Low-Rank Tensor Completion Based on Bivariate Equivalent Minimax-Concave Penalty

Hongbing Zhang, Xinyi Liu, Hongtao Fan, Yajing Li, Yinlin Ye

Abstract—Low-rank tensor completion (LRTC) is an important problem in computer vision and machine learning. The minimax-concave penalty (MCP) function as a non-convex relaxation has achieved good results in the LRTC problem. To make all the constant parameters of the MCP function as variables so that furtherly improving the adaptability to the change of singular values in the LRTC problem, we propose the bivariate equivalent minimax-concave penalty (BEMCP) theorem. Applying the BEMCP theorem to tensor singular values leads to the bivariate equivalent weighted tensor 1-norm (BEWTGN) theorem, and we analyze and discuss its corresponding properties. Besides, to facilitate the solution of the LRTC problem, we give the proximal operators of the BEMCP theorem and BEWTGN. Meanwhile, we propose a BEMCP model for the LRTC problem, which is optimally solved based on alternating direction multiplier (ADMM). Finally, the proposed method is applied to the data restorations of multispectral image (MSI), magnetic resonance imaging (MRI) data recovery and color image/video (CI/CV) processing, and the experimental results demonstrate that it outperforms the state-of-arts methods.

Index Terms—Bivariate equivalent minimax-concave penalty (BEMCP), bivariate equivalent weighted tensor 1-norm (BEWTGN), low-rank tensor completion (LRTC).

1. INTRODUCTION

With the rapid development of information technology, researchers increasingly encounter real data with high dimensions and complex structures. Tensors, as high-dimensional generalizations of vectors and matrices, can better represent the complex properties of high-dimensional data and play an increasingly important role in many applications, such as color image/video (CI/CV) processing, hyperspectral/multispectral image (HSI/MSI) processing, magnetic resonance imaging (MRI) data recovery, background subtraction, video rain stripe removal, and signal reconstruction.

Low-rank tensor completion (LRTC) is an important issue in tensor recovery research. The general method expresses the low-rank tensor completion problem as follows:

\[
\min_{X} \text{rank}(X) \quad \text{s.t.} \quad P_{\Omega}(X - Y) = 0, \tag{1}
\]

where \( Y \in \mathbb{R}^{i_1 \times i_2 \times \cdots \times i_d} \) is the observation, \( \Omega \) is the index set for the known entries, and \( P_{\Omega}(Y) \) is a projection operator that keeps the entries of \( Y \) in \( \Omega \) and sets all others to zero, \( \text{rank}(X) \) defines the tensor rank of \( X \). In fact, tensors differ from matrices in that the definition of their rank is not unique. In the past decades, the most popular definitions of rank are CANDECOMP/PARAFAC/CP rank based on CP decomposition \([19], [20]\) and Tucker rank based on Tucker decomposition \([21], [22]\), as well as tubal rank and multi-rank based on t-SVD \([23]\). Solving the CP rank problem of tensors is a NP-hard \([24]\), which is not conducive to better application. The calculation of Tucker rank requires data to be folded and unfolded, which will cause structural damage to data. Compared with CP rank and Tucker rank, the tubal rank and multiple rank obtained based on t-SVD can better maintain the data structure, but its tensor-tensor product limitation prevents it from being applied to higher order cases. Recently, Zheng et al. \([25]\) proposed a new form of rank (N-tubal rank) based on tubal rank, which adopts a new unfold method of higher-order tensors into third-order tensors in various directions. This approach makes good use of the properties of tensor tubal rank but also enables t-SVD to be applied to higher-order cases. Therefore, because of the excellent properties of N-tubal rank, we will also consider N-tubal rank to construct the model we propose in this paper.

Undoubtedly, the development of computationally efficient algorithms to solve problem (1) is of great practical value. However, the rank optimization problem in problem (1) will lead to NP-hard problem \([24]\), which will seriously affect the efficiency of solving the problem. In this regard, researchers turn to its convex relaxation or non-convex relaxation forms. For convex relaxation, although it is easier to solve, it will produce biased estimates \([26]\). It is relatively difficult to solve for non-convex relaxations, but leads to more accurate results \([27], [28], [29]\). Recently, the minimax-concave penalty function as a non-convex relaxation has achieved good results in the LRTC problem \([30], [31]\). Further research on this function has profound implications for the LRTC problem, and its definition is as follows:

**Definition 1 (Minimax-concave penalty (MCP) function \([32]\)):** Let \( \lambda > 0, \gamma > 1 \). The minimax-concave penalty function \( h_{\gamma, \lambda} : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0} \) is defined as

\[
h_{\gamma, \lambda}(y) = \begin{cases} 
\lambda |y| - \frac{\gamma^2}{2\gamma}, & |y| \leq \gamma \lambda, \\
\frac{|y|^2}{2\gamma}, & |y| > \gamma \lambda. 
\end{cases}
\tag{2}
\]

It is not difficult to find that the MCP function contains two constant parameters, i.e., \( \lambda \) and \( \gamma \). It is considered that...
in the LRTC problem when the MCP function acts on the tensor singular values, its parameters are fixed. But the tensor singular values will change with the iteration update. Recently, an equivalent MCP (EMCP) method was proposed in [33], which is a novel non-convex relaxation method based on the traditional MCP method. The EMCP transforms the parameter \( \lambda \) in MCP into a variable form through the equivalence theorem so that it can adapt to the change of tensor singular values. It is worth noting that the MCP function contains two parameters, i.e., \( \lambda \) and \( \gamma \), which, in reality, affect each other, and the nonconvex relaxations produced by different \( \lambda \) and \( \gamma \) vary widely. A key idea is that it is especially important to expect to turn both parameters into variables at the same time, resulting in a more efficient equivalence theorem. Motivated by this, we propose a new structural equivalence theorem, i.e., the bivariate equivalent minimal-max-concave penalty (BEMCP) theorem, that allows \( \lambda \) and \( \gamma \) to be transformed into variables at the same time. The difference between MCP, EMCP, and BEMCP can be seen from the Table.[1]

**TABLE I**

| Variable Case for Three Methods |
|---------------------------------|
| Method  | \( \lambda \) | \( \gamma \) |
| MCP     | \( \times \)   | \( \times \)   |
| EMCP    | \( \checkmark \) | \( \times \)   |
| BEMCP   | \( \checkmark \) | \( \checkmark \) |

The symbols "\( \checkmark \)" and "\( \times \)" indicate whether \( \lambda \) and \( \gamma \) are variables or not.

To sum up, the main contributions of our paper are:

Firstly, a new structural equivalence theorem called BEMCP theorem is proposed, which turns two constant parameters \( \lambda \) and \( \gamma \) into variables at the same time and further improves adaptability of tensor singular value change in the LRTC problem. Applying the BEMCP theorem to tensor singular values leads to the bivariate equivalent weighted tensor \( \Gamma \)-norm (BEWTGN) theorem, and the corresponding properties are analyzed and discussed. Furthermore, to solve the established model based on this new theorem, the corresponding proximal operators of the BEMCP theorem and BEWTGN are proposed.

Secondly, for the LRTC problem, we propose a new model, i.e., BEMCP model based on N-tubal rank. Furthermore, we design an efficient alternating direction multiplier method (ADMM) algorithm [34], [35] to optimally solve these problems. On this basis, the closed solution of each variable update is deduced, so that the algorithm can be executed efficiently.

