,
\]

where \( W^{(k)} = QR \) and \( R^* = Q_1R_1 \) are the QR factorizations of \( W^{(k)} \) and \( R^* \), with \( Q \in \mathbb{C}^{n_r \times 2r}, R \in \mathbb{C}^{2r \times n_r}, R_1 \in \mathbb{C}^{2r \times 2r}, \) and \( Q_1 \in \mathbb{C}^{n_r \times 2r} \). Thus, the SVD of \( W^{(k)} \) can be obtained from the SVD of \( R_1 \), and computing the SVD of \( W^{(k)} \), will use \( O(r^3) \) floating point operations instead of \( O(N^3) \) flops [20].

### 4. Analysis of the FBRT Algorithm

In this section, we will study theoretically the performance of the FBRT algorithm, and the following Theorem 2 and 3 show the theoretical guarantees in both noise-free and noisy cases.

**Theorem 2.** Let the sequence \( \{X^{(i)}\} \) be defined by the FBRT algorithm with \( \mathcal{A}(X) = \mathcal{A}(X) = b \) and \( \text{rank}(X) = r \). If the thresholding operator in the FBRT algorithm is the BRT operator with a band parameter \( c \) and the following constant

\[
\rho := \frac{\sqrt{4 + 2c^2r}}{\delta_{2r} + \delta_{3r} + \sqrt{4 + 2c^2r} \delta_{2r+1} \max(\mathcal{X}/\mathcal{X})/\sigma_{\min}(\mathcal{X})},
\]

is less than \( 1 \), we have

\[
\| X^{(i)} - X \|_F \leq \rho^i \| X^{(0)} - X \|_F.
\]

Particularly, \( \rho < 1 \) can be satisfied if

\[
\delta_{3r} \leq \frac{1}{2\sqrt{4 + 2c^2r} + (4 + 2c^2r)\sqrt{r} \sigma_{\max}(\mathcal{X})/\sigma_{\min}(\mathcal{X})}.
\]
The proof of Theorem 2 is given in Appendix, and we discuss condition of Theorem 2, which shows the performance of the FBRT algorithm in the AMRM problem.

Remark 3. According to Theorem 2, the performance of the FBRT algorithm depends on $\sigma_{\text{max}}(X)/\sigma_{\text{min}}(X)$, i.e., the condition number of the original $r$-rank matrix $X$. It plays an important role for the projection operator $P_δ$.  

In the noisy case, the error tolerance $\eta$ is a significant parameter. According to Theorem 1, we know that there exists a gap between the solution of model (1) and the original low-rank matrix. Thus, we always assume that the model error is small.

Theorem 3. Assume that $\|X\|_F \geq \mu \eta$, where rank $(X) = r$ and $\eta = \|e\|_2$ is error tolerance. Let the sequence $\{X^{(n)}\}$ be defined by the FBRT algorithm with $\|a(X) - b\|_2 \leq \eta$. If the thresholding operator in the FBRT algorithm is the BRT operator with a band parameter $c$ and $a$ satisfies

$$\delta_3 \leq \frac{\mu}{2\sqrt{4 + 2c^2r} \left(\mu + 1\right) + \left(\sqrt{4 + 2c^2r} / \sigma_{\text{min}}(X)\right) \left(\|X\|_F - \mu \eta\right)}$$

(17)

then the sequence $\{X^{(n)}\}$ must satisfy one of the following results:

(i) There exists a positive integer $n$, such that

$$\|X^{(n)} - X\|_F \leq c' \eta,$$

(18)

where $c' = 4(\mu + 1)\sqrt{2 + c^2r}$.

(ii) For any positive integer $n$, we have

$$\|X^{(n)} - X\|_F \leq \rho\|X^{(n)} - X\|_F + \frac{\sqrt{4 + 2c^2r} \left(\mu + 1\right) + \delta_2 r + 1}{1 - \rho} \eta,$$

(19)

where $\rho = \left(\mu / (\mu + 1)\right) < 1$.

The proof of this theorem is also presented in Appendix. First of all, if $\delta_3r$ satisfies the condition of Theorem 3, $\delta_3 \leq \mu / (\mu + 1) < 1$. Besides, when the error tolerance $\eta$ is not too large, the result of Theorem 3 illustrates that there exists a positive integer $n$, such that $X^{(n)}$ will be close to the original matrix $X$. Since $\mathfrak{A}$ is a bounded linear operator, it implies that $\mathfrak{A}(X^{(n)})$ is close to $\mathfrak{A}(X)$. In the mean time, the fact that $\mathfrak{A}(X^{(n)})$ is close to $\mathfrak{A}(X)$ also implies that $X^{(n)}$ is close to the original matrix $X$, according to Theorem 1. On the contrary, the parameter $\mu$ in this theorem is a key parameter, and performance of algorithm will get better as $\|X\|_F - \mu \eta$ is close to 0. Meanwhile, the performance of algorithm also depends on the condition number of the original $r$-rank matrix $X$, which is similar to Theorem 2.

