

Improved Robust Low-Rank Regularization Tensor Completion

Xiangyi Wang¹, Wei Jiang^{1,2}

¹School of Mathematics, Liaoning Normal University, Dalian, China ²College of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou, China Email: 1245315345@qq.com

How to cite this paper: Wang, X.Y. and Jiang, W. (2022) Improved Robust Low-Rank Regularization Tensor Completion. *Open Access Library Journal*, 9: e9425. https://doi.org/10.4236/oalib.1109425

Received: October 6, 2022 Accepted: November 7, 2022 Published: November 10, 2022

Copyright © 2022 by author(s) and Open Access Library Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0). http://creativecommons.org/licenses/by/4.0/

Abstract

In recent years, there are many surrogates for tensor tubal rank. In this paper, we propose a hybrid norm consisting of the weighted nuclear norm and the weighted Frobenius norm (WTNFN) of a tensor. The WTNFN is a surrogate for tensor tubal rank, and studies the weighted tensor nuclear and Frobenius norm minimization (WTNFNM) problem. The aim is to enhance the stability of the solution and improve the shortcomings of the traditional method of minimizing approximate rank functions based on the tensor nuclear norm (TNN) in the field of low-rank tensor recovery. Based on this definition, we build a novel model for typical tensor recovery problems *i.e.* the weighted tensor nuclear and Frobenius tensor completion (WTNFNTC). The experimental results on synthetic data and real data show that the stability and recovery effects of the model are improved compared with related algorithms.

Subject Areas

Machine Learning

Keywords

Low-Rank Recovery, T-SVD, Weighted Tensor Nuclear, Weighted Tensor Frobenius Norm

1. Introduction

With the development of modern computer information technology, various data have emerged in many application fields such as mobile Internet, statistics, and psychometrics. These data are not only huge in scale, but also have the remarkable characteristics of extremely high dimensionality and complex structure. As a natural representation of multi-dimensional data, tensors are the ge-

neralization of 1-order vectors and 2-order matrices in higher-order forms [1].

In recent years, the research interest in tensors has expanded to many fields: e.g., computer vision [2], machine learning [3], pattern recognition [4], signal processing and image processing [5] [6]. For example, in a video recommendation system, user rating data can be expressed as a 3-order tensor of "user \times video \times time" [7]; a color image is a 3-way object with column, row and color modes. Compared with vectors and matrices, in these applications, tensors can better retain the intrinsic structure information of high-dimensional spatial data. But the tensor of interest is frequently low-rank, or approximately so [8], and hence has a much lower-dimensional structure. This motivates the low-rank tensor approximation and recovery problem. The core of the low-rank tensor recovery model is the tensor rank minimization.

In matrix processing, the rank minimization is NP-hard and is difficult to solve, the problem is generally relaxed by substitutively minimizing the nuclear norm of the estimated matrix, which is a convex relaxation of minimizing the matrix rank [9]. However, as a high-level extension of vectors and matrices, whether there is also a method similar to the matrix LMRA for low-rank tensor approximation and how to extend the low-rank definition from a matrix to a tensor is still a question worth studying. Because the notions of the tensor nuclear norm, and even tensor rank, are still ambiguous or nonuniform.

The most two popular tensor rank definitions are the

CANDECOMP/PARAFAC(CP)-rank, which is related to the CANDECOMP/ PARAFAC decomposition [10], and Tucker-rank, which is corresponding to the Tucker decomposition [11] [12]. The CP rank is defined as the smallest number of rank one tensor decomposition, where the computation is generally NP-hard [13] and its convex relaxation is intractable. For the Tucker rank minimization, Liu *et al.* [14] first proposed the sum of nuclear norms (SNN) of unfolding matrices of a tensor to approximate the sum of Tucker rank of a tensor for tensor completion. Unfortunately, albeit the tucker rank and its convex relaxation are tractable and widely used, its direct unfolding and folding operation tend to damage intrinsic structure information of the tensor and SNN is not the convex envelope of the tucker rank according to [15].

Recently, Kilmer *et al.* [16] focused on a new tensor decomposition paradigm, namely tensor singular value decomposition (T-SVD), which is similar to the singular value decomposition of matrix. By using T-SVD, Kilmer *et al.* [16] defined the concept of tensor tubal rank, which is based on tensor-product (t-product) and its algebraic framework. T-SVD allows the matrix analysis to be extended to the tensor, while avoiding the inherent information loss in matricization or flattening of the tensor. The tensor tubal rank can well characterize the intrinsic structure of a tensor, and thus, solving low-tubal-rank related problems starts to attract more and more reasonable research interest these years.

This work focuses on the low-tubal-rank tensor recovery problem under the T-SVD framework. For the underlying tensor with low-tubal-rank assumption,

it can be recovered by solving the following optimization:

$$\min_{\boldsymbol{\mathcal{X}}} f(\boldsymbol{\mathcal{X}}) + \tau \cdot rank_t(\boldsymbol{\mathcal{X}}) \tag{1}$$

where $\tau > 0$ is a regularization parameter, $rank_t(\cdot)$ denotes the tensor tubal rank.

It has been proved that solving (1) is NP-hard. The commonly used convex relaxation for the rank function is the tensor nuclear norm. In the recent work [17], Lu *et al.* first defined the tensor average rank and deduced that a tensor always has low average rank if it has low tubal rank and then proposed a novel tensor nuclear norm that is proven to be the convex envelope of the tensor average rank within the unit ball of the tensor spectral norm, which performed much better than other types of tensor nuclear norm in many tasks. This methodology is called as tensor nuclear norm minimization (TNNM). The nuclear norm of a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, denoted by $\|\mathcal{X}\|_{\star}$. The TNNM [17] approach has been attracting significant attention due to its rapid development in both theory and implementation. A general TNNM problem can be formulated as

$$\min_{\mathcal{X}} f(\mathcal{X}) + \tau \cdot \left\| \mathcal{X} \right\|_{+} \tag{2}$$

However, even though tensor nuclear norm relaxation is becoming a popular scheme for the low-tubal-tensor recovery problem, it is also associated with shortcoming. The tensor nuclear norm suffers from the major limitation that all singular values are simultaneously minimized, which implies that large singular values are penalized more heavily than small ones. Under this circumstance, method based on nuclear norms usually fails to find a good solution. Even worse, the resulting global optimal solution may deviate significantly from the ground truth. In order to break through the limitations of the tensor nuclear norm relaxation method, there is currently work to solve the low-tubal-rank recovery problem by employing a weighted tensor nuclear norm and nonconvex relaxation strategy. The essence of relaxation is to overcome the unbalanced punishment of different singular values. Essentially, they will keep the larger singular values larger and shrink the smaller singular values. Hence, the method has already been well studied and applied in many low-tubal-rank tensor recovery problems, and in real applications, weighted tensor nuclear norm relaxation methods usually perform better than tensor nuclear norm relaxation. In addition, for low-rank representation, the Frobenius norm has been shown to perform similarly to the nuclear norm [18] [19].

This paper mainly studies the tensor completion problem. Different from the existing nonconvex tensor completion models, this paper also considers the stability and low rank of the solution of the tensor recovery model, and proposes a relatively stable low-rank nonconvex approximation of tensors. Specifically, our contributions can be summarized in the following three points.

• First, Similar to the tensor nuclear norm definition form, we show that the square of the tensor Frobenius norm can be defined as the sum of its singular value squares. Based on the above definition, assign different weights to sin-

gular values of a tensor. We propose a surrogate of the tensor tubal rank, *i.e.*, the WTNFN.

- Second, to optimize the WTNFN-based minimization Problem, we extend the weighted tensor nuclear and Frobenius norm singular value thresholding (WTNFNSVT) operator which is given with careful derivations. For the tensor in the complex field, and demonstrate that it is the exact solution to the WTNFN-based minimization problem. The bounded convergence is also demonstrated with strict proof.
- Third, we adapt the proposed weighted tensor nuclear and Frobenius norm model to different computer vision tasks. In tensor completion the model outperforms TNN methods in both PSNR index and visual perception. The proposed WTNFN-TC models also achieve better performance than traditional algorithms, especially, the low-rank learning methods designed for these tasks. The global optimal solution obtained above is relatively stable compared to existing weighted nuclear norm models.

The rest of this paper proceeds as follows: Section II briefly introduces some notations, preliminaries, and some preliminary works; Section III analyzes the optimization of the WTNFNM models, including the basic WTNFN operator in detail, and the convergences of the proposed methods are specifically analyzed with strict formulation; Section IV applies the WTNFNM model to tensor completion and compares the proposed algorithm with numerical experiments demonstrate the effectiveness of the proposed methods compared with other state-of-the-art methods; finally the conclusions and future work are given in Section V.

2. Notations and Preliminaries

2.1. Notations

We denote tensors by boldface Euler script letters, e.g., \mathcal{A} . Matrices are denoted by boldface capital letters, e.g., A, vectors are denoted by boldface lowercase letters, e.g., a, and scalars are denoted by lowercase letters, e.g., a. Real and complex numbers are in \mathbb{R} and \mathbb{C} fields respectively. For a 3-way tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we denote $\mathcal{A}(i,:,:)$, $\mathcal{A}(:,i,:)$, and $\mathcal{A}(:,:,i)$ as the *i*-th horizontal, lateral and frontal slice. Generally, the frontal slice $\mathcal{A}(:,:,i)$ can also be denoted as $A^{(i)}$. We denote its (i, j, k)-th entry as $\mathcal{A}(i, j, k)$ or \mathcal{A}_{ijk} . The tube is denoted as $\langle A, B \rangle = \operatorname{Tr}(A^*B)$, where $\operatorname{Tr}(\cdot)$ stands for the trace, and A^{T} is the conjugate transpose of A, and the inner product of any two tensors \mathcal{A} and \mathcal{B} in $\mathbb{C}^{n_1 \times n_2 \times n_3}$, $\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i=1}^{n_3} \langle A^{(i)}, B^{(i)} \rangle$. For any $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$, the complex conjugate of \mathcal{A} is denoted as

For any $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$, the complex conjugate of \mathcal{A} is denoted as $conj(\mathcal{A})$ which takes the complex conjugate of each entry of \mathcal{A} . $\lceil b \rceil$ is the nearest integer greater than or equal to b. Now we consider applying the Discrete Fourier Transformation (DFT) on tensors. $\overline{\mathcal{A}}$ is the result of the Discrete Fourier Transformation of \mathcal{A} along the third dimension, that is,

$$\overline{\mathbf{A}} = fft(\mathbf{A}, [], 3) \tag{3}$$

We denote $\overline{A} \in \mathbb{C}^{n_i n_3 \times n_2 n_3}$ as a block diagonal matrix with its *i*-th block on the diagonal as the *i*-th frontal slice $\overline{A}^{(i)}$ of \mathcal{A} , *i.e.*,

$$\overline{A} = \text{bdiag}\left(\overline{A}\right) = \begin{bmatrix} \overline{A}^{(1)} & & \\ & \overline{A}^{(2)} & & \\ & & \ddots & \\ & & & \overline{A}^{(n_3)} \end{bmatrix}$$
(4)