Thirdly, three different types of data, i.e., MSI, MRI, and CV, are used to verify the effectiveness and efficiency of proposed method. Extensive numerical experiments demonstrate that the results obtained by our method have clear advantages over the comparative method in both visual and quantitative values.

The summary of this article is as follows: In Section II, some preliminary knowledge and background of the tensors are given. The theorems about BEMCP and its properties are presented in Section III. In Section IV, we give the corresponding proximal operators and proofs of the BEMCP theorem and BEWTGN. The main results, including the proposed model and algorithm, are shown in Section V. The results of extensive experiments and discussions are presented in Section VI. Conclusions are drawn in Section VII.

**II. PRELIMINARIES**

**A. Tensor Notations and Definitions**

In this section, we give some basic notations and briefly introduce some definitions used throughout the paper. Generally, a lowercase letter and an uppercase letter denote a vector \( y \) and a matrix \( Y \), respectively. An \( N \)-th order tensor is denoted by a calligraphic uppercase letter \( Y \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N} \) and \( Y_{i_1,i_2,\ldots,i_N} \) is its \((i_1,i_2,\ldots,i_N)\)-th element. The Frobenius norm of a tensor is defined as \( \|Y\|_F = (\sum_{i_1,i_2,\ldots,i_N} Y_{i_1,i_2,\ldots,i_N}^2)^{1/2} \). For a three order tensor \( Y \in \mathbb{R}^{I_1 \times I_2 \times I_3} \), we use \( Y \) to denote along each tube of \( Y \), i.e., \( Y = f f t(Y, \{1\}, 3) \). The inverse DFT is computed by command \( f f t(Y, \{1\}, 3) \). More often, the frontal slice \( Y(\cdot,;i) \) is denoted compactly as \( Y^{(i)} \). The Hadamard product is the elementwise tensor product. Given tensors \( A \) and \( B \), both of size \( I_1 \times I_2 \times \cdots \times I_N \), their Hadamard product is denoted by \( A \circ B \). And the elementwise tensor division is denoted by \( \frac{A}{B} \). The set of real numbers greater than \( b \) real numbers is denoted as \( \mathbb{R}_{>b} = \{u \in \mathbb{R} \mid u > b \} \).

**Definition 2 (Mode-\( k \) k \( 1 \) k \( 2 \) slices)**: For an \( N \)-th order tensor \( Y \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N} \), its mode-\( k \) k \( 1 \) k \( 2 \) slices \( Y^{(k_1,k_2)} \), \( 1 \leq k_1 < k_2 \leq N, k_1, k_2 \in \mathbb{Z} \) are two-dimensional sections, defined by fixing all but the mode-\( k \) and the mode-\( k_2 \) indexes.

**Definition 3 (Tensor Mode-\( k \) Unfolding and Folding)**: For an \( N \)-th order tensor \( Y \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N} \), its mode-\( k \) Unfolding is a three order tensor denoted by \( Y_{(k_1,k_2)} \in \mathbb{R}^{I_k \times I_1 \times I_2 \times \cdots \times I_{k-1} \times I_{k+1} \times \cdots \times I_N} \), the frontal slices of which are the lexicographic orderings of the mode-\( k \) k \( 1 \) k \( 2 \) slices of \( Y \). Mathematically, the \((i_1,i_2,\ldots,i_N)\)-th element of \( Y \) maps to the \((i_k,i_{k+1},\ldots,i_N)\)-th element of \( Y^{(k_1,k_2)} \), where

\[
\begin{align*}
\sum_{s=1, s \neq k_1, s \neq k_2}^{N} (i_s - 1) = J_s \quad \text{with} \quad J_s = \prod_{m=1, m \neq k_1, m \neq k_2}^{N-1} I_m.
\end{align*}
\]

The mode-\( k \) Unfolding operator and its inverse operation are respectively represented as \( Y^{(k_1,k_2)} := t - \text{unfold}(Y, k_1, k_2) \) and \( Y := t - \text{fold}(Y^{(k_1,k_2)}, k_1, k_2) \).

For a three order tensor \( Y \in \mathbb{R}^{I_1 \times I_2 \times I_3} \), the block circulation operation is defined as

\[
bcirc(Y) := \left[ \begin{array}{cccc}
Y^{(1)} & Y^{(0)} & \cdots & Y^{(2)} \\
Y^{(2)} & Y^{(1)} & & \cdots \\\n\vdots & \vdots & \ddots & \vdots \\
Y^{(1)} & Y^{(0)} & \cdots & Y^{(1)}
\end{array} \right] \in \mathbb{R}^{I_1 \times I_2 \times I_3}.
\]
The block diagonalization operation and its inverse operation are given by

\[
bdia(Y) := \begin{pmatrix}
Y^{(1)} \\
Y^{(2)} \\
\vdots \\
Y^{(n)}
\end{pmatrix} \in \mathbb{R}^{l_1 \times l_2 \times \cdots \times l_n},
\]

\[
bdfold(bdia(Y)) := Y.
\]

The block vectorization operation and its inverse operation are defined as

\[
bvec(Y) := \begin{pmatrix}
Y^{(1)} \\
Y^{(2)} \\
\vdots \\
Y^{(n)}
\end{pmatrix} \in \mathbb{R}^{l_1 l_2 \cdots l_n},
\]

\[
bfold(bvec(Y)) := Y.
\]

**Definition 4 (T-product [36]):** Let \( A \in \mathbb{R}^{l_1 \times l_2 \times \cdots \times l_n} \) and \( B \in \mathbb{R}^{l_3 \times l_4 \times \cdots \times l_n} \). Then the t-product \( A \ast B \) is defined to be a tensor of size \( I_1 \times I_2 \times I_3 \).

\[
A \ast B := bfold(bcirc(A)bvec(B)).
\]

Since that circular convolution in the spatial domain is equivalent to multiplication in the Fourier domain, the T-product between two tensors \( C = A \ast B \) is equivalent to

\[
\tilde{C} = bdfold(bdiag(\tilde{A})bdiag(\tilde{B})).
\]

**Definition 5 (Tensor conjugate transpose [36]):** The conjugate transpose of a tensor \( A \in \mathbb{C}^{l_1 \times l_2 \times \cdots \times l_n} \) is the tensor \( A^H \in \mathbb{C}^{l_3 \times l_4 \times \cdots \times l_n} \) obtained by conjugate transposing each of the frontal slices and then reversing the order of transposed frontal slices 2 through 1.

**Definition 6 (Identity tensor [36]):** The identity tensor \( I \in \mathbb{R}^{l_1 \times l_2 \times \cdots \times l_n} \) is the tensor whose first frontal slice is the \( I_1 \times I_2 \) identity matrix, and whose other frontal slices are all zeros.

It is clear that \( bcirc(I) \) is the \( I_1 \times I_2 \times I_3 \) identity matrix. So it is easy to get \( A \ast I = A \) and \( I \ast A = A \).

**Definition 7 (Orthogonal tensor [36]):** A tensor \( Q \in \mathbb{R}^{l_1 \times l_2 \times l_3} \) is orthogonal if it satisfies

\[
Q \ast Q^H = Q^H \ast Q = I.
\]

**Definition 8 (F-diagonal tensor [36]):** A tensor is called f-diagonal if each of its frontal slices is a diagonal matrix.

**Theorem 1 (t-SVD [36]):** Let \( X \in \mathbb{R}^{l_1 \times l_2 \times l_3} \) be a three order tensor, then it can be factored as

\[
X = U \ast S \ast V^H,
\]

where \( U \in \mathbb{R}^{l_1 \times l_2 \times l_3} \) and \( V \in \mathbb{R}^{l_2 \times l_3 \times l_1} \) are orthogonal tensors, and \( S \in \mathbb{R}^{l_1 \times l_2 \times l_3} \) is an f-diagonal tensor.