5. Numerical Demonstration

In this section, we present some empirical observations of FBRT algorithms with two thresholding operator:

$$h_{\text{r,atan}}(u) = \begin{cases} u - \tau + \frac{2\tau}{\pi} \arctan(u - \tau), & \text{if } u \geq \tau, \\ 0, & \text{if } u < \tau, \end{cases} \quad \text{(Atan)}$$

(20)

and compare them with some state-of-the-art methods (singular value thresholding (SVD) algorithm [2], singular value pursuit (SVP) algorithm [16], half norm fixed point (Half) algorithm [22], and Riemann gradient descent (RGrad) algorithm [20]).

Here, we define two quantities to quantify performance of the algorithm: $\text{SR} = d/n_1 n_2$, i.e., the number of measurements divided by the number of entries of the matrix, which denotes the sampling ratio, and $\text{OS} = d / (r(n_1 + n_2 - r))$ is the oversampling ratio, i.e., the ratio between the number of sampled entries and the "true dimensionality" of an $n_1 \times n_2$ matrix of rank $r$. In fact, if $\text{OS} < 1$, we cannot recover the original low-rank matrix because there are always an infinite number of matrices of rank $r$ with the given entries [26]. We use the relative error (RE) in the Frobenius norm

$$\text{RE} := \frac{\|M - X_{\text{opt}}\|_F}{\|M\|_F},$$

(21)

to evaluate the closeness of $X_{\text{opt}}$ to $M$, where $X_{\text{opt}}$ is the "optimal" solution of the algorithm and $M$ is the original low-rank matrix.

\[alg1\]

**Algorithm 1:** Fast band restricted thresholding (FBRT).
5.1. Empirical Phase Transition. In this section, we test how many measurements are needed to recover a low rank matrix. For the sake of simplicity, we set \( n = m \) and generate rank \( r \) matrix \( X = LR \), where \( L \in \mathbb{R}^{m \times r} \) and \( R \in \mathbb{R}^{r \times n} \) and the components of \( L \) and \( R \) is sampled from the standard normal distribution. The stopping criterion is as follows:

\[
\frac{\|X^k - X^{k-1}\|_F}{\|X^{k-1}\|_F} \leq 10^{-8}.
\]

Simulations of FBRT algorithms are repeated for 10 times, and numerical results are reported in Table 2. We consider an algorithm to successfully recover the low rank matrix \( X \) if the “optimal” solution of this algorithm with \( \text{RE} \leq 0.01 \). Furthermore, we denote \( r_{\text{min}} \) as the largest rank such that the corresponding max \( \text{RE} \leq 0.01 \) and \( r_{\text{max}} \) as the smallest rank such that the corresponding min \( \text{RE} \geq 0.01 \). \( OS_{\text{max}} \) is computed via \( r_{\text{min}} \) and \( OS_{\text{min}} \) is computed via \( r_{\text{max}} \). Figure 2 shows the empirical of the tested algorithms on the (SR and OS), where we calculate oversampling ratio OS via \( r = (r_{\text{min}} + r_{\text{max}})/2 \).

Table 2 and Figure 2 show that these two FBRT algorithms can affect recover rank \( r \) matrices, where OS is close to 1, and there is not much difference between two empirical phase transition curves.

5.2. Comparison with the State-of-the-Art Algorithms in Noise-Free Case. In this section, we consider the noise-free case and test the performances of FBRT algorithms on matrix completion problems. We also generate a low rank matrix in the way of Section 5.1. The stopping criterion is as follows:

\[
\frac{\|X^k - X^{k-1}\|_F}{\|X^{k-1}\|_F} \leq 10^{-8}.
\]

Simulations are repeated for 10 times, and numerical results are reported in Tables 3 and 4 and Figure 3.

In Table 3, we show the comparison experiments with SVT, SVP, and Half. We conduct the tests for \( n \in \{400, 1000\} \), \( r = 50 \), and OS = 2 and record the average, maximum, and minimum values of relative errors and running time, respectively. Based on Table 3, we can observe that the relative errors of FBRT algorithms are smaller, and the running time is also faster. In the meantime, the running time of FBRT algorithms increases slowly as \( n \) increases. We explore the trends, displayed in Figure 3, between the relative error and the running time of the first time algorithm operation results. In Table 4, we show the comparison experiments with RGrad. We conduct the tests in two cases: \( n = 100, r = 19, SR = 0.4 \) and \( n = 1000, r = 50, OS = 2 \). Based on Table 4, we see that the running time of different

### Table 2: Numerical results of two FBRT algorithms: FBRT-Atan and FBRT-firm \((a = 3)\) with fixed \( m = n = 300 \). For any SR, the maximum of 10 relative errors is less than 0.01 when \( r \leq r_{\text{min}} \) and the minimum of 10 relative errors is greater than or equal to 0.01 when \( r > r_{\text{max}} \).