Next, we define the block circulant matrix bcirc(A) reshapes tensor $A \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ into a block circulant matrix with size $n_1 n_3 \times n_2 n_3$, *i.e.*,

$$\operatorname{bcirc}(\mathcal{A}) = \begin{bmatrix} A^{(1)} & A^{(n_{3})} & \cdots & A^{(2)} \\ A^{(2)} & A^{(1)} & \cdots & A^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ A^{(n_{3})} & A^{(n_{3}-1)} & \cdots & A^{(1)} \end{bmatrix}$$
(5)

2.2. T-Product and T-SVD

For tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, unfold (\mathcal{A}) unfolds tensor \mathcal{A} as follows:

unfold
$$(\mathcal{A}) = \begin{bmatrix} \mathbf{A}^{(1)} \\ \mathbf{A}^{(2)} \\ \vdots \\ \mathbf{A}^{(n_3)} \end{bmatrix}$$
, fold $(unfold(\mathcal{A})) = \mathcal{A}$ (6)

According to [17], t-product is equivalent to the familiar matrix multiplication in the Fourier domain.*i.e.*, C = A * B is equivalent to $\overline{C} = \overline{A} \cdot \overline{B}$. Then the tensor product of two third-order tensors can be defined as follows:

Definition 1 (T-product [16]) Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{B} \in \mathbb{R}^{n_1 \times m \times n_3}$. Then the t-product $\mathcal{C} = \mathcal{A} * \mathcal{B}$ is defined to be a tensor of size $n_1 \times m \times n_3$

$$\mathcal{C} = \mathcal{A} * \mathcal{B} = \operatorname{fold}(\operatorname{bcirc}(\mathcal{A}) \cdot \operatorname{unfold}(\mathcal{B}))$$
(7)

Definition 2 (Conjugate Transpose [16]) The conjugate transpose of a tensor of $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ size $n_1 \times n_2 \times n_3$ is a $n_2 \times n_1 \times n_3$ tensor \mathcal{A}^{T} obtained by conjugate transposing each of the frontal slice and then reversing the order of transposed frontal slices 2 through n_3 . If tensor is $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, Conjugate transpose is also recorded as \mathcal{A}^{T} .

Definition 3 (Identity Tensor [16]) The identity tensor $\mathcal{I} \in \mathbb{R}^{n \times n \times n_3}$ is the tensor whose first frontal slice is an $n \times n$ identity matrix, and whose other frontal slices are all zeros.

Definition 4 (Orthogonal Tensor [16]) A tensor $\mathcal{Q} \in \mathbb{R}^{n \times n \times n_3}$ is orthogonal if it satisfies

$$\boldsymbol{\mathcal{Q}}^{\mathrm{T}} \ast \boldsymbol{\mathcal{Q}} = \boldsymbol{\mathcal{Q}} \ast \boldsymbol{\mathcal{Q}}^{\mathrm{T}} = \boldsymbol{\mathcal{I}}$$
(8)

Definition 5 (F-diagonal Tensor [16]) A tensor is called *F*-diagonal if each of its frontal slices is a diagonal matrix.

As stated above, tensor-SVD can be defined.

Theorem 1 (T-SVD [17]) Let $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. Then it can be factorized as

$$\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^{\mathrm{T}}$$
⁽⁹⁾

where $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$, $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogona tensors, and $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is an *F*-diagonal tensor.

It is worth noting that T-SVD can be efficiently computed based on the matrix SVD in the Fourier domain. The details of T-SVD are described in Algorithm 1.

Algorithm 1 T-SVD
Input: The tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$.
Output: T-SVD components \mathcal{U}, \mathcal{S} , and \mathcal{V} of \mathcal{A} .
1. Compute $\bar{\mathcal{A}} = fft(\mathcal{A}, [], 3)$.
2. Compute each frontal slice of $\bar{\mathcal{U}}, \bar{\mathcal{S}}$ and $\bar{\mathcal{V}}$ from $\bar{\mathcal{A}}$ by
for $i=1,,\lceil rac{n_3+1}{2} ceil$ do
$[ar{oldsymbol{U}}^{(i)},ar{oldsymbol{S}}^{(i)},ar{oldsymbol{V}}^{(i)}]=SVD(ar{oldsymbol{A}}^{(i)})$;
end for
for $i = \lceil rac{n_3+1}{2} ceil,, n_3$ do
$ar{oldsymbol{U}}^{(i)}=ar{conj}ar{oldsymbol{U}}^{(n_3+2-i)};$
$ar{oldsymbol{S}}^{(i)}=ar{oldsymbol{S}}^{(n_3+2-i)};$
$ar{m{V}}^{(i)}=conjar{m{V}}^{(n_3+2-i)};$
end for
3. Compute \mathcal{U}, \mathcal{S} and \mathcal{V} by
$ar{\mathcal{U}} = ifft(\mathcal{U}, [], 3), \ ar{\mathcal{S}} = ifft(\mathcal{S}, [], 3), ar{\mathcal{V}} =$
$ifft(oldsymbol{\mathcal{V}},[],3).$

Definition 6 (Tensor Tubal Rank [20] [21]) For $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, the tensor tubal rank, denoted as $rank_t(\mathcal{A})$, is defined as the number of nonzero singular tubes of \mathcal{S} , where \mathcal{S} is from the T-SVD of $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^{\mathrm{T}}$. The tensor tubal rank is equivalent to the number of nonzero singular values of \mathcal{A} . That is

$$rank_{i}\left(\mathcal{A}\right) = \#\left\{i, \mathcal{S}\left(i, i, :\right) \neq 0\right\} = \#\left\{i, \mathcal{S}\left(i, i, 1\right) \neq 0\right\}.$$
(10)

where $\mathcal{S}(i,i,1) = \frac{1}{n_3} \sum_{j=1}^{n_3} \overline{\mathcal{S}}(i,i,j)$, it uses the property induced by inverse DFT.

Denote $\sigma(\mathcal{A}) \in \mathbb{R}^{n \times n_3}$ with $\sigma_{ij}(\mathcal{A}) = \overline{\mathcal{S}}(i, i, j)$, therefore, the Tensor unclear norm and Frobenius norm tensor are defined as follows.

Definition 7 (Tensor Nuclear Norm (TNN) [22]) For tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, based on the result about the tensor nuclear norm derived from the T-product. we have the tensor nuclear norm (TNN) is defined to be

$$\left\|\boldsymbol{\mathcal{A}}\right\|_{\star} = \frac{1}{n_3} \sum_{i=1}^{m} \sum_{j=1}^{n_3} \sigma_{ij}\left(\boldsymbol{\mathcal{A}}\right)$$
(11)

Theorem 2 (Tensor Frobenius Norm (TFN)) Given a tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, if $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$ the following hold:

$$\left\|\boldsymbol{\mathcal{A}}\right\|_{F} = \frac{1}{\sqrt{n_{3}}} \sum_{i=1}^{m} \sum_{j=1}^{n_{3}} \sqrt{\sigma_{ij}^{2}\left(\boldsymbol{\mathcal{A}}\right)}$$
(12)

3. Proposed Algorithm Scheme

Based on the definition of tensor nuclear norm and tensor Frobenius norm in

Definition 8 and Theorem 2, we present a new nonconvex regularizer which is defined as the sum of the weighted tensor nuclear norm and the square of the weighted tensor Frobenius norm (WTNFN) of $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$

$$\begin{aligned} \left\| \boldsymbol{\mathcal{X}} \right\|_{\text{WTNFN}} &= \left\| \boldsymbol{\mathcal{X}} \right\|_{\boldsymbol{W},\star} + \gamma \left\| \boldsymbol{\mathcal{X}} \right\|_{\boldsymbol{W},F}^{2} \\ &= \frac{1}{n_{3}} \sum_{i=1}^{m} \sum_{j=1}^{n_{3}} w_{ij} \sigma_{ij} \left(\boldsymbol{\mathcal{X}} \right) + \frac{\gamma}{n_{3}} \sum_{i=1}^{m} \sum_{j=1}^{n_{3}} w_{ij} \sigma_{ij}^{2} \left(\boldsymbol{\mathcal{X}} \right) \end{aligned}$$
(13)

where $\mathbf{W} = (w_{ij})_{n \times n_3}$ is the weight matrix, $n = \min(n_1, n_2)$, γ is a positive parameter.

3.1. Problem Formulation

Consider the general weighted tensor nuclear and Frobenius norm framework for tensor tubal rank minimization problems

$$\min_{\boldsymbol{\mathcal{X}}} f(\boldsymbol{\mathcal{X}}) + \tau \cdot \left\| \boldsymbol{\mathcal{X}} \right\|_{\text{WTNFN}}$$
(14)

where $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\tau > 0$ is a positive parameter. And the loss function $f(\mathcal{X})$ satisfy the following assumptions:

f: ℝ<sup>n₁×n₂×n₃ → ℝ⁺ is continuously differentiable with Lipschitz continuous gradient, *i.e.*, for any X, Y ∈ ℝ<sup>n₁×n₂×n₃
</sup></sup>

$$\left\|\nabla f\left(\boldsymbol{\mathcal{X}}\right) - \nabla f\left(\boldsymbol{\mathcal{Y}}\right)\right\|_{F} \le L(f)\left\|\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{Y}}\right\|$$
(15)

where L(f) > 0 is the Lipschitz constant.

• *f* is coercive, *i.e.*, $f(\mathcal{X}) \to \infty$ when $\mathcal{X} \to \infty$.

To enhance low rank, we need to design a scheme to keep the weights of large singular values sufficiently small and the weights of small singular values sufficiently large, which will lead to a nearly unbiased low rank approximation. Intuitively, one can set each weight, which the entries of matrix $w_{ij} \in \mathbb{R}^+$ to be inversely proportional to the corresponding singular value $\sigma_{ij}(\mathcal{X})$, which will penalize large singular values less and overcome the unfair penalization of different singular values.