**Definition 9 (Tensor tubal-rank and multi-rank [23]):** The tubal-rank of a tensor \( Y \in \mathbb{R}^{l_1 \times l_2 \times l_3} \), denoted as \( rank_t(Y) \), is defined to be the number of non-zero singular tubes of \( S \), where \( S \) comes from the t-SVD of \( Y : Y = U \ast S \ast V^H \). That is

\[
rank_t(Y) = \#\{i : S(i, \ldots, :) \neq 0\}.
\]

The tensor multi-rank of \( Y \in \mathbb{R}^{l_1 \times l_2 \times l_3} \) is a vector, denoted as \( rank_v(Y) \in \mathbb{R}^{l_1 \times l_2 \times l_3} \), with the \( i \)-th element equals to the rank of \( i \)-th frontal slice of \( Y \).

**Definition 10 (Tensor nuclear norm (TNN)):** The tensor nuclear norm of a tensor \( Y \in \mathbb{R}^{l_1 \times l_2 \times l_3} \), denoted as \( \|Y\|_{TNN} \), is defined as the sum of the singular values of all the frontal slices of \( Y \), i.e.,

\[
\|Y\|_{TNN} := \sum_{i=1}^{l_1} \|Y^{(i)}\|.
\]

where \( Y^{(i)} \) is the \( i \)-th frontal slice of \( Y \), with \( \bar{Y} = f_{tt}(Y) \).

**Definition 11 (N-tubal rank [23]):** The N-tubal rank of an \( N \)-th order tensor \( Y \in \mathbb{R}^{l_1 \times l_2 \times \cdots \times l_N} \) is defined as a vector, the elements of which contain the tubal rank of all mode-\( k_1 k_2 \) unfolding tensors, i.e.,

\[
N \cdot rank_t(Y) := \{ rank_t(Y_{(1)}), \ldots, rank_t(Y_{(N)}) \}.
\]

**Theorem 2 (Equivalent minimax-concave penalty (EMCP) [23]):** Let \( \lambda \in \mathbb{R}_{\geq 0}, \gamma > 1 \) and \( \gamma \in \mathbb{R} \). The MCP \( h_{\gamma,\lambda} : \mathbb{R} \to \mathbb{R}_{\geq 0} \) is the solution of the following optimization problem:

\[
h_{\gamma,\lambda}(y) = \min_{y \in \mathbb{R}_{\geq 0}} \left( \gamma |y| + \frac{\gamma}{2} (\gamma - \lambda)^2 \right).
\]

**III. BIVARIATE EQUIVALENT MINIMAX-CONCAVE PENALTY**

In this section, we will construct and obtain the BEMCP theorem, which turns \( \lambda \) and \( \gamma \) into variables at the same time. Then the BEWTGN theorem is applied to tensor singular values to deduce the BEWTGN theorem, and its corresponding properties are analyzed and established.

**Theorem 3 (Bivariate Equivalent Minimax-Concave Penalty (BEMCP)):** Let \( \lambda, \gamma \in \mathbb{R}_{>0}, \gamma \in \mathbb{R}_{>1} \) and \( \gamma \in \mathbb{R} \). The MCP \( h_{\gamma,\lambda} : \mathbb{R} \to \mathbb{R}_{>0} \) is the solution of the following optimization problem:

\[
h_{\gamma,\lambda}(y) = \min_{\nu \geq 0} \frac{2\nu |y| + (\nu - \lambda)^2}{2\gamma}.
\]

**Proof:** Consider the following function

\[
h(y, \nu) = \frac{2\nu |y| + (\nu - \lambda)^2}{2\gamma}.
\]

Let \( \nu^* \) denote the first-order critical point of \( h(y, \nu) \), i.e.,

\[
\nu^* = \arg \min_{\nu} h(y, \nu).
\]

Since \( h(y, \nu) \) is differentiable with respect to \( \nu \), setting

\[
\frac{\partial h(y, \nu)}{\partial \nu} \big|_{\nu=\nu^*} = 0
\]

gives

\[
\nu^* = \begin{cases} 
\frac{\lambda y - |y|}{\gamma}, & |y| \leq \gamma \lambda, \\
0, & |y| \geq \gamma \lambda.
\end{cases}
\]

The BEMCP is given by \( h_{\gamma,\lambda}(y) = h(y, \nu^*) \). Substituting for \( \nu^* \) in \( h(y, \nu) \), we get

\[
h_{\gamma,\lambda}(y) = \begin{cases} 
\frac{\lambda |y| - \gamma^2}{2\gamma}, & |y| \leq \gamma \lambda, \\
\frac{\gamma^2}{2\gamma}, & |y| \geq \gamma \lambda.
\end{cases}
\]
Definition 12 (Weighted Tensor Γ-norm (WTGN)): The tensor Γ-norm of \( \mathbf{Y} \in \mathbb{R}^{I_1 \times I_2 \times I_3} \), denoted by \( \| \mathbf{Y} \|_{\Gamma,A} \), is defined as follows:

\[
\| \mathbf{Y} \|_{\Gamma,A} = \sum_{i=1}^{I_1} \sum_{j=1}^{R_i} h_{i,(i,j)}(\sigma_j(\mathbf{Y}^{(i)})),
\]

where \( R = \min(I_1, I_2), \Lambda \in \mathbb{R}_{\geq 0}^{I_1 \times R}, \Gamma \in \mathbb{R}_{\geq 1}^{I_1 \times R}, \) and \( R = \min(I_1, I_2) \).

Theorem 4 (Bivariate Equivalent Weighted Tensor Γ-norm (BEWTGN)): For a third-order tensor \( \mathbf{Y} \in \mathbb{R}^{I_1 \times I_2 \times I_3} \). Let \( \nu, \Lambda \in \mathbb{R}_{\geq 0}^{I_1 \times R}, \Gamma \in \mathbb{R}_{\geq 1}^{I_1 \times R} \), and \( R = \min(I_1, I_2) \).

Proof: The proof is similar to that of Theorem 3 since the objective function in Eq. (15) is non-negative and separable.

Remark 1: In particular, when the third dimension \( I_3 \) of the third-order tensor \( \mathbf{Y} \) is 1, the BEWTGN can degenerate into the form of the bivariate equivalent matrix Γ-norm.

Remark 2: Unlike the nuclear norm penalty, the WTGN (14), and the BEWTGN (15) do not satisfy the triangle inequality. Some important properties of the BEWTGN itself are presented below.

Proposition 1: The BEWTGN defined in (15) satisfies the following properties:

(a) Non-negativity: The BEWTGN is non-negative, i.e., \( \| \mathbf{Y} \|_{\Gamma,A} \geq 0 \). The equality holds if and only if \( \mathbf{Y} \) is the null tensor.

(b) Concavity: \( \| \mathbf{Y} \|_{\Gamma,A} \) is concave in the modulus of the singular values of \( \mathbf{Y} \).

(c) Upper bound: The BEWTGN is upper-bounded by the weighted nuclear norm, i.e., \( \| \mathbf{Y} \|_{\Gamma,A} \leq \| \mathbf{Y} \|_{\Lambda,A} \).

(d) Nuclear norm property: The BEWTGN approaches the weighted nuclear norm asymptotically, i.e., \( \lim_{\Gamma \to \infty} \| \mathbf{Y} \|_{\Gamma,A} = \| \mathbf{Y} \|_{\Lambda,A} \).