<table>
<thead>
<tr>
<th>( z )</th>
<th>( r_{\text{min}} )</th>
<th>( r_{\text{max}} )</th>
<th>Min RE</th>
<th>Max RE</th>
<th>Min RE</th>
<th>Max RE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>11</td>
<td>13</td>
<td>1.0260e-5</td>
<td>0.0023</td>
<td>2.1594e-5</td>
<td>0.0068</td>
</tr>
<tr>
<td>0.2</td>
<td>26</td>
<td>28</td>
<td>6.1794e-6</td>
<td>6.5852e-5</td>
<td>8.8405e-6</td>
<td>2.2534e-4</td>
</tr>
<tr>
<td>0.3</td>
<td>43</td>
<td>45</td>
<td>7.6326e-6</td>
<td>9.6796e-6</td>
<td>8.2757e-5</td>
<td>3.0298e-4</td>
</tr>
<tr>
<td>0.4</td>
<td>61</td>
<td>62</td>
<td>7.7705e-6</td>
<td>1.8222e-5</td>
<td>7.8113e-6</td>
<td>8.3863e-6</td>
</tr>
<tr>
<td>0.5</td>
<td>80</td>
<td>81</td>
<td>7.0491e-6</td>
<td>7.3872e-6</td>
<td>1.1833e-5</td>
<td>2.2862e-5</td>
</tr>
<tr>
<td>0.6</td>
<td>102</td>
<td>103</td>
<td>7.4470e-6</td>
<td>2.3292e-5</td>
<td>1.9080e-5</td>
<td>6.4072e-5</td>
</tr>
<tr>
<td>0.7</td>
<td>127</td>
<td>128</td>
<td>7.7453e-6</td>
<td>5.2701e-5</td>
<td>1.8729e-5</td>
<td>1.3984e-4</td>
</tr>
<tr>
<td>0.8</td>
<td>156</td>
<td>158</td>
<td>6.7983e-6</td>
<td>7.4203e-6</td>
<td>5.5031e-5</td>
<td>2.1316e-4</td>
</tr>
<tr>
<td>0.9</td>
<td>196</td>
<td>197</td>
<td>9.9473e-6</td>
<td>4.7302e-4</td>
<td>4.6863e-5</td>
<td>1.7010e-4</td>
</tr>
</tbody>
</table>

![Empirical phase transition](image-url)
algorithms is about the same, but the relative errors of FBRT algorithms are smaller than the relative error of RGrad. Therefore, we can find that FBRT algorithms perform better than others in noise-free case.

5.3. Comparison with the State-of-the-Art Algorithms in Noisy Case. In this section, we consider the noisy case and compare the performances of FBRT algorithms with RGrad algorithm on the image inpainting problems. Here,
we test these algorithms on the grayscale image: $419 \times 400$ intracranial venous image (IVI), and we obtain approximated low-rank image truncated from IVI with rank $r = 35$. We generate the noised image with normal distribution by

$$\text{imnoise}(\text{image}, \text{"gaussian"}, 0, \sigma^2),$$

where $\sigma^2$ is the variance of normal distribution, and generate the sample image from noised image with SR = 0.4. We consider the variance $\sigma^2 \in [0.01, 0.001]$ and iterations are

![Image](image-url)
Table 5: Numerical results of four algorithms on image inpainting (approximated low-rank IVI) with different ranks (noisy case).

<table>
<thead>
<tr>
<th>RE</th>
<th>RGrad</th>
<th>FBRT-Atan</th>
<th>FBRT-firm</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2 = 0.01$</td>
<td>0.4577 (156)</td>
<td>0.4108 (114)</td>
<td>0.4017 (121)</td>
</tr>
<tr>
<td>$\sigma^2 = 0.001$</td>
<td>0.1262 (524)</td>
<td>0.1251 (514)</td>
<td>0.1250 (578)</td>
</tr>
</tbody>
</table>

150 and 550, respectively. The original image, its approximated low-rank image, its noised image, and its recovering images of different algorithms and displayed in Figure 4, respectively. Figure 4(a) is original IVI with full rank. Figure 4(b) is a approximated low-rank image truncated from Figure 4(a) with rank $r = 35$. Figure 4(c) is a noised image of Figure 4(b) with $\sigma^2 = 0.01$. Figures 4(d)–4(f) are the recovering image via different algorithms, respectively. Numerical results for image inpainting are reported in Table 5. Comparing these comparison experiments, we find that FBRT algorithms perform much better that RGrad algorithm in image inpainting noisy case.