3.2. Methodology

We solve problem (14) by generating an iterative sequence $\mathcal{X}^{(k)}$, where $\mathcal{X}^{(k+1)}$ is the minimizer of a weighted tensor nuclear and Frobenius norm regularized problem that comes from the linearized approximation of the objective function $f(\mathcal{X})$ at $\mathcal{X}^{(k)}$. To realize this, for the loss function $f(\mathcal{X})$, we approximate it as a quadratic function

$$f(\boldsymbol{\mathcal{X}}) = f(\boldsymbol{\mathcal{X}}^{(k)}) + \left\langle \nabla f(\boldsymbol{\mathcal{X}}^{(k)}), \boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{X}}^{(k)} \right\rangle + \frac{\boldsymbol{\mu}^{(k)}}{2} \left\| \boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{X}}^{(k)} \right\|_{F}^{2}$$
(16)

Since optimizing $f(\mathcal{X})$ directly is hard, we minimize its approximation (16) instead. Hence, we solve the following relaxed problem to update $\mathcal{X}^{(k+1)}$

$$\mathcal{X}^{(k+1)} = \arg\min_{\mathcal{X}} f\left(\mathcal{X}^{(k)}\right) + \left\langle \nabla f\left(\mathcal{X}^{(k)}\right), \mathcal{X} - \mathcal{X}^{(k)} \right\rangle + \frac{\mu^{(k)}}{2} \left\| \mathcal{X} - \mathcal{X}^{(k)} \right\|_{F}^{2} + \tau \cdot \left\| \mathcal{X}^{(k)} \right\|_{WTNFN}$$
(17)

By ignoring constant terms and combining others, (17) can be expressed equivalently as

$$\boldsymbol{\mathcal{X}}^{(k+1)} = \frac{\boldsymbol{\mu}^{(k)}}{2\tau} \left\| \boldsymbol{\mathcal{X}} - \left(\boldsymbol{\mathcal{X}}^{(k)} - \frac{1}{\boldsymbol{\mu}^{(k)}} \nabla f\left(\boldsymbol{\mathcal{X}}^{(k)}\right) \right) \right\|_{F}^{2} + \tau \cdot \left\| \boldsymbol{\mathcal{X}}^{(k)} \right\|_{WTNFN}$$
(18)

Thus, consider $\boldsymbol{\mathcal{Y}}^{(k)} = \boldsymbol{\mathcal{X}}^{(k)} - \frac{1}{\mu^{(k)}} \nabla f(\boldsymbol{\mathcal{X}}^{(k)})$ the above formula is equivalent

to the following, which is a nonconvex proximal operator problem. Next, we will prove (18) has a closed-form solution by exploiting the special structure of it.

3.3. WTNFNSVT Operator

The weighted tensor nuclear and Frobenius norm singular value thresholding (WTNFNSVT) operator is defined as follows to compute the proximal minimizer of the tensor weighted nuclear and Frobenius norm minimization.

Theorem 3 (WTNFNSVT Operator) Let $\mathcal{Y} = \mathcal{U} * \mathcal{S} * \mathcal{V}^{T}$ be the T-SVD of $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. For any $\tau > 0$, $\gamma > 0$, the solution of the problem

$$\min_{\boldsymbol{\mathcal{X}}} \|\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{Y}}\|_{F}^{2} + \tau \cdot \|\boldsymbol{\mathcal{X}}\|_{\text{WTNFN}}$$
(19)

is given by $\mathcal{D}_{W,\gamma}$, which is defined the WTNFNSVT operator as

$$\mathcal{D}_{W,\gamma} = \mathcal{U} * \mathcal{S}_{W,\gamma} * \mathcal{V}^{\mathrm{T}}$$
⁽²⁰⁾

where
$$\boldsymbol{\mathcal{S}}_{\boldsymbol{W},\boldsymbol{\gamma}} = ifft \left(\left(\left(1 + \gamma \tau \boldsymbol{\mathcal{W}} \right)^{-1} \left(\boldsymbol{\overline{\mathcal{S}}} - \frac{\tau}{2} \boldsymbol{\mathcal{W}} \right) \right)_{+}, [], 3 \right), \boldsymbol{\mathcal{W}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$$
 is

F-diagonal whose diagonal entries of the *j*-th frontal slice are equal to the *j*-th column of weight matrix W. The entries of weight matrix W satisfies $0 \le w_{1j} \le w_{2j} \le \cdots \le w_{nj}$, $n = \min[n_1, n_2]$, $j = 1, 2, \cdots, n_3$.

Theorem 2 shows that the WTNFNSVT operator $\mathcal{D}_{W,\gamma}$ gives a close-from of the proximal minimizer of the WTNFN minimization with monotone nonnegative weights, which is a natural extension of the weighted matrix SVT. The feasibility of this extension is guaranteed by the T-SVD framework. Using Algorithm 1, we show how to compute $\mathcal{D}_{W,\gamma}$ efficiently in Algorithm 2.

Algorith	m 2 W]	INFN:	SVT				
Input:	Tensor	${\mathcal Y}$	\in	\mathbb{R}^{n}	$_{1} \times n_{2} \times n_{3}$,	W	=
$[oldsymbol{w}_1,oldsymbol{u}]$	$\boldsymbol{v}_2,\cdots, \boldsymbol{w}_n$	$_{n_3}] \in$	$\mathbb{R}^{n \times n}$	³ .			
Output:	$\mathcal{D}_{oldsymbol{W},\gamma}(\mathcal{Y})$	2)					
1. Com	pute $ar{\mathcal{Y}}=$	fft(${m {\cal Y}}, [], 3$	3).			
2. Com	pute each	front	al slice	e of $ar{\mathcal{U}}$, $ar{\mathcal{S}}$ and $ar{\mathcal{V}}$	from J	₽ by
for	$i = 1, \cdots$	$, \left\lceil \frac{n_3}{2} \right\rceil$	<u>-1</u>]do				
[Ū	$ar{J}^{(i)},ar{oldsymbol{S}}^{(i)},$	$[ar{ar{V}}^{(i)}]$	= SV	$^{\prime}D(ar{m{Y}}$	$^{(i)});$		
Ī	$\dot{\mathbf{D}}^{(i)} =$	$ar{U}^{(0)}$	(i) .	((1 +)	$-2\gamma \boldsymbol{w}_i)^{(i)}$	$^{-1)}(ar{m{S}}^{(i)}$) _
au Diag	$(\boldsymbol{w}_i)))_+$	$\cdotar{V}^{(i)}$;				
end	for						
for	$i = \left\lceil \frac{n_3 + n_3}{2} \right\rceil$	$\frac{1}{2}$],,	n_3 do				
Ī	$\hat{\mathbf{D}}^{(i)} = con$	$njar{D}^{(r)}$	$a_3 + 2 - i$);			
end	for						
3. Com	pute ${\cal D}_W$	$,_{\gamma}(\mathcal{Y})$	by ${\cal D}$	$\mathbf{W}_{,\gamma}(\mathbf{C})$	$(\mathcal{V}) = iff$	$t(\bar{\mathcal{D}}, [], \bar{\mathcal{D}})$	3).

Obviously, when $\gamma = 0$ the weighted tensor nuclear norm minimization (WTNNM) was a special case of this problem at that time. And secondly, when all weights were set the same, tensor nuclear norm minimization (TNNM) was a special case of this problem. Therefore, the solution of this model covers the solutions of traditional TNNM and WTNNM.

3.4. The Setting of Weighting Matrix

In previous sections, we proposed to utilize the WTNFNM model and its solving method. By introducing the weight vector, the WTNFNM model improves the flexibility of the original TNNM model. However, the weight vector itself also brings more parameters in the model. Appropriate setting of the weights plays a crucial role in the success of the proposed WTNFNM model.

Inspired by the weighted unclear norm proposed by Gu *et al.* [23], a similar weighting method can be used in the weighted tensor unclear norm and Frobenius norm, that is

$$w_{ij}^{k+1} = \frac{C}{\sigma_{ij}\left(\boldsymbol{\mathcal{X}}^{(k)}\right) + \varepsilon}$$
(21)

where $\sigma_{ij}(\boldsymbol{\mathcal{X}}^{(k)})$ is the singular value of the tensor $\boldsymbol{\mathcal{X}}$ in the *k*-th iteration, w_{ij}^{k+1} is the corresponding weight of $\sigma_{ij}(\boldsymbol{\mathcal{X}}^{(k)})$. Because (21) is monotonically decreasing with the singular values, the non-descending order of weights with respect to singular values will be kept throughout the reweighting process. According to the Remark 1 in [23], we can get Corollary 1 as follow:

Corollary 1. Let $\mathcal{Y} = \mathcal{U} * \mathcal{S} * \mathcal{V}^{\mathrm{T}}$ be the T-SVD of tensor $\mathcal{Y} \in \mathbb{R}^{n_{1} \times n_{2} \times n_{3}}$, $\sigma_{ij}(\mathcal{Y}) := \overline{\mathcal{S}}(i, i, j)$ denotes singular value of the tensor \mathcal{Y} . If the regularization parameter *C* is positive and the positive value ε is small enough. By using the reweighting formula $w_{ij}^{k+1} = \frac{C}{\sigma_{ij}(\mathcal{X}^{(k)}) + \varepsilon}$, and initial estimate $\mathcal{X}^{(0)} = \mathcal{Y}$.

The reweighted problem

$$\min_{\mathcal{X}} \left\| \mathcal{X} - \mathcal{Y} \right\| + \tau \cdot \left(\left\| \mathcal{X} \right\|_{w,\star} + \gamma \left\| \mathcal{X} \right\|_{w,F}^{2} \right)$$
(22)

has a closed-form solution

$$\hat{\boldsymbol{\mathcal{X}}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}}_{w} * \boldsymbol{\mathcal{V}}^{\mathrm{T}}$$
(23)

where $\boldsymbol{\mathcal{S}}_{w}$ is a *F*-diagonal tensor, $\sigma_{ij}(\boldsymbol{\mathcal{X}}) \coloneqq \boldsymbol{\overline{\mathcal{S}}}_{w}(i,i,j)$. Consider $c_{1} = \tau \left(\sigma_{j}\left(\boldsymbol{\overline{Y}}^{(i)}\right) - \varepsilon - \gamma \boldsymbol{\mathcal{C}}\right)$ and $c_{2} = \tau^{2} \left(\sigma_{j}\left(\boldsymbol{\overline{Y}}^{(i)}\right) - \varepsilon - \gamma \boldsymbol{\mathcal{C}}\right)^{2} - 4 \left(2\boldsymbol{\mathcal{C}} - \tau\varepsilon\sigma_{j}\left(\boldsymbol{\overline{Y}}^{(i)}\right)\right)$. We get $\sigma_{ij}\left(\boldsymbol{\hat{\mathcal{X}}}\right) = \begin{cases} 0 & \text{if } c_{2} > 0 \\ \frac{c_{1} + \sqrt{c_{2}}}{2} & \text{if } c_{2} > 0 \end{cases}$ (24)

The proof of Corollary 1 can be found in supplementary material. Corollary 1 shows that although a reweighting strategy (21) is used, we do not need to itera-

tively perform the thresholding and weight calculation operations. In each iteration of both the WTNFNM models to solve the typical tensor recovery problems, and the tensor completion algorithms, such a reweighting strategy is performed on the WTNFNSVT subproblem (step 2 in Algorithms 2) to adjust weights based on current $\mathcal{X}^{(k)}$. In implementation, we initialize $\mathcal{X}^{(0)}$ as the observation tensor \mathcal{Y} . The above weight setting strategy greatly facilitates the WTNFNM calculations. Note that there remain two parameters *C* and γ in the WTNFNM implementation. In all of our experiments, we set it by experience for certain tasks. Please see the following sections for details.

3.5. Convergence Analysis

In this section, we give the convergence analysis for (18). Suppose that the sequence $\mathcal{X}^{(k)}$ is generated by Algorithm 2. For the simplicity of notation, we denote σ_{ij}^k as the singular value of \mathcal{X} in the *k*-th iteration, and w_{ij}^k is the corresponding weight of σ_{ij}^k .