(e) Unitary invariance: The BEWTGN is unitary invariant, i.e., \( \| \mathbf{U} \ast \mathbf{Y} \ast \mathbf{V} \|_{w,s} = \| \mathbf{Y} \|_{\Gamma,A} \) for any unitary tensors \( \mathbf{U} \) and \( \mathbf{V} \).

Proof: Let

\[
p(\mathbf{Y}) = \| \mathbf{Y} \|_{\Gamma,A} + \frac{1}{2} \left( \frac{\nu - \Lambda \ast \Gamma}{\Gamma \ast \Gamma} \right)^2.
\]

Since \( p(\mathbf{Y}) \) is the sum of two non-negative functions, \( \| \mathbf{Y} \|_{\Gamma,A} \geq 0 \). The equality holds if \( \| \mathbf{Y} \|_{\Gamma,A} = 0 \), i.e., \( \mathbf{Y} = 0 \) or \( \nu = 0 \), the latter being the trivial solution.

(b) The function \( p(\mathbf{Y}) \) is separable of \( \mathbf{Y} \), i.e.,

\[
p(\mathbf{Y}) = \sum_{i=1}^{I_1} \sum_{j=1}^{R_i} \left( \frac{\nu_{(i,j)} - \Lambda_{(i,j)} \Gamma_{(i,j)}}{2 \Gamma_{(i,j)}} \right)^2.
\]
IV. PROXIMAL OPERATORS FOR THE BEMCP THEOREM AND BEWTGN

In this section, to solve the model established based on the BEMCP theorem, we present the proximal operators for the BEMCP theorem and BEWTGN.

**Theorem 5 (Proximal operator for the BEMCP):** Consider the BEMCP given in Eq. (8). Its proximal operator denoted by $P_{\gamma,\lambda}: \mathbb{R} \to \mathbb{R}$, $y \in \mathbb{R}_{>1}$, $\lambda \in \mathbb{R}_{>0}$ and defined as follows:

$$P_{\gamma,\lambda}(y) = \arg \min_{y} \left\{ \frac{1}{2} (g - y)^2 + h_{\gamma,\lambda}(g) \right\}$$

is given by

$$P_{\gamma,\lambda}(y) = \min \left\{ |y|, \max \left( \frac{\gamma (|y| - \lambda)}{\gamma - 1}, 0 \right) \right\} \text{sign}(y).$$

**Proof:** Let

$$P_{\gamma,\lambda}(y) = \arg \min_{y} \left\{ \frac{1}{2} (g - y)^2 + \frac{2\gamma |g| + (\nu - \lambda y^2)}{2\gamma} \right\},

= \arg \min_{y} \left\{ \frac{1}{2} (g - y)^2 + h(g, \nu^*) \right\},$$

where

$$\nu^* = \arg \min_{\nu \geq 0} h(g, \nu) = \max(\lambda y - |g|, 0).$$

We must now determine $P_{\gamma,\lambda}(y)$ for a given $\nu^*$. This is derived considering various values that $\nu^*$ can take:

1. Case 1: $\nu^* = y = 0$. Correspondingly, $h(g, \nu^*) = \frac{4g}{g - 2}$, and then

$$P_{\gamma,\lambda}(y) = \arg \min_{y} \left\{ \frac{1}{2} (g - y)^2 + \frac{2\gamma |g| + (\nu^* - \lambda y^2)}{2\gamma} \right\} = \frac{\gamma (|y| - \lambda)}{\gamma - 1} \text{ sign}(y).$$

The condition $\nu^* = y$ translates to $|g| > \lambda y$, therefore,

$$P_{\gamma,\lambda}(y) = y, \quad \text{for } |g| > \lambda y.$$  

(22)

2. Case 2: $0 < \nu^* < \lambda y$. By definition of $\nu^*$, we have that

$$\nu^* = \lambda y - |g|, \quad |g| < \lambda y.$$  

(23)

Further, considering $0 < g < \lambda y$ gives $\nu^* = \lambda y - g$, which result in

$$P_{\gamma,\lambda}(y) = \arg \min_{y} \left\{ \frac{1}{2} (g - y)^2 + \frac{2\gamma |g| + (\nu^* - \lambda y^2)}{2\gamma} \right\},

= \frac{\gamma (|y| - \lambda)}{\gamma - 1} \text{ for } \lambda y < y < \lambda y.$$  

Similarly, considering $-\lambda y < g < 0$ gives $\nu^* = \lambda y + g$, which result in

$$P_{\gamma,\lambda}(y) = \arg \min_{y} \left\{ \frac{1}{2} (g - y)^2 + \frac{2\gamma |g| + (\nu^* - \lambda y^2)}{2\gamma} \right\},

= \frac{\gamma (|y| - \lambda)}{\gamma - 1} (-\lambda y < y < -\lambda y).$$  

Combining the two expressions brings

$$P_{\gamma,\lambda}(y) = \frac{\gamma}{\gamma - 1} (|y| - \lambda) \text{sign}(y), \quad \text{for } \lambda < |y| < \lambda y.$$  

(24)

3. Case 3: $\nu^* = \lambda y$. This induces $g = 0$, which implies

$$P_{\gamma,\lambda}(y) = \frac{\gamma}{\gamma - 1} (|y| - \lambda) \text{sign}(y), \quad \text{for } \lambda < |y| < \lambda y.$$  

(24)

$\nu^* = \lambda y$ in the definition of $P_{\gamma,\lambda}(y)$ gives

$$0 = \arg \min_{\nu} \left\{ \frac{1}{2} (g - y)^2 + \lambda |g| \right\}.$$  

(25)

Considering the subdifferential, we get $0 \in -y + \lambda \partial |g|$, or equivalently, $|g| \leq \lambda$. Therefore,

$$P_{\gamma,\lambda}(y) = y, \quad \text{for } |y| \leq \lambda.$$  

(26)

Combining equations (22), (24), and (25) yields

$$P_{\gamma,\lambda}(y) = \min \left\{ |y|, \max \left( \frac{\gamma (|y| - \lambda)}{\gamma - 1}, 0 \right) \right\} \text{sign}(y).$$  

(26)

**Theorem 6 (Proximal operator for the BEWTGN):** Consider the BEWTGN given in Eq. (15). Its proximal operator denoted by $S_{\Gamma,\Lambda}: \mathbb{R} \to \mathbb{R}$, $Y \in \mathbb{R}^{R \times \Lambda}$, $\Gamma \in \mathbb{R}^{R \times \Lambda}$, $\Lambda \in \mathbb{R}^{R \times \Lambda}$, $R = \min \{L_1, L_2\}$ and defined as follows:

$$S_{\Gamma,\Lambda}(Y) = \arg \min_{Y} \left\{ \frac{1}{2} \| Y - Y \|_F^2 + \| L \|_{\Gamma,\Lambda} \right\},$$

(27)

is given by

$$S_{\Gamma,\Lambda}(Y) = U \ast S_1 \ast \mathcal{V}^H.$$  

(28)

where $U$ and $V$ are derived from the t-SVD of $Y = U \ast S_2 \ast \mathcal{V}^H$. More importantly, the $ith$ front slice of DFT of $S_1$ and $S_2$, i.e., $S_1(i) = \sigma(\vec{L}(i))$ and $S_2(i) = \sigma(\vec{L}(i))$, has the following relationship:

$$\sigma(\vec{L}(i)) = \min \left\{ \sigma(\vec{S}(i)), \max \left\{ \frac{\Gamma(i)}{\Gamma(i)} - 1, 0 \right\} \right\}.$$  

