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

Fujun Zhao, Jigen Peng, Kai Sun, and Angang Cui

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

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:

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

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, AMRM 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:

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

or the corresponding unconstrained optimization problem:

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

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

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

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 \(\lambda\)


(II) It is hard to design the thresholding operator $\Phi(\cdot)$ for penalties $G(\cdot)$ except some special penalties, such as Schatten $p$-norm ($p = 0, 1, 1/2, 2/3$) [2, 16–18] and SCAD [12]

(III) Every iteration of the iterative thresholding algorithms needs to calculate the SVD, and if the size of the original matrix is large, the iterative thresholding algorithms will be time consuming

Two papers that relate to our work need to be overviewed here. Chartrand [19] firstly studied the way to compute the penalty $G(\cdot)$ via a given thresholding operator. However, his work did not show the stability guarantees of thresholding algorithms. Wei et al. [20] proposed an iterative hard algorithms with subspace projections, which had relatively lower computation complexity. However, his work did not examine the performance in the noisy case. Taking into account these work, we try to design a new algorithm and overcome these shortcomings to a certain extent.

In this paper, we design a family of new fast band restricted thresholding (FBRT) algorithms to solve model (1) and elaborate the theoretical guarantees in both noise-free and noisy cases. In fact, many related works [17, 18, 21, 22] focus on the stability of the optimal solution of the model. For the effect of noise on the algorithm, an empirical explanation is often given through numerical experiments. At the end of this paper, we give some simple examples to roughly verify our theoretical results.

The remainder of this paper is organized in the following. In Section 2, we introduce some basic preliminary results and notations. In Section 3, we describe the fast band restricted thresholding (FBRT) algorithm. In Section 4, we prove the convergence of the FBRT algorithm. In Section 5, we display the numerical simulations, and then conclude this paper in Section 6. All the proofs are presented in Appendix.

2. Preliminaries

In this section, we introduce some notations and useful results to facilitate the following research.

2.1. Notations. We denote $\mathcal{A}$ as a linear operator from $\mathbb{C}^{n_1 \times n_2}$ to $\mathbb{C}^d$. For any $X \in \mathbb{C}^{n_1 \times n_2}$, $\mathcal{A}(X)$ is $\langle X, A_i \rangle_F \delta_i$, where $\langle X, A_i \rangle_F = \text{tr}(X A_i^*)$ is a Frobenius inner product. Thus, we naturally obtain the dual operator $\mathcal{A}^*$ as follows:

$$\mathcal{A}^* : \mathbb{C}^d \rightarrow \mathbb{C}^{n_1 \times n_2}$$

$$b \rightarrow \sum_{i=1}^d b_i A_i.$$  (5)

We also denote $\overline{X}$ as the original matrix with rank $(\overline{X}) = r$ and $e = b - \mathcal{A}(\overline{X})$ as the noise with error tolerance $\eta = \|e\|_2$. For any subspaces $\Omega_1, \Omega_2 \subset \mathbb{C}^{n_1 \times n_2}$, we obtain the sum of subspaces $\Omega_1 \oplus \Omega_2$, and we also rewrite $\Omega_1 \cap (\Omega_2)^c$ as $\Omega_1 \setminus \Omega_2$.

2.2. Some Useful Results. In this section, we introduce some useful results, which show the relationship between the original low-rank matrix and the solution of model (1). For this purpose, we review the definition of RIC, which plays an important role in the following research.

**Definition 1** (see [23]). Given a linear operator $\mathcal{A} : \mathbb{C}^{n_1 \times n_2} \rightarrow \mathbb{C}^d$ and $1 \leq r \leq \min(m, n)$, the restricted isometry constant $\delta_r$ is the smallest nonnegative number, such that, for any matrix $X \in \mathbb{C}^{n_1 \times n_2}$, with rank $(X) \leq r$, we have the following inequality:

$$(1 - \delta_r)\|X\|_2^2 \leq \|\mathcal{A}(X)\|_2^2 \leq (1 + \delta_r)\|X\|_2^2,$$  (6)

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

For a given linear map $\mathcal{A}$, it is difficult to calculate the restricted isometry constant $\delta_r$, but for Gaussian random linear map, there has been a result that the number of measurements $d \geq c_r \frac{r n_1 n_2 \log (n_1 n_2)}{2}$ ($c_r$ only depends on $\delta$) 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 of 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., $\|\mathcal{A}(Z) - b\|_2, \|\mathcal{A}(Z'') - b\|_2 \leq \eta$). If $\text{rank}(Z) = \text{rank}(Z'') = r$ and $\delta_{2r} < 1$, we can obtain a stability claim of the form

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

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

then $Z$ is the unique lowest rank solution.

Taking into account the abovementioned results, it implies that the original low-rank matrix $\overline{X}$ 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 review 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 HFFPA [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 definition.
Definition 2. A function $h_\tau$ of $\mathbb{R}^n$ with parameter $\tau > 0$ is called as band restricted thresholding (BRT) function, if it satisfies

(I) $h_\tau (u) = 0$, with $u \leq \tau$

(II) $h_\tau (u) \leq h_\tau (v)$, with $\tau < u \leq v$

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

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

Given a BRT function $h_\tau$, we can obtain a thresholding operator $\mathcal{H}_\tau (\cdot)$ which is defined as

$$
\mathcal{H}_\tau (X) = \sum_{i=1}^n h_\tau (\sigma_i) u_i v_i^*,
$$

where $\sum_{i=1}^n \sigma_i u_i v_i^*$ is the SVD of $X$, and we can also have the corresponding thresholding algorithm described as

$$
W^{(k)} = X^{(k-1)} + 2\epsilon \epsilon (b - \mathcal{A}(X^{(k-1)})),
$$

$$
X^{(k)} = \mathcal{H}_\tau (W^{(k)}),
$$

where $\tau$ is the threshold parameter. In fact, every iteration of the abovementioned algorithm needs to compute the SVD of $W^{(k)}$. Thus, if $N = (n_1 + n_2)$ is large, this algorithm will be computationally expensive. To improve the efficiency of algorithm, we propose to orthogonally project $W^{(k)}$ onto a subspace $S_k$, which was first applied in [20]. $S_k$ is the tangential space of the rank $r$ matrix manifold at $X^{(k)}$, is described as

$$
S_k := \{U_k P + Q V_k^* : P \in C^{m \times n}, Q \in C^{n \times r}\},
$$

where $U_k \in C^{n \times n}$ and $V_k \in C^{m \times r}$ are the left and right singular vectors of $X^{(k)}$. Furthermore, for any matrix $Z \in C^{n \times n}$, $\mathcal{P}_k (Z)$ is defined as

$$
\mathcal{P}_k (Z) = U_k U_k^* Z + Z V_k V_k^* - U_k U_k^* Z V_k V_k^* = (I - U_k U_k^*) Z (I - V_k V_k^*).
$$

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 $\|\mathcal{A}(X) - b\|_F \leq \epsilon$. Taking into account Theorem 1, $X^{(k)}$ cannot be far from $X$, if $\mathcal{A}$ satisfies a certain RIC. On the contrary, it is an important situation of classified discussion that $X^{(0)}$ is close to $X$, and the detail will be shown in the following section.

It is worth noting that $X^{(k)} + 2\epsilon \epsilon (b - \mathcal{A}(X^{(k)}))$ 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

$$
W^{(k)} = QR = Q^* R^* Q^*,
$$

where $W^{(k)} = QR$ and $R^* = Q_1 R_1$ are the QR factorizations of $W^{(k)}$ and $R^*$, with $Q \in C^{n \times 2r}, R \in C^{2r \times n}, R_1 \in C^{2r \times 2r}$, and $Q_1 \in C^{n \times 2r}$. 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 $\{X^{(k)}\}$ be defined by the FBRT algorithm with $\mathcal{A}(X) = \mathcal{A}(X) = b$ and rank($X$) = $r$. If the thresholding operator in the FBRT algorithm is the BRT operator with a band parameter $c$ and the following constant

$$
\rho := \sqrt{4 + 2c^2} \left( \delta_{2r} + \frac{\sqrt{4 + 4c^2 r} \delta_{r+1}}{\sigma_{\min}(X)} \|X\|_F \right),
$$

is less than 1, we have

$$
\|X^{(0)} - X\|_F \leq \rho^n \|X^{(0)} - X\|_F.
$$

Particularly, $\rho < 1$ can be satisfied if

$$
\frac{1}{2r} \leq \frac{1}{2 \sqrt{4 + 2c^2} + (4 + 2c^2 r)^{1/(r+1)}} \sigma_{\max}(X)/\sigma_{\min}(X).
$$