6. Conclusions

In this paper, we proposed the FBRT algorithm and developed the theoretical guarantees in both noise-free and noisy case. The numerical demonstration showed that this algorithm is effective. However, it is important to estimate error tolerance $\eta$, which has a great influence on FBRT algorithm. In the low SNR cases, more priori assumptions of the original low rank matrix are necessary, and we will study it in our future work. On the contrary, the condition number of the original matrix is also a significant parameter. As we know, phase retrieval [5, 6] and blind deconvolution [27, 28] problems can be transformed into a rank one matrix recovery problem. Since the condition number of a rank one matrix is always equal to one, it may be worth to study phase retrieval and blind deconvolution problems via the FBRT algorithm.

Appendix

Proof of Theorems 2 and 3

Now, let us turn to the proof of Theorems 2 and 3. Before Theorems 2 and 3, we need to denote some notations and introduce some lemmas.

We denote $I$, $I^o$, and $I^n$ as subspaces of $C^{n_1 \times n_2}$, which are described, respectively, as

$$ I = \{ U \U^* : U \in C^{n_1 \times r} \}, $$

$$ I^o = \{ U (V_n)^* : U \in C^{n_1 \times r} \}, $$

$$ I^n = \{ U (V_{n'})^* : U \in C^{n_1 \times (r+1)} \}, $$

where $\U \in C^{n_1 \times n_2}$ is the right singular vectors of $X$, $V_n \in C^{r \times n_2}$ is the right singular vectors of $X^{(n)}$, and $V_{n'} \in C^{(r+1) \times n_2}$ is the right singular vectors corresponding to the first $r + 1$ singular values of $W^{(n)}$. According to FBRT algorithm, it holds that $I^o \subset I^n$. For any matrix $M \in C^{n_1 \times n_2}$ and matrix subspaces $I, J \subset C^{n_1 \times n_2}$, $M_I$ represents $\mathcal{P}_I (M)$, where $\mathcal{P}_I (\cdot)$ is an orthogonal projection onto $I$, and we also denote $I + J$ as the sum of two subspaces and rewrite $I \cap (J)^\perp$ as $I \backslash J$.

**Lemma A.1.** Let $\sigma_{r+1}^n$ be the $(r + 1)$th singular value of the matrix $W^{(n)}$. Then,

$$ \sigma_{r+1}^n \leq \left\| (W^{(n)} - X)_{I^n} \right\|_2 \leq \left\| (W^{(n)} - X)_{I^n} \right\|_F. $$

**Proof.** Because of the definitions of $\sigma_{r+1}^n$ and $I^n$, it holds that

$$ \sigma_{r+1}^n = \min_{\dim(\ker(W_{I^n}^{(n)})) = n_2 - r} \max_{\dim(M \cap \ker(W_{I^n}^{(n)}))) = 0} \left\| W^{(n)}_{I^n} M \right\|_F $$

where the second equality is based on the definition of $I^n$. In the mean time, we project $X$ onto a subspace $I^n$ and obtain $\hat{X}_{I^n}$ described as

$$ \hat{X}_{I^n} = X_{I^n} = V_{n'}^\perp (V_n)^* . $$

Thus, rank $\left( \hat{X}_{I^n} \right) \cap \ker (X) = r$, and we obtain a subspace ker($\hat{X}_{I^n}$) with dim (ker($\hat{X}_{I^n}$)) $\geq n_2 - r$. Furthermore, it holds that

$$ \sigma_{r+1}^n = \min_{\dim(M \cap \ker(W_{I^n}^{(n)}))) = 0} \max_{\dim(M \cap \ker(W_{I^n}^{(n)}))) = 0} \left\| W^{(n)}_{I^n} M \right\|_F \leq \max_{\dim(M \cap \ker(W_{I^n}^{(n)}))) = 0} \left\| W^{(n)}_{I^n} M \right\|_F $$

$$ \leq \max_{\dim(M \cap \ker(W_{I^n}^{(n)}))) = 0} \left\| W^{(n)}_{I^n} \right\|_2 \leq \left\| (W^{(n)} - X)_{I^n} \right\|_2. $$

$$ \| (\mathcal{A}^* (e))_I \|_2 \leq \sqrt{1 + \delta} \| e \|_2. $$

**Proof.** In fact, it holds that

$$ \| (\mathcal{A}^* (e))_I \|_2 = \langle \mathcal{A}^* (e), (\mathcal{A}^* (e))_I \rangle_F = \langle e, \mathcal{A} (\mathcal{A}^* (e))_I \rangle $$

$$ \leq \| e \|_2 \| \mathcal{A}^* (\mathcal{A} (e))_I \|_2 \leq \| e \|_2 \sqrt{1 + \delta} \| (\mathcal{A}^* (e))_I \|_F. $$

and divide through by $\| (\mathcal{A}^* (e))_I \|_F$ to complete the proof.