(28)

**Proof:** Let $\mathcal{Y} = U \ast S_2 \ast \mathcal{V}^H$ and $\mathcal{L} = W \ast S_1 \ast \mathcal{R}^H$ be the t-SVD of $\mathcal{Y}$ and $\mathcal{L}$, respectively. Consider

$$S_{\Gamma,\Lambda}(Y) = \arg \min_{Y} \left\{ \frac{1}{2} \| Y - Y \|_F^2 + \| L \|_{\Gamma,\Lambda} \right\},

= \arg \min_{Y} \left\{ \frac{1}{2} \| W \ast S_1 \ast \mathcal{R}^H - U \ast S_2 \ast \mathcal{V}^H \|_F^2 + \| L \|_{\Gamma,\Lambda} \right\},

= \arg \min_{Y} \left\{ \sum_{i=1}^{R} \frac{1}{2} \| \tilde{W}(i) \ast \tilde{S}(i) \ast \tilde{R}(i) - \tilde{U}(i) \|_F^2 + \| L \|_{\Gamma(i)} \right\},$$

(29)

It can be found that (29) is separable and can be divided into $I_3$ sub-problems. For the $ith$ sub-problem:

$$\arg \min_{\tilde{L}(i)} \left\{ \frac{1}{2} \| \tilde{W}(i) \ast \tilde{S}(i) \ast \tilde{R}(i) - \tilde{U}(i) \ast \tilde{S}(i) \ast \tilde{V}(i) \|_F^2 + \| \tilde{L}(i) \|_{\Gamma(i),\Lambda(i)} \right\},$$

(30)
Invoking von Neumann’s trace inequality \[38\], we can write
\[
\arg \min_{\tilde{L}^{(i)}} \frac{1}{2} \| \tilde{W}^{(i)} \star \tilde{S}_1^{(i)} + \tilde{R}^{(i)H} - \tilde{U}^{(i)} \star \tilde{S}_2^{(i)} + \tilde{V}^{(i)H} \|_F^2 \\
+ \| \tilde{L}^{(i)} \|_{\Gamma(\alpha, \Lambda)}^2 \\
\geq \arg \min_{\tilde{S}^{(i)}} \frac{1}{2} \text{Tr}(\tilde{S}_1^{(i)} \tilde{S}_1^{(i)H}) + \frac{1}{2} \text{Tr}(\tilde{S}_2^{(i)} \tilde{S}_2^{(i)H}) \\
+ \text{Tr}(\tilde{S}_2^{(i)} \tilde{S}_1^{(i)H}) + \| \tilde{L}^{(i)} \|_{\Gamma(\alpha, \Lambda)}^2 \\
= \arg \min_{\alpha, \Lambda(\tilde{L}^{(i)})} \frac{1}{2} \| \sigma(\tilde{L}^{(i)}) - \sigma(\tilde{Y}^{(i)}) \|_F^2 + \| \tilde{L}^{(i)} \|_{\Gamma(\alpha, \Lambda)}^2 .
\]
The equality holds when \( \tilde{W}^{(i)} = \tilde{U}^{(i)} \) and \( \tilde{R}^{(i)} = \tilde{V}^{(i)} \). Therefore, the optimal solution to (29) is obtained by solving the problem given below:
\[
\sigma(\tilde{L}^{(i)}) = \min \left\{ \sigma(\tilde{Y}^{(i)}), \max \left\{ \frac{\Gamma(\alpha, \Lambda(\tilde{L}^{(i)}))}{\Gamma(\alpha, \Lambda)}, 0 \right\} \right\}.
\]

V. THE BEMCP MODELS AND SOLVING ALGORITHMS

In this section, in order to better verify the superiority of our BEMCP theorem, before giving the BEMCP model based on the N-tubal rank, we first review the models of traditional MCP and EMCP based on the N-tubal rank. And the traditional MCP model based on N-tubal rank is called the NCMP model.

\[
\min_{\chi} \sum_{1 \leq l_1 \leq l_2 \leq N} \alpha_{l_1 l_2} \| \chi_{l_1 l_2} \|_{\gamma, \Lambda}, \quad \text{s.t.} \quad \mathcal{P}_{\Omega}(\chi - \Omega) = 0.
\]

Applying different MCP non-convex functions to singular values of problem (31) leads to the following model.

A. The NMCP model

Using the MCP function, we get the following optimization model:

\[
\min_{\chi} \sum_{1 \leq l_1 \leq l_2 \leq N} \alpha_{l_1 l_2} \| \chi_{l_1 l_2} \|_{\gamma, \Lambda}, \quad \text{s.t.} \quad \mathcal{P}_{\Omega}(\chi - \Omega) = 0.
\]

Under the framework of alternation direction method of multipliers (ADMM) \[34\], \[35\], \[39\], the easy-to-implement optimization strategy could be provided to solve (32). We introduce a set of tensors \( \{ \gamma_{l_1 l_2} = X_{l_1 l_2} \}_{1 \leq l_1 \leq l_2 \leq N} \) and transfer optimization problem (32), in its augmented Lagrangian form, as follows:

\[
J(\chi, \gamma, Q) = \sum_{1 \leq l_1 \leq l_2 \leq N} \alpha_{l_1 l_2} \| \gamma_{l_1 l_2} \|_{\gamma, \Lambda} + \frac{\rho_{l_1 l_2}}{2} \| \chi - \gamma_{l_1 l_2} + Q_{l_1 l_2} \|_F^2 \\
\text{s.t.} \quad \mathcal{P}_{\Omega}(\chi - \Omega) = 0,
\]

where \( \{ Q_{l_1 l_2} \}_{1 \leq l_1 \leq l_2 \leq N} \) are tensor Lagrangian multiplier sets; \( \{ \rho_{l_1 l_2} \}_{1 \leq l_1 \leq l_2 \leq N} > 0 \) are the augmented Lagrangian parameters; \( \alpha_{l_1 l_2} \geq 0 \) are N-tubal rank weights and \( \sum_{1 \leq l_1 \leq l_2 \leq N} \alpha_{l_1 l_2} = 1 \). Besides, variables \( \chi, \gamma, Q \) are updated alternately in the order of \( \gamma \rightarrow \chi \rightarrow Q \). For convenience, we mark the updated variable as \( (\cdot)^+ \). The update equations are acquired in the following.

B. The EMCP model

Similarly, using the EMCP Theorem, we get the following optimization model:

\[
\min_{\chi} \sum_{1 \leq l_1 \leq l_2 \leq N} \alpha_{l_1 l_2} \| \chi_{l_1 l_2} \|_{\gamma, \Lambda(\gamma)} = 0.
\]

Under the framework of the alternation direction method of multipliers (ADMM), the easy-to-implement optimization strategy could be provided to solve (37). We introduce a set of tensors \( \{ \gamma_{l_1 l_2} = X_{l_1 l_2} \}_{1 \leq l_1 \leq l_2 \leq N} \) and transfer optimization problem (37), in its augmented Lagrangian form, as follows:

\[
J(\chi, \gamma, W, Q) = \sum_{1 \leq l_1 \leq l_2 \leq N} \alpha_{l_1 l_2} \| \gamma_{l_1 l_2} \|_{\gamma, \Lambda(\gamma)} + \frac{\rho_{l_1 l_2}}{2} \| \chi - \gamma_{l_1 l_2} + Q_{l_1 l_2} \|_F^2 \\
\text{s.t.} \quad \mathcal{P}_{\Omega}(\chi - \Omega) = 0.
\]
where $\mathbf{Y}$ and $\mathbf{Q}$ are tensor sets; $\bar{\Lambda}$ and $\Lambda$ are matrix sets; 
$\{Y_{i_1i_2} = X_{i_1i_2}\}_{1 \leq i_1 < i_2 \leq N}$; 
$\{Q_{i_1i_2}\}_{1 \leq i_1 < i_2 \leq N}$ are Lagrangian multipliers; 
$\{\Lambda, W\}_{1 \leq i_1 < i_2 \leq N} \in \mathbb{R}^{p \times N}$ are MCP variable and weight sets, respectively; 
$\{\rho_{i_1i_2}\}_{1 \leq i_1 < i_2 \leq N} > 0$ are the augmented Lagrangian parameters; 
$\alpha_{i_1i_2} \geq 0$ are weights and 
$\sum_{1 \leq i_1 < i_2 \leq N} \alpha_{i_1i_2} = 1$.