Theorem 4. In problem (14), assume that $f(\mathcal{X})$ satisfy the assumptions and ∇f is Lipschitz continuous. Consider $F(\mathcal{X}) = f(\mathcal{X}) + \tau \cdot \|\mathcal{X}\|_{WTNFN}$, and the parameter $\mu^{(k)} \geq \frac{L(f)}{1-\tau\alpha}$ for any $k \geq 0$, $0 < \alpha < 1$. Then the sequence $\{\mathcal{X}^{(k)}\}$ and $\{F(\mathcal{X}^{(k)})\}$ generated by WTNFNM satisfies the following properties:

1) $F(\boldsymbol{\mathcal{X}}^{(k)})$ is monotonically decreasing. Indeed,

$$F\left(\boldsymbol{\mathcal{X}}^{(k)}\right) - F\left(\boldsymbol{\mathcal{X}}^{(k+1)}\right) \leq \frac{\boldsymbol{\mu}^{(k)} - L(f)}{1 - 2\tau} \left\| \boldsymbol{\mathcal{X}}^{(k+1)} - \boldsymbol{\mathcal{X}}^{(k)} \right\|_{F}^{2}$$
(25)

2) The sequence $\{\boldsymbol{\mathcal{X}}^{(k)}\}$ is always bounded. Moreover,

$$\lim_{k \to +\infty} \left(\boldsymbol{\mathcal{X}}^{(k+1)} - \boldsymbol{\mathcal{X}}^{(k)} \right) = 0$$

3) Any accumulation point $\hat{\mathcal{X}}$ of the sequence $\{\mathcal{X}^{(k)}\}$ is a critical point of $F(\mathcal{X})$.

Proof. (1) First, with the definition of $\mathcal{X}^{(k+1)}$, *i.e.*,

$$\boldsymbol{\mathcal{X}}^{(k+1)} = \arg\min_{\boldsymbol{\mathcal{X}}} \frac{\boldsymbol{\mu}^{(k)}}{2\tau} \left\| \boldsymbol{\mathcal{X}} - \left(\boldsymbol{\mathcal{X}}^{(k)} - \frac{1}{\mu} \nabla f\left(\boldsymbol{\mathcal{X}}^{(k)} \right) \right) \right\|_{F}^{2} + \left\| \boldsymbol{\mathcal{X}} \right\|_{WTNFN}$$
(26)

We have

$$\begin{split} \boldsymbol{X}^{(k+1)} &= \arg\min_{\boldsymbol{\mathcal{X}}} \frac{1}{n_{3}} \sum_{i=1}^{m} \sum_{j=1}^{n_{3}} \left(w_{ij}^{(k)} \sigma_{ij} + w_{ij}^{(k)} \sigma_{ij}^{2} \right) + \frac{\mu^{(k)}}{2\tau} \left\| \boldsymbol{\mathcal{X}} - \left(\boldsymbol{\mathcal{X}}^{(k)} - \frac{1}{\mu} \nabla f\left(\boldsymbol{\mathcal{X}}^{(k)} \right) \right) \right\|_{F}^{2} \\ &= \arg\min_{\boldsymbol{\mathcal{X}}} \frac{1}{n_{3}} \sum_{i=1}^{m} \sum_{j=1}^{n_{3}} \left(w_{ij}^{(k)} \sigma_{ij} + w_{ij}^{(k)} \sigma_{ij}^{2} \right) + \left\langle \nabla f\left(\boldsymbol{\mathcal{X}}^{(k)} \right), \boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{X}}^{(k)} \right\rangle \\ &+ \frac{\mu^{(k)}}{2\tau} \left\| \boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{X}}^{(k)} \right\|_{F}^{2} \end{split}$$

Then

$$\begin{split} &\frac{1}{n_{3}}\sum_{i=1}^{m}\sum_{j=1}^{n_{3}}\left(w_{ij}^{(k)}\sigma_{ij}^{(k+1)}+w_{ij}^{(k)}\left(\sigma_{ij}^{(k+1)}\right)^{2}\right)\\ &+\left\langle\nabla f\left(\boldsymbol{\mathcal{X}}^{(k)}\right),\boldsymbol{\mathcal{X}}^{(k+1)}-\boldsymbol{\mathcal{X}}^{(k)}\right\rangle+\frac{\mu^{(k)}}{2\tau}\left\|\boldsymbol{\mathcal{X}}^{(k+1)}-\boldsymbol{\mathcal{X}}^{(k)}\right\|_{F}^{2}\\ &\leq\frac{1}{n_{3}}\sum_{i=1}^{m}\sum_{j=1}^{n_{3}}\left(w_{ij}^{(k)}\sigma_{ij}^{(k)}+w_{ij}^{(k)}\left(\sigma_{ij}^{(k)}\right)^{2}\right)\\ &+\left\langle\nabla f\left(\boldsymbol{\mathcal{X}}^{(k)}\right),\boldsymbol{\mathcal{X}}^{(k)}-\boldsymbol{\mathcal{X}}^{(k)}\right\rangle+\frac{\mu^{(k)}}{2\tau}\left\|\boldsymbol{\mathcal{X}}^{(k)}-\boldsymbol{\mathcal{X}}^{(k)}\right\|_{F}^{2} \end{split}$$

which can be rewritten as

.

$$\left\langle \nabla f\left(\boldsymbol{\mathcal{X}}^{(k)}\right), \boldsymbol{\mathcal{X}}^{(k+1)} - \boldsymbol{\mathcal{X}}^{(k)} \right\rangle \\ \leq \frac{1}{n_3} \sum_{i=1}^{m} \sum_{j=1}^{n_3} \left(w_{ij}^{(k)} \left(\sigma_{ij}^{(k)} - \sigma_{ij}^{(k+1)}\right) + \left(\left(\sigma_{ij}^{(k)}\right)^2 - \left(\sigma_{ij}^{(k+1)}\right)^2 \right) \right) - \frac{\mu^{(k)}}{2\tau} \left\| \boldsymbol{\mathcal{X}}^{(k+1)} - \boldsymbol{\mathcal{X}}^{(k)} \right\|_F^2 \right)^2 \right\}$$
(27)

Second, note that *f* is Lipschitz gradient continuous. According to [[24] Lemma 2.11], we have

$$f\left(\boldsymbol{\mathcal{X}}^{(k+1)}\right) - f\left(\boldsymbol{\mathcal{X}}^{(k)}\right) \leq \left\langle \nabla f\left(\boldsymbol{\mathcal{X}}^{(k)}\right), \boldsymbol{\mathcal{X}}^{(k+1)} - \boldsymbol{\mathcal{X}}^{(k)}\right\rangle + \frac{L(f)}{2} \left\|\boldsymbol{\mathcal{X}}^{(k+1)} - \boldsymbol{\mathcal{X}}^{(k)}\right\|_{F}^{2}$$
(28)

Finally, we combine (27) and (28), it obtains that

$$F\left(\boldsymbol{\mathcal{X}}^{(k+1)}\right) - F\left(\boldsymbol{\mathcal{X}}^{(k)}\right)$$

$$= \frac{1}{n_3} \sum_{i=1}^{m} \sum_{j=1}^{n_3} \left(w_{ij}^{(k+1)} \left(\sigma_{ij}^{(k+1)} - \left(\sigma_{ij}^{(k+1)} \right)^2 \right) + w_{ij}^{(k)} \left(\sigma_{ij}^{(k)} - \left(\sigma_{ij}^{(k)} \right)^2 \right) \right)$$

$$+ f\left(\boldsymbol{\mathcal{X}}^{(k+1)}\right) - f\left(\boldsymbol{\mathcal{X}}^{(k)}\right)$$

$$\leq \frac{1}{n_3} \sum_{i=1}^{m} \sum_{j=1}^{n_3} \left(w_{ij}^{(k)} \left(\sigma_{ij}^{(k)} - \sigma_{ij}^{(k+1)} \right) + \left(\sigma_{ij}^{(k)} \right)^2 - \left(\sigma_{ij}^{(k+1)} \right)^2 \right)$$

$$+ \left\langle \nabla f\left(\boldsymbol{\mathcal{X}}^{(k)}\right), \boldsymbol{\mathcal{X}}^{(k+1)} - \boldsymbol{\mathcal{X}}^{(k)} \right\rangle + \frac{L\left(f\right)}{2\tau} \left\| \boldsymbol{\mathcal{X}}^{(k+1)} - \boldsymbol{\mathcal{X}}^{(k)} \right\|_F^2$$

which leads

$$F\left(\boldsymbol{\mathcal{X}}^{(k+1)}\right) + \frac{\mu^{(k)} - L(f)}{2\tau} \left\| \boldsymbol{\mathcal{X}}^{(k+1)} - \boldsymbol{\mathcal{X}}^{(k)} \right\|_{F}^{2} \leq F\left(\boldsymbol{\mathcal{X}}^{(k)}\right)$$
(29)

Since for any integer k, if $\mu^{(k)} \ge \frac{L(f)}{1-\alpha\tau}$ we have $\frac{\mu^{(k)}-L(f)}{2\tau} \ge \frac{\alpha}{2}\mu^{(k)}$ always held. Then the monotone search criterion (25) is always satisfied, which makes the $F(\boldsymbol{\mathcal{X}}^{(k)})$ is monotonically decreasing.

2) It's easy to obtain that $F(\mathcal{X}^{(k)})$ is bounded. Then, we know that $\{\mathcal{X}^{(k)}\}$ is bounded as well from the assumption. Besides, summing up the inequality (27) for all $k \ge 0$, it leads that

$$F\left(\boldsymbol{\mathcal{X}}^{(0)}\right) \geq \sum_{k=0}^{+\infty} \frac{\mu^{(k)} - L(f)}{2\tau} \left\| \boldsymbol{\mathcal{X}}^{(k+1)} - \boldsymbol{\mathcal{X}}^{(k)} \right\|_{F}^{2}$$
$$\geq \frac{2 - 2\alpha\tau}{\alpha L(f)} \sum_{k=0}^{+\infty} \left\| \boldsymbol{\mathcal{X}}^{(k+1)} - \boldsymbol{\mathcal{X}}^{(k)} \right\|_{F}^{2}$$

Hence, we get

$$\sum_{k=0}^{+\infty} \left\| \mathcal{X}^{(k+1)} - \mathcal{X}^{(k)} \right\|_{F}^{2} \le \frac{\mu^{(k)} - L(f)}{2\tau} < +\infty$$
(30)

From (30), it yields that

$$\lim_{k \to +\infty} \left(\boldsymbol{\mathcal{X}}^{(k+1)} - \boldsymbol{\mathcal{X}}^{(k)} \right) = 0$$
(31)

3) Because of the boundedness of the sequence $\{\mathcal{X}^{(k)}\}$, there exists a subsequence $\{\mathcal{X}^{(k_t)}\}$ such that it converges to an accumulation point $\hat{\mathcal{X}}$, *i.e.*, $\lim_{r \to +\infty} \mathcal{X}^{(k_t)} = \hat{\mathcal{X}}$. Based on that convergence result (31), we have

 $\lim_{t \to +\infty} \mathcal{X}^{(k_i)} = \hat{\mathcal{X}} \text{, which implies that } \lim_{t \to +\infty} \sigma_{ij} \left(\mathcal{X}^{(k_i)} \right) = \sigma_{ij} \left(\hat{\mathcal{X}} \right) \text{ for all } i \text{ and } j. \text{ Note that the subsequence } \left\{ \mathcal{X}^{(k_i)} \right\} \text{ is bounded obviously, according to the } \begin{bmatrix} [25] \text{ Proposition 2.1.5} \end{bmatrix}, \text{ there exists } \hat{w}_{ij} \in w_{ij} \left(\sigma_{ij} \left(\hat{\mathcal{X}} \right) + \sigma_{ij}^2 \left(\hat{\mathcal{X}} \right) \right) \text{ such that } \lim_{t \to +\infty} w_{ij}^{(k_i)} = \hat{w}_{ij} \text{ for all } i \text{ and } j. \text{ Denote}$

$$T\left(\boldsymbol{\mathcal{X}},\boldsymbol{W}\right) = \frac{1}{n_3} \sum_{i=1}^{m} \sum_{j=1}^{n_3} w_{ij} \left(\sigma_{ij}\left(\boldsymbol{\mathcal{X}}\right) + \sigma_{ij}^2\left(\boldsymbol{\mathcal{X}}\right)\right)$$
(32)

Since $\mathcal{X}^{(k_t)}$ is optimal to problem (30), there exists $\mathcal{Z}^{(k_t)} \in T\left(\mathcal{X}^{(k_t)}, \mathbf{W}^{(k_t)}\right)$ such that

$$\boldsymbol{\mathcal{Z}}^{(k_i)} \in T\left(\boldsymbol{\mathcal{X}}^{(k_i)}, \boldsymbol{W}^{(k_i)}\right)$$
(33)

such that

$$\boldsymbol{\mathcal{Z}}^{(k_t+1)} + \nabla f\left(\boldsymbol{\mathcal{X}}^{(k_t)}\right) + \boldsymbol{\mu}^{(k_t)}\left(\boldsymbol{\mathcal{X}}^{(k_t+1)} - \boldsymbol{\mathcal{X}}^{(k_t)}\right) = \boldsymbol{\mathcal{O}}$$
(34)

Let $t \to +\infty$ in the above equation, then we obtain that there exists $\hat{\mathcal{Z}} \in \partial T(\hat{\mathcal{X}}, \hat{W})$

$$\mathcal{O} = \hat{\mathcal{Z}} + \nabla f\left(\hat{\mathcal{X}}\right) \in \partial F\left(\hat{\mathcal{X}}\right)$$
(35)

Thus $\hat{\mathcal{X}}$ is a critical point of problem (14). This completes the proof.

4. Experiments

In this section, we conducted a series of experiments on both synthetic and real-world data to demonstrate the superiority of the proposed model. we empirically demonstrated that the WTNFN regularization algorithm leads to more stable solutions than WTNN (with $\gamma = 0$). Moreover, it could achieve better performance when compared to the previous algorithm.

We solved the low rank tubal tensor completion problem by the following program using the Discrete Fourier Transform based tensor nuclear norm based on the proposed weighted tensor nuclear and Frobenius norm

$$\min_{\boldsymbol{\mathcal{X}}} \tau \left\| \boldsymbol{\mathcal{X}} \right\|_{\text{WTNFN}} + \frac{1}{2} \left\| \mathcal{P}_{\Omega} \left(\boldsymbol{\mathcal{X}} \right) - \mathcal{P}_{\Omega} \left(\boldsymbol{\mathcal{M}} \right) \right\|_{F}^{2},$$
(36)

where $\mathcal{X}, \mathcal{M}, \Omega$ were respectively the observed data and the underlying recover result, a binary support indicator tensor. Zeros in Ω indicate the missing entries in the observed tensor, and $\mathcal{P}_{\Omega}: \mathbb{R}^{n_1 \times n_1 \times n_3} \to \mathbb{R}^{n_1 \times n_1 \times n_3}$ was a linear opera-

tor that keeps the entries in unchanged and sets those outside zeros. It was easy to know that the gradient of the squared loss function was a Lipschitz continuous function with a Lipschitz constant L(f)=1. The quality of the restored image was evaluated in terms of the peak-signal-to-noise ratio (PSNR) defined by

$$PSNR = 10 \log_{10} \left(\frac{\|\mathcal{M}\|_{\infty}^2}{\frac{1}{n_1 n_2 n_3} \|\hat{\mathcal{X}} - \mathcal{M}\|_F^2} \right)$$
(37)

4.1. Stability of the WTNFN Regularization Model

In this section, we show that WTNFN model leads to more stable solutions and provides improved experimental results than WTNN for color images. In our experiments, the testing images shown in **Figure 1** were the "Im1" and "Im5" with size $300 \times 300 \times 3$. For each image, pixels were sampled randomly with sampling rates of 30%, 50%, and 70%. These two images were used to illustrate the effectiveness of the Frobenius norm in the WTNFN.

We first applied the WTNFN with different γ values to "Im1" and "Im5". A suitable value of should γ result to a stabilized solution and meanwhile result to a solution of high PSNR value. We denote WTNN the WTNN regularization model with $\gamma = 0$, WTNFN1 with γ is about 10⁻⁶, WTNFN2 with a lager γ about 10⁻⁵. Figure 1 shows the PSNR values of sampling rates of 30%, 50%, and 70%, respectively, for recovering images "Im1" and "Im5" 10 times. Comparing the WTNFN with different values of γ , we can find that the PSNR values of WTNFN1 and WTNFN2 are more stable than that of WTNN, which indicates that the Frobenius norm in the WTNFN penalty can induce stable solutions. In addition, we found that WTNFN11 and WTNFN2 behave very similarly for the values of PSNR for the same image at the same sample rate. So we should choose moderate value of γ to balance the stability and the quality of the restored image. The results in Figure 1 show that setting $\gamma = 10^{-6}$ to make sure that always achieves good results.

4.2. Synthetic Experiments

We conducted an experiment to demonstrate the practical applicability of the WTNFN heuristic for recovering low-rank tensors. We generated a tensor by $\mathcal{X} = \mathcal{M} * \mathcal{N}$, where the entries of $\mathcal{M} \in \mathbb{R}^{n \times r \times n}$ and $\mathcal{N} \in \mathbb{R}^{r \times n \times n}$ were independently sampled from standard Gaussian distribution. The tubal rank of tensor \mathcal{X} was defined as *r*. We solve (36) and obtain the solution $\hat{\mathcal{X}}$. Then we reported to the tubal rank r_t of $\hat{\mathcal{X}}$ and Relative Error was defined as

Relative Error :=
$$\frac{\left\|\hat{\mathcal{X}} - \mathcal{X}\right\|_{F}}{\left\|\mathcal{X}\right\|_{F}}$$
(38)

We simply considered tensors of size $n \times n \times n$, with n = [50, 100, 200]. To



Figure 1. Performance of the WTNFN on images "Im1" and "Im5" for different randomly sampled sets. (a) Results for sampling rate 30% "Im1". (b) Results for sampling rate 30% "Im5". (c) Results for sampling rate 50% "Im1". (d) Results for sampling rate 50% "Im5". (e) Results for sampling rate 70% "Im1". (f) Results for sampling rate 70% "Im5".

create the known entries, we sampled $m = pn^3$ items uniformly from \mathcal{X} , where p was the sampling rate. A reference quantity that is defined as $d_r = r(2n-r)n$ in the literature [26]. We set the parameter $\tau = 0.01$, $\gamma = 10^{-6}$ for model (36)

in experiment. Table 1 respectively gives the recovery results for the linear transform DFT. It can be seen that the WTNFN program (36) gives the correct rank estimation of \mathcal{X} in all cases and the relative errors are small. Therefore, these results do a good job of validating the correct recovery guarantees of our algorithm.

Next, we studied the recovery phenomenon with different tubal ranks of \mathcal{X} and different sampling rates of p. We generated a 50 × 50 × 50 tensor \mathcal{X} similarly to the previous experiment. The range of variations for tubal ranks r and sampling rates p were $r = [1, \dots, 20]$ and $p = [0.01, \dots, 0.99]$. The tubal rank was gradually increased in steps of 1. The sampling rate was gradually increased in steps of 0.01. We conducted an experiment for each (p, r) pair. The algorithms stopped when the Relative Error was less than 10⁻⁵. Figure 2 reports the percentage of success for each pair of WTNFN and TNN for the linear transform DFT, in which the yellow region reflects the full successful recovery, and the blue region reflects the full failed recovery. The bigger the yellow area, the stronger ability to recover. It can be seen that the WTNFN method has a much larger yellow area than the TNN method. The recovery phenomena can be explained by the fact that WTNFN has a stronger ability to recovery than TNN. In a word, the WTNFN minimization model is more robust and efficient than the classic TNN minimization model.

4.3. Real World Experiments

Due to coding and transmission difficulties, photos may be partially damaged or obscured by text or branding. Matrix completion methods, of course, can be used to recover the missing information in these images. We considered the

п	r	$\frac{m}{d_r}$	р	r _t	$\frac{\left\ \hat{\boldsymbol{\mathcal{X}}}-\boldsymbol{\mathcal{X}}\right\ _{F}}{\left\ \boldsymbol{\mathcal{X}}\right\ _{F}}$
50	3	4	0.47	3	6.64e-07
50	5	3	0.57	5	8.50e-07
50	8	2	0.59	8	5.54e-07
50	10	2	0.72	10	2.69e-07
100	5	4	0.39	5	1.55e-06
100	10	3	0.57	10	8.28e-07
100	15	2	0.56	15	1.01e-06
100	20	2	0.72	20	5.99e-07
200	5	4	0.20	5	4.2e-06
200	10	3	0.29	10	3.26e-06
200	20	2	0.38	20	2.94e-06
200	30	2	0.56	30	1.92e-06

 Table 1. Exact tensor completion on random data.



Figure 2. Comparison of the recovery capacity varying tubal rank *r* and sampling rate *p* of the WTNFN (a) and TNN (b). The *x*-axis represents the sampling rate p, ranging from 0.01 to 0.99, the *y*-axis represents the tubal rank r, ranging from 1 to 20.

WFNFN to color image recovery with partially missing pixels, as color images with three channels (Red, Green, Blue) can be modeled as third-order low-tubal rank tensors. The Peak Signal-to-Noise Ratio (PSNR) is a widely used measure for assessing image quality. To compare the performance of different approaches, we utilize the PSNR value of the recovered image. Larger values of PSNR indicate better image restoration performance.