**Lemma A.3.** Let $M_1, M_2 \in C^{n_1 \times n_2}$ and $I$ be a linear subspace of $C^{n_1 \times n_2}$. If $r_1 = \max \{ \text{rank}(M): M = \kappa_1 M_1 + \kappa_2 M_2, \kappa_1, \kappa_2 \in \mathbb{C} \}$ and $r_2 = \max \{ \text{rank}(M): M \in \text{span} [I, M_1] \}$, we have

$$ \langle (I d - \delta A^* (M_2)) \rangle_F \leq \delta_1 \| M_1 \| \| M_2 \|_F, $$

$$ \| ((I d - \delta A^* (M_1)) \|_F \leq \delta_1 \| M_1 \|_F. $$

8 Mathematical Problems in Engineering
Proof. To the first inequality, let \( S = \text{span}\{M_1, M_2\} \); thus, \((M_1)_S = M_1, (M_2)_S = M_2 \) and \( r_1 = \max(\text{rank}(M); \ M \in S) \). Therefore, we have

\[
\|\langle M_1, (I - \mathcal{A}^* \mathcal{A}) (M_2) \rangle_F \|_F = \|\langle M_1, M_2 \rangle_F - \langle \mathcal{A} (M_1), \mathcal{A} (M_2) \rangle_F \|_F \\
= \|\langle (M_1)_S, (M_2)_S \rangle_F - \langle \mathcal{A} (M_1)_S, \mathcal{A} (M_2)_S \rangle_F \|_F \\
\leq \|\langle (M_1)_S, (I - \mathcal{A}^* \mathcal{A}) (M_2)_S \rangle \|_F \leq R_1 \|\mathcal{A} (M_1)_S \|_F, \quad (A.9)
\]

and divide through by \( \|((I - \mathcal{A}^* \mathcal{A}) (M_2))_F \|_F \) to complete the proof.

\[
\text{Lemma A.4 (see [20]). Let } X^{(i)} = U_i \Sigma_i V_i^* \text{ be a rank } r \text{ matrix and } S_i \text{ be the tangent space of the rank } r \text{ matrix manifold at } X^{(i)}. \text{ Let } X \text{ be another rank } r \text{ matrix. Then,}
\]

\[
\|I - \mathcal{P}_{S_i}(X)\|_F \leq \frac{1}{\sigma_{\min}(X)} \|X - X^{(i)}\|_F. \quad (A.11)
\]

\[
\|X^{(n)} - X\|_F^2 \\
= \|\langle X^{(n)} \rangle_{F^*} - \langle X \rangle_{F^*} \|_F^2 + \|\langle X \rangle_{F^*} \|_F^2 \\
\leq 2 \|\langle W^{(n)} \rangle_{F^*} - \langle X \rangle_{F^*} \|_F^2 + \|W^{(n)}\|_{F^*}^2 + \|W^{(n)}\|_{F^*}^2 \\
+ \|\langle W^{(n)} \rangle_{F^*} - \langle X \rangle_{F^*} \|_F^2 + \|\langle X \rangle_{F^*} \|_F^2 \\
= 2 \|\langle W^{(n)} \rangle_{F^*} - \langle X \rangle_{F^*} \|_F^2 + \|\langle X \rangle_{F^*} \|_F^2 \\
= J_1 + J_2 + J_3, \quad (A.14)
\]

since the space \((C_{n \times m}, \| \cdot \|_F)\) is an Euclidean space and \((X^{(n)})_{F^*} = 0 \). In the following, we will bound \( J_1, J_2 \), and \( J_3 \) one by one.

According to the property of the orthogonal projection, we obtain \( J_1 \leq 2 \|W^{(n)} - X\|_{F^*}^2 \).

Besides, bound of \( J_3 \) begins with the following equations:

\[
\|W^{(n)}\|_{F^*}^2 = \|W^{(n)} - (X^{(n)})_{F^*}^2 + \|W^{(n)}\|_{F^*}^2 \\
= \|W^{(n)} - (X^{(n)})_{F^*}^2 + \|W^{(n)}\|_{F^*}^2 \\
= \|W^{(n)} - (X^{(n)})_{F^*}^2 + \|W^{(n)}\|_{F^*}^2. \quad (A.15)
\]
Since the definition of $I^n$, $(W^{(n)})_\tau$ is the best rank $r$ approximation of $W^{(n)}$. Thus, it implies that
\[
J_3 \leq 2\| (W^{(n)})_{r_1}^\tau - X \|_F^2 \\
\leq 2\| (W^{(n)} - X)_{r_1}^\tau \|_F^2, \tag{A.16}
\]
where $X_{r_1}^\tau = 0$.