Besides, variables $X$, $\mathbf{Y}$, $\bar{\Lambda}$, $\Lambda$, $W$, $\mathbf{Q}$ are updated alternately in the order of $\mathbf{Y} \rightarrow W \rightarrow \bar{\Lambda} \rightarrow X \rightarrow Q$. The update equations are acquired in the following.

**Update $\mathbf{Y}$:** Fix other variables, and the corresponding optimization is as follows:

$$
\mathbf{Y}^+_{i_1i_2} = \arg\min_{\mathbf{Y}_{i_1i_2}} \frac{\rho_{i_1i_2}}{2} \|X - \mathbf{Y}_{i_1i_2} + \mathbf{Q}_{i_1i_2}\|^2_F.
$$

Invoking Theorem $\mathbf{6}$ the solution to the above optimization is given by:

$$
\mathbf{Y}^+_{i_1i_2} = \mathbf{S}_{\gamma\rho_{i_1i_2}}(X + \mathbf{Q}_{i_1i_2}),
$$

where $\mathbf{S}$ denotes the proximal operator defined in $[28]$. 

**Update $W$:** Retaining only those components in $\mathbf{Y}^+_{i_1i_2}$ in $\mathbf{38}$ that depend on $W_{i_1i_2}$, we write

$$
W^+_{i_1i_2} = \arg\min_{W_{i_1i_2}} \frac{\rho_{i_1i_2}}{2} \|X - \mathbf{Y}^+_{i_1i_2} + \mathbf{Q}_{i_1i_2}\|^2_F,
$$

which has the following closed-form solution:

$$
W^+_{i_1i_2} = \max(\bar{\Lambda}_{i_1i_2}, \mathbf{0}).
$$

**Update $\bar{\Lambda}$:** The update for $\bar{\Lambda}_{i_1i_2}$ has the following closed-form solution:

$$
\bar{\Lambda}^+_{i_1i_2} = \arg\min_{\bar{\Lambda}_{i_1i_2}} \|\bar{W}_{i_1i_2} - \bar{\Lambda}_{i_1i_2}\|^2_F,
$$

**Update $X$:** The closed form of $X$ can be acquired by setting the derivative of $[38]$ to zero. We can now update $X$ by the following equation:

$$
X^{+} = \mathcal{P}_{\Omega_{\mathbf{Y}}}(-\frac{\sum_{1 \leq i_1 < i_2 \leq N} \rho_{i_1i_2} (\mathbf{Y}_{i_1i_2} - \mathbf{Q}_{i_1i_2})}{\sum_{1 \leq i_1 < i_2 \leq N} \rho_{i_1i_2}}) + \mathcal{P}_{\Omega}(\mathbf{Z}).
$$

**Update $Q$:** Finally, multipliers $Q_{i_1i_2}$ are updated as follows:

$$
Q^{+}_{i_1i_2} = Q_{i_1i_2} + \rho_{i_1i_2}(X - \mathbf{Y}_{i_1i_2}).
$$

The optimization steps of the EMCP formulation are listed in Algorithm $\mathbf{2}$.

### C. The BEMCP model

Using the BEMCP Theorem, we get the following optimization model:

$$
\min_X \sum_{1 \leq i_1 < i_2 \leq N} \alpha_{i_1i_2} \|\mathbf{X}_{i_1i_2}\|_{\Gamma_{i_1i_2}} \text{ s.t. } \mathcal{P}_{\Omega}(X - \mathbf{Z}) = \mathbf{0}.
$$

Under the framework of the ADMM, the easy-to-implement optimization strategy could be provided to solve $[44]$. We introduce a set of tensors $\{\mathbf{Y}_{i_1i_2} = X_{i_1i_2}\}_{1 \leq i_1 < i_2 \leq N}$ and transform optimization problem $[44]$, in its augmented Lagrangian form, as follows:

$$
J(X, \mathbf{Y}, \Lambda, \Gamma, \nu, \mathbf{Q}) = \sum_{1 \leq i_1 < i_2 \leq N} \alpha_{i_1i_2} \|\mathbf{Y}_{i_1i_2}\|_{\Gamma_{i_1i_2}} \text{ s.t. } \mathcal{P}_{\Omega}(X - \mathbf{Z}) = \mathbf{0},
$$

where $\mathbf{Y}$ and $\mathbf{Q}$ are tensor sets; $\Lambda, \Gamma, \nu$ are matrix sets; 
$\{\mathbf{Y}_{i_1i_2} = X_{i_1i_2}\}_{1 \leq i_1 < i_2 \leq N}$; 
$\{Q_{i_1i_2}\}_{1 \leq i_1 < i_2 \leq N}$ are Lagrangian multipliers; 
$\{\Lambda, \Gamma, \nu\}_{1 \leq i_1 < i_2 \leq N} \in \mathbb{R}^{p \times N}$ are MCP variable sets; 
$\{\rho_{i_1i_2}\}_{1 \leq i_1 < i_2 \leq N} > 0$ are the augmented Lagrangian parameters; 
$\alpha_{i_1i_2} \geq 0$ are weights and 
$\sum_{1 \leq i_1 < i_2 \leq N} \alpha_{i_1i_2} = 1$.

Besides, variables $X, \mathbf{Y}, \Lambda, \Gamma, \nu, \mathbf{Q}$ are updated alternately in the order of $\mathbf{Y} \rightarrow \nu \rightarrow \Lambda \rightarrow \Gamma \rightarrow X \rightarrow \mathbf{Q}$. The update equations are derived in the following.

**Update $\mathbf{Y}$:** Fix other variables, and the corresponding optimization are as follows:

$$
\mathbf{Y}^+_{i_1i_2} = \arg\min_{\mathbf{Y}_{i_1i_2}} \frac{\rho_{i_1i_2}}{2} \|X - \mathbf{Y}_{i_1i_2} + \mathbf{Q}_{i_1i_2}\|^2_F + \frac{1}{\nu_{i_1i_2}} \|\mathbf{Y}_{i_1i_2} - \Lambda_{i_1i_2}\|^2_F + \frac{1}{\nu_{i_1i_2}} \|\mathbf{Y}_{i_1i_2} - \Gamma_{i_1i_2}\|^2_F.
$$

Calling Theorem $\mathbf{6}$ the solution to the above optimization is given by:

$$
\mathbf{Y}^+_{i_1i_2} = \mathbf{S}_{\gamma\rho_{i_1i_2}}(X + \mathbf{Q}_{i_1i_2}),
$$

where $\mathbf{S}$ denotes the proximal operator defined in $[28]$. 