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 $P_k$.

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 $\eta = \|e\|_1$ is error tolerance. Let the sequence $\{X^{(n)}\}$ be defined by the FBRT algorithm with $\|\mathcal{A}(X) - b\|_2 \leq \eta$. If the thresholding operator in the FBRT algorithm is the BRT operator with a band parameter $c$ and $\mathcal{A}$ satisfies

$$\delta_{3r} \leq \frac{\mu}{2\sqrt{4 + 2\sigma^2 r}} \left(\mu + 1\right) \left(1 + \frac{4\sqrt{4 + 2\sigma^2 r}}{\sigma_{\text{min}}(X)}\right) \left(\|X\|_F - \mu \eta\right),$$

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

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

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

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

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

$$\|X^{(n)} - X\|_F \leq \rho \|X^{(0)} - X\|_F + \frac{\sqrt{4 + 2\sigma^2 r} \sqrt{1 + \delta_{2r+1}} \eta}{1 - \rho},$$

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

The proof of this theorem is also presented in Appendix. First of all, if $\delta_{3r}$ satisfies the condition of Theorem 3, $\delta_{3r} \leq \mu/(\mu + 1) < 1$. 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 $\mathcal{A}$ is a bounded linear operator, it implies that $\mathcal{A}(X^{(n)})$ is close to $\mathcal{A}(X)$. In the mean time, the fact that $\mathcal{A}(X^{(n)})$ is close to $\mathcal{A}(X)$ 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:

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

$$h_{r,\text{firm}}(u) = \begin{cases} \frac{a}{a - 1} (x - u), & \text{if } \tau \leq u \leq a \cdot \tau, \\ 0, & \text{if } u < \tau, \quad \text{(firm)}, \end{cases}$$

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

Here, we define two quantities to quantify performance of the algorithm: 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 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 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

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

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.
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 \(r > r_{\text{max}}\).

\[
\begin{array}{cccccc}
\hline
z & \text{FBRT-Atan} & & & \text{FBRT-firm} & \\
& \text{Min RE} & \text{Max RE} & \text{Min RE} & \text{Max RE} \\
\hline
0.1 & r_{\text{min}} = 11 & 1.0260e-5 & 0.0023 & r_{\text{min}} = 11 & 2.5949e-5 \\
& r_{\text{max}} = 13 & 0.0079 & 0.6592 & r_{\text{max}} = 12 & 2.7247e-4 \\
0.2 & r_{\text{min}} = 26 & 6.1794e-6 & 6.5852e-5 & r_{\text{min}} = 27 & 8.8405e-6 \\
& r_{\text{max}} = 28 & 3.3154e-4 & 0.5499 & r_{\text{max}} = 28 & 3.6679e-4 \\
0.3 & r_{\text{min}} = 43 & 7.6326e-6 & 9.6796e-6 & r_{\text{min}} = 44 & 8.2757e-5 \\
& r_{\text{max}} = 45 & 0.0017 & 0.5188 & r_{\text{max}} = 45 & 0.0032 \\
0.4 & r_{\text{min}} = 61 & 7.7705e-6 & 1.8222e-5 & r_{\text{min}} = 61 & 7.8113e-6 \\
& r_{\text{max}} = 62 & 6.8547e-5 & 0.4236 & r_{\text{max}} = 63 & 0.0011 \\
0.5 & r_{\text{min}} = 80 & 7.0491e-6 & 7.3872e-6 & r_{\text{min}} = 81 & 1.1833e-5 \\
& r_{\text{max}} = 81 & 1.1140e-5 & 0.3924 & r_{\text{max}} = 83 & 0.0059 \\
0.6 & r_{\text{min}} = 102 & 7.4470e-6 & 2.3292e-5 & r_{\text{min}} = 103 & 1.9080e-5 \\
& r_{\text{max}} = 103 & 2.2913e-5 & 0.3510 & r_{\text{max}} = 105 & 0.0016 \\
0.7 & r_{\text{min}} = 127 & 7.7453e-6 & 5.2701e-5 & r_{\text{min}} = 128 & 1.8729e-5 \\
& r_{\text{max}} = 128 & 5.1908e-5 & 0.2978 & r_{\text{max}} = 129 & 1.5370e-4 \\
0.8 & r_{\text{min}} = 156 & 6.7983e-6 & 7.4203e-6 & r_{\text{min}} = 158 & 5.5031e-5 \\
& r_{\text{max}} = 158 & 1.3876e-4 & 0.2559 & r_{\text{max}} = 159 & 3.2228e-4 \\
0.9 & r_{\text{min}} = 196 & 9.9473e-6 & 4.7502e-4 & r_{\text{min}} = 197 & 4.6863e-5 \\
& r_{\text{max}} = 197 & 2.5510e-4 & 0.1592 & r_{\text{max}} = 198 & 0.0011 \\
\hline
\end{array}
\]

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 \(R \in \mathbb{R}^{r \times r}\) and the components of \(L\) and \(R\) is sampled from the standard normal distribution. The stopping criterion is as follows:

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

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 RE \(\leq 0.01\) and \(r_{\text{max}}\) as the smallest rank such that the corresponding min RE \(\geq 0.01\). OS \(= r_{\text{min}} \text{ max} \) is computed via \(r_{\text{min}}\), and OS \(= r_{\text{min}} \text{ max} \) 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 \(r = \left(\left(r_{\text{min}} + r_{\text{max}}\right)\right)/2\).

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 in the way of Section 5.1. The stopping criterion is as follows:

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

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: \(n = 100, r = 19, \text{SR} = 0.4\) and \(n = 1000, r = 50, \text{OS} = 2\). Based on Table 4, we see that the running time of different
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 × 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
I, J, respectively, as described, respectively, as 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 $I$, $I^*$, and $I^n$ as subspaces of $C^{n \times n}$, which are described, respectively, as

$$
I = \{ U \mathbf{V}^{*} : U \in C^{n \times r} \},
$$

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

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

where $\mathbf{V}$ is the $r$ right singular vectors of $\mathbf{X}$, $V_n \in C^{r \times n}$ is the right singular vectors of $\mathbf{X}^{(n)}$, and $V_n^* \in C^{(r+1) \times n}$ is the right singular vectors corresponding to the first $r + 1$ singular values of $W^{(n)}$. According to FBRT algorithm, it holds that $I^* \subseteq I^n$. For any matrix $M \in C^{n \times n}$ and matrix subspaces $I, J \subset C^{n \times n}$, $M_j$ represents $\mathcal{P}_j(M)$, where $\mathcal{P}_j(\cdot)$ 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 $I \setminus J$.

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

$$
\sigma_{r+1}^n \leq \left( \frac{\|W^{(n)} - \mathbf{X}\|_F}{\|W^{(n)} - \mathbf{X}\|_F} \right)_{I^n}. \tag{A.2}
$$

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

$$
\sigma_{r+1}^n = \min_{\dim(M) \leq r} \max_{U \in M} \left\{ \|W^{(n)} U\|_2 \right\}, \tag{A.3}
$$

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

$$
\mathbf{X}_{I^n} = \mathbf{X} V_n (V_n^*)^T. \tag{A.4}
$$

Thus, $\text{rank} (\mathbf{X}_{I^n}) \leq \text{rank} (\mathbf{X}) = r$, and we obtain a subspace $\ker (\mathbf{X}_{I^n})$ with $\dim (\ker (\mathbf{X}_{I^n})) \geq n - r$. Furthermore, it holds that

$$
\sigma_{r+1}^n = \min_{\dim(M) \leq r} \max_{U \in M} \left\{ \|W^{(n)} U\|_2 \right\}, \tag{A.5}
$$

Lemma A.2. Let $e \in C^d$ and $S$ be a linear subspace of $C^{n \times n}$. If $r = \max \{ \text{rank} (M) : M \in S \}$, we have

$$
\| (\mathcal{A}^* (e))_S \|_F \leq \sqrt{1 + \delta} \| e \|_2. \tag{A.6}
$$

Proof. In fact, it holds that

$$
\| (\mathcal{A}^* (e))_S \|_F = \| (\mathcal{A}^* (e), (\mathcal{A}^* (e))_S) \|_F = \langle (\mathcal{A}^* (e)), (\mathcal{A}^* (e)_S) \rangle \leq \langle e, \mathcal{A}^* (\mathcal{A}^* (e))_S \rangle \leq \| e \|_2 \sqrt{1 + \delta} \| (\mathcal{A}^* (e))_S \|_F, \tag{A.7}
$$

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

Lemma A.3. Let $M_1, M_2 \in C^{n \times n}$, and $I$ be a linear subspace of $C^{n \times n}$. If $r_1 = \max \{ \text{rank} (M) : M = k_1 M_1 + k_2 M_2, k_1, k_2 \in C \}$ and $r_2 = \max \{ \text{rank} (M) : M \in \text{span} \{ I, M_1 \} \}$, we have

$$
\| (I d - \mathcal{A}^* (\mathcal{A})) (M_2) \|_F \leq \delta \| M_1 \|_F \| M_2 \|_F, \tag{A.8}
$$

$$
\| (I d - \mathcal{A}^* (\mathcal{A})) (M_1) \|_F \leq \delta \| M_1 \|_F \| M_1 \|_F.
$$
\textbf{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

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

To the second inequality, we have

\[
\|((I - \mathbb{A}^* \mathbb{A}) (M_2))\|_F^2
= \left| \langle ((I - \mathbb{A}^* \mathbb{A}) (M_2)) \rangle_F, (I - \mathbb{A}^* \mathbb{A}) (M_1) \rangle_F \right|
\leq \delta_r \|((I - \mathbb{A}^* \mathbb{A}) (M_2))\|_F \|M_1\|_F,
\] (A.10)

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

\textbf{Lemma A.4 (see [20]).} Let \( X^{(l)} = U_l \Sigma V_l^* \) be a rank \( r \) matrix with the tangent space \( S_l \). Let \( X \) be another rank \( r \) matrix. Then, \( X^{(l+1)} = \mathcal{P}_{S_l} (X) \) can be bounded as

\[
\|\mathcal{P}_{S_l} \mathbb{A} (I - \mathcal{P}_{S_l}) (X) \|_F \leq \delta_r \|I - \mathcal{P}_{S_l}\|_F \|X\|_F.
\] (A.12)

\textbf{Lemma A.4 and A.5} can be found in [20] and the details of the proofs are omitted here. Based on lemmas above, we start to prove Theorems 2 and 3.

\textbf{Proof of Theorem 2.} To prove inequality (15), we just need to prove

\[
\|X^{(n+1)} - X\|_F \leq \rho \|X^{(n)} - X\|_F.
\] (A.13)

First of all, we have the following inequality:

\[
\|X^{(n)} - X\|_F^2
= \|X^{(n)} - (X)_F\|_F^2 + \|X\|_F^2 - (X)_F^2
\leq 2 \left( \|X^{(n)} - (X)_F\|_F^2 + \|W^{(n)}\|_F^2 \right)
\leq 2 \left( \|X^{(n)}\|_F^2 + \|W^{(n)}\|_F^2 \right)
= 2 \left( \|X^{(n)}\|_F^2 + \|W^{(n)}\|_F^2 \right)
= J_1 + J_2 + J_3,
\]

since the space \((C_n^{m,n}, \|\cdot\|_F)\) is an Euclidean space and \((X^{(n)})_{T,F} = 0\). 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 \( J_1 \leq 2 \|W^{(n)} - (X)\|_{T,F}^2 \).

Besides, bound of \( J_1 \) begins with the following equations:

\[
\|W^{(n)}\|_F^2
= \|W^{(n)} - (X)\|_{T,F}^2 + \|W^{(n)}\|_{T,F}^2
= \|W^{(n)}\|_{F,F}^2 + \|W^{(n)}\|_{T,F}^2
= \|W^{(n)}\|_{F,F}^2 + \|W^{(n)}\|_{T,F}^2
= \|W^{(n)}\|_{F,F}^2 + \|W^{(n)}\|_{T,F}^2.
\] (A.15)
Since the definition of $I^n$, $(W^{(n)})_r$ is the best rank $r$ approximation of $W^{(n)}$. Thus, it implies that
\[
J_3 \leq 2 \left\| (W^{(n)})_r - X \right\|_F^2 = 2 \left\| (W^{(n)} - X)_r \right\|_F^2 \leq 2 \left\| (W^{(n)} - X)_{r+1} \right\|_F^2,
\] (A.16)
where $X_{r+1} = 0$.

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

Because the thresholding function is a BRT function with a band parameter $c$, it holds that
\[
J_2 \leq 2c^2 r (\sigma_{r+1})^2 \leq 2c^2 r \left\| (W^{(n)})_r - (X)_r \right\|_F^2 \leq 2c^2 r \left\| (W^{(n)} - X)_{r+1} \right\|_F^2,
\] (A.18)
where the second inequality follows from Lemma A.1. Therefore, we have
\[
\left\| (X^{(n)} - X) \right\|_F^2 \leq J_1 + J_2 + J_3 \leq (4 + 2c^2 r) \left\| (W^{(n)} - X)_{r+1} \right\|_F^2.
\] (A.19)

In the meantime, we have
\[
W^{(n+1)} = X^{(n)} + \mathcal{P}_{S^n}(b - \mathcal{A}(X^{(n)})).
\]
Thus, it implies that
\[
\left\| W^{(n+1)} - X \right\|_F \leq \left\| X^{(n)} - X + \mathcal{P}_{S^n}(b - \mathcal{A}(X^{(n)})) \right\|_F
\leq \left\| X^{(n)} - X - \mathcal{P}_{S^n}(b - \mathcal{A}(X^{(n)})) \right\|_F.
\]
\[
\leq \left\| X^{(n)} - X - \mathcal{P}_{S^n}(b - \mathcal{A}(X^{(n)})) \right\|_F
\leq \left\| (I - \mathcal{P}_{S^n})\left( X^{(n)} - X \right) \right\|_F.
\]
\[
\leq \left\| (I - \mathcal{P}_{S^n})\left( X^{(n)} - X \right) \right\|_F + \left\| \mathcal{A}(b - \mathcal{A}(X^{(n)})) \right\|_F.
\] (A.20)
where the second equality follows from the fact $(I - \mathcal{P}_{S^n})(X^{(n)}) = 0$.

Bound of $J_1$: it is a fact that RIC can be represented as
\[
\delta_{2r} = \sup_{0 < \text{rank}(X) < 2r} \frac{\left\| \mathcal{A}(X) \right\|_F^2 - \left\| X \right\|_F^2}{\left\| X \right\|_F^2}.
\]
\[
\geq \sup_{0 < \text{rank}(X)} \frac{\left\| \left( \mathcal{P}_{S^n} - \mathcal{P}_{S^n}^* \mathcal{A}^* \mathcal{P}_{S^n} \right)(X) \right\|_F^2}{\left\| X \right\|_F^2}.
\] (A.21)

where the second equal sign is based on the fact that $\mathcal{P}_{S^n} - \mathcal{P}_{S^n}^* \mathcal{A}^* \mathcal{P}_{S^n}$ is a self-adjoint operator. Thus, the first term $J_1'$ can be bounded as
\[
J_1' \leq \delta_{2r} \left\| (X^{(n)} - X) \right\|_F.
\] (A.22)

Bound of $J_1'$: according to Lemma A.4, the second term $J_1'$ can be bounded as
\[
J_1' \leq \frac{1}{\sigma_{\min}(X)} \left\| X^{(n)} - X \right\|_F^2.
\] (A.23)

Bound of $J_2'$: according to Lemma A.5, the third term $J_2'$ can be bounded as
\[
J_2' \leq \delta_{3r} \left\| (I - \mathcal{P}_{S^n})X \right\|_F \leq \delta_{3r} \left\| (I - \mathcal{P}_{S^n})(X^{(n)} - X) \right\|_F.
\] (A.24)

where the equality follows from the fact $(I - \mathcal{P}_{S^n})(X) = 0$.

Therefore, we have
\[
\left\| W^{(n+1)} - X \right\|_F \leq \left( \delta_{2r} + \delta_{3r} + \frac{\left\| X^{(n)} - X \right\|_F^2}{\sigma_{\min}(X)} \right) \left\| X - X \right\|_F.
\] (A.25)

Combining (A.19) and (A.25), it holds that
\[
\left\| X^{(n+1)} - X \right\|_F \leq \sqrt{4 + 2c^2 r} \left( \delta_{2r} + \delta_{3r} + \frac{\left\| X^{(n)} - X \right\|_F^2}{\sigma_{\min}(X)} \right) \left\| X^{(n)} - X \right\|_F.
\] (A.26)

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

If \( \rho < 1 \), we use mathematical induction to prove (A.13). When \( n = 0 \), we obtain

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

and inequality (A.13) holds true.

Furthermore, if

\[
\delta_{3r} \leq \frac{1}{2\sqrt[4]{4 + 2c^2r} + (4 + 2c^2r) \sqrt{r} \sigma_{\max}(X)/\sigma_{\min}(X)},
\]
(A.31)
it implies that

\[
\rho \leq \sqrt[4]{4 + 2c^2r} \delta_{2r} \left( 2 + \frac{\sqrt{4 + 2c^2r} \sqrt{r} \sigma_{\max}(X)}{\sigma_{\min}(X)} \right) < 1,
\]
(A.32)

where \( \delta_{2r} \leq \delta_{2r+1} \leq \delta_{3r} \) and \( \|X\|_F \leq \sqrt{r} \sigma_{\max}(X) \).

Proof of Theorem 3. If there exists a positive integer \( n \) such that

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

and inequality (A.13) holds true.

If inequality (A.13) holds true for \( n \leq k - 1 \) (where \( k \) is a positive integer), then when \( n = k \), we obtain

\[
\|X^{(n)} - X\|_F \leq 4(\mu + 1) \sqrt{2 + c^2r} \eta,
\]
(A.33)

the result of Theorem 3 holds true. If

\[
\|X^{(n)} - X\|_F > 4(\mu + 1) \sqrt{2 + c^2r} \eta,
\]
(A.34)

for any positive integer \( n \), it implies that

\[
2 \left( \|X^{(n)} - X\|_F - (\mu + 1) \sqrt{4 + 2c^2r} \sqrt{1 + \delta_{2r+1} \eta} \right) \geq \|X^{(n)} - X\|_F,
\]
(A.35)

since \( \delta_{2r+1} \leq \delta_{3r} < 1 \).

To prove inequality (19), we firstly need to prove

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

\[
\left\| W^{(n+1)} - \overline{X} \right\|_F = \left\| X^{(n)} - \overline{X} + \mathcal{S}_n \mathcal{A}^* (b - \mathcal{A} (X^{(n)})) \right\|_F
\]

\[
= \left\| X^{(n)} - \overline{X} + \mathcal{S}_n \mathcal{A}^* (\mathcal{A} (X) + e - \mathcal{A} (X^{(n)})) \right\|_F
\]

\[
\leq \left\| X^{(n)} - \overline{X} - \mathcal{S}_n \mathcal{A}^* (X^{(n)} - \overline{X}) \right\|_F + \left\| \mathcal{S}_n \mathcal{A}^* (e) \right\|_F
\]

\[
\leq \left\| (\mathcal{S}_n - \mathcal{S}_n \mathcal{A}^* \mathcal{A} \mathcal{S}_n) (X^{(n)} - \overline{X}) \right\|_F + \left\| (1 - \mathcal{S}_n) (X^{(n)} - \overline{X}) \right\|_F
\]

\[
+ \left\| \mathcal{S}_n \mathcal{A}^* (1 - \mathcal{S}_n) (X^{(n)} - \overline{X}) \right\|_F + \left\| \mathcal{S}_n \mathcal{A}^* (e) \right\|_F
\]

\[
= J_1 + J_2 + J_3 + J_4.
\]

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

\[
\left\| \mathcal{S}_n \mathcal{A}^* (e) \right\|_F \leq \sqrt{1 + \delta_{2r}} \left\| e \right\|_2.
\]

Thus, we obtain

\[
\left\| W^{(n+1)} - \overline{X} \right\|_F \leq \left( \delta_{2r} + \delta_{3r} + \frac{\left\| X^{(n)} - \overline{X} \right\|_F}{\sigma_{\min} (X)} \right) \left\| X^{(n)} - \overline{X} \right\|_F
\]

\[
+ \sqrt{1 + \delta_{2r}} \left\| e \right\|_2.
\]

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

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

\[
+ \sqrt{4 + 2c^2 \tau} \sqrt{1 + \delta_{2r}} \left\| e \right\|_2
\]

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

\[
+ \sqrt{4 + 2c^2 \tau} \sqrt{1 + \delta_{2r+1}} \left\| e \right\|_2.
\]

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

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

now we use mathematical induction to prove (A.43).

When \( n = 0 \), we obtain

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

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

\[ \|X^{(k+1)} - X\|_F \leq \rho \|X^{(k)} - X\|_F + \sqrt{4 + 2c^2r\sqrt{1 + \delta_{2r+1}}} \eta. \] (A.47)
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
[1] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," Foundations of Computational Mathematics, vol. 9, no. 6, pp. 717–772, 2009.

[2] E. J. Candès, J. K停滞, and Z. Shen, "A singular value thresholding algorithm for matrix completion," SIAM Journal on Optimization, vol. 20, no. 4, pp. 1956–1982, 2008.

[3] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" Journal of the ACM, vol. 58, no. 3, 2000.

[4] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 1, pp. 171–184, 2013.

[5] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," SIAM Review, vol. 57, no. 2, pp. 225–251, 2015.

[6] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: theory and algorithms," IEEE Transactions on Information Theory, vol. 61, no. 4, pp. 1985–2007, 2015.

[7] Y. Yu, J. Peng, and S. Yue, "A new nonconvex approach to low-rank matrix completion with application to image inpainting," Multidimensional Systems and Signal Processing, vol. 30, no. 1, pp. 145–174, 2019.

[8] Y. Gao, X. Han, and M. Ma, "Recovery of low-rank matrices based on the rank null space properties," International Journal of Wavelets, Multiresolution and Information Processing, vol. 15, no. 4, p. 1750032, 2017.

[9] F. Zhao, J. Peng, and A. Cui, "Design strategy of thresholding operator for low-rank matrix recovery problem," Signal Processing, vol. 171, Article ID 107510, 2020.

[10] I. E. Frank and J. H. Friedman, "A statistical view of some chemometrics regression tools," Technometrics, vol. 35, no. 2, pp. 109–135, 1993.

[11] C.-H. Zhang, "Nearly unbiased variable selection under minimax concave penalty," The Annals of Statistics, vol. 38, no. 2, pp. 894–942, 2010.

[12] J. Fan and R. Li, "Variable selection via nonconcave penalized likelihood and its oracle properties," Journal of the American Statistical Association, vol. 96, no. 456, pp. 1348–1360, 2001.

[13] J. H. Friedman, "Fast sparse regression and classification," International Journal of Forecasting, vol. 28, no. 3, pp. 722–738, 2012.

[14] D. Geman and C. Yang, "Nonlinear image recovery with quadratic regularization," IEEE Transactions on Image Processing, vol. 4, no. 7, pp. 932–946, 1995.