In addition, the thresholding algorithm is a kind of algorithm for singular value of the matrix. It is easy to obtain that
\[
J_2 = 2\| \Sigma^{(n)} - H_{\sigma_{r_1}} (\Sigma^{(n)}) \|_F^2 = 2 \sum_{i=1}^r (\sigma_i^2 - h_{\sigma_{r_1}} (\sigma_i^2))^2. \tag{A.17}
\]

Because the thresholding function is a BRT function with a band parameter $c$, it holds that
\[
J_2 \leq 2c^2 r (\sigma^2) \sum_{i=1}^r (\sigma_i^2 - h_{\sigma_{r_1}} (\sigma_i^2))^2 \leq 2c^2 r \| (W^{(n)})_{r_1}^\tau - (X)_{r_1}^\tau \|_F^2, \tag{A.18}
\]
where the second inequality follows from Lemma A.1. Therefore, we have
\[
\| (X^n - X) \|_F^2 \leq J_1 + J_2 \leq (4 + 2c^2 r) \| (W^{(n)} - X)_{r_1}^\tau \|_F^2. \tag{A.19}
\]

In the meantime, we have $W^{(n+1)} = X^n + \mathcal{P}_{S_n} d^* (b - d^* (X^n))$. Thus, it implies that
\[
\| W^{(n+1)} - X \|_F = \| (X^n - X + \mathcal{P}_{S_n} d^* (b - d^* (X^n))) \|_F \\
\leq \| (X^n - X - \mathcal{P}_{S_n} d^* (X^n - X)) \|_F \] \\
\leq \| (I - \mathcal{P}_{S_n}) d^* (X^n - X) \|_F \] \\
\leq \| (I - \mathcal{P}_{S_n}) d^* (I - \mathcal{P}_{S_n}) (X^n - X) \|_F \] \\
\leq \| (I - \mathcal{P}_{S_n}) (X^n - X) \|_F, \tag{A.24}
\]
where the equality follows from the fact $(I - \mathcal{P}_{S_n}) (X) = 0$.

Therefore, we have
\[
\| W^{(n+1)} - X \|_F \leq \delta_{2r} + \delta_{2r} + \frac{\| (X^n - X) \|_F}{\sigma_{\min} (X)} \| (X^n - X) \|_F. \tag{A.25}
\]

Combining (A.19) and (A.25), it holds that
\[
\| X^{(n+1)} - X \|_F \leq \sqrt{4 + 2c^2 r} (\delta_{2r} + \delta_{2r} + \frac{\| (X^n - X) \|_F}{\sigma_{\min} (X)}) \| (X^n - X) \|_F. \tag{A.26}
\]

Considering $X^{(0)} = H_{\sigma_{r_1}} (W^{(0)}) = H_{\sigma_{r_1}} (d^* (b))$ and (A.19), it implies that
\[
\| X^{(0)} - X \|_F \leq \frac{\| (W^{(0)} - X) \|_F}{\sigma_{\min} (X)} \| (W^{(0)} - X) \|_F \] \\
\leq \sqrt{4 + 2c^2 \delta_{2r+1}} \| X \|_F. \tag{A.27}
\]
where the second inequality follows from Lemma A.3. Define
\[ \rho := \sqrt{4 + 2c^2r} \left( \delta_{2r} + \delta_3 + \frac{\sqrt{4 + 2c^2r} \delta_{2r+1}}{\sigma_{\min}(X)} \right). \] (A.28)

If \( \rho < 1 \), we use mathematical induction to prove (A.13). When \( n = 0 \), we obtain
\[
\|X^{(0)} - \Xi\|_F \leq \sqrt{4 + 2c^2r} \left( \delta_{2r} + \delta_3 + \frac{\|X^{(0)} - \Xi\|_F}{\sigma_{\min}(X)} \right) \|X^{(0)} - \Xi\|_F.
\] (A.29)

and inequality (A.13) holds true.

Furthermore, if
\[
\|X^{(k+1)} - \Xi\|_F \leq \sqrt{4 + 2c^2r} \left( \delta_{2r} + \delta_3 + \frac{\|X^{(k)} - \Xi\|_F}{\sigma_{\min}(X)} \right) \|X^{(k)} - \Xi\|_F
\]
\[
\leq \sqrt{4 + 2c^2r} \left( \delta_{2r} + \delta_3 + \frac{\rho \|X^{(k)} - \Xi\|_F}{\sigma_{\min}(X)} \right) \|X^{(k)} - \Xi\|_F \leq \cdots
\]
\[
\leq \sqrt{4 + 2c^2r} \left( \delta_{2r} + \delta_3 + \frac{\|X^{(0)} - \Xi\|_F}{\sigma_{\min}(X)} \right) \|X^{(k)} - \Xi\|_F
\]
\[
\leq \sqrt{4 + 2c^2r} \left( \delta_{2r} + \delta_3 + \frac{\sqrt{4 + 2c^2r} \delta_{2r+1}\|X^{(0)} - \Xi\|_F}{\sigma_{\min}(X)} \right) \|X^{(k)} - \Xi\|_F
\]
\[
= \rho \|X^{(k)} - \Xi\|_F,
\] (A.30)

and inequality (A.13) holds true. Furthermore, if
\[ \delta_3 \leq \frac{1}{2\sqrt{4 + 2c^2r + (4 + 2c^2r)\sqrt{r} \sigma_{\max}(X)/\sigma_{\min}(X)}}, \] (A.31)
it implies that
\[ \rho \leq \sqrt{4 + 2c^2r} \delta_3 \left( 2 + \frac{\sqrt{4 + 2c^2r} \sqrt{r} \sigma_{\max}(X)}{\sigma_{\min}(X)} \right) < 1, \] (A.32)
where \( \delta_{2r} \leq \delta_{2r+1} \leq \delta_3 \) and \( \|X\|_F \leq \sqrt{r} \sigma_{\max}(X). \)