**Update $\nu$:** Retaining only those components in $\mathbf{Y}^+_{i_1i_2}$ in $[45]$ that depend on $u_{i_1i_2}$, we write

$$
u^+_{i_1i_2} = \arg\min_{\nu_{i_1i_2}} \frac{1}{\nu_{i_1i_2}} \|\mathbf{Y}_{i_1i_2} - \Lambda_{i_1i_2}\|^2_F + \frac{1}{\nu_{i_1i_2}} \|\mathbf{Y}_{i_1i_2} - \Gamma_{i_1i_2}\|^2_F,
$$

which has the following closed-form solution:

$$
\nu^+_{i_1i_2} = \max(\Lambda_{i_1i_2}, \mathbf{0}) - \sigma(\mathbf{Y}^+_{i_1i_2}), \mathbf{e},
$$

**Algorithm 2 EMCP**

**Input:** An incomplete tensor $\mathbf{Z}$, the index set of the known elements $\Omega$, convergence criteria $\epsilon$, maximum iteration number $K$.

**Initialization:** $X^0 = \mathbf{Z}_\Omega$, $\mathbf{Y}^0_{i_1i_2} = X^0$, $\rho^0_{i_1i_2} > 0$, $\mu > 1$.

while not converged and $k < K$ do

Updating $\mathbf{Y}^k_{i_1i_2}$ via $[39]$;

Updating $\mathbf{W}^k_{i_1i_2}$ via $[40]$;

Updating $\bar{\Lambda}^k_{i_1i_2}$ via $[41]$;

Updating $\Lambda^k_{i_1i_2}$ via $[42]$;

Updating the multipliers $Q^k_{i_1i_2}$ via $[43]$;

$\mu^k_{i_1i_2} = \mu\mu^{k-1}$, $k = k + 1$;

Check the convergence conditions $\|X^{k+1} - X^{k}\|_{\infty} \leq \epsilon$.
end while

return $X^{k+1}$.

**Output:** Completed tensor $X^{k+1}$.
where the element values of vector $\vec{e}$ are all small values close to 0, which will avoid the situation where $\Lambda$ becomes 0 and $\Gamma$ cannot be solved.

**Update $\Lambda$:** The update for $\Lambda_{l_1l_2}$ has the following closed-form solution:

$$
\Lambda_{l_1l_2}^+ = \arg \min_{\Lambda_{l_1l_2}} \frac{1}{\|\Gamma_{l_1l_2}\|_F^2} \left\| \frac{u_{l_1l_2} - \Lambda_{l_1l_2} \Gamma_{l_1l_2}}{\Gamma_{l_1l_2}} \right\|_F^2 = \frac{u_{l_1l_2}^2}{\Gamma_{l_1l_2}^2}. \quad (48)
$$

**Update $\Gamma$:** The update for $\Gamma_{l_1l_2}$ has the following closed-form solution:

$$
\Gamma_{l_1l_2}^+ = \arg \min_{\Gamma_{l_1l_2}} \frac{1}{\|\Gamma_{l_1l_2}\|_F^2} \left\| \frac{Y_{l_1l_2} - \Lambda_{l_1l_2}^+ \Gamma_{l_1l_2}}{\Gamma_{l_1l_2}} \right\|_F^2. \quad (49)
$$

Problem (49) is element-wise separable, and by proposition we have the following results:

$$
\|Y\|_{1l_1l_2} = \sum_{i=1}^{l_1} \sum_{j=1}^{l_2} 2v_{(i,j)}\sigma_f(\tilde{Y}^{(i)}) + (v_{(i,j)} - \Lambda_{(i,j)}^\Gamma_{(i,j)})^2.
$$

Then, we consider the case of one of the elements individually:

$$
\frac{2v_{(i,j)}\sigma_f(\tilde{Y}^{(i)}) + (v_{(i,j)} - \Lambda_{(i,j)}^\Gamma_{(i,j)})^2}{2\Gamma_{(i,j)}}.
$$

The closed form of $\Gamma_{(i,j)}$ can be derived by setting the derivative of (50) to zero:

$$
(\Lambda_{(i,j)}^\Gamma_{(i,j)})^2 - \frac{2v_{(i,j)}\sigma_f(\tilde{Y}^{(i)}) - u_{(i,j)}^2}{2\Gamma_{(i,j)}} = 0.
$$

So, $\Gamma_{(i,j)}$ is updated by the following:

$$
\Gamma_{(i,j)}^+ = \sqrt{\frac{2v_{(i,j)}^2\sigma_f(\tilde{Y}^{(i)}) + u_{(i,j)}^2}{\Lambda_{(i,j)}^2}}. \quad (51)
$$

**Update $X$:** The closed form of $X$ can be derived by setting the derivative of (45) to zero. We can now update $X$ by the following equation:

$$
X^+ = \mathcal{P}_\Omega \left( \frac{\sum_{i:l_1<l_2<l_{\Omega}} \rho_{l_1l_2} (Y_{l_1l_2} - \frac{Q_{l_1l_2}}{\rho_{l_1l_2}})}{\sum_{i:l_1<l_2<l_{\Omega}} \rho_{l_1l_2}} \right) + \mathcal{P}_\Omega (Z). \quad (52)
$$

**Update $Q$:** Finally, multipliers $Q_{l_1l_2}$ are updated as follows:

$$
Q_{l_1l_2}^+ = \rho_{l_1l_2} - \rho_{l_1l_2} (X - Y_{l_1l_2}). \quad (53)
$$

The optimization steps of BEMCP formulation are listed in Algorithm 3. The main per-iteration cost lies in the update of $Y_{l_1l_2}$, which requires computing t-SVD. The per-iteration complexity is $O(LE(\sum_{i:l_1<l_2<l_{\Omega}} \log (l_{l_1l_2}) + \min (l_1, l_2)))$, where $LE = \prod_{i=1}^{N} l_i$ and $l_{l_1l_2} = LE(l_1l_2)$.

**VI. EXPERIMENTS**

We evaluate the performance of the proposed LRTC methods. We employ the peak signal-to-noise rate (PSNR) value, the structural similarity (SSIM) value [40], the feature similarity (FSIM) value [41], and erreur relative globale adimensionnelle de synthèse (ERGAS) value [42] to measure the quality of the recovered results. The PSNR, SSIM, and FSIM values are the bigger the better, and the ERGAS value is the smaller the better. All tests are implemented on the Windows 10 platform and MATLAB (R2019a) with an Intel Core i7-10875H 2.30 GHz and 32 GB of RAM.

In this section, we test three kinds of real-world data: MSI, MRI, and CV. The method for testing the data is purely random sampling. The comparative LRTC methods are as follows: HaLRTC [43], LRTC-vI [44] represent state-of-the-art for the Tucker-decomposition-based methods; and TNN [45], PSTNN [46], FTNN [47], WSTNN [25] represent state-of-the-art for the t-SVD-based methods. Since the TNN, PSTNN, and FTNN methods only apply to three-order tensors, in all four-order tensor tests, we reshape them into three-order tensors and then test the performance of these methods.