We considered four methods for image recovery and compare their performance: 1) TNN-DFT [26]: tensor nuclear norm based discrete Fourier transform tensor completion model; 2) TNNR [27]: tensor truncated nuclear norm regularization; 3) L_p [22]: tensor completion by nonconvex L_p function regularization; 4) TCTF [28]: tensor completion by tensor factorization.

a) Random Mask: We first conducted experiments to deal with a relatively easy tensor completion problem. We randomly selected 5 images with a size of $300 \times 300 \times 3$ from the LJU dataset, and randomly sampled 50% of pixels for each image.

Five recovered image examples with sampling rate of 50% by different methods are shown in "Fig.". In the magnified views of the comparable sections in color boxes, it reveals that our recovered images have a clearer shape, fuller color, and brightness than the other four state-of-the-art tensor completion methods. Furthermore, we display the PSNR values achieved by various approaches on five images in **Figure 3** when the sampling rate is 50%. It can demonstrate that the WTNFN-based algorithms can achieve much higher PSNR values compared with the other four state-of-the-art tensor completion methods quantitatively for the majority of the images.

b) Text Mask: Text removal is difficult because the pixels covered by the text are not spread evenly throughout the image, and the text may obscure essential textural information. We start by looking at the text's location and treating the



Figure 3. Five images of natural scenes.

Table 2. PSNR values of the various methods on text mask for image recovery.

ID	Images	TNN-DFT	TNNR	L_p	TCTF	WTNFN
1	Im1	37.87	30.17	31.15	28.88	38.99
2	Im2	36.24	28.56	24.00	21.14	37.68
3	Im3	39.91	25.41	30.43	24.96	40.73
4	Im4	34.07	29.32	25.37	20.19	35.97
5	Im5	38.34	25.17	26.81	23.28	39.82



Figure 4. Comparison of the PSNR values obtained by TNNDTF, TNNR, L_p , TCTF and WTNFN on the tested images in **Figure 5**

items that correspond to it as missing values. As a result, the text removal problem can be thought of as a tensor completion problem in general. In this experiment, we apply our method to estimate the images from text mask.

As previously noted, we compared our results to the TNN-DFT, TNNR, L_p , and TCTF methods. PSNR measurements are used to assess performance. The recovered images from the five tensor completion algorithms are shown in **Figure 6**. On all five image sequences, We rounded the experimental results to two decimal places. **Table 2** provides the PSNR values of the compared approaches. We



Figure 5. Performance comparison for image recovery on some sample images. (a) original image; (b) observed image; (c) recovered images by TNN-DFT, (d) recovered images by TNNR, (e) recovered images by L_p , (f) recovered images by TCTF, (g) recovered images by WTNFN.

can see that the WTNFN has absolute superiority, as evidenced by its far higher PSNR values.

5. Conclusions

In this work, we have presented a novel method called the weighted nuclear and Frobenius norm regularization for low-tubal rank tensor recovery. We proposed the solving approach for the weighted nuclear and Frobenius norm proximal (WTNFNSVT) operator. The WTNFN model was subsequently extended to tensor completion tasks, and efficient algorithms were developed to solve these using the derived WTNFNSVT operator. Influenced by the weighted nuclear for automatic weight setting. We also developed a rational scheme for automatic weight setting, which yields closed from solutions of the WTNFN operator. This approach can shorten the calculation time.

On both synthetic and actual visual datasets, we ran many experiments. In comparison to state-of-the-art tensor completion approaches based on the tenor nuclear norm or tensor factorization, the experimental findings indicate the



Figure 6. Comparison of tensor completion algorithms for text removal problem. (a) original image; (b) observed image; (c) recovered images by TNN-DFT (d) recovered images by TNNR, (e) recovered images by L_p , (f) recovered images by TCTF, (g) recovered images by WTNFN.

advantages of the WTNFN-based algorithms.

Acknowledgements

This work is supported by Innovative talents support plan of colleges and universities in Liaoning Province (2021).

Conflicts of Interest

The authors declare no conflicts of interest.

References

- Tucker, L.R. (1963) Implications of Factor Analysis of Three-Way Matrices for Measurement of Change. *Problems in Measuring Change*, 15, 122-137. <u>https://doi.org/10.1108/09534810210423008</u>
- [2] Vasilescu, M.A.O. and Terzopoulos, D. (2002) Multilinear Analysis of Image Ensembles: TensorFaces. In: Heyden, A., et al., Eds., European Conference on Computer Vision, Springer, Berlin, 447-460. <u>https://doi.org/10.1007/3-540-47969-4_30</u>

- [3] Sidiropoulos, N.D., De Lathauwer, L., Fu, X., *et al.* (2017) Tensor Decomposition for Signal Processing and Machine Learning. *IEEE Transactions on Signal Processing*, 65, 3551-3582. <u>https://doi.org/10.1109/TSP.2017.2690524</u>
- [4] Lai, Z., Wong, W.K., Jin, Z., et al. (2012) Sparse Approximation to the Eigensubspace for Discrimination. *IEEE Transactions on Neural Networks and Learning Sys*tems, 23, 1948-1960. <u>https://doi.org/10.1109/TNNLS.2012.2217154</u>
- [5] De Lathauwer, L. and De Moor, B. (1998) From Matrix to Tensor: Multilinear Algebra and Signal Processing. Institute of Mathematics and Its Applications Conference Series 67, Oxford University Press, Oxford, 1-16.
- [6] Cichocki, A., Mandic, D., De Lathauwer, L., et al. (2015) Tensor Decompositions for Signal Processing Applications: From Two-Way to Multiway Component Analysis. *IEEE Signal Processing Magazine*, **32**, 145-163. <u>https://doi.org/10.1109/MSP.2013.2297439</u>
- [7] Song, Q., Huang, X., Ge, H., et al. (2017) Multi-Aspect Streaming Tensor Completion. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, 13-17 August 2017, 435-443. https://doi.org/10.1145/3097983.3098007
- [8] Kolda, T.G. and Bader, B.W. (2009) Tensor Decompositions and Applications. SIAM Review, 51, 455-500. https://doi.org/10.1137/07070111X
- [9] Fazel, M. (2002) Matrix Rank Minimization with Applications. PhD Thesis, Stanford University, Redwood City.
- Kiers, H.A. (2000) Towards a Standardized Notation and Terminology in Multiway Analysis. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 14, 105-122. https://doi.org/10.1002/1099-128X(200005/06)14:3<105::AID-CEM582>3.0.CO;2-I
- [11] Tucker, L.R. (1966) Some Mathematical Notes on Three-Mode Factor Analysis. *Psychometrika*, **31**, 279-311. https://doi.org/10.1007/BF02289464
- Kapteyn, A., Neudecker, H. and Wansbeek, T. (1986) An Approach Ton-Mode Components Analysis. *Psychometrika*, 51, 269-275. https://doi.org/10.1007/BF02293984
- [13] Hillar, C.J. and Lim, L.-H. (2013) Most Tensor Problems Are NP-Hard. Journal of the ACM, 60, Article No. 45. https://doi.org/10.1145/2512329
- [14] Liu, J., Musialski, P., Wonka, P. and Ye, J. (2013) Tensor Completion for Estimating Missing Values in Visual Data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **35**, 208-220. <u>https://doi.org/10.1109/TPAMI.2012.39</u>
- [15] Romera-Paredes, B. and Pontil, M. (2013) A New Convex Relaxation for Tensor Completion. 27 th Annual Conference on Neural Information Processing Systems, Lake Tahoe, 5-10 December 2013, 2967-2975.
- [16] Kilmer, M.E. and Martin, C.D. (2011) Factorization Strategies for Third-Order Tensors. *Linear Algebra and Its Applications*, 435, 641-658. https://doi.org/10.1016/j.laa.2010.09.020
- [17] Lu, C., Feng, J., Liu, W., Lin, Z. and Yan, S. (2020) Tensor Robust Principal Component Analysis with a New Tensor Nuclear Norm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **42**, 925-938.
- [18] Peng, X., Lu, C., Yi, Z. and Tang, H. (2018) Connections between Nuclear-Norm and Frobenius-Norm-Based Representations. *IEEE Transactions on Neural Networks* and Learning Systems, 29, 218-224. <u>https://doi.org/10.1109/TNNLS.2016.2608834</u>
- [19] Zhang, H., Yi, Z. and Peng, X. (2014) fLRR: Fast Low-Rank Representation Using

Frobenius-Norm. *Electronics Letters*, **50**, 936-938. https://doi.org/10.1049/el.2014.1396

- [20] Kilmer, M.E., Braman, K., Hao, N. and Hoover, R.C. (2013) Third-Order Tensors as Operators on Matrices: A Theoretical and Computational Framework with Applications in Imaging. *SIAM Journal on Matrix Analysis and Applications*, 34, 148-172. https://doi.org/10.1137/110837711
- [21] Zhang, Z., Ely, G., Aeron, S., Hao, N. and Kilmer, M. (2014) Novel Methods for Multilinear Data Completion and De-Noising Based on Tensor-SVD. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 3842-3849.
- [22] Wang, H., Zhang, F., Wang, J., et al. (2022) Generalized Nonconvex Approach for Low-Tubal-Rank Tensor Recovery. *IEEE Transactions on Neural Networks and Learning Systems*, **33**, 3305-3319. <u>https://doi.org/10.1109/TNNLS.2021.3051650</u>
- [23] Gu, S., Xie, Q., Meng, D., et al. (2017) Weighted Nuclear Norm Minimization and Its Applications to Low Level Vision. International Journal of Computer Vision, 121, 183-208. https://doi.org/10.1007/s11263-016-0930-5
- [24] Clarke, F.H. (1990) Optimization and Nonsmooth Analysis. Society for Industrial and Applied Mathematics, Philadelphia. <u>https://doi.org/10.1137/1.9781611971309</u>
- [25] Beck, A. and Teboulle, M. (2009) A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *SIAM Journal on Imaging Sciences*, 2, 183-202. <u>https://doi.org/10.1137/080716542</u>
- [26] Lu, C., Peng, X. and Wei, Y. (2019) Low-Rank Tensor Completion with a New Tensor Nuclear Norm Induced by Invertible Linear Transforms. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, 16-20 June 2019, 5996-6004. <u>https://doi.org/10.1109/CVPR.2019.00615</u>
- [27] Xue, S., Qiu, W., Liu, F., et al. (2018) Low-Rank Tensor Completion by Truncated Nuclear Norm Regularization. 2018 24th International Conference on Pattern Recognition (ICPR) IEEE, Beijing, 20-24 August 2018, 2600-2605. https://doi.org/10.1109/ICPR.2018.8546008
- [28] Zhou, P., Lu, C., Lin, Z., et al. (2017) Tensor Factorization for Low-Rank Tensor Completion. IEEE Transactions on Image Processing, 27, 1152-1163. https://doi.org/10.1109/TIP.2017.2762595

Appendix A. Proof of Theorem 2

Lemma 1. Given a third-order tensor $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ if $\mathcal{U} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is orthogonal, *i.e.*, $\mathcal{U} * \mathcal{U}^T = \mathcal{U}^T * \mathcal{U} = \mathcal{I}$, that is

$$\left\|\boldsymbol{\mathcal{U}}\ast\boldsymbol{\mathcal{A}}\right\|_{F}=\left\|\boldsymbol{\mathcal{A}}\ast\boldsymbol{\mathcal{U}}\right\|_{F}=\left\|\boldsymbol{\mathcal{X}}\right\|_{F}$$
(39)

The theorem 2 can be proved by lemma 1. *Proof* By lemma 1 and $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^{T}$ can get

$$\left\|\boldsymbol{\mathcal{A}}\right\|_{F} = \left\|\boldsymbol{\mathcal{S}}\right\|_{F} = \frac{1}{\sqrt{n_{3}}} \sum_{i=1}^{n} \sum_{j=1}^{n_{3}} \sqrt{\sigma_{ij}^{2}\left(\boldsymbol{\mathcal{A}}\right)}$$
(40)

End of proof.

Appendix B. Proof of Theorem 3

Proof of Theorem 3 requires the following Lemma 2.

Lemma 2 (*Proximal Operator for WNFN*) For a matrix $X \in \mathbb{C}^{m \times n} (m > n)$, denote by $Y = U\Sigma^{T}V$ the singular value decomposition of matrix Y, and $0 \le w_1 \le w_2 \le \cdots \le w_n$, the following minimization problem

$$\min_{\boldsymbol{X}} \left\| \boldsymbol{X} - \boldsymbol{Y} \right\|_{F}^{2} + \tau \cdot \left(\left\| \boldsymbol{X} \right\|_{w,\star} + \gamma \left\| \boldsymbol{X} \right\|_{w,F}^{2} \right)$$
(41)

The solution is $prox_{\|\cdot\|_{w,*} + \|\cdot\|_{w,F}^2}$,

$$prox_{\|\cdot\|_{w,\star}+\|\cdot\|_{w,F}^{2}} = US_{w}(\Sigma)V^{\mathrm{T}}$$

$$(42)$$

where $S_w(\Sigma) = \max\left\{ \left(1 + \gamma \tau w_i\right)^{-1} \left(\Sigma_{ii} - \frac{2}{\tau} w_i\right), 0 \right\}$

B1. Proof of Lemma 1

Proof In order to prove Lemma 2, we first show the following theorem for the proximal operator of weighted nuclear and Frobenius norm for real-valued matrices. Then, we extend the result to prove Lemma 2 for complex-valued matrices.

For any $X, Y \in \mathbb{R}^{m \times n} (m > n)$, denote by $\overline{U} \overline{U} \overline{V}^{T}$ and $U \Sigma V$ the singular value decomposition of matrix X and Y, where

$$\boldsymbol{\Sigma} = \begin{pmatrix} \operatorname{diag}\left(\sigma_{1}, \sigma_{2}, \cdots, \sigma_{n}\right) \\ 0 \end{pmatrix} \in \mathbb{R}^{m \times n} \text{, and } \boldsymbol{D} = \begin{pmatrix} \operatorname{diag}\left(d_{1}, d_{2}, \cdots, d_{n}\right) \\ 0 \end{pmatrix} \text{ are the di-}$$

agonal singular value matrices. Based on the property of Frobenius norm, the following derivations hold:

$$\|\boldsymbol{X} - \boldsymbol{Y}\|_{F}^{2} + \tau \cdot \left(\|\boldsymbol{X}\|_{w,*} + \gamma \|\boldsymbol{X}\|_{w,F}^{2}\right)$$

= $\operatorname{Tr}\left(\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y}\right) - 2\operatorname{Tr}\left(\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{X}\right) + \operatorname{Tr}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right) + \tau \cdot \left(\sum_{i}^{n} w_{i}d_{i} + \gamma \sum_{i}^{n} w_{i}d_{i}^{2}\right)$
= $\sum_{i}^{n} \sigma_{i}^{2} - 2\operatorname{Tr}\left(\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{X}\right) + \sum_{i}^{n} d_{i}^{2} + \tau \cdot \left(\sum_{i}^{n} w_{i}d_{i} + \gamma \sum_{i}^{n} w_{i}d_{i}^{2}\right)$

Based on the von Neumann trace inequality, we know that $\operatorname{Tr}(X^{T}Y)$ achieves its upper bound $\sum_{i}^{n} \sigma_{i} d_{i}$ if $U = \overline{U}$, $V = \overline{V}$. Then we have

$$\begin{split} \min_{\boldsymbol{X}} \|\boldsymbol{X} - \boldsymbol{Y}\|_{F}^{2} + \tau \cdot \left(\|\boldsymbol{X}\|_{w,*} + \gamma \|\boldsymbol{X}\|_{w,F}^{2}\right) \\ \Leftrightarrow \min_{\boldsymbol{D}} \sum_{i}^{n} \sigma_{i}^{2} - 2\sum_{i}^{n} \sigma_{i} d_{i} + \sum_{i}^{n} d_{i}^{2} + \tau \cdot \left(\sum_{i}^{n} w_{i} d_{i} + \gamma \sum_{i}^{n} w_{i} d_{i}^{2}\right) \\ \text{s.t. } d_{1} \geq d_{2} \geq \cdots \geq d_{n} \geq 0 \\ \Leftrightarrow \min_{\boldsymbol{D}} \sum_{i}^{n} \left(\sigma_{i} - d_{i}\right)^{2} + \tau \cdot \left(w_{i} d_{i} + \gamma w_{i} d_{i}^{2}\right) \\ \text{s.t. } d_{1} \geq d_{2} \geq \cdots \geq d_{n} \geq 0 \end{split}$$

From the above derivation, we can see that the optimal solution of the WNFNP problem is $X^* = UDV^T$, where D is the optimum of the constrained quadratic optimization problem.

The following proof holds the Lemma 2 on the complex number field.

Let $Y = Y_1 + iY_2 \in \mathbb{C}^{n \times m}$ be an arbitrary complex-valued matrix and its SVD be $Y = (U + iP)\Sigma(V + iQ)^T$, where $Y_1 \in \mathbb{R}^{n \times m}$, $Y_2 \in \mathbb{R}^{n \times m}$, $U \in \mathbb{R}^{n \times n}$,

 $\boldsymbol{P} \in \mathbb{R}^{n \times m}$, $\Sigma \in \mathbb{R}^{n_0 \times n_0}$, $\boldsymbol{V} \in \mathbb{R}^{m \times m}$, $\boldsymbol{Q} \in \mathbb{R}^{m \times m}$, and $n_0 = \{n, m\}$. Then, it is easy to verify that the SVD of the matrix

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{Y}_1 & \boldsymbol{Y}_2 \\ -\boldsymbol{Y}_2 & \boldsymbol{Y}_1 \end{bmatrix} \in \mathbb{R}^{2m \times 2m}$$
(43)

can be expressed as

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{U} & \boldsymbol{P} \\ -\boldsymbol{P} & \boldsymbol{U} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma} & \\ \boldsymbol{\Sigma} \end{bmatrix} \begin{bmatrix} \boldsymbol{V} & -\boldsymbol{Q} \\ \boldsymbol{Q} & \boldsymbol{V} \end{bmatrix}^{\mathrm{T}}$$
(44)

Let
$$\boldsymbol{X} = \boldsymbol{L}_1 + i\boldsymbol{L}_2 \in \mathbb{C}^{n \times m}$$
, $\boldsymbol{B} = \begin{bmatrix} \boldsymbol{L}_1 & \boldsymbol{L}_2 \\ -\boldsymbol{L}_2 & \boldsymbol{L}_1 \end{bmatrix} \in \mathbb{R}^{2n \times 2m}$, $\tilde{\boldsymbol{w}} = \begin{bmatrix} \boldsymbol{w}^{\mathrm{T}}, \boldsymbol{w}^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{2n_0}_+$,

and the functions

$$h_{1}(\boldsymbol{X}) = \|\boldsymbol{X} - \boldsymbol{Y}\|_{F}^{2} + \tau \cdot \left(w \|\boldsymbol{X}\|_{\star} + \gamma w \|\boldsymbol{X}\|_{F}^{2}\right)$$
(45)

$$h_{2}\left(\boldsymbol{B}\right) = \frac{1}{2} \left\|\boldsymbol{A} - \boldsymbol{B}\right\|_{F}^{2} + \tau \cdot \left(\frac{\tilde{w}}{2} \left\|\boldsymbol{A}\right\|_{\star} + \gamma \frac{\tilde{w}}{2} \left\|\boldsymbol{A}\right\|_{F}^{2}\right)$$
(46)

From the decomposition formula of the above matrix, $h_1(X) = h_2(B)$, According to Lemma 1, one minimizer to the following minimization problem

$$\min_{\boldsymbol{B}} \frac{1}{2} \|\boldsymbol{A} - \boldsymbol{B}\|_{F}^{2} + \tau \cdot \left(\frac{\tilde{w}}{2} \|\boldsymbol{A}\|_{\star} + \gamma \frac{\tilde{w}}{2} \|\boldsymbol{A}\|_{F}^{2}\right)$$
(47)

The optimal solution is $\boldsymbol{B}^* = \boldsymbol{U}\boldsymbol{S}_{\bar{w},\gamma}\boldsymbol{V}^{\mathrm{T}}$. Since $h_1(\boldsymbol{X}) = h_2(\boldsymbol{B})$, the function $h_2(\boldsymbol{B}^*)$ the same minimum value as $h_2(\boldsymbol{X}^*)$ with $\boldsymbol{X}^* = \boldsymbol{U}\boldsymbol{S}_{w,\gamma}\boldsymbol{V}^{\mathrm{T}}$. End of proof.

B2. Proof of Theorem 3

Proof Theorem 3 describes one minimizer to the following minimization problem

$$\min_{\boldsymbol{\mathcal{X}}} \left\| \boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{Y}} \right\|_{F}^{2} + \tau \cdot \left\| \boldsymbol{\mathcal{X}} \right\|_{\text{WTNFN}} \Leftrightarrow \min_{\overline{\boldsymbol{X}}^{(i)}} \frac{1}{n_{3}} \sum_{i=1}^{n_{3}} \left(\left\| \overline{\boldsymbol{X}}^{(i)} - \overline{\boldsymbol{Y}}^{(i)} \right\|_{F}^{2} \right) + \tau \cdot \left\| \overline{\boldsymbol{X}}^{(i)} \right\|_{\text{WTNFN}}$$
(48)

From Lemma 2, the optimal solution of the above formula can be expressed as

$$\overline{\boldsymbol{X}}^{(i)\star} = \overline{\boldsymbol{U}}^{(i)} \overline{\boldsymbol{S}}^{(i)}_{\boldsymbol{w},\boldsymbol{\gamma}} \left(\overline{\boldsymbol{S}}^{(i)} \right) \overline{\boldsymbol{V}}^{(i)\mathrm{T}}$$
(49)

where $\overline{S}_{w,\gamma}^{(i)}(\overline{S}^{(i)}) = \max\left\{ \left(1 + \gamma w_{i,j}\right)^{-1} (\overline{S}^{(i)} - w_{i,j}), 0 \right\}$, denotes by $\overline{Y}^{(i)} = \overline{U}^{(i)} \overline{S}^{(i)} \overline{V}^{(i)T}$ the singular value decomposition of matrix $\overline{Y}^{(i)}$. So denotes \overline{U} , $\overline{S}_{w,\gamma}$, \overline{V} as three tensors. Their frontal slices are $\overline{U}^{(i)}$, $\overline{S}_{w,\gamma}^{(i)}$, $\overline{V}^{(i)}$. Performing the inverse fast Fourier transform to get $\mathcal{U} = ifft(\overline{U}, [], 3)$, $\mathcal{S}_{w,\gamma} = ifft(\overline{S}_{w,\gamma}, [], 3)$, $\mathcal{V} = ifft(\overline{V}, [], 3)$, the optimal solution as follow

$$\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{US}}_{\boldsymbol{w},\boldsymbol{\mathcal{Y}}} \boldsymbol{\mathcal{V}}^{\mathrm{T}}$$
(50)

End of proof.

B3. Proof of Corollary 1

Proof From Lemma 1, we can see that the weighted near-end operator problem:

$$\min_{\boldsymbol{\mathcal{X}}} \left\| \boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{Y}} \right\|_{F}^{2} + \tau \cdot \left\| \boldsymbol{\mathcal{X}} \right\|_{\text{WTNFN}} = \min_{\overline{\boldsymbol{X}}^{(i)}} \frac{1}{n_{3}} \sum_{i=1}^{n_{3}} \left(\left\| \overline{\boldsymbol{X}}^{(i)} - \overline{\boldsymbol{Y}}^{(i)} \right\|_{F}^{2} \right) + \tau \cdot \left\| \overline{\boldsymbol{X}}^{(i)} \right\|_{\text{WTNFN}}$$
(51)

The closed form solution is

$$\overline{\boldsymbol{X}}^{(i)\star} = \overline{\boldsymbol{U}}^{(i)} \overline{\boldsymbol{S}}^{(i)}_{\boldsymbol{w},\boldsymbol{\gamma}} \left(\overline{\boldsymbol{S}}^{(i)} \right) \overline{\boldsymbol{V}}^{(i)\mathrm{T}}$$
(52)

where

$$\overline{S}_{w,\gamma}^{(i)} = \begin{pmatrix} \operatorname{diag}\left(\sigma_{1}\left(\overline{X}^{(i)\star}\right), \cdots, \sigma_{\min\{n_{1},n_{2}\}}\left(\overline{X}^{(i)\star}\right)\right) \\ 0 \end{pmatrix}$$
(53)

According to the updated rules of the algorithm, in each iteration, the weight $\left\{w_{i,j}^{(k)}\right\}$ are in a non-descending order, and the singular values $\sigma_j(\bar{X}^{(i)})$ can be updated independently in a non-increasing order as

$$\sigma_{j}\left(\overline{\boldsymbol{X}}_{k}^{(i)}\right) = \max\left\{\left(1 + \gamma w_{i,j}^{(k-1)}\right)^{-1} \left(\sigma_{j}\left(\overline{\boldsymbol{Y}}^{(i)}\right) - \frac{2}{\tau} w_{i,j}^{(k-1)}\right), 0\right\}$$
$$= \max\left\{\left(1 + \gamma C\left(\sigma_{j}\left(\overline{\boldsymbol{X}}_{k-1}^{(i)}\right) + \varepsilon\right)^{-1}\right)^{-1}, \left(\sigma_{j}\left(\overline{\boldsymbol{Y}}^{(i)}\right) - \frac{2}{\tau} C\left(\sigma_{j}\left(\overline{\boldsymbol{X}}_{k-1}^{(i)}\right) + \varepsilon\right)^{-1}\right), 0\right\}$$
(54)

Based on the prior condition that the weights are initialized as

$$w_{i,j}^{(0)} = \frac{C}{\sigma_j\left(\overline{\boldsymbol{Y}}^{(i)}\right) + \varepsilon}, \text{ then we have}$$

$$\sigma_j\left(\overline{\boldsymbol{X}}_1^{(i)}\right) = \max\left\{ \left(1 + \gamma w_{i,j}^{(0)}\right)^{-1} \left(\sigma_j\left(\overline{\boldsymbol{Y}}^{(i)}\right) - \frac{2}{\tau} w_{i,j}^{(0)}\right), 0\right\}$$

$$= \max\left\{ \left(1 + \gamma C \left(\sigma_j\left(\overline{\boldsymbol{Y}}^{(i)}\right) + \varepsilon\right)^{-1}\right)^{-1}, \left(\sigma_j\left(\overline{\boldsymbol{Y}}^{(i)}\right) - -\frac{2}{\tau} C \left(\sigma_j\left(\overline{\boldsymbol{Y}}^{(i)}\right) + \varepsilon\right)^{-1}\right), 0\right\} \quad (55)$$

$$\leq \sigma_j\left(\overline{\boldsymbol{Y}}^{(i)}\right)$$

Since $\gamma > 0, C > 0$. Then in the next iteration, we have

$$\sigma_{j}\left(\overline{\boldsymbol{X}}_{2}^{(i)}\right) = \max\left\{\left(1 + \gamma w_{i,j}^{(1)}\right)^{-1} \left(\sigma_{j}\left(\overline{\boldsymbol{Y}}^{(i)}\right) - \frac{2}{\tau} w_{i,j}^{(1)}\right), 0\right\}$$

$$= \max\left\{\left(1 + \gamma \boldsymbol{C}\left(\sigma_{j}\left(\overline{\boldsymbol{X}}_{2}^{(i)}\right) + \varepsilon\right)^{-1}\right)^{-1}, \left(\sigma_{j}\left(\overline{\boldsymbol{Y}}^{(i)}\right) - \frac{2}{\tau} \boldsymbol{C}\left(\sigma_{j}\left(\overline{\boldsymbol{X}}_{2}^{(i)}\right) + \varepsilon\right)^{-1}\right), 0\right\}$$

$$\leq \max\left\{\left(1 + \gamma \boldsymbol{C}\left(\sigma_{j}\left(\overline{\boldsymbol{Y}}^{(i)}\right) + \varepsilon\right)^{-1}\right)^{-1}, \left(\sigma_{j}\left(\overline{\boldsymbol{Y}}^{(i)}\right) - \frac{2}{\tau} \boldsymbol{C}\left(\sigma_{j}\left(\overline{\boldsymbol{Y}}^{(i)}\right) + \varepsilon\right)^{-1}\right), 0\right\}$$

$$= \sigma_{j}\left(\overline{\boldsymbol{X}}_{1}^{(i)}\right)$$
(56)

Then, the weights in the first and second iteration satisfy:

$$w_{i,j}^{(2)} = \boldsymbol{C} \left(\boldsymbol{\sigma}_j \left(\boldsymbol{\bar{X}}_2^{(i)} \right) + \boldsymbol{\varepsilon} \right)^{-1} \ge \boldsymbol{C} \left(\boldsymbol{\sigma}_j \left(\boldsymbol{\bar{X}}_1^{(i)} \right) + \boldsymbol{\varepsilon} \right)^{-1} = w_{i,j}^{(1)}$$
(57)

Hence, in the I iteration, suppose that, holds for all the singular values, then the inequality

$$\sigma_{j}\left(\overline{\boldsymbol{X}}_{I+1}^{(i)}\right) = \max\left\{\left(1 + \gamma \boldsymbol{C}\left(\sigma_{j}\left(\overline{\boldsymbol{X}}_{I}^{(i)}\right) + \varepsilon\right)^{-1}\right)^{-1}, \left(\sigma_{j}\left(\overline{\boldsymbol{Y}}^{(i)}\right) - \frac{2}{\tau}\boldsymbol{C}\left(\sigma_{j}\left(\overline{\boldsymbol{X}}_{I}^{(i)}\right) + \varepsilon\right)^{-1}\right), 0\right\}$$

$$\leq \max\left\{\left(1 + \gamma \boldsymbol{C}\left(\sigma_{j}\left(\overline{\boldsymbol{X}}_{I-1}^{(i)}\right) + \varepsilon\right)^{-1}\right)^{-1}, \left(\sigma_{j}\left(\overline{\boldsymbol{Y}}^{(i)}\right) - \frac{2}{\tau}\boldsymbol{C}\left(\sigma_{j}\left(\overline{\boldsymbol{X}}_{I-1}^{(i)}\right) + \varepsilon\right)^{-1}\right), 0\right\}$$

$$= \sigma_{j}\left(\overline{\boldsymbol{X}}_{I}^{(i)}\right)$$
(58)

In summary, $\forall j = 1, 2, \dots, \min\{n_1, n_2\}$, $i = 1, 2, 3, \dots, n_3$, $\sigma_j(\overline{X}_k^{(i)}) \leq \sigma_j(\overline{X}_{k-1}^{(i)})$ and $w_{i,j}^{(k)} \geq w_{i,j}^{(k-1)}$. Because the singular value $\{\sigma_j(\overline{X}_k^{(i)})\}$ is bounded below by 0, then the sequence is convergent. And sequence $\{\sigma_j(\overline{X}_k^{(i)})\}$ converges to a non-negative number $\sigma_j(\overline{X}_k^{(i)*})$.

$$\sigma_{j}\left(\overline{\boldsymbol{X}}^{(i)\star}\right) = \lim_{k \to \infty} \sigma_{j}\left(\overline{\boldsymbol{X}}_{k}^{(i)}\right)$$
$$= \max\left\{\left(1 + \gamma \lim_{k \to \infty} w_{i,j}^{(k)}\right)^{-1} \left(\sigma_{j}\left(\overline{\boldsymbol{Y}}^{(i)}\right) - \frac{2}{\tau} \lim_{k \to \infty} w_{i,j}^{(k)}\right), 0\right\}$$
$$= \left(1 + \gamma C \left(\sigma_{j}\left(\overline{\boldsymbol{X}}^{(i)\star}\right) + \varepsilon\right)^{-1}\right)^{-1} \left(\sigma_{j}\left(\overline{\boldsymbol{Y}}^{(i)}\right) - \frac{2}{\tau} C \left(\sigma_{j}\left(\overline{\boldsymbol{X}}^{(i)\star}\right) + \varepsilon\right)^{-1}\right)$$
(59)

Then we have

$$\sigma_{j}\left(\bar{\boldsymbol{X}}^{(i)\star}\right) = \begin{cases} 0 & \text{if } c_{2} > 0\\ \frac{c_{1} + \sqrt{c_{2}}}{2} & \text{if } c_{2} > 0 \end{cases}$$
(60)

where $c_2 = \tau^2 \left(\sigma_j \left(\overline{\mathbf{Y}}^{(i)} \right) - \varepsilon - \gamma \mathbf{C} \right)^2 - 4 \left(2\mathbf{C} - \tau \varepsilon \sigma_j \left(\overline{\mathbf{Y}}^{(i)} \right) \right)$ and $c_1 = \tau \left(\sigma_j \left(\overline{\mathbf{Y}}^{(i)} \right) - \varepsilon - \gamma \mathbf{C} \right).$

End of proof.