Proof of Theorem 3. If there exists a positive integer \( n \) such that
\[
\|X^{(n)} - \Xi\|_F \leq 4(\mu + 1) \sqrt{2 + c^2r} \eta,
\] (A.33)
the result of Theorem 3 holds true. If
\[
\|X^{(n)} - \Xi\|_F > 4(\mu + 1) \sqrt{2 + c^2r} \eta,
\] (A.34)
for any positive integer \( n \), it implies that
\[
2 \left( \|X^{(n)} - \Xi\|_F - (\mu + 1) \sqrt{4 + 2c^2r} \sqrt{1 + \delta_{2r+1} \eta} \right) \geq \|X^{(n)} - \Xi\|_F,
\] (A.35)
since \( \delta_{2r+1} \leq \delta_3 < 1 \).

To prove inequality (19), we firstly need to prove
\[
\|X^{(n+1)} - \Xi\|_F \leq \rho \|X^{(n)} - \Xi\|_F + \sqrt{4 + 2c^2r} \sqrt{1 + \delta_{2r+1} \eta}.
\] (A.36)
where $\rho = (\mu/(\mu + 1)) < 1$. Similar to the discussion of Theorem 2, inequality (A.19) holds. Meanwhile, we also have

$$
\|W^{(n+1)} - \bar{X}\|_F = \|X^{(n)} - \bar{X} + P_{S_n} \sigma^* (b - \sigma(t))\|_F \\
\leq \|X^{(n)} - \bar{X} - P_{S_n} \sigma^* (X^{(n)} - \bar{X})\|_F + \|P_{S_n} \sigma^* (e)\|_F \\
\leq \|P_{S_n} \sigma^* (I - P_{S_n}) (X^{(n)} - \bar{X})\|_F + \|P_{S_n} \sigma^* (e)\|_F \\
= J_1 + J_2 + J_3 + J_4. 
$$

(A.37)

In the meantime, similar to the proof of Theorem 2, we can also obtain bounds of $J_1, J_2, J_3$, and according to Lemma A.2, bound of $J_4$ can be described as

$$
\|P_{S_n} \sigma^* (e)\|_F \leq \sqrt{1 + \delta_2 \|e\|_2}. 
$$

(A.38)

Thus, we obtain

$$
\|W^{(n+1)} - \bar{X}\|_F \leq \left(\delta_2 + \frac{\|X^{(n)} - \bar{X}\|_F}{\sigma_{\min}(\bar{X})}\right) \|X^{(n)} - \bar{X}\|_F \\
+ \sqrt{1 + \delta_2 \|e\|_2}. 
$$

(A.39)

Combining (A.19) and (A.39), it holds that

$$
\|X^{(n+1)} - \bar{X}\|_F \leq \sqrt{4 + 2c^2r} \left(\delta_2 + \delta_3 + \frac{\|X^{(n)} - \bar{X}\|_F}{\sigma_{\min}(\bar{X})}\right) \|X^{(n)} - \bar{X}\|_F \\
+ \sqrt{4 + 2c^2r} \|e\|_2 \\
\leq \sqrt{4 + 2c^2r} \left(\delta_2 + \delta_3 + \frac{\|X^{(n)} - \bar{X}\|_F}{\sigma_{\min}(\bar{X})}\right) \|X^{(n)} - \bar{X}\|_F \\
+ \sqrt{4 + 2c^2r} \|e\|_2. 
$$

(A.40)

Therefore, we have (A.35), it implies that

$$
\|X^{(n+1)} - \bar{X}\|_F \leq \sqrt{4 + 2c^2r} \left(\delta_2 + \delta_3 + \frac{2}{\sigma_{\min}(\bar{X})}\right) \|X^{(n)} - \bar{X}\|_F \\
+ \sqrt{4 + 2c^2r} \|e\|_2. 
$$

(A.41)

Considering $X^{(0)} = H_{\sigma_{\max}} (W^{(0)}) = H_{\sigma_{\max}} (\sigma^* (b))$ and (A.19), it implies that

$$
\|X^{(0)} - \bar{X}\|_F \leq \sqrt{4 + 2c^2r} \left(\delta_2 + \delta_3 + \frac{2}{\sigma_{\min}(\bar{X})}\right) \|X^{(n)} - \bar{X}\|_F \\
+ \sqrt{4 + 2c^2r} \|e\|_2. 
$$