**A. MSI completion**

We test 32 MSIs in the dataset CAVE [4]. All testing data are of size $256 \times 256 \times 31$. In Fig[1] we randomly select three from 32 MSIs, bringing the different sampling rates and band visible results. The individual MSI names and their corresponding bands are written in the caption of Fig.1. As shown from Fig[1], the visual effect of the BEMCP is better than the contrast method at all three sampling rates. To further highlight the superiority of our method, the average quantitative results of 32 MSIs are listed in Table[1]. It can be seen that the three methods proposed in this paper have a great improvement compared to the WSTNN method. The PSNR value at both 10% and 20% sampling rate is at least 1.5dB higher than the WSTNN method, and even reaches 5dB at 5% sampling rate. Besides, the results show that the PSNR value of the BEMCP method is 0.2db higher than that of the EMCP method at all three sampling rates. This indicates that the BEMCP method is better than the univariate EMCP method. And compared with the NMCP method that directly uses the MCP function, our improvement is more prominent. More experimental results are available in Appendix A.

1http://www.cs.columbia.edu/CAVE/databases/multispectral/
TABLE II
THE AVERAGE PSNR, SSIM, FSIM AND ERGAS VALUES FOR 32 MSIs TESTED BY OBSERVED AND THE NINE UTILIZED LRCT METHODS.

| Method  | 5% PSNR | 5% SSIM | 5% FSIM | 5% ERGAS | 10% PSNR | 10% SSIM | 10% FSIM | 10% ERGAS | 20% PSNR | 20% SSIM | 20% FSIM | 20% ERGAS | Time(s) |
|---------|---------|---------|---------|----------|---------|---------|---------|----------|---------|---------|---------|----------|---------|
| Observed | 15.438  | 0.153   | 0.644   | 845.339  | 15.673  | 0.194   | 0.646   | 822.808  | 16.185  | 0.269   | 0.651   | 775.716  | 0.000   |
| HalRTC  | 25.307  | 0.747   | 0.837   | 298.634  | 29.385  | 0.856   | 0.894   | 184.887  | 35.038  | 0.930   | 0.946   | 105.307  | 26.636  |
| TNN     | 25.350  | 0.713   | 0.817   | 289.617  | 33.114  | 0.880   | 0.918   | 127.987  | 40.201  | 0.964   | 0.972   | 58.856   | 85.578  |
| LRTC-I  | 25.886  | 0.800   | 0.835   | 276.943  | 30.725  | 0.890   | 0.906   | 162.443  | 35.516  | 0.949   | 0.957   | 94.262   | 449.471 |
| PSTNN   | 18.708  | 0.474   | 0.650   | 574.925  | 23.211  | 0.683   | 0.782   | 352.958  | 34.315  | 0.924   | 0.942   | 116.434  | 91.420  |
| FTNN    | 32.645  | 0.899   | 0.924   | 131.419  | 37.151  | 0.954   | 0.963   | 78.977   | 43.023  | 0.984   | 0.987   | 41.714   | 503.684 |
| WSTNN   | 31.431  | 0.806   | 0.911   | 208.954  | 40.143  | 0.981   | 0.981   | 33.010   | 47.049  | 0.995   | 0.995   | 24.974   | 125.812 |
| NMCP    | 37.474  | 0.962   | 0.962   | 70.020   | 42.673  | 0.987   | 0.987   | 39.156   | 48.599  | 0.996   | 0.996   | 20.403   | 177.065 |
| EMCP    | 37.602  | 0.960   | 0.961   | 69.286   | 43.507  | 0.987   | 0.987   | 35.883   | 49.668  | 0.995   | 0.996   | 17.996   | 194.907 |
| BEMCP   | 37.939  | 0.963   | 0.963   | 66.775   | 43.734  | 0.988   | 0.988   | 34.998   | 49.904  | 0.996   | 0.996   | 17.551   | 204.299 |

Fig. 1. (a) Original image. (b) Observed image. (c) HalRTC. (d) TNN. (e) LRTC-I. (f) PSTNN. (g) FTNN. (h) WSTNN. (i) NMCP. (j) EMCP. (k) BEMCP. SR: top row is 5%, middle row is 10% and last row is 20%. The rows of MSIs are in order: beads, feathers, jelly_beans. The corresponding bands in each row are: 31, 12, 20.

B. MRI completion

We test the performance of the proposed method and the comparative method on MR\textsuperscript{2} data with the size of 181×217×181. First, we demonstrate the visual effect recovered by MRI data at sampling rates of 5%, 10% and 20% in Fig.\textsuperscript{4}. Our method is clearly superior to the comparative methods. Then, we list the average quantitative results of frontal sections of MRI restored by all methods at different sampling rates in Table IV. At the sampling rate of 5% and 10%, the PSNR value of all three methods is at least 3db higher than that of the WSTNN method. When the sampling rate is 5%, the PSNR value of our method is 0.5dB higher than it. In addition, at the sampling rate of 5% and 10%, the PSNR value of all three methods is at least 3db higher than that of the WSTNN method. More experimental results are available in Appendix B.

D. Discussions

1) Model Analysis: The two parameters \(\alpha\) and \(\rho\) are included in our proposed BEMCP model. In addition, the initial values of the variables \(\nu\), \(\Lambda\), and \(\Gamma\) also strongly influence on the efficient solution of the BEMCP model. The parameter values of \(\alpha\) and \(\rho\) follow the settings in the N-tubal rank [25]. The initial value of the variable \(\nu_0\) is \(\nu_0 = \Lambda_0 \bullet \Gamma_0\). The specific data affects the initial value design of \(\Lambda\) and \(\Gamma\). The optimal initial value design can be found in Appendix C.

2) Convergence Behaviours: We take the LRCT of MSI, MRI, CV data as examples to illustrate the convergence behavior of the three algorithms under 5% sampling rate. We have drawn \(|X^{k+1} - X^k|_\infty\) for each iteration in Fig.\textsuperscript{4}. It can be seen that our algorithm converges stably and quickly.

VII. CONCLUSION

This paper proposes a new structural equivalence theorem called the BEMCP theorem, which converts two constant

\footnotesize{\textsuperscript{2}http://brainweb.bic.mni.mcgill.ca/brainweb/selection_normal.html
\textsuperscript{3}http://trace.eas.asu.edu/yuv/}
parameters in the MCP function into variables so that it can be adaptively updated with the iterative update of tensor singular values in the algorithm. Compared with other existing methods, the effect is remarkable, and it is slightly better than the same type of the NMCP and EMCP methods. On this basis, we give the BEMCP model for solving the LRCT problem. Extensive experiments show that our method can achieve better visual and numerical quantitative results than the comparison methods.

REFERENCES
[1] Y.-M. Huang, H.-Y. Yan, Y.-W. Wen, and X. Yang, “Rank minimization with applications to image noise removal,” Information Sciences, vol. 429, pp. 147–163, 2018.
[2] B. Madathil and S. N. George, “Twist tensor total variation regularized-reweighted nuclear norm based tensor completion for video missing area recovery,” Information Sciences, vol. 423, pp. 376–397, 2018.
[3] Y. Wang, D. Meng, and M. Yuan, “Sparse recovery: from vectors to tensors,” National Science Review, vol. 5, no. 5, pp. 756–767, 2018.
[4] X.-L. Zhao, W.-H. Xu, T.-X. Jiang, Y. Wang, and M. K. Ng, “Deep plug-and-play prior for low-rank tensor completion,” Neurocomputing, vol. 400, pp. 137–149, 2020.
[5] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, “Fusing Hyperspectral and Multispectral Images via Coupled Sparse Tensor Factorization,” IEEE Transactions on Image Processing, vol. 27, no. 8, pp. 4118–4130, 2018.
[6] X. Fu, W.-K. Ma, J. M. Bioucas-Dias, and T.-H. Chan, “Semiblind Hyperspