

## A Novel Nonconvex Rank Approximation with Application to the Matrix Completion

Jin-Liang Xiao, Ting-Zhu Huang\*, Zhong-Cheng Wu  
and Liang-Jian Deng

*School of Mathematical Sciences, University of Electronic Science  
and Technology of China, Chengdu 611731, PR. China.*

*Received 12 October 2023; Accepted (in revised version) 14 August 2024.*

---

**Abstract.** The matrix rank approximation has shown high effectiveness in the matrix rank minimization (MRM) problem, which aims to recover the underlying low-rank structure from the observed matrix by imposing the rank constraint. The nuclear norm, serving as a convex surrogate of matrix rank, is employed in the MRM problem by shrinking singular values of the observed entry. However, this substitution treats each singular value equally, which is virtually  $\ell_1$ -norm penalty of the singular value vector. Theoretically, the rank function of the matrix can be considered as  $\ell_0$ -norm of its singular values. Consequently, minimizing the nuclear norm frequently results in biased solutions in various applications. In this article, we first propose a novel nonconvex rank approximation, named tight and flexible rank (TFR) approximation, to describe rank function effectively. Specifically, the TFR approximation can more tightly approach the rank function and exhibit greater flexibility in handling diverse singular values, as compared to existing nonconvex rank approximations. Furthermore, we apply TFR approximation to matrix completion and develop a solving algorithm with guaranteed convergence based on the framework of proximal alternating minimization. Extensive experiments reveal that the proposed matrix completion model with TFR approximation outperforms several existing state-of-the-art convex and nonconvex methods.

**AMS subject classifications:** 68U10, 65K10, 15A83

**Key words:** Rank minimization, low-rank matrix completion, rank approximation.

---

### 1. Introduction

Matrices are widely used in various fields, including computer vision and machine learning [10, 27, 43–45, 47, 60, 61], where some of their features, especially the low-rank property, can be utilized [36, 39, 56, 63]. It is worth noting that many related tasks — e.g. matrix completion [5], compressive sensing [11], and image denoising [13, 38, 70], can be

---

\*Corresponding author. Email addresses: jinliang\_xiao@163.com (J.-L. Xiao), tingzhuhuang@126.com (T.-Z. Huang), wuzhch97@163.com (Z.-C. Wu), liangjian.deng@uestc.edu.cn (L.-J. Deng)

described as the matrix rank minimization (MRM) problem, which consists in finding an approximate low-rank matrix from its degraded observation by the rank constraint [29]. Mathematically, this rank minimization problem can be formulated as follows:

$$\begin{aligned} & \min_{\mathbf{X} \in \mathbb{R}^{m \times n}} \text{rank}(\mathbf{X}) \\ & \text{s.t. } \mathcal{A}(\mathbf{X}) = \mathbf{B}, \end{aligned} \tag{1.1}$$

where  $\mathbf{X} \in \mathbb{R}^{m \times n}$  is the underlying matrix,  $\mathcal{A} : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^{m \times n}$  a linear map, and  $\mathbf{B} \in \mathbb{R}^{m \times n}$  the observed matrix. The choice of  $\mathcal{A}$  depends on the specific application. Since the direct solution of the problem (1.1) is NP-hard [6], it usually depends on replacing the discontinuous rank function by an appropriate matrix rank approximation [20, 21, 62]. The later can be represented as

$$\begin{aligned} & \min_{\mathbf{X} \in \mathbb{R}^{m \times n}} \Psi(\mathbf{X}) \\ & \text{s.t. } \mathcal{A}(\mathbf{X}) = \mathbf{B}, \end{aligned} \tag{1.2}$$

where  $\Psi(\mathbf{X})$  is the rank approximation of matrix  $\mathbf{X}$ . Because the matrix rank is the number of non-zero singular values — i.e.  $\ell_0$ -norm of the singular value vector,  $\Psi(\mathbf{X})$  is usually defined as a function of singular values [24, 71]. Thus, the MRM problem (1.1) is often approached by minimizing an appropriate rank approximation that penalizes singular values. This method is widely-used in various applications [40, 48, 65].

As the tightest convex approximation of the matrix rank, the nuclear norm is defined as the sum of the singular values. This transforms the minimization of the matrix rank into a constraint on singular values of the underlying matrix [15]. Candès and Recht [7] proved that low-rank structures can be extracted from the degraded matrix by minimizing the nuclear norm with a high probability. The application of the nuclear norm further demonstrates the effectiveness of the rank approximation defined by singular values [25]. Virtually, the singular value provides quantifiable information of the matrix. For example, larger singular values usually contain significant information about textures and edges [59]. However, the nuclear norm treats the singular values of the matrix equally. As a result, the nuclear norm shrinks the same value for each singular value. Numerous studies — e.g. [33, 50, 68], show that usually such a uniform shrinkage leads to a restricted performance.

Theoretically, the rank function of the matrix is  $\ell_0$ -norm of its singular values, and the relationship between the nuclear norm and the rank of matrices can be seen as the relationship between  $\ell_0$ -norm and  $\ell_1$ -norm of singular value vectors [37], cf. Section 2 for more details. Clearly, there exists a distance between two norms for the constraint on singular values, which limits the performance of the nuclear norm. Note that there are many nonconvex matrix rank approximations aimed to better describe the  $\ell_0$ -norm for singular values, [35, 40]. In particular, Hu *et al.* [24] proposed the truncated nuclear norm, defined as the sum of the smaller singular values. Dong *et al.* [11] achieved promising results on compressive sensing by using the nonconvex logdet function as the surrogate of the rank function. Kang *et al.* [26] successfully applied the logdet function to recommender system via matrix completion. Nie *et al.* [34] utilized the Schatten p-norm for low-rank matrix restoration. Chen *et al.* [8] proposed the logarithmic norm to induce a sparsity-driven

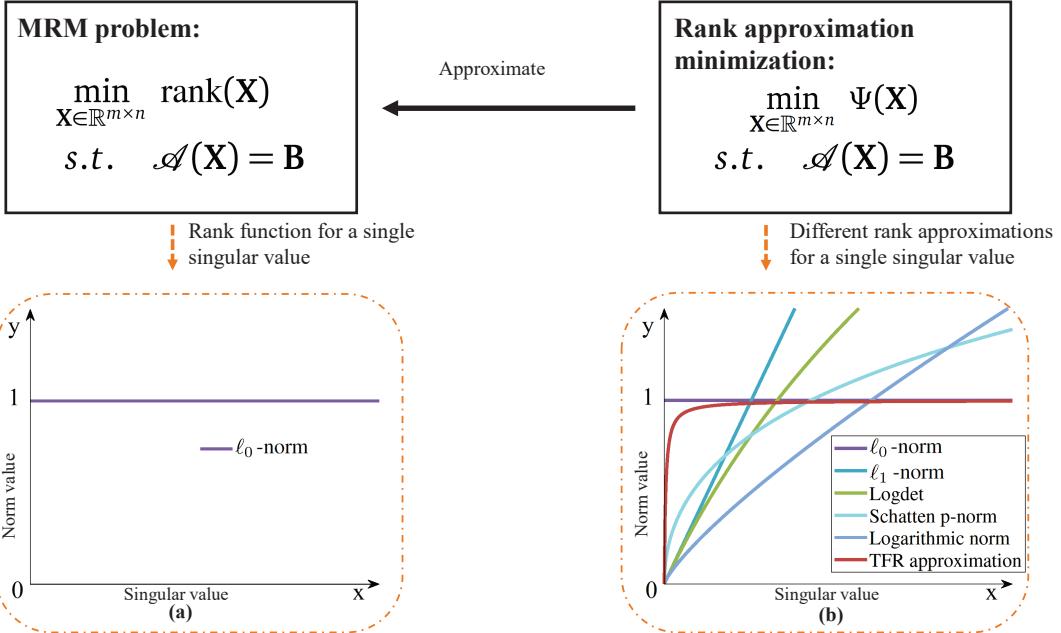


Figure 1: The relationship between MRM problem and rank approximation minimization. (a) The rank function for singular values. (b) The similarity comparison between different matrix rank approximations and  $\ell_0$ -norm. It is evident that our TFR approximation is closer to  $\ell_0$ -norm compared with the convex and nonconvex rank surrogates — i.e.  $\ell_1$ -norm [5], logdet function [11], Schatten p-norm [34], and logarithmic norm [8].

surrogate of the rank function. However, as Fig. 1(b) shows, neither of these approaches produces an appropriate approximation of the  $\ell_0$ -norm. This observation provides the impetus to explore a more stringent approximation of the matrix rank function, one that aligns notably well with the  $\ell_0$ -norm characteristic of the singular value vector. Furthermore, in practical applications, the real image data typically exhibit approximate low-rank structures rather than an absolute low-rank profile. This means that there are still slight perturbations around  $\ell_0$ -norm according to specific data [33]. Consequently, the rank approximation has to be flexible while approaching  $\ell_0$ -norm to achieve excellent results [53]. Hence, it is imperative for the rank approximation to contemplate both flexibility and approximation [42].

In this paper, we first propose a nonconvex tight and flexible rank (TFR) approximation. On the one hand, unlike the current matrix rank approximations, the TFR approximation is closer to  $\ell_0$ -norm of singular values. On the other hand, as displayed in Fig. 2, the proposed TFR approximation can flexibly treat different singular values by the appropriate parameter selection, which is essential in practical applications [22, 53]. In addition, the TFR proximal problem of singular values — i.e. (3.8), is the critical step in minimizing the TFR approximation. To address this proximal problem, we design a convergent algorithm — viz. Algorithm 3.1. Finally, to verify the practical potential of the proposed TFR approximation, we give a new matrix completion model with the TFR approximation. Subsequently, an al-

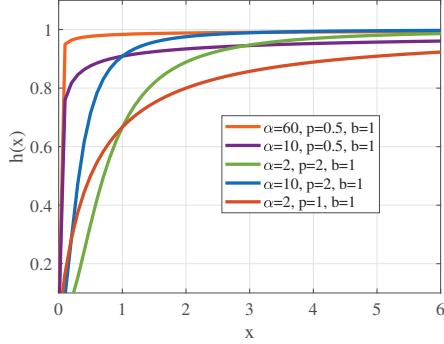


Figure 2: The proposed TFR approximation with different parameters. It is clear that the TFR approximation can flexibly treat different singular values by parameter adjustment.

gorithm based on the proximal alternating minimization (PAM) framework [4] is developed to solve the model, and the convergence guarantee of the algorithm — i.e. Algorithm 4.1, is provided. Experiments on synthetic and real data demonstrate the effectiveness of the novel TFR approximation in matrix completion.

The main contributions of this paper can be summarized as follows:

1. We propose a novel TFR approximation, which can achieve a tighter approximation of the matrix rank and flexibly penalize different singular values, to well describe  $\ell_0$ -norm for singular values. Also, we give a convergent algorithm to solve the TFR proximal problem of singular values.
2. We build a new matrix completion model with the TFR approximation and develop a PAM-based algorithm with convergence guarantee to solve it.
3. Extensive experiments demonstrate that the proposed matrix completion model achieves excellent performance compared to many state-of-the-art methods, which further verifies the great practical potential of the TFR approximation.

The rest of this paper is organized as follows. Section 2 provides some preliminaries. Section 3 introduces the TFR approximation. Section 4 applies TFR approximation to matrix completion and provides the model, algorithm, and convergence guarantee. Section 5 conducts the numerical experiments to verify the effectiveness. Finally, Section 6 draws conclusions.

## 2. Preliminaries

In this section, we elaborate on the MRM problem. The main goal of this article is to solve the MRM problem (1.1). The key step is to solve the following unconstrained optimization problem:

$$\min_{\mathbf{X}} \lambda \operatorname{rank}(\mathbf{X}) + \frac{1}{2} \|\mathbf{X} - \mathbf{Y}\|_F^2, \quad (2.1)$$

where  $\mathbf{Y}$  is a known matrix and  $\lambda$  a positive balanced parameter. This problem occurs in many practical tasks — cf. [54]. To clearly clarify the rank constraint and its relation to the nuclear norm, we present the following theorem, whose proof can be found in Appendix.

**Theorem 2.1.** *Let  $\mathbf{Y} = \mathbf{U}\text{diag}(\sigma_1, \sigma_2, \dots, \sigma_m)\mathbf{V}^T$  be the SVD of a matrix  $\mathbf{Y} \in \mathbb{R}^{m \times n}$ ,  $m \leq n$ , and  $d_i$  denotes the  $i$ -th singular value of  $\mathbf{X}$ . The optimum to the problem*

$$\min_{\mathbf{X}} \lambda \text{rank}(\mathbf{X}) + \frac{1}{2} \|\mathbf{X} - \mathbf{Y}\|_F^2$$

*can be expressed as  $\mathbf{X} = \mathbf{U}\text{diag}(\hat{d}_1, \hat{d}_2, \dots, \hat{d}_m)\mathbf{V}^T$ , where  $(\hat{d}_1, \hat{d}_2, \dots, \hat{d}_m)$  is the solution to the following optimization problem:*

$$\min_{d_1, d_2, \dots, d_m} \lambda \sum_{i=1}^m \|d_i\|_0 + \frac{1}{2} (d_i - \sigma_i)^2, \quad (2.2)$$

where  $\|\cdot\|_0$  means  $\ell_0$ -norm.

According to Theorem 2.1, the rank constraint on the matrix  $\mathbf{X}$  can be transformed into  $\ell_0$ -norm constraint on its singular values. However, direct solving (2.2) is difficult since  $\ell_0$ -norm is discontinuous. To overcome this drawback, the nuclear norm defined as the sum of the singular values is used to relax the  $\ell_0$ -norm. Following Theorem 2.1, the minimization of nuclear norm is similarly equal to the  $\ell_1$ -norm constraint on singular values — i.e.

$$\min_{\mathbf{X}} \lambda \|\mathbf{X}\|_* + \frac{1}{2} \|\mathbf{X} - \mathbf{Y}\|_F^2 \Leftrightarrow \min_{d_1, d_2, \dots, d_m} \lambda \sum_{i=1}^m \|d_i\|_1 + \frac{1}{2} (d_i - \sigma_i)^2, \quad (2.3)$$

where  $\|\mathbf{X}\|_*$  and  $\|d_i\|_1$  denote nuclear norm of the matrix  $\mathbf{X}$  and  $\ell_1$ -norm of the singular value  $d_i$ , respectively. It is clear that the nuclear norm minimization problem (2.3) is to replace  $\ell_0$ -norm by  $\ell_1$ -norm [31]. Due to the difference between the two norms, the nuclear norm cannot describe the constraint of  $\ell_0$ -norm for singular values. Subsequently, to overcome this drawback, some nonconvex rank approximations — e.g. logdet function [11], Schatten p-norm [34], and logarithmic norm [8], are proposed and achieve better results. Nonetheless, these rank approximations lack the necessary flexibility and cannot closely approximate the  $\ell_0$ -norm. The critical point of the MRM problem is finding an appropriate function that can obtain a good resemblance of  $\ell_0$ -norm. Besides, due to the nonconvexity of these approximations, designing an effective algorithm for solving the nonconvex problem is challenging [30]. It is still necessary to develop an algorithm in order to solve the nonconvex rank approximation minimization problem — i.e. the problem (1.2) for nonconvex rank approximation  $\Psi(\mathbf{X})$ , cf. refs. [9, 18, 66].

### 3. A TFR Approximation

Bearing in mind the above concern, we propose a novel nonconvex TFR approximation, which is defined as follows.

**Definition 3.1** (TFR Approximation). *The TFR approximation for the matrix  $\mathbf{X}$  is defined as*

$$\|\mathbf{X}\|_{\text{TFR}} := \sum_{i=1}^m h(d_i), \quad (3.1)$$

where  $\mathbf{X} \in \mathbb{R}^{m \times n}$ ,  $m \leq n$ ,  $d_i$ ,  $i = 1, 2, \dots, m$  denotes  $i$ -th singular value of  $\mathbf{X}$ , and

$$h(x) = \frac{bx^p}{1/\alpha + x^p}, \quad p > 0. \quad (3.2)$$

The TFR approximation (3.1) constrains the matrix  $\mathbf{X}$  by the function  $h(x)$ , i.e. (3.2) and the TFR approximation is defined as

$$\begin{aligned} & \min_{\mathbf{X}} \|\mathbf{X}\|_{\text{TFR}} \\ \text{s.t. } & \mathcal{A}(\mathbf{X}) = \mathbf{B}. \end{aligned} \quad (3.3)$$

It is equivalent to (1.2) when  $\Psi(\mathbf{X})$  is TFR approximation. Compared to other rank replacements, this TFR approximation provides a tighter substitute for  $\ell_0$ -norm. In addition, TFR approximation can flexibly constrain the singular values. As displayed in Fig. 2, the height, the slope, and the tightness of TFR approximation are adjusted by  $b$ ,  $p$ , and  $\alpha$ , respectively. Similar to the nuclear norm, the TFR approximation minimization (3.3) can be effectively solved by the following proximal operation:

$$\arg \min_{\mathbf{X}} \lambda \sum_{i=1}^m h(d_i) + \frac{1}{2} \|\mathbf{X} - \mathbf{Y}\|_F^2. \quad (3.4)$$

Note that because of the lack of the convexity, it is difficult construct an effective solution by the nonconvex rank approximation — cf. [67]. In the rest of this section, we concentrate on developing a convergent algorithm for the approximal problem (3.4), starting with a lemma, which shows an important property — i.e. the monotonicity of the TFR approximation.

**Lemma 3.1.** *For the function  $h$  in (3.2), let*

$$\text{Prox}_h(y) := \arg \min_x \lambda h(x) + \frac{1}{2}(x - y)^2.$$

*Then the function  $\text{Prox}_h(\cdot)$  is monotone in the sense that if  $y_1 > y_2$  and  $x_i \in \text{Prox}_h(y_i)$ ,  $i = 1, 2$ , then  $x_1 \geq x_2$ .*

*Proof.* Since the function  $h$  is bounded below, the proximal operator  $\text{Prox}_h(\cdot)$  is bounded. The optimality of  $x_i$ ,  $i = 1, 2$  yields

$$\lambda h(x_2) + \frac{1}{2}(x_2 - y_1)^2 \geq \lambda h(x_1) + \frac{1}{2}(x_1 - y_1)^2, \quad (3.5)$$

$$\lambda h(x_1) + \frac{1}{2}(x_1 - y_2)^2 \geq \lambda h(x_2) + \frac{1}{2}(x_2 - y_2)^2. \quad (3.6)$$

Summing (3.5) and (3.6), we obtain

$$(x_2 - y_1)^2 + (x_1 - y_2)^2 \geq (x_1 - y_1)^2 + (x_2 - y_2)^2,$$

which implies

$$(x_1 - x_2)(y_1 - y_2) \geq 0.$$

Hence, if  $y_1 > y_2$ , then  $x_1 \geq x_2$ .  $\square$

Based on Lemma 3.1, we can give the following Theorem 3.1.

**Theorem 3.1.** *Let  $\mathbf{Y} = \mathbf{U}\text{diag}(\sigma_1, \sigma_2, \dots, \sigma_m)\mathbf{V}^T$  be the SVD of  $\mathbf{Y}$ . Then an optimal solution to (3.4) is*

$$\mathbf{X}^* = \mathbf{U}\text{diag}(d_1^*, d_2^*, \dots, d_m^*)\mathbf{V},$$

where  $d_1^* \geq d_2^* \geq \dots \geq d_m^*$ , and

$$d_i^* \in \text{Prox}_h(d_i) = \arg \min_{d_i \geq 0} \lambda h(d_i) + \frac{1}{2} \|d_i - \sigma_i\|_F^2.$$

*Proof.* Let  $d_1 \geq d_2 \geq \dots \geq d_m \geq 0$  be the singular values of  $\mathbf{X}$ . According to the von Neumanns trace inequality [14], we have

$$\begin{aligned} \|\mathbf{X} - \mathbf{Y}\|_F^2 &= \text{Tr}(\mathbf{X}^T \mathbf{X}) - 2\text{Tr}(\mathbf{X}^T \mathbf{Y}) + \text{Tr}(\mathbf{Y}^T \mathbf{Y}) \\ &= \sum_{i=1}^m d_i^2 - 2\text{Tr}(\mathbf{X}^T \mathbf{Y}) + \sum_{i=1}^m \sigma_i^2 \\ &\geq \sum_{i=1}^m d_i^2 - 2 \sum_{i=1}^m d_i \sigma_i + \sum_{i=1}^m \sigma_i^2 \\ &= \sum_{i=1}^m (d_i - \sigma_i)^2. \end{aligned}$$

Note that the above inequality becomes equality if  $\mathbf{X}$  admits the singular value decomposition representation  $\mathbf{X} = \mathbf{U}\text{diag}(d_1, d_2, \dots, d_m)\mathbf{V}^T$ , where  $\mathbf{U}$  and  $\mathbf{V}$  are the left and right orthonormal matrices in the SVD of  $\mathbf{Y}$  [14]. In this case, the problem (3.4) can be written as

$$\min_{d_1 \geq d_2 \geq \dots \geq d_m \geq 0} \lambda \sum_{i=1}^m h(d_i) + \frac{1}{2} (d_i - \sigma_i)^2. \quad (3.7)$$

Since  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_m$  and by Lemma 3.1, the function  $\text{Prox}_h(\cdot)$  is monotone, there exist  $d_i^* \in \text{Prox}_h(\sigma_i)$ ,  $i = 1, 2, \dots, m$  such that  $d_1^* \geq d_2^* \geq \dots \geq d_m^*$ . Such a choice of  $d_i^*$  is optimal to (3.7). Therefore,  $\mathbf{U}\text{diag}(d_1^*, d_2^*, \dots, d_m^*)\mathbf{V}$  is optimal to (3.4).  $\square$

According to Lemma 3.1 and Theorem 3.1, the crucial step in the proximal problem (3.4) is to resolve the TFR proximal problem for singular values as follows:

$$\text{Prox}_h(y) = \arg \min_x \lambda h(x) + \frac{1}{2} (x - y)^2. \quad (3.8)$$

**Theorem 3.2.** Let

$$f_y(x) := \lambda h(x) + \frac{1}{2}(x - y)^2,$$

$$\hat{x}^y := \max \{x \mid \nabla f_y(x) = 0, 0 \leq x \leq y\}.$$

If the function  $h$  in (3.2) is concave and its gradient  $\nabla h$  is convex, then

$$x^* = \arg \min_{x=0, \hat{x}^y} f_y(x)$$

is optimal to (3.8).

**Remark 3.1.** The proof of Theorem 3.2 is given in Appendix. Theorem 3.2 assumes some conditions for  $h$ , which are easily satisfied by appropriate parameter choice — cf. Fig. 2.

Theorem 3.2 and Algorithm 3.1 allow to determine the solution to the problem (3.8). The next theorem guarantees the convergence of the corresponding sequence.

---

**Algorithm 3.1** Solution of (3.8).

---

**Input:**  $y \geq 0$ .

**Parameter:**  $\alpha$ ,  $b$ ,  $\lambda$ , and  $p$ .

**Output:** optimal solution  $x^*$ .

```

1: if  $\nabla h(y) = 0$  then
2:   Return  $\hat{x}^y = y$ ,
3: else
4:   Initialization  $x_0 = y$  and  $k = 0$ .
5:   while not converge do
6:     Update  $x_{k+1} = y - \lambda \nabla h(x_k)$ .
7:     if  $x_{k+1} < 0$  then
8:       Return  $\hat{x}^y = 0$ ,
9:       break.
10:    end if
11:     $k = k + 1$ .
12:  end while
13: end if
14: Compare  $f_y(0)$  and  $f_y(\hat{x}^y)$  to identify  $x^*$ .
```

---

**Theorem 3.3.** Let  $x_0 = y \geq 0$  and  $\{x_k\}$  be the sequence produced by Algorithm 3.1. Then, under the notations of Theorem 3.2, we have

$$\lim_{k \rightarrow +\infty} x_k = \hat{x}^y.$$

*Proof.* First, for any  $x > \hat{x}^y$ , we prove that

$$y - \lambda \nabla h(x) < x.$$

Assume that there exists  $\tilde{x} > \hat{x}^y$  such that  $y - \nabla h(\tilde{x}) > \tilde{x}$ . Because

$$\nabla h(x) = \begin{cases} \frac{b}{\alpha(1/\alpha + x)^2}, & p = 1, \\ \frac{pb}{\alpha x^{1-p}(1/\alpha + x^p)^2}, & \text{otherwise.} \end{cases}$$

It is easily seen that  $\nabla h(x)$  is continuous, decreasing, and positive. Since  $\nabla h(y) > 0$ , for  $y > \hat{x}^y$  we have

$$y - \lambda \nabla h(y) < y.$$

Therefore, there are  $\hat{x} > \hat{x}^y$  and  $\hat{x} \in (\min(y, \tilde{x}), \max(y, \tilde{x}))$  such that  $y - \lambda h(\hat{x}) = \hat{x}$ . This contradicts the definition of  $\hat{x}^y$ . Thus,

$$x_{k+1} = y - \lambda \nabla h(x_k) < x_k.$$

Since  $\{x_k\}$  is bounded below by  $\hat{x}^y$ , the sequence  $\{x_k\}$  converges to an  $\bar{x} \geq \hat{x}^y$ . Passing to the limit in the equation

$$x_{k+1} = y - \lambda \nabla h(x_k)$$

as  $k$  tends to  $\infty$ , we get  $\bar{x} = y - \lambda \nabla h(\bar{x})$ , i.e.  $\lim_{k \rightarrow +\infty} x_k = \hat{x}^y$ .  $\square$

#### 4. Application to Matrix Completion

To verify the validity of the TFR approximation, we use it in matrix completion problem — a significant application of the rank approximations [16]. Matrix completion refers to completing the missing matrix by the low-rank property of matrices [2]. More exactly, the matrix completion problem has the form

$$\begin{aligned} & \min_{\mathbf{X}} \operatorname{rank}(\mathbf{X}) \\ & \text{s.t. } \mathcal{P}_{\Omega}(\mathbf{X}) = \mathcal{P}_{\Omega}(\mathbf{M}), \end{aligned} \tag{4.1}$$

where  $\mathbf{M}$  is the degraded matrix,  $\mathbf{X}$  is the underlying matrix,  $\mathcal{P}_{\Omega}(\cdot)$  is a projector, and  $\mathcal{P}_{\Omega}(\mathbf{X}) = \mathcal{P}_{\Omega}(\mathbf{M})$  means the values of  $\mathbf{X}$  and  $\mathbf{M}$  in the area  $\Omega$  are equal. Obviously, (4.1) is a specific MRM problem (1.1). In this section, we propose a novel matrix completion model. Considering the non-convexity of the proposed TFR approximation, we develop an algorithm based on the PAM framework [4] and provide its convergence analysis.

##### 4.1. The model and algorithm

Based on the proposed TFR approximation, we construct a new nonconvex model for the matrix completion. It can be formulated as follows:

$$\begin{aligned} & \min_{\mathbf{X}} \|\mathbf{X}\|_{\text{TFR}} \\ & \text{s.t. } \mathcal{P}_{\Omega}(\mathbf{X}) = \mathcal{P}_{\Omega}(\mathbf{M}). \end{aligned} \tag{4.2}$$

The above model (4.2) can be transformed into the unconstrained format

$$\min_{\mathbf{X}} \|\mathbf{X}\|_{\text{TFR}} + \mathcal{I}_{\Phi}(\mathbf{X}), \quad (4.3)$$

where

$$\mathcal{I}_{\Phi} = \begin{cases} 0, & \mathbf{X} \in \Phi, \\ \infty, & \text{otherwise,} \end{cases}$$

and  $\Phi := \{\mathbf{X} \mid \mathcal{P}_{\Omega}(\mathbf{X}) = \mathcal{P}_{\Omega}(\mathbf{M})\}$ . Since it is difficult to solve the problem (4.3) directly, we introduce an auxiliary variable  $\mathbf{W} = \mathbf{X}$ . Using the half quadratic splitting (HQS) technique [17], we can rewrite the problem (4.3) in the form

$$\min_{\mathbf{X}, \mathbf{W}} \|\mathbf{W}\|_{\text{TFR}} + \frac{\mu}{2} \|\mathbf{X} - \mathbf{W}\|_F^2 + \mathcal{I}_{\Phi}(\mathbf{X}), \quad (4.4)$$

where  $\mu$  is a positive penalty parameter. Under the PAM framework [4, 49], the problem (4.4) is effectively solved by updating each variable alternately — i.e.

$$\begin{aligned} \mathbf{X}^{k+1} &= \arg \min_{\mathbf{X}} \mathcal{I}_{\Phi}(\mathbf{X}) + \frac{\mu}{2} \|\mathbf{X} - \mathbf{W}^k\|_F^2 + \frac{\rho}{2} \|\mathbf{X} - \mathbf{X}^k\|_F^2, \\ \mathbf{W}^{k+1} &= \arg \min_{\mathbf{W}} \|\mathbf{W}\|_{\text{TFR}} + \frac{\mu}{2} \|\mathbf{X}^{k+1} - \mathbf{W}\|_F^2 + \frac{\rho}{2} \|\mathbf{W} - \mathbf{W}^k\|_F^2, \end{aligned}$$

where  $\rho$  is the penalty parameter.

**X sub-problem:** The X sub-problem at  $(k+1)$ -th iteration has the form

$$\min_{\mathbf{X}} \mathcal{I}_{\Phi}(\mathbf{X}) + \frac{\mu}{2} \|\mathbf{X} - \mathbf{W}^k\|_F^2 + \frac{\rho}{2} \|\mathbf{X} - \mathbf{X}^k\|_F^2. \quad (4.5)$$

It can be exactly solved by

$$\mathbf{X}^{k+1} = \mathcal{P}_{\Omega}(\mathbf{M}) + \mathcal{P}_{\Omega^C} \left( \frac{\mu \mathbf{W}^k + \rho \mathbf{X}^k}{\mu + \rho} \right), \quad (4.6)$$

where  $\Omega^C$  refers to the complement of  $\Omega$ .

**W sub-problem:** According to (4.4), the W sub-problem at  $(k+1)$ -th iteration has the form

$$\min_{\mathbf{W}} \|\mathbf{W}\|_{\text{TFR}} + \frac{\mu}{2} \|\mathbf{X}^{k+1} - \mathbf{W}\|_F^2 + \frac{\rho}{2} \|\mathbf{W} - \mathbf{W}^k\|_F^2. \quad (4.7)$$

It can be reduced to the problem

$$\min_{\mathbf{W}} \|\mathbf{W}\|_{\text{TFR}} + \frac{\mu + \rho}{2} \left\| \frac{\mu \mathbf{X}^{k+1} + \rho \mathbf{W}^k}{\mu + \rho} - \mathbf{W} \right\|_F^2. \quad (4.8)$$

Theorem 3.1 shows that the solution of (4.8) has the form

$$\mathbf{W}^{k+1} = \tilde{\mathbf{U}} \text{diag} (\sigma_1(\mathbf{W}^{k+1}), \dots, \sigma_m(\mathbf{W}^{k+1})) \tilde{\mathbf{V}}^T, \quad (4.9)$$

where  $\tilde{\mathbf{U}}\text{diag}(\tilde{\sigma}_1, \tilde{\sigma}_2, \dots, \tilde{\sigma}_m)\tilde{\mathbf{V}}^T$  is the SVD of  $(\mu\mathbf{X}^{k+1} + \rho\mathbf{W}^k)/(\mu + \rho)$ , and  $\sigma_i(\mathbf{W}^{k+1})$  is the optimum to the optimization problems

$$\sigma_i(\mathbf{W}) = \text{Prox}_h(\tilde{\sigma}_i), \quad i = 1, 2, \dots, m. \quad (4.10)$$

The problems (4.10) can be effectively solved by Algorithm 3.1. The solution process is summarized in Algorithm 4.1, where the relative change (RC) and the number of iterations  $k_{mit}$  are used as the termination condition. The real change is defined by

$$\text{RC} = \frac{\|\mathbf{X}^{k+1} - \mathbf{X}^k\|_F}{\|\mathbf{X}^k\|_F}. \quad (4.11)$$

---

**Algorithm 4.1** PAM-Based Solver for Matrix Completion Model (4.2).

---

**Input:** Observed image  $\mathbf{M}$ .

**Parameter:**  $\mu, k_{mit}, \varepsilon, \alpha, b, \rho, p$ .

**Output:**  $\mathbf{X}$ .

- 1: Initialization  $k = 0, \mathbf{X}^0 = \mathbf{M}$ , and  $\mathbf{W}^0 = \mathbf{0}$ .
  - 2: **while**  $k < k_{mit}$  and  $\text{RC} > \varepsilon$  **do**
  - 3:   Update  $\mathbf{X}^{k+1}$  via (4.6).
  - 4:   Update  $\mathbf{W}^{k+1}$  via (4.9).
  - 5:    $k = k + 1$ .
  - 6: **end while**
- 

## 4.2. Convergence of Algorithm 4.1

Here we present sufficient conditions for the convergence of Algorithm 4.1, but let us first recall the definitions of semi-algebraic sets, semi-algebraic functions, and Kurdyka-Łojasiewicz (KL) functions.

**Definition 4.1** (Semi-Algebraic Set, cf. Attouch *et al.* [1]). A subset  $S$  of  $\mathbb{R}^n$  is called *real semi-algebraic* if there exists a finite number of real polynomial functions  $A_{ij}, B_{ij} : \mathbb{R}^n \rightarrow \mathbb{R}$  such that

$$S = \bigcup_{j=1}^p \bigcap_{i=1}^q \{x \in \mathbb{R}^n : A_{ij} = 0, B_{ij} < 0\}.$$

**Definition 4.2** (Semi-Algebraic Function, cf. Attouch *et al.* [1]). A function  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  is called *semi-algebraic* if its graph  $\{(x, y) \in \mathbb{R}^{n+1} : f(x) = y\}$  is a semi-algebraic subset of  $\mathbb{R}^{n+1}$ .

By  $\text{dom}(\partial f)$  we denote the domain of  $\partial f$ .

**Definition 4.3** (KL property, cf. Attouch *et al.* [1]). We say that a function  $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$  satisfies the *KL* property at  $x \in \text{dom}(\partial f)$  if there exist  $\eta \in (0, +\infty]$ , a neighborhood  $U$  of  $x$  and a continuous concave function  $\varphi : [0, \eta] \rightarrow \mathbb{R}_+$  such that

- \*  $\varphi(0) = 0$ ,
- \*  $\varphi$  is  $C^1$  on  $(0, \eta)$ ,
- \* for all  $v \in (0, \eta)$ ,  $\varphi'(v) > 0$ ,
- \* for all  $s$  in  $U \cap \{f(x) < f(s) < f(x) + \eta\}$ , the following KŁ inequality holds:

$$\varphi'(f(s) - f(x)) \text{dist}(0, \partial f(s)) \geq 1,$$

where  $\text{dist}(0, \partial f(s)) := \|\partial f(s) - 0\|$ .

**Remark 4.1.** A proper lower semicontinuous function is called KŁ function if the function is semi-algebraic and the KŁ inequality holds at any point  $x \in \text{dom}(\partial f)$  [3].

Let us study the convergence of Algorithm 4.1. For convenience, we denote  $F(\mathbf{W}, \mathbf{X}) = \|\mathbf{W}\|_{\text{TFR}} + (\mu/2)\|\mathbf{X} - \mathbf{W}\|_F^2 + \mathcal{I}_\Phi(\mathbf{X})$ .

**Lemma 4.1.** *The function  $F(\mathbf{W}, \mathbf{X})$  is a KŁ function.*

*Proof.* Following [46, 49],  $h(x)$  is a real analytic function, and the TFR approximation is the finite sum of  $h(x)$  for all singular values, which indicates that the TFR approximation is a real analytic function. The set  $\Phi := \{\mathbf{X} \mid \mathcal{P}_\Omega(\mathbf{X}) = \mathcal{P}_\Omega(\mathbf{M})\}$  is semi-algebraic. Thus, the indicator function  $\mathcal{I}_\Phi$  is semi-algebraic function according to Definitions 4.1 and 4.2. We can further obtain that  $F(\mathbf{W}, \mathbf{X})$  is a KŁ function since it is the finite sum of real analytic and semi-algebraic functions.  $\square$

**Lemma 4.2** (Sufficient Decrease Condition). *Let  $\{\mathbf{W}^k, \mathbf{X}^k\}$  be the sequence produced by Algorithm 4.1. Then, we have*

$$F(\mathbf{W}^k, \mathbf{X}^k) - F(\mathbf{W}^{k+1}, \mathbf{X}^{k+1}) \geq \frac{\rho}{2}\|\mathbf{W}^{k+1} - \mathbf{W}^k\|_F^2 + \frac{\rho}{2}\|\mathbf{X}^{k+1} - \mathbf{X}^k\|_F^2.$$

*Proof.* Since  $\mathbf{X}^{k+1}$  is produced by minimizing (4.5), we have

$$\begin{aligned} & \mathcal{I}_\Phi(\mathbf{X}^{k+1}) + \frac{\mu}{2}\|\mathbf{X}^{k+1} - \mathbf{W}^k\|_F^2 + \frac{\rho}{2}\|\mathbf{X}^{k+1} - \mathbf{X}^k\|_F^2 \\ & \leq \mathcal{I}_\Phi(\mathbf{X}^k) + \frac{\mu}{2}\|\mathbf{X}^k - \mathbf{W}^k\|_F^2 + \frac{\rho}{2}\|\mathbf{X}^k - \mathbf{X}^k\|_F^2. \end{aligned} \quad (4.12)$$

Similarly, from (4.7), we can obtain

$$\begin{aligned} & \|\mathbf{W}^{k+1}\|_{\text{TFR}} + \frac{\mu}{2}\|\mathbf{X}^{k+1} - \mathbf{W}^{k+1}\|_F^2 + \frac{\rho}{2}\|\mathbf{W}^{k+1} - \mathbf{W}^k\|_F^2 \\ & \leq \|\mathbf{W}^k\|_{\text{TFR}} + \frac{\mu}{2}\|\mathbf{X}^{k+1} - \mathbf{W}^k\|_F^2 + \frac{\rho}{2}\|\mathbf{W}^k - \mathbf{W}^k\|_F^2. \end{aligned} \quad (4.13)$$

Combining (4.12) and (4.13), we can easily get that

$$F(\mathbf{W}^k, \mathbf{X}^k) - F(\mathbf{W}^{k+1}, \mathbf{X}^{k+1}) \geq \frac{\rho}{2}\|\mathbf{W}^{k+1} - \mathbf{W}^k\|_F^2 + \frac{\rho}{2}\|\mathbf{X}^{k+1} - \mathbf{X}^k\|_F^2.$$

The proof is complete.  $\square$

**Lemma 4.3.** Assume  $\{\mathbf{W}^k, \mathbf{X}^k\}$  is the sequence produced by Algorithm 4.1. Then, we have

$$\|\partial F(\mathbf{W}^{k+1}, \mathbf{X}^{k+1})\|_F \leq (\rho + \mu)(\|\mathbf{W}^{k+1} - \mathbf{W}^k\|_F + \|\mathbf{W}^{k+1} - \mathbf{W}^k\|_F).$$

*Proof.* According to the first-order optimal conditions, we have

$$\begin{aligned} 0 &\in \mu(\mathbf{X}^{k+1} - \mathbf{W}^k) + \rho(\mathbf{X}^{k+1} - \mathbf{X}^k), \\ 0 &\in \psi'(\mathbf{W}^{k+1}) - \mu(\mathbf{X}^{k+1} - \mathbf{W}^{k+1}) + \rho(\mathbf{W}^{k+1} - \mathbf{W}^k), \end{aligned}$$

where  $\psi(\mathbf{W}) = \|\mathbf{W}\|_{\text{TFR}}$ . Then, we can build the following inequality:

$$\begin{aligned} &\|\partial F(\mathbf{W}^{k+1}, \mathbf{X}^{k+1})\|_F \\ &\leq \|\partial_{\mathbf{W}} F(\mathbf{W}^{k+1}, \mathbf{X}^{k+1}) - \psi'(\mathbf{W}^{k+1}) + \mu(\mathbf{X}^{k+1} - \mathbf{W}^{k+1}) - \rho(\mathbf{W}^{k+1} - \mathbf{W}^k)\|_F \\ &\quad + \|\partial_{\mathbf{X}} F(\mathbf{W}^{k+1}, \mathbf{X}^{k+1}) - \mu(\mathbf{X}^{k+1} - \mathbf{W}^k) - \rho(\mathbf{X}^{k+1} - \mathbf{X}^k)\|_F \\ &= \rho\|\mathbf{W}^{k+1} - \mathbf{W}^k\|_F + \|\mu(\mathbf{W}^{k+1} - \mathbf{W}^k) - \rho(\mathbf{X}^{k+1} - \mathbf{X}^k)\|_F \\ &\leq (\rho + \mu)(\|\mathbf{W}^{k+1} - \mathbf{W}^k\|_F + \|\mathbf{X}^{k+1} - \mathbf{X}^k\|_F). \end{aligned}$$

The proof is complete.  $\square$

**Lemma 4.4** (Bounded Condition). Let  $\{\mathbf{W}^k, \mathbf{X}^k\}$  be the sequence produced by Algorithm 4.1. Assuming  $F(\mathbf{W}, \mathbf{X}) \rightarrow +\infty$  when  $\|(\mathbf{W}, \mathbf{X})\|_F \rightarrow +\infty$ , we have the sequence  $\{\mathbf{W}^k, \mathbf{X}^k\}$  is bounded.

*Proof.* Since  $\mathbf{X}^{k+1}$  is obtained by (4.6),  $\mathcal{I}_{\Phi}(\mathbf{X}^{k+1}) \equiv 0$ . According to Lemma 4.2, the sequence  $\{F(\mathbf{W}^k, \mathbf{X}^k)\}$  is sufficiently decreased. Thus, we have  $0 \leq F(\mathbf{W}^k, \mathbf{X}^k) \leq F(\mathbf{W}^0, \mathbf{X}^0)$ . According to the condition  $F(\mathbf{W}, \mathbf{X}) \rightarrow +\infty$  when  $\|(\mathbf{W}, \mathbf{X})\|_F \rightarrow +\infty$ , we can get that the sequence  $\{\mathbf{W}^k, \mathbf{X}^k\}$  is bounded.  $\square$

**Theorem 4.1.** Assume that  $F(\mathbf{W}, \mathbf{X}) \rightarrow +\infty$  when  $\|(\mathbf{W}, \mathbf{X})\|_F \rightarrow +\infty$ . Then the sequence  $\{\mathbf{W}^k, \mathbf{X}^k\}$  produced by Algorithm 4.1 converges to a critical point (local minimum point)  $\{\mathbf{W}^*, \mathbf{X}^*\}$  of the function  $F(\mathbf{W}, \mathbf{X})$ .

*Proof.* According to Lemmas 4.1-4.4 and the finite length theorem — cf. [3, Theorem 1], the sequence  $\{\mathbf{W}^k, \mathbf{X}^k\}$  converges to a critical point (local minimum point)  $\{\mathbf{W}^*, \mathbf{X}^*\}$  of  $F(\mathbf{W}, \mathbf{X})$ .  $\square$

## 5. Numerical Experiments

In this section, we present the results of experiments on synthetic and real data to validate the effectiveness of the proposed model. The parameters of the methods employed are fine-tuned according to the author recommendations. For the proposed method, we choose the parameters  $\alpha, b, p$ , and  $\mu$  in the intervals  $[10^{-3}, 1]$ ,  $[10^2, 10^5]$ ,  $[1.1, 1.3]$ , and  $[10^1, 10^4]$ , respectively. All calculations are carried out in Matlab (R2020a) on the same computer with

32Gb RAM, Intel(R) Core(TM) i7-8700K CPU @3.70GHz, and NVIDIA GeForce GTX 1080. For synthetic data, we evaluate the relative square error (RSE) as

$$\text{RSE} = \frac{\|\mathbf{X} - \hat{\mathbf{X}}\|_F}{\|\mathbf{X}\|_F},$$

where  $\mathbf{X}$  is the underlying matrix and  $\hat{\mathbf{X}}$  denotes the matrix obtained by the provided algorithm. For the experiments on real data, the peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM) [41] are utilized to give the quantitative assessment. In all experiments of the proposed method, we set  $k_{mit} = 100$ ,  $\varepsilon = 10^{-8}$  as the termination condition of Algorithm 4.1.

### 5.1. Synthetic data

To demonstrate the effectiveness of the proposed TFR approximation, we compare it with a convex nuclear norm and two nonconvex rank approximations — i.e. logdet norm and logarithmic norm, on the synthetic data. These convex and nonconvex rank approximations are borrowed from [5], [26], and [8], respectively. The rank  $r$  of the matrix  $\mathbf{X} \in \mathbb{R}^{m \times n}$  is generated by  $\mathbf{M}\mathbf{N}$ , where  $\mathbf{M} \in \mathbb{R}^{m \times r}$ , and  $\mathbf{N} \in \mathbb{R}^{r \times n}$  are randomly produced by a Gaussian distribution  $\mathcal{N}(0, 1)$ . We complete the observed matrix with respect to the matrix rank and the sampling rate (SR). For each matrix rank and SR, the experiment is repeated ten times. When the RSE of the test result is no more than  $10^{-5}$ , the test is considered successful. Fig. 3 displays the success rate with the fixed matrix rank and SR. One can observe that these nonconvex rank approximations indeed improve the success rate compared with the convex approach. Among them, the TFR approximation achieves the best performance, which further verifies the great flexibility of the TFR approximation.

### 5.2. Real data

In this part, we compare the proposed TFR matrix completion model with several state-of-the-art matrix completion models including SVT [5], RegL1 [69], Top-N [26], OP [55], LRIN [19], and LRMF [8] on the real data — i.e. on grey and color images.

#### 5.2.1. Grey images

For grey data, we employ the methods on Monarch, Parrots, and House with the SR = 0.4, 0.6 and 0.8, respectively. The size of these images is  $256 \times 256$ . The quantitative results are shown in Table 1. One can observe that our proposed model achieves the best performance on images with different sampling rates (SRs). Although LRIN [19] can produce a good result in quicker time, both the performance and the time metric are weaker than the proposed method. Since these rank approximations are applied to the entire image, rank approximations need to deal with different singular values simultaneously. Table 1 shows that our TFR approximation is flexible when dealing with singular values and the running time of the proposed model is less than most algorithms. Besides, Fig. 4 displays the visual

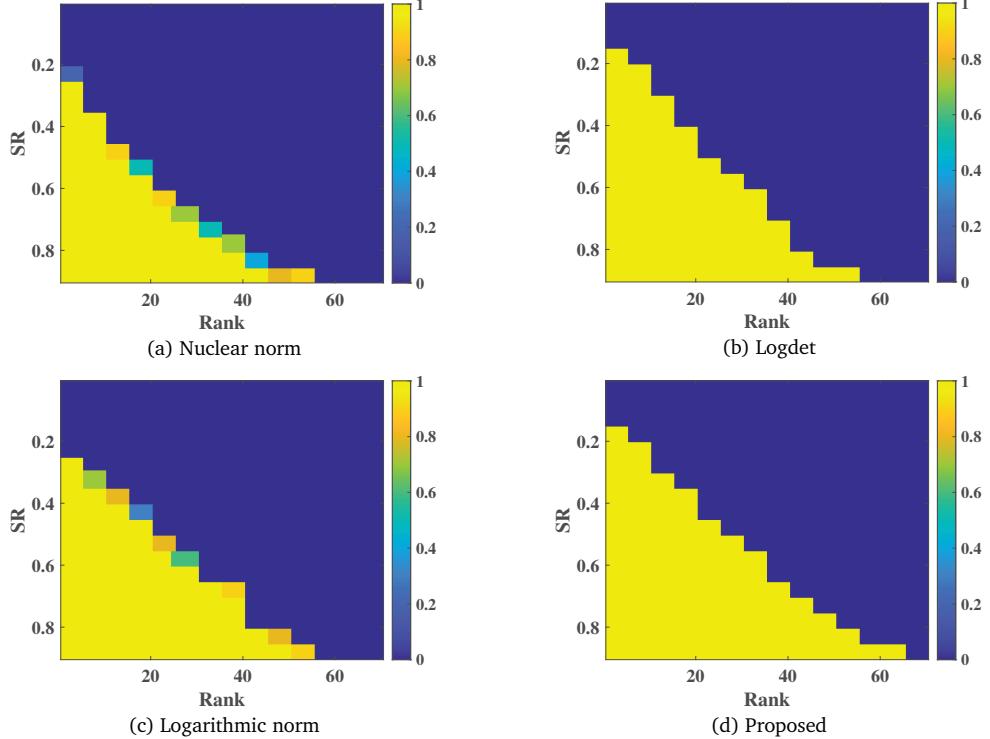


Figure 3: The success rates of the nuclear norm [5], logdet surrogate [11], logarithmic norm [8], and the proposed TFR approximation for synthetic data with different matrix ranks and sampling rates. It is worth to be remarked that the yellow area represents the range of success. Clearly, the TFR approximation is more robust.

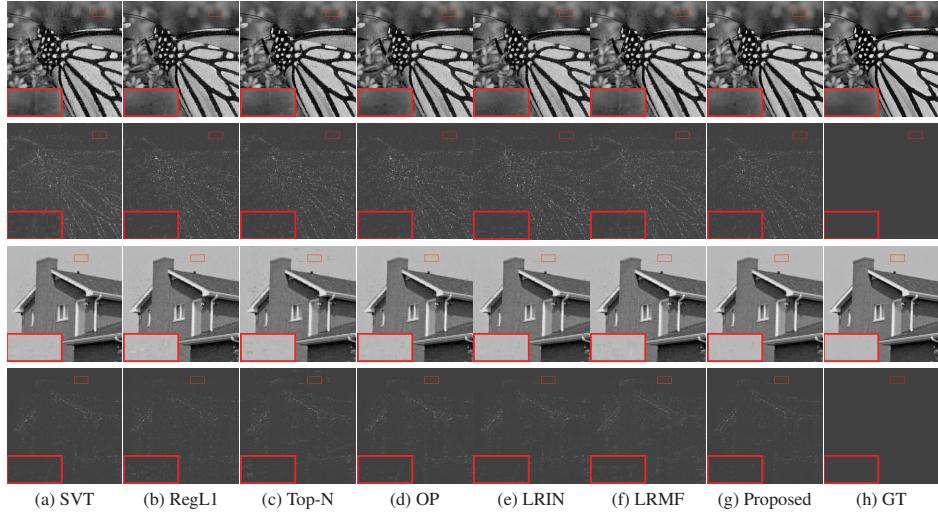


Figure 4: Matrix completion results on grey image Monarch and House with  $SR = 0.6$ . The first row is the visual comparisons, and the second row is the corresponding residual images and added 0.3 for better visualization. From left to right are the observed image, results with different methods, and the ground truth (GT).

Table 1: Quantitative results on grey images (Monarch, Parrots, and House) with SR = 0.4, 0.6, and 0.8, respectively. (Bold: best; Underline: second best).

Image	SR	0.4			0.6			0.8		
		Method	PSNR	SSIM	Time(s)	PSNR	SSIM	Time (s)	PSNR	SSIM
	Observed	8.45	0.098	—	10.19	0.160	—	13.19	0.273	—
	SVT [5]	21.52	0.599	4.190	24.93	0.744	4.335	28.69	0.848	4.085
	RegL1 [69]	21.78	0.623	<b>0.657</b>	25.98	0.813	3.806	<u>31.88</u>	<u>0.934</u>	3.416
	Top-N [26]	21.52	0.619	<u>0.936</u>	<u>26.30</u>	<u>0.816</u>	<b>0.737</b>	31.77	0.931	<b>0.280</b>
	OP [55]	21.62	<u>0.631</u>	4.064	25.89	0.808	3.446	31.83	0.933	2.78
	LRIN [19]	21.74	0.630	1.147	25.96	0.809	<u>0.934</u>	31.68	<u>0.934</u>	1.227
	LRMF [8]	<u>22.11</u>	0.630	1.865	26.22	0.794	1.641	30.92	0.900	2.526
	Proposed	<b>22.95</b>	<b>0.690</b>	0.966	<b>27.02</b>	<b>0.842</b>	1.176	<b>32.94</b>	<b>0.944</b>	<u>0.910</u>
	Observed	7.26	0.054	—	9.02	0.082	—	11.96	0.137	—
	SVT [5]	23.85	0.687	4.451	27.09	0.794	4.344	30.30	0.862	4.049
	RegL1 [69]	24.70	<u>0.743</u>	<u>0.986</u>	28.60	0.865	0.994	33.83	0.954	3.420
	Top-N [26]	24.25	0.733	0.997	28.41	0.862	<b>0.507</b>	33.21	0.947	<b>0.310</b>
	OP [55]	24.56	0.739	0.997	28.60	<u>0.872</u>	3.038	33.82	0.953	2.492
	LRIN [19]	24.64	0.735	1.149	<u>28.63</u>	0.870	1.376	<u>33.95</u>	<u>0.955</u>	1.695
	LRMF [8]	<u>24.81</u>	0.735	1.395	28.60	0.847	2.372	32.82	0.918	3.546
	Proposed	<b>25.37</b>	<b>0.772</b>	<b>0.930</b>	<b>29.47</b>	<b>0.887</b>	<u>0.965</u>	<b>34.31</b>	<b>0.962</b>	<u>0.882</u>
	Observed	7.10	0.045	—	8.88	0.069	—	11.88	0.116	—
	SVT [5]	27.04	0.725	4.527	30.29	0.828	4.412	31.82	0.856	4.186
	RegL1 [69]	27.82	0.768	<b>0.774</b>	32.04	0.846	<u>0.706</u>	36.82	0.932	1.029
	Top-N [26]	27.61	0.763	1.245	31.94	0.885	<b>0.580</b>	36.49	0.956	<b>0.305</b>
	OP [55]	27.45	0.768	4.301	32.85	<u>0.906</u>	2.992	36.97	0.960	4.015
	LRIN [19]	27.62	0.766	1.270	<u>32.89</u>	0.903	1.610	<u>38.52</u>	<u>0.967</u>	2.171
	LRMF [8]	<u>28.31</u>	<u>0.778</u>	1.993	<u>32.55</u>	0.862	2.193	<u>36.27</u>	<u>0.929</u>	4.601
	Proposed	<b>29.11</b>	<b>0.801</b>	<u>1.228</u>	<b>33.67</b>	<b>0.920</b>	0.966	<b>38.92</b>	<b>0.974</b>	<u>0.896</u>
	Ideal value	$+\infty$	1	0	$+\infty$	1	0	$+\infty$	1	0

comparison and residual maps of different methods on Monarch and House with SR = 0.6. It is obvious that nonconvex rank approximations have better visual results than the convex rank approximation. These red boxes of Fig. 4 show that, compared with other approaches, our method can preserve more details. For example, RegL1 [69] and Top-N [26] produce some artifacts (see the red box of House). Different from the above methods, our model can achieve a smooth result.

### 5.2.2. Color images

For the color images, the results on Lily, Fence, and Leaves with the SR = 0.4, 0.6, and 0.8 are shown in Table 2. The size of these images is  $256 \times 256 \times 3$ . We can see that our method obtains the best results than other convex and nonconvex rank approximations. The visual quality comparisons of the different methods for Fence and Leaves with SR = 0.6 are shown in Fig. 5. Compared with other approaches, our method can effectively deal

Table 2: Quantitative results on color images (Lily, Fence, and Leaves) with SR = 0.4, 0.6, and 0.8, respectively. (Bold: best; Underline: second best).

Image	SR	0.4			0.6			0.8		
		Method	PSNR	SSIM	Time(s)	PSNR	SSIM	Time (s)	PSNR	SSIM
	Observed	8.59	0.128	—	10.32	0.220	—	13.55	0.243	—
	SVT [5]	23.39	0.824	13.03	26.41	0.902	13.12	28.80	0.813	12.53
	RegL1 [69]	<u>23.73</u>	<u>0.844</u>	5.408	<u>27.44</u>	<u>0.928</u>	5.758	<u>32.52</u>	<u>0.928</u>	10.97
	Top-N [26]	23.55	0.829	<b>2.717</b>	<u>27.57</u>	<u>0.928</u>	<b>1.510</b>	32.49	0.923	<b>0.811</b>
	OP [55]	23.62	0.843	11.84	27.39	0.927	9.679	32.49	0.928	7.928
	LRIN [19]	23.69	0.626	<u>2.721</u>	27.42	0.803	<u>1.857</u>	32.48	0.926	<u>2.830</u>
	LRMF [8]	23.57	0.822	8.140	27.13	0.915	6.715	31.81	0.896	13.64
	Proposed	<b>24.50</b>	<b>0.869</b>	2.926	<b>28.12</b>	<b>0.937</b>	2.843	<b>33.07</b>	<b>0.935</b>	3.063
	Observed	6.71	0.097	—	8.46	0.169	—	11.66	0.225	—
	SVT [5]	24.69	0.850	12.86	27.16	0.901	12.98	28.77	0.827	24.99
	RegL1 [69]	25.11	0.870	5.074	28.47	0.931	4.151	<u>33.21</u>	<u>0.944</u>	19.28
	Top-N [26]	<u>25.44</u>	<u>0.882</u>	<b>2.266</b>	<u>28.70</u>	<u>0.937</u>	<u>1.782</u>	32.84	0.943	<b>2.078</b>
	OP [55]	25.02	0.871	10.89	28.45	0.931	8.951	33.19	0.943	13.29
	LRIN [19]	25.08	0.733	<u>2.458</u>	28.48	0.853	<b>1.584</b>	33.26	0.942	<u>2.134</u>
	LRMF [8]	25.03	0.849	5.954	27.98	0.913	7.282	32.23	0.911	19.59
	Proposed	<b>25.91</b>	<b>0.890</b>	2.822	<b>29.19</b>	<b>0.940</b>	3.070	<b>33.76</b>	<b>0.948</b>	4.541
	Observed	4.73	0.126	—	6.48	0.221	—	9.49	0.326	—
	SVT [5]	18.59	0.622	13.23	23.00	0.786	12.87	27.45	0.881	20.55
	RegL1 [69]	18.66	0.629	5.159	23.38	0.804	5.545	29.33	<u>0.937</u>	15.22
	Top-N [26]	18.47	<u>0.641</u>	<b>1.965</b>	23.53	<u>0.805</u>	<b>1.863</b>	29.33	0.926	<u>2.380</u>
	OP [55]	18.47	0.620	12.17	23.25	0.800	9.599	29.84	0.936	12.78
	LRIN [19]	18.75	0.611	4.669	23.48	0.803	<u>2.509</u>	30.00	0.936	<b>1.791</b>
	LRMF [8]	<u>18.80</u>	0.625	9.337	<u>23.94</u>	<u>0.805</u>	8.046	<u>30.01</u>	0.927	18.79
	Proposed	<b>19.57</b>	<b>0.667</b>	<u>2.775</u>	<b>24.55</b>	<b>0.835</b>	2.994	<b>30.92</b>	<b>0.948</b>	4.937
	Ideal value	$+\infty$	1	0	$+\infty$	1	0	$+\infty$	1	0

with the edges — cf. the red box in Fig. 5. More exactly, for the image Fence, we can see that OP [55] and LRMF [8] over-smooth the edge of fence. Our method can preserve the detail of fence well. Besides, for the running time, Top-N [26] and LRIN [19] consume less time. However, their performance is limited. The proposed method can achieve powerful results in less time.

### 5.3. Discussion

#### 5.3.1. Rigorous comparison

In this part, we discuss the rigorous comparison of matrix completion methods, which are run on the data at SR = 0.1-0.9. As displayed in Fig. 6, many methods vary in their ability to handle images with different SRs. For example, LRMF [8] can achieve good results when the SR is 0.3-0.6. OP [55] is stable for different conditions. LRIN [19] demonstrates

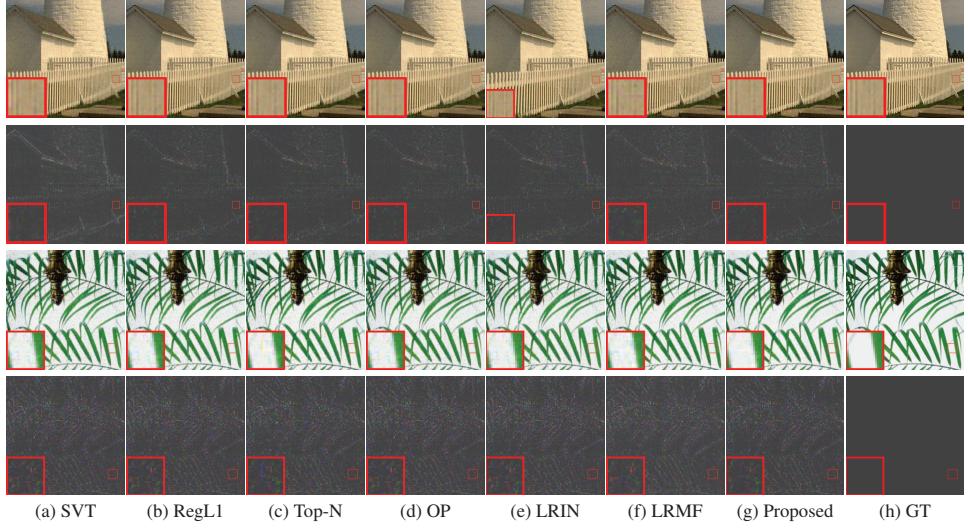


Figure 5: Matrix completion results on color image Fence and Leaves with  $SR = 0.6$ . The first row is the visual comparisons, and the second row is the corresponding residual images and added 0.3 for better visualization. From left to right are the observed image, results with different methods, and the ground truth (GT).

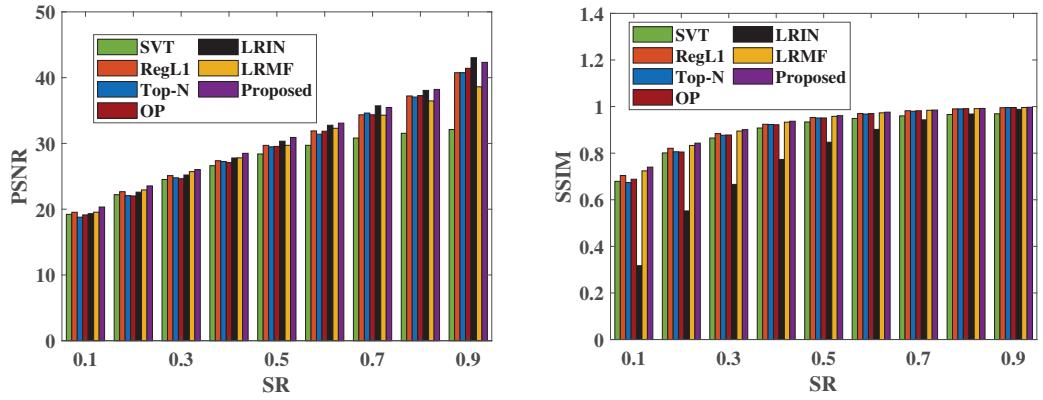


Figure 6: PSNR and SSIM for rigorous comparison with SRs from 0.1 to 0.9. (data: House).

strong performance according to the PSNR metric. However, its performance is notably constrained when evaluated by using the SSIM metric. On the other hand, the proposed method obtains the best results on the data with different SRs. The reason may be that for different conditions — e.g. different distributions of singular values, the TFR approximation used can flexibly shrink singular values by the appropriate parameter choice.

### 5.3.2. Parameter analysis

There are four main parameters in our matrix completion model, including the penalty parameter  $\mu$  and the parameters of the TFR approximation  $\alpha, b$ , and  $p$ . We adjust one

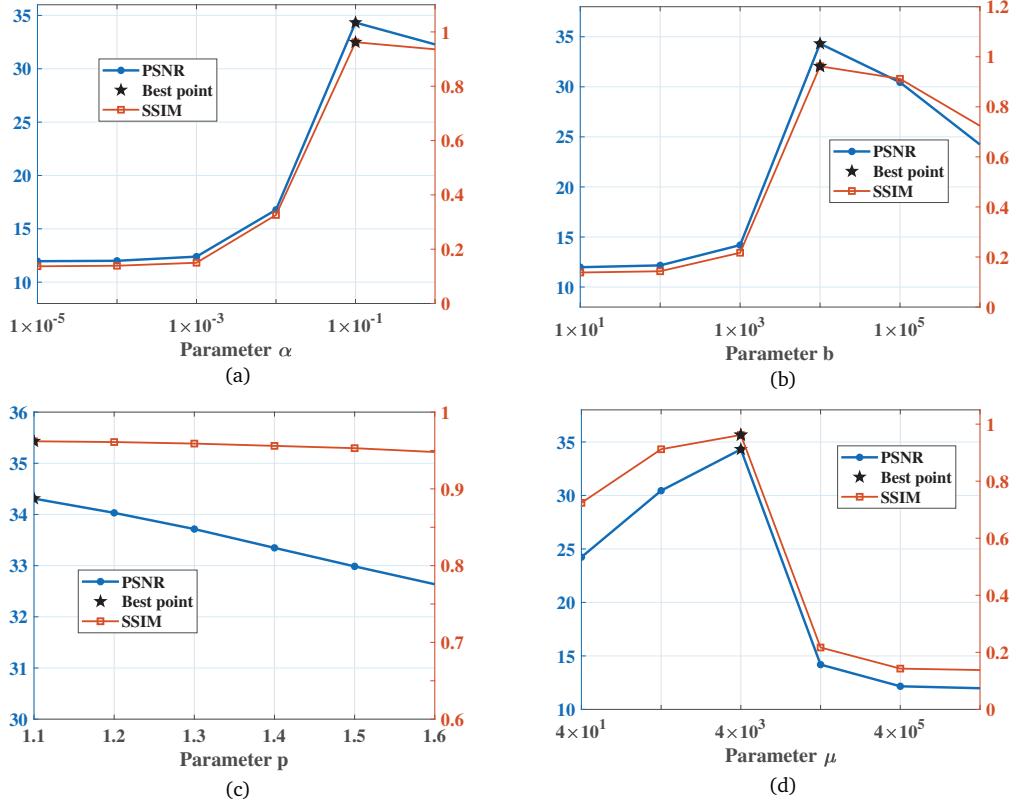


Figure 7: Robustness analysis (using PSNR and SSIM as metrics) for the four parameters (a)  $\alpha$ , (b)  $b$ , (c)  $p$ , and (d)  $\mu$ . (data: House SR = 0.8).

parameter at a time, and the others are fixed. Figs. 7(a)-7(d) show the four parameter curves on the image Parrots with the SR = 0.6. From Fig. 7, one can easily observe that the proposed model is more robust for the parameter  $p$  than the others, and other parameters are relatively sensitive. Thus, according the four parameter curves, we choose  $\alpha = 0.1$ ,  $b = 1 \times 10^4$ ,  $p = 1.1$ , and  $\mu = 4 \times 10^3$  in this experiment.

### 5.3.3. Numerical convergence analysis

In Section 4.2, we have proved the convergence of the matrix completion model with a condition — i.e.  $F(\mathbf{W}, \mathbf{X}) \rightarrow +\infty$  when  $\|(\mathbf{W}, \mathbf{X})\|_F \rightarrow +\infty$  (see Theorem 4.1). In this part, we provide numerical analysis of the proposed algorithm. To realize it, we calculate the RC — i.e. (4.11), in each iteration of the algorithm. As shown in Fig. 8, we plot the RC curves of the proposed model on the data with two different SRs. Experimental results demonstrate the rapid convergence of the algorithm. Specifically, when the number of iterations exceeds 20, the RC curves basically flatten out, which means the results computed by the proposed algorithm tend to be stable. Thus, it is sufficient to choose  $k_{mit} = 100$  as the termination condition of Algorithm 4.1.

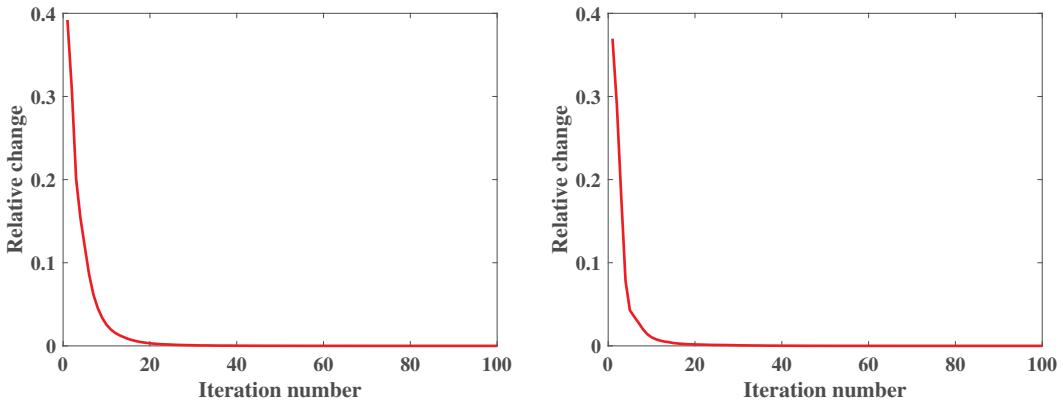


Figure 8: The RC curves of the matrix completion model on the image House with the SR = 0.4 and 0.6, respectively.

### 5.3.4. Experiment on Hankel matrix completion

In this part, we test the performance of the proposed method on the low-rank approximation of the Hankel matrix. Specifically, we use the Hankel matrix  $\mathbf{H} \in \mathbb{R}^{10 \times 10}$ , i.e.

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & & 1 & 0 \\ \vdots & \ddots & & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix}.$$

Following the settings of [19], the rank  $r$  approximation of the Hankel matrix  $\mathbf{H}$ , denoted by  $\mathbf{X}$ , is obtained by choosing the first  $r$  components of singular value decomposition. The index set of known entries consists of the positions where the values are positive. Table 3 shows the performance of different methods on the different rank  $r$  approximation cases. The TNN-based method SVT [5] is not effective for these cases. The low-rank inducing norm performs well under different rank conditions. LRMF [8] is effective only for the low-rank case. The proposed method can achieve excellent results, particularly when the rank is very low.

Table 3: The relative square error (RSE) performance of different methods on Hankel matrix completion. (Bold: best; Underline: second best).

Methods	$r = 2$	$r = 4$	$r = 6$	$r = 8$
SVT [5]	0.016	0.085	0.100	0.080
LRIN [19]	$2.6 \times 10^{-9}$	$1.6 \times 10^{-7}$	$5.4 \times 10^{-7}$	$3.1 \times 10^{-9}$
LRMF [8]	$3.2 \times 10^{-7}$	0.015	0.220	0.244
Proposed	$6.9 \times 10^{-10}$	$3.0 \times 10^{-4}$	$1.7 \times 10^{-3}$	$1.4 \times 10^{-3}$

## 6. Conclusion

In this article, we explore the essence of the current nonconvex matrix rank approximations in the MRM problem, namely, narrowing the gap with  $\ell_0$ -norm. Based on observations, we propose a novel TFR approximation, which can well describe the properties of the matrix rank and flexibly deal with the shrinkage of different singular values. Algorithm 3.1 is developed for solving TFR proximal problem of singular values. We then apply TFR approximation to matrix completion and develop Algorithm 4.1 with convergence guarantee to solve the new matrix completion model. Extensive numerical experiments show the advantages of our model over other convex and nonconvex rank approximation-based methods. This approach provides favorable results compared with the state-of-the-art methods. Note that the TFR approximation would be helpful in other low-level tasks, such as image fusion [28, 51, 52, 57], image denoising [12, 32, 64], tensor completion [23], and image inpainting [58].

## Acknowledgments

This research is supported by the Natural Science Foundation of Sichuan Province (Grant 2024NSFSC0038), by the NSFC (Grants 12171072, 12271083), and by the National Key Research and Development Program of China (Grant 2020YFA0714001).

## Appendix A. Proof of Theorem 2.1

*Proof.* Following [22], for any  $\mathbf{X} \in \mathbb{R}^{m \times n}$ , we denote by  $\overline{\mathbf{U}}\mathbf{D}\overline{\mathbf{V}}^T$  the SVD of  $\mathbf{X}$ , where  $\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_m)$  and  $d_1 \geq d_2 \geq \dots \geq d_m \geq 0$ . Using the properties of the Frobenius norm, we write

$$\begin{aligned} & \lambda \text{rank}(\mathbf{X}) + \frac{1}{2} \|\mathbf{X} - \mathbf{Y}\|_F^2 \\ &= \frac{1}{2} \|\mathbf{Y}\|_F^2 - \text{Tr}(\mathbf{Y}^T \mathbf{X}) + \frac{1}{2} \|\mathbf{X}\|_F^2 + \lambda \text{rank}(\mathbf{X}) \\ &= -\text{Tr}(\mathbf{Y}^T \mathbf{X}) + \frac{1}{2} \sum_{i=1}^m (\sigma_i^2 + d_i^2 + 2\lambda \|d_i\|_0), \end{aligned}$$

where  $\text{Tr}(\mathbf{Y}^T \mathbf{X})$  is the trace of  $\mathbf{Y}^T \mathbf{X}$ . Then,

$$\begin{aligned} & \min_{\overline{\mathbf{U}}, \mathbf{D}, \overline{\mathbf{V}}^T} \left\{ -\text{Tr}(\mathbf{Y}^T \mathbf{X}) + \frac{1}{2} \sum_{i=1}^m (\sigma_i^2 + d_i^2 + 2\lambda \|d_i\|_0) \right\} \\ & \Leftrightarrow \min_{\mathbf{D}} \left\{ -\max_{\overline{\mathbf{U}}, \overline{\mathbf{V}}^T} \text{Tr}(\mathbf{Y}^T \mathbf{X}) + \frac{1}{2} \sum_{i=1}^m (\sigma_i^2 + d_i^2 + 2\lambda \|d_i\|_0) \right\}. \end{aligned}$$

Based on von Neumanns trace inequality [14], it is obvious that  $\text{Tr}(\mathbf{Y}^T \mathbf{X})$  achieves its upper bound  $\sum_{i=1}^m (\sigma_i d_i)$  if  $\overline{\mathbf{U}} = \mathbf{U}$  and  $\overline{\mathbf{V}} = \mathbf{V}$ . Consequently,

$$\min_{\mathbf{D}} \frac{1}{2} \sum_{i=1}^m (\sigma_i^2 - 2\sigma_i d_i + d_i^2) + \lambda \|d_i\|_0 \Leftrightarrow \min_{d_1, d_2, \dots, d_m} \frac{1}{2} \sum_{i=1}^m (\sigma_i - d_i)^2 + \lambda \|d_i\|_0$$

s.t.  $d_1 \geq d_2 \geq \dots \geq d_m \geq 0.$

From the above derivation, the optimal solution of (2.1) is  $\mathbf{U}\text{diag}(\hat{d}_1, \hat{d}_2, \dots, \hat{d}_m)\mathbf{V}^T$ , where  $(\hat{d}_1, \hat{d}_2, \dots, \hat{d}_m)$  is the solution to (2.2).  $\square$

## Appendix B. Proof of Theorem 3.2

*Proof.* Let  $G_1^y(x) = y - x$  and  $G_2(x) = \lambda \nabla h(x)$  be two functions defined on  $[0, +\infty]$ . The properties of  $h(x)$  are related to the value of  $p$ . We prove the theorem in two cases.

**Case 1.**  $p \neq 1$ . Let

$$\bar{y} = \sup \{y \mid G_1^y(x) \cap G_2(x) = \emptyset\},$$

and

$$x_2^{\bar{y}} = \inf \{x \mid (x, q) \text{ be the intersection point of } G_1^y(x) \text{ and } G_2(x)\}.$$

When  $y > \bar{y}$ , there are two intersection points between  $G_1^y(x)$  and  $G_2(x)$ , denoted as  $(x_1^y, q_1^y)$  and  $(x_2^y, q_2^y)$ , where  $x_2^y > x_1^y$ . Since in this case,

$$\nabla h(x) = \frac{pb}{ax^{1-p}(1/\alpha + x^p)^2},$$

we note that  $G_2(0) = +\infty$ .

First, we consider  $y \leq \bar{y}$ . According to the definition of  $\bar{y}$ , there at most one intersection point of  $G_1^y(x)$  and  $G_2(x)$ . We have

$$\nabla f_y(x) = G_2(x) - G_1^y(x) \geq 0.$$

Therefore, the global minimum of  $f_y(x)$  is the value at  $x = 0$ .

Second, we consider  $y > \bar{y}$ . Then there are two intersection points of  $G_1^y(x)$  and  $G_2(x)$  and two conditions — viz.

(i) If there exists  $y > \bar{y}$  such that  $f_y(0) = f_y(x_2^y)$ , denote

$$y^* = \inf \{y \mid y > \bar{y}, f_y(0) = f_y(x_2^y)\}.$$

When  $y > y^*$ , let  $y = y^* + \epsilon$  for some  $\epsilon > 0$ . We have

$$\begin{aligned} f_y(x_2^{y^*}) - f_y(0) &= \frac{1}{2} (x_2^{y^*} - y^* - \epsilon)^2 + \lambda h(x_2^{y^*}) - \frac{1}{2} (y^* + \epsilon)^2 \\ &= \frac{1}{2} (x_2^{y^*} - y^*)^2 - \frac{1}{2} (y^*)^2 - \epsilon x_2^{y^*} + \lambda h(x_2^{y^*}) \end{aligned}$$

$$\begin{aligned}
&= f_{y^*}(x_2^{y^*}) - f_{y^*}(0) - \epsilon x_2^{y^*} \\
&= -\epsilon x_2^{y^*} < 0.
\end{aligned}$$

Because  $f_y$  is decreasing on  $[x_2^{y^*}, x_2^y]$ , we have

$$f_y(0) > f_y(x_2^{y^*}) \geq f_y(x_2^y).$$

Hence, if  $y > y^*$ , then  $f_y(x_2^y)$  is the global minimum of  $f_y(x)$ . When  $\bar{y} < y \leq y^*$ , we prove that  $f_y(0) \leq f_y(x_2^y)$  by contradiction. Suppose that there exists  $y^0$  such that  $\bar{y} < y^0 < y^*$  and  $f_{y^0}(0) > f_{y^0}(x_2^{y^0})$ . Since  $f_{\bar{y}}$  is strictly increasing on  $(0, x_2^{\bar{y}})$ , we have  $f_{\bar{y}}(x_2^{\bar{y}}) > f_{\bar{y}}(0)$ . Since

$$\begin{aligned}
f_{\bar{y}}(x_2^{\bar{y}}) &> f_{\bar{y}}(0), \\
f_{y^0}(x_2^{y^0}) &< f_{y^0}(0),
\end{aligned}$$

we have

$$\begin{aligned}
\lambda h(x_2^{\bar{y}}) - \lambda x_2^{\bar{y}} \nabla h(x_2^{\bar{y}}) - \frac{1}{2}(x_2^{\bar{y}})^2 &> 0, \\
\lambda h(x_2^{y^0}) - \lambda x_2^{y^0} \nabla h(x_2^{y^0}) - \frac{1}{2}(x_2^{y^0})^2 &< 0.
\end{aligned}$$

According to the intermediate value theorem, there exists  $\tilde{x}$  such that  $x_2^{\bar{y}} < \tilde{x} < x_2^{y^0}$  and

$$\lambda h(\tilde{x}) - \lambda \tilde{x} \nabla h(\tilde{x}) - \frac{1}{2}(\tilde{x})^2 = 0.$$

Let  $\tilde{y} = \lambda \nabla h(\tilde{x}) + \tilde{x}$ . Note that  $(\tilde{x}, \tilde{y} - \tilde{x})$  is the intersection point of  $G_1^{\tilde{y}}(x)$  and  $G_2(x)$  such that  $f_{\tilde{y}}(\tilde{x}) = f_{\tilde{y}}(0)$ . Since  $x_2^{\bar{y}} < \tilde{x} < x_2^{y^0}$  and  $\nabla h$  is convex and nonincreasing, we get that  $\bar{y} < \tilde{y} < y^0 < y^*$ , which contradicts the minimality of  $y^*$ .

- (ii) Since  $f_{\bar{y}}$  is increasing on  $(0, x_2^{\bar{y}})$ , we have  $f_{\bar{y}}(x_2^{\bar{y}}) > f_{\bar{y}}(0)$ . We now show that  $f_y(x_2^y) \geq f_y(0)$  for all  $y > \bar{y}$ . Suppose this is not true, and there exists  $y$  such that  $y > \bar{y}$  and  $f_y(x_2^y) < f_y(0)$ . It follows from the relations

$$\begin{aligned}
f_{\bar{y}}(x_2^{\bar{y}}) &> f_{\bar{y}}(0), \\
f_y(x_2^y) &< f_y(0),
\end{aligned}$$

that

$$\begin{aligned}
\lambda h(x_2^{\bar{y}}) - \lambda x_2^{\bar{y}} \nabla h(x_2^{\bar{y}}) - \frac{1}{2}(x_2^{\bar{y}})^2 &> 0, \\
\lambda h(x_2^y) - \lambda x_2^y \nabla h(x_2^y) - \frac{1}{2}(x_2^y)^2 &< 0.
\end{aligned}$$

Therefore, by the intermediate value theorem, there exists  $\tilde{x}$  such that  $x_2^{\bar{y}} < \tilde{x} < x_2^y$  and

$$\lambda h(\tilde{x}) - \lambda \tilde{x} \nabla h(\tilde{x}) - \frac{1}{2}(\tilde{x})^2 = 0.$$

Let  $\tilde{y} = \lambda \nabla h(\tilde{x}) + \tilde{x}$ . Note that  $(\tilde{x}, \tilde{y} - \tilde{x})$  is the intersection point of  $G_1^{\tilde{y}}(x)$  and  $G_2(x)$  such that  $f_{\tilde{y}}(\tilde{x}) = f_{\tilde{y}}(0)$ . Since  $x_2^{\bar{y}} < \tilde{x} < x_2^y$  and  $\nabla h$  is convex and nonincreasing, we get that  $\bar{y} < \tilde{y} < y$ , which contradicts  $f_y(0) \neq f_y(x_2^y)$  for all  $y > \bar{y}$ . Thus, for all  $y > \bar{y}$ , 0 is the minimum of  $f_y(x)$  on  $[0, y]$ . Similarly, when  $y \leq \bar{y}$ , we have  $\nabla f_y(x) = G_2(x) - G_1^y(x) \geq 0$ . Hence, the global minimum of  $f_y(x)$  is the value at  $x = 0$ .

**Case 2.**  $p = 1$ . In this case,  $G_2(0) < +\infty$ . We think about the following two conditions:

(i) Suppose that

$$G_1^{\lambda \nabla h(0)}(x) = \lambda \nabla h(0) - x \leq \lambda \nabla h(x)$$

for all  $x$  on  $(0, \lambda \nabla h(0))$ . For all  $y \leq \lambda \nabla h(0)$ , we have  $\nabla f_y(x) = G_2(x) - G_1^y(x) \geq 0$ . Thus, when  $y \leq \lambda \nabla h(0)$ , the minimum point of  $f_y(x)$  is  $f_y(0)$ . For all  $y > \lambda \nabla h(0)$ ,  $G_1^y(x) = y - x$  and  $G_2(x)$  have only one intersection point denoted as  $(x^y, q^y)$ . Then,  $f_y$  is decreasing on  $(0, x^y)$  and increasing on  $(x^y, y)$ . Hence, when  $y > \lambda \nabla h(0)$ , the minimum of  $f_y(x)$  is the value at  $x = x^y$ .

(ii) Suppose there exists  $0 < \hat{x} < \lambda \nabla h(0)$  such that

$$G_1^{\lambda \nabla h(0)}(\hat{x}) = \lambda \nabla h(0) - \hat{x} > \lambda \nabla h(\hat{x}).$$

Then,  $G_1^y(x) = y - x$  and  $G_2(x)$  have two intersection points — i.e.  $(0, \lambda \nabla h(0))$  and  $(x_2^{\lambda \nabla h(0)}, q_2^{\lambda \nabla h(0)})$ . Note that  $f_{\lambda \nabla h(0)}$  is strictly decreasing on  $(0, x_2^{\lambda \nabla h(0)})$ , we have

$$f_{\lambda \nabla h(0)}(x_2^{\lambda \nabla h(0)}) < f_{\lambda \nabla h(0)}(0).$$

Also, we denote

$$\bar{y} = \sup \{y \mid G_1^y(x) \cap G_2(x) = \emptyset\}.$$

Since  $f_{\bar{y}}$  is strictly increasing on  $(0, x_2^{\bar{y}})$ , we have  $f_{\bar{y}}(x_2^{\bar{y}}) > f_{\bar{y}}(0)$ . Since

$$\begin{aligned} f_{\bar{y}}(x_2^{\bar{y}}) &> f_{\bar{y}}(0), \\ f_{\lambda \nabla h(0)}(x_2^{\lambda \nabla h(0)}) &< f_{\lambda \nabla h(0)}(0), \end{aligned}$$

we have

$$\begin{aligned} \lambda h(x_2^{\bar{y}}) - x_2^{\bar{y}} \lambda \nabla h(x_2^{\bar{y}}) - \frac{1}{2}(x_2^{\bar{y}})^2 &> 0, \\ \lambda h(x_2^{\lambda \nabla h(0)}) - x_2^{\lambda \nabla h(0)} \lambda \nabla h(x_2^{\lambda \nabla h(0)}) - \frac{1}{2}(x_2^{\lambda \nabla h(0)})^2 &< 0. \end{aligned}$$

By the intermediate value theorem, there exists  $\tilde{x}$  such that  $x_2^{\bar{y}} < \tilde{x} < x_2^{\lambda \nabla h(0)}$  and

$$\lambda h(\tilde{x}) - \tilde{x} \lambda \nabla h(\tilde{x}) - \frac{1}{2}(\tilde{x})^2 = 0.$$

Let  $\tilde{y} = \lambda \nabla h(\tilde{x}) + \tilde{x}$ . Note that  $(\tilde{x}, \tilde{y} - \tilde{x})$  is the intersection point of  $G_1^{\tilde{y}}(x)$  and  $G_2(x)$  such that  $f_{\tilde{y}}(\tilde{x}) = f_{\tilde{y}}(0)$ . Since  $x_2^{\bar{y}} < \tilde{x} < x_2^{\lambda \nabla h(0)}$  and  $\nabla h$  is convex and nonincreasing, we get that  $\bar{y} < \tilde{y} < \lambda \nabla h(0)$ . Next, we set

$$y^* = \inf \{y \mid \bar{y} < y < \lambda \nabla h(0), f_y(0) = f_y(x_2^y)\}.$$

Given  $\bar{y} < y \leq \lambda \nabla h(0)$ , we can see that  $f_y$  is increasing on  $(0, x_1^y)$ , decreasing on  $(x_1^y, x_2^y)$  and increasing on  $(x_2^y, y)$ . Thus, 0 and  $x_2^y$  are two local minimum points of  $f_y(x)$  on  $[0, y]$ .

Next, for  $y^* < y \leq \lambda \nabla h(0)$ , let  $y = y^* + \epsilon$  for some  $\epsilon > 0$ . We have

$$\begin{aligned} f_y(x_2^{y^*}) - f_y(0) &= \frac{1}{2}(x_2^{y^*} - y^* - \epsilon)^2 + \lambda h(x_2^{y^*}) - \frac{1}{2}(y^* + \epsilon)^2 \\ &= \frac{1}{2}(x_2^{y^*} - y^*)^2 - \frac{1}{2}(y^*)^2 - \epsilon x_2^{y^*} + \lambda h(x_2^{y^*}) \\ &= f_{y^*}(x_2^{y^*}) - f_{y^*}(0) - \epsilon x_2^{y^*} \\ &= -\epsilon x_2^{y^*} < 0. \end{aligned}$$

Because  $f_y$  is decreasing on  $(x_2^{y^*}, x_2^y)$ , we have  $f_y(0) > f_y(x_2^{y^*}) \geq f_y(x_2^y)$ . Thus, when  $y > y^*$ ,  $f_y(x_2^y)$  is the global minimum of  $f_y(x)$ . Then for all  $\bar{y} < y \leq y^*$ , we prove  $f_y(0) \leq f_y(x_2^y)$  by contradiction. We suppose there exists  $y$  such that  $f_y(0) > f_y(x_2^y)$ . It follows from

$$\begin{aligned} f_{\bar{y}}(x_2^{\bar{y}}) &> f_{\bar{y}}(0), \\ f_y(x_2^y) &< f_y(0), \end{aligned}$$

that

$$\begin{aligned} \lambda h(x_2^{\bar{y}}) - x_2^{\bar{y}} \lambda \nabla h(x_2^{\bar{y}}) - \frac{1}{2}(x_2^{\bar{y}})^2 &> 0, \\ \lambda h(x_2^y) - x_2^y \lambda \nabla h(x_2^y) - \frac{1}{2}(x_2^y)^2 &< 0. \end{aligned}$$

By the intermediate value theorem, there exists  $\tilde{x}$  such that  $x_2^{\bar{y}} < \tilde{x} < x_2^y$  and

$$\lambda h(\tilde{x}) - \tilde{x} \lambda \nabla h(\tilde{x}) - \frac{1}{2}(\tilde{x})^2 = 0.$$

Let  $\tilde{y} = \lambda \nabla h(\tilde{x}) + \tilde{x}$ .  $(\tilde{x}, \tilde{y} - \tilde{x})$  is the intersection point of  $G_1^{\tilde{y}}(x)$  and  $G_2(x)$  such that  $f_{\tilde{y}}(\tilde{x}) = f_{\tilde{y}}(0)$ . Since  $x_2^{\bar{y}} < \tilde{x} < x_2^y$  and  $\nabla h$  is convex and nonincreasing, we get that  $\bar{y} < \tilde{y} < y \leq y^*$ , which contradicts the minimality of  $y^*$ .

If  $y \leq \bar{y}$ , we have

$$\nabla f_y(x) = G_2(x) - G_1^y(x) \geq 0,$$

so the global minimum point of  $f_y(x)$  on  $[0, y]$  is 0. Also, when  $y > \lambda \nabla h(0)$ ,  $G_1^y = y - x$  and  $G_2(x)$  have only one intersection point  $(x^y, q^y)$ . Then, we can get that  $f_y$  is decreasing on  $(0, x^y)$  and increasing on  $(x^y, y)$ . Thus, the global minimum point of  $f_y(x)$  is  $x^y$ .

The above considerations show that the solution of  $\text{Prox}_h(y)$  is the largest intersection point of  $G_1^y(x)$  and  $G_2(x)$  — i.e.

$$\hat{x}^y = \max \{x \mid \nabla f_y(x) = 0, 0 \leq x \leq y\},$$

when  $y$  is larger than a certain threshold. For other choices of  $y$ ,  $0 \in \text{Prox}_h(y)$ . It can be seen that we only need to compute the largest local minimum  $\hat{x}^y$  and compare the values of the objective function  $f_y(x)$  at 0 and  $\hat{x}^y$ .  $\square$

**Matlab Code.** The code is available at [https://github.com/Jin-liangXiao/TFR\\_code](https://github.com/Jin-liangXiao/TFR_code)

## References

- [1] H. Attouch, J. Bolte, P. Redont and A. Soubeiran, *Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality*, Math. Oper. Res. **35**, 438–457 (2010).
- [2] S. Bellavia, J. Gondzio and M. Porcelli, *A relaxed interior point method for low-rank semidefinite programming problems with applications to matrix completion*, J. Sci. Comput. **89**, 46 (2021).
- [3] J. Bolte, S. Sabach and M. Teboulle, *Proximal alternating linearized minimization for nonconvex and nonsmooth problems*, Math. Program. **146**, 459–494 (2014).
- [4] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Mach. Learn. **3**, 1–122 (2011).
- [5] J.-F. Cai, E.J. Candès and Z. Shen, *A singular value thresholding algorithm for matrix completion*, SIAM J. Optim. **20**, 1956–1982 (2010).
- [6] E.J. Candès, X. Li, Y. Ma and J. Wright, *Robust principal component analysis?*, J. ACM **58**, 1–37 (2011).
- [7] E.J. Candès and B. Recht, *Exact matrix completion via convex optimization*, Found. Comput. Math. **6**, 717–772 (2009).
- [8] L. Chen, X. Jiang, X. Liu and Z. Zhou, *Logarithmic norm regularized low-rank factorization for matrix and tensor completion*, IEEE Trans. Image Process. **30**, 3434–3449 (2021).
- [9] L.-J. Deng, R. Glowinski and X.-C. Tai, *A new operator splitting method for the Euler elastica model for image smoothing*, SIAM J. Imaging Sci. **12**, 1190–1230 (2019).
- [10] J. Ding and N.H. Rhee, *Computing solutions of the Yang-Baxter-Like matrix equation for diagonalisable matrices*, East Asian J. Appl. Math. **5**, 75–84 (2015).
- [11] W. Dong, G. Shi, X. Li, Y. Ma and F. Huang, *Compressive sensing via nonlocal low-rank regularization*, IEEE Trans. Image Process. **23**, 3618–3632 (2014).
- [12] H.-X. Dou, T.-Z. Huang, L.-J. Deng, X.-L. Zhao and J. Huang, *Directional  $l(0)$  sparse modeling for image stripe noise removal*, Remote Sens. **10**, 361 (2018).

- [13] Y. Duan, Q. Zhong, X.-C. Tai and R. Glowinski, *A fast operator-splitting method for Beltrami color image denoising*, J. Sci. Comput. **92**, 89 (2022).
- [14] K. Fan and A.J. Hoffman, *Some metric inequalities in the space of matrices*, Proc. Amer. Math. Soc. **6**, 111–116 (1955).
- [15] M. Fazel, *Matrix Rank Minimization with Applications*, PhD Thesis, Stanford University, 2002.
- [16] Q. Gao, P. Zhang, W. Xia, D. Xie, X. Gao and D. Tao, *Enhanced tensor RPCA and its application*, IEEE Trans. Pattern Anal. Mach. Intell. **43**, 2133–2140 (2021).
- [17] D. Geman and G. Reynolds, *Constrained restoration and the recovery of discontinuities*, IEEE Trans. Pattern Anal. Mach. Intell. **14**, 367–383 (1992).
- [18] C. Grussler and P. Giselsson, *Local convergence of proximal splitting methods for rank constrained problems*, in: *Proceedings of IEEE 56th Annual Conference on Decision and Control*, 702–708 (2017).
- [19] C. Grussler and P. Giselsson, *Low-rank inducing norms with optimality interpretations*, SIAM J. Optim. **28**, 3057–3078 (2018).
- [20] C. Grussler and P. Giselsson, *Efficient proximal mapping computation for low-rank inducing norms*, J. Optim. Theory Appl. **192**, 168–194 (2022).
- [21] C. Grussler, A. Rantzer and P. Giselsson, *Low-rank optimization with convex constraints*, IEEE Trans. Autom. Control **63**, 4000–4007 (2018).
- [22] S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng and L. Zhang, *Weighted nuclear norm minimization and its applications to low level vision*, Int. J. Comput. Vis. **121**, 183–208 (2017).
- [23] H. He, C. Ling and W. Xie, *Tensor completion via a generalized transformed tensor t-product decomposition without t-svd*, J. Sci. Comput. **93**, 47 (2022).
- [24] Y. Hu, D. Zhang, J. Ye, X. Li and X. He, *Fast and accurate matrix completion via truncated nuclear norm regularization*, IEEE Trans. Pattern Anal. Mach. Intell. **35**, 2117–2130 (2012).
- [25] Z.-F. Jin, Z. Wan, Y. Jiao and X. Lu, *An alternating direction method with continuation for nonconvex low rank minimization*, J. Sci. Comput. **66**, 849–869 (2016).
- [26] Z. Kang, C. Peng and Q. Cheng, *Top-n recommender system via matrix completion*, Proc. AAAI Conf. Artif. Intell. **30**, 1 (2016).
- [27] X. Li, H. Zhang and R. Zhang, *Matrix completion via non-convex relaxation and adaptive correlation learning*, IEEE Trans. Pattern Anal. Mach. Intell. **45**, 1981–1991 (2022).
- [28] J. Liu, L.-J. Deng, F. Fang and T. Zeng, *A Rudin-Osher-Fatemi model-based pansharpening approach using RKHS and AHF representation*, East Asian J. Appl. Math. **9**, 13–27 (2019).
- [29] Q. Liu, Z. Lai, Z. Zhou, F. Kuang and Z. Jin, *A truncated nuclear norm regularization method based on weighted residual error for matrix completion*, IEEE Trans. Image Process. **25**, 316–330 (2015).
- [30] C. Lu, J. Tang, S. Yan and Z. Lin, *Nonconvex nonsmooth low rank minimization via iteratively reweighted nuclear norm*, IEEE Trans. Image Process. **25**, 829–839 (2016).
- [31] C. Lu, C. Zhu, C. Xu, S. Yan and Z. Lin, *Generalized singular value thresholding*, in: *Proceedings of the AAAI Conference on Artificial Intelligence* **29**(1), (2015).
- [32] L. Ma, T. Zeng and G. Li, *Hybrid variational model for texture image restoration*, East Asian J. Appl. Math. **7**, 629–642 (2017).
- [33] F. Nie, Z. Hu and X. Li, *Matrix completion based on non-convex low-rank approximation*, IEEE Trans. Image Process. **28**, 2378–2388 (2018).
- [34] F. Nie, H. Huang and C. Ding, *Low-rank matrix recovery via efficient schatten p-norm minimization*, in: *Proceedings of the 26 AAAI Conference on Artificial Intelligence* **26**, 655–661 (2012).
- [35] C. Olsson, M. Carlsson and E. Bylow, *A non-convex relaxation for fixed-rank approximation*, in: *IEEE International Conference on Computer Vision Workshop*, pp. 1809–1817 (2017).
- [36] W. Qin, H. Wang, F. Zhang, J. Wang, X. Luo and T. Huang, *Low-rank high-order tensor comple-*

- tion with applications in visual data*, IEEE Trans. Image Process. **31**, 2433–2448 (2022).
- [37] B. Recht, M. Fazel and P.A. Parrilo, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM Rev. **52**, 471–501 (2010).
- [38] Y. Shan, D. Hu and Z. Wang, *A novel truncated norm regularization method for multi-channel color image denoising*, IEEE Trans. Circuits Syst. Video Technol. **34**, 8427–8441 (2024).
- [39] C. Shi, Z. Huang, L. Wan and T. Xiong, *Low-rank tensor completion based on log-det rank approximation and matrix factorization*, J. Sci. Comput. **80**, 1888–1912 (2019).
- [40] H. Wang, F. Zhang, J. Wang, T. Huang, J. Huang and X. Liu, *Generalized nonconvex approach for low-tubal-rank tensor recovery*, IEEE Trans. Neural Netw. Learn. Syst. **33**, 3305–3319 (2021).
- [41] Z. Wang, A.C. Bovik, H.R. Sheikh and E.P. Simoncelli, *Image quality assessment: From error visibility to structural similarity*, IEEE Trans. Image Process. **13**, 600–612 (2004).
- [42] Z. Wang, Y. Liu, X. Luo, J. Wang, C. Gao, D. Peng and W. Chen, *Large-scale affine matrix rank minimization with a novel nonconvex regularizer*, IEEE Trans. Neural Netw. Learn. Syst. **33**, 4661–4675 (2021).
- [43] Y. Wei and Y. Qu, *Perturbation bound for the eigenvalues of a singular diagonalizable matrix*, East Asian J. Appl. Math. **4**, 88–94 (2014).
- [44] R. Wen, L.-J. Deng, Z.-C. Wu, X. Wu and G. Vivone, *A novel spatial fidelity with learnable nonlinear mapping for panchromatic sharpening*, IEEE Trans. Geosci. Remote Sens. **61**, 1–15 (2023).
- [45] C. Wu, X. Guo, Y. Gao and Y. Xue, *A general non-Lipschitz infimal convolution regularized model: Lower bound theory and algorithm*, SIAM J. Imaging Sci. **15**, 1499–1538 (2022).
- [46] Z.-C. Wu, T.-Z. Huang, L.-J. Deng, H.-X. Dou and D. Meng, *Tensor wheel decomposition and its tensor completion application*, Proc. Adv. Neural Inf. Process. Syst. **35**, 27008–27020 (2022).
- [47] Z.-C. Wu, T.-Z. Huang, L.-J. Deng, J.-F. Hu and G. Vivone, *Vo+ net: An adaptive approach using variational optimization and deep learning for panchromatic sharpening*, IEEE Trans. Geosci. Remote Sens. **60**, 1–16 (2021).
- [48] Z.-C. Wu, T.-Z. Huang, L.-J. Deng, J. Huang, J. Chanussot and G. Vivone, *Lrtcfpan: Low-rank tensor completion based framework for pansharpening*, IEEE Trans. Image Process. **32**, 1640–1655 (2023).
- [49] Z.-C. Wu, T.-Z. Huang, L.-J. Deng and G. Vivone, *A framelet sparse reconstruction method for pansharpening with guaranteed convergence*, Inverse Probl. Imaging **17**, 1277–1300 (2023).
- [50] J.-L. Xiao, T.-Z. Huang, L.-J. Deng and H.-X. Dou, *A novel  $\ell_0$  minimization framework of tensor tubal rank and its multi-dimensional image completion application*, Inverse Probl. Imaging **18**, 1366–1389 (2024).
- [51] J.-L. Xiao, T.-Z. Huang, L.-J. Deng, Z.-C. Wu and G. Vivone, *A new context-aware details injection fidelity with adaptive coefficients estimation for variational pansharpening*, IEEE Trans. Geosci. Remote Sens. **60**, 1–15 (2022).
- [52] J.-L. Xiao, T.-Z. Huang, L.-J. Deng, Z.-C. Wu, X. Wu and G. Vivone, *Variational pansharpening based on coefficient estimation with nonlocal regression*, IEEE Trans. Geosci. Remote. Sens. **61**, 1–15 (2023).
- [53] T. Xie, S. Li and B. Sun, *Hyperspectral images denoising via nonconvex regularized low-rank and sparse matrix decomposition*, IEEE Trans. Image Process. **29**, 44–56 (2019).
- [54] Q. Xie, Q. Zhao, D. Meng, Z. Xu, S. Gu, W. Zuo and L. Zhang, *Multispectral images denoising by intrinsic tensor sparsity regularization*, in: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1692–1700 (2016).
- [55] H. Xu, C. Caramanis and S. Sanghavi, *Robust PCA via outlier pursuit*, IEEE Trans. Inf. Theory **58**, 3047–3064 (2012).
- [56] T. Xu, T.-Z. Huang, L.-J. Deng, J.-L. Xiao, C. Broni-Bediako, J. Xia and N. Yokoya, *A coupled ten-*

- sor double-factor method for hyperspectral and multispectral image fusion*, IEEE Trans. Geosci. Remote Sens. **62**, 1–17 (2024).
- [57] T. Xu, T.-Z. Huang, L.-J. Deng and N. Yokoya, *An iterative regularization method based on tensor subspace representation for hyperspectral image super-resolution*, IEEE Trans. Geosci. Remote Sens. **60**, 1–16 (2022).
- [58] J. Xue, Y. Zhao, W. Liao and J.C.-W. Chan, *Hyper-Laplacian regularized nonlocal low-rank matrix recovery for hyperspectral image compressive sensing reconstruction*, Inf. Sci. **501**, 406–420 (2019).
- [59] J. Xue, Y. Zhao, W. Liao and J. C.-W. Chan, *Nonconvex tensor rank minimization and its applications to tensor recovery*, Inf. Sci. **503**, 109–128 (2019).
- [60] Y. Yang, Y. Li, C. Wu and Y. Duan, *A convergent iterative support shrinking algorithm for non-Lipschitz multi-phase image labeling model*, J. Sci. Comput. **96**, 47 (2023).
- [61] J. Yuan, J. Shi and X.-C. Tai, *A convex and exact approach to discrete constrained tv-l1 image approximation*, East Asian J. Appl. Math. **1**, 172–186 (2011).
- [62] C. Zeng, *Proximal linearization methods for schatten p-quasi-norm minimization*, Numer. Math. **153**, 213–248 (2023).
- [63] C. Zeng, *Rank properties and computational methods for orthogonal tensor decompositions*, J. Sci. Comput. **94**, 6 (2023).
- [64] C. Zeng, C. Wu and R. Jia, *Non-Lipschitz models for image restoration with impulse noise removal*, SIAM J. Imaging Sci. **12**, 420–458 (2019).
- [65] H. Zeng, X. Xie and J. Ning, *Hyperspectral image denoising via global spatial-spectral total variation regularized nonconvex local low-rank tensor approximation*, Signal Process. **178**, 107805 (2021).
- [66] H. Zhang, C. Gong, J. Qian, B. Zhang, C. Xu and J. Yang, *Efficient recovery of low-rank matrix via double nonconvex nonsmooth rank minimization*, IEEE Trans. Neural Netw. Learn. Syst. **30**, 2916–2925 (2019).
- [67] H. Zhang, J. Qian, B. Zhang, J. Yang, C. Gong and Y. Wei, *Low-rank matrix recovery via modified schatten-p norm minimization with convergence guarantees*, IEEE Trans. Image Process. **29**, 3132–3142 (2019).
- [68] H. Zhang, B. Wen, Z. Zha, B. Zhang, Y. Tang, G. Yu and W. Du, *Accelerated palm for nonconvex low-rank matrix recovery with theoretical analysis*, IEEE Trans. Circuits Syst. Video Technol. **34**, 2304–2317 (2024).
- [69] Y. Zheng, G. Liu, S. Sugimoto, S. Yan and M. Okutomi, *Practical low-rank matrix approximation under robust l 1-norm*, in: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1410–1417 (2012).
- [70] Q. Zhong, R.W. Liu and Y. Duan, *Spatially adapted first and second order regularization for image reconstruction: From an image surface perspective*, J. Sci. Comput. **92**, 33 (2022).
- [71] X. Zhong, L. Xu, Y. Li, Z. Liu and E. Chen, *A nonconvex relaxation approach for rank minimization problems*, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, **29**, 1 (2015).