(A.42)

where the third inequality follows from Lemmas A.2 and A.3. To prove inequality (A.36), we just need to prove the following inequality:

$$
\sqrt{4 + 2c^2r} \left(\delta_2 + \delta_3 + \frac{2}{\sigma_{\min}(\bar{X})}\right) \|X^{(n)} - \bar{X}\|_F \\
+ \sqrt{4 + 2c^2r} \|e\|_2 < \rho. 
$$

(A.43)

Based on the fact that
\[ \frac{\mu}{2 \sqrt{4 + 2c^2 r (\mu + 1)(1 + (\sqrt{4 + 2c^2 r} / \sigma_{\min}(X))(\|X\|_F - \mu \eta))} } \]

(A.44) now we use mathematical induction to prove (A.43). When \( n = 0 \), we obtain

\[ \sqrt{4 + 2c^2 r} \left( \delta_{2r} + \delta_{3r} + \frac{2}{\sigma_{\min}(X)} \left( \|X^{(0)} - X\|_F - (\mu + 1) \sqrt{4 + 2c^2 r} \sqrt{1 + \delta_{2r+1}} \eta \right) \right) \]

\[ \leq \sqrt{4 + 2c^2 r} \left( 2\delta_{3r} + \frac{2}{\sigma_{\min}(X)} \left( \|X^{(0)} - X\|_F - (\mu + 1) \sqrt{4 + 2c^2 r} \sqrt{1 + \delta_{2r+1}} \eta \right) \right) \]

\[ \leq \sqrt{4 + 2c^2 r} \left( 2\delta_{3r} + \frac{2\sqrt{4 + 2c^2 r}}{\sigma_{\min}(X)} \left( \delta_{3r} \|X\|_F - \mu \sqrt{1 + \delta_{2r+1}} \eta \right) \right) \]

\[ \leq \sqrt{4 + 2c^2 r} \left( 2\delta_{3r} + \frac{2\sqrt{4 + 2c^2 r}}{\sigma_{\min}(X)} \left( \delta_{3r} \|X\|_F - \mu \delta_{3r} \eta \right) \right) \]

\[ = \delta_{3r} \sqrt{4 + 2c^2 r} \left( 2 + \frac{2\sqrt{4 + 2c^2 r}}{\sigma_{\min}(X)} (\|X\|_F - \mu \eta) \right) \]

\[ \leq \rho, \]

where the fourth inequality sign is based on \( \sqrt{1 + \delta_{2r+1}} \geq 1 \geq \delta_{3r} \). Thus, inequality (A.43) holds true. If inequality (A.43) holds true for \( n \leq k - 1 \) (where \( k \) is a positive integer), it implies that

\[ \sqrt{4 + 2c^2 r} \left( \delta_{2r} + \delta_{3r} + \frac{2}{\sigma_{\min}(X)} \left( \|X^{(k)} - X\|_F - (\mu + 1) \sqrt{4 + 2c^2 r} \sqrt{1 + \delta_{2r+1}} \eta \right) \right) \]

\[ \leq \sqrt{4 + 2c^2 r} \left( \delta_{2r} + \delta_{3r} + \frac{2}{\sigma_{\min}(X)} \left( \frac{\rho}{\|X\|_F} \|X^{(k-1)} - X\|_F - (\mu + 1) \sqrt{4 + 2c^2 r} \sqrt{1 + \delta_{2r+1}} \eta \right) \right) \]

\[ = \sqrt{4 + 2c^2 r} \left( \delta_{2r} + \delta_{3r} + \frac{2}{\sigma_{\min}(X)} \rho \left( \|X^{(k-1)} - X\|_F - (\mu + 1) \sqrt{4 + 2c^2 r} \sqrt{1 + \delta_{2r+1}} \eta \right) \right) \]

\[ \leq \sqrt{4 + 2c^2 r} \left( \delta_{2r} + \delta_{3r} + \frac{2}{\sigma_{\min}(X)} \left( \|X^{(k-1)} - X\|_F - (\mu + 1) \sqrt{4 + 2c^2 r} \sqrt{1 + \delta_{2r+1}} \eta \right) \right) \]

\[ \leq \cdots \leq \sqrt{4 + 2c^2 r} \left( \delta_{2r} + \delta_{3r} + \frac{2}{\sigma_{\min}(X)} \left( \|X^{(0)} - X\|_F - (\mu + 1) \sqrt{4 + 2c^2 r} \sqrt{1 + \delta_{2r+1}} \eta \right) \right) \leq \rho, \]

\[ \|X^{(n+1)} - X\|_F \leq \rho \|X^{(n)} - X\|_F + \sqrt{4 + 2c^2 r} \sqrt{1 + \delta_{2r+1}} \eta. \] (A.46)

Then, when \( n = k \), we can obtain


and inequality (A.43) holds true.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundations of China (11771347, 91730306, 41390454, and 11271297).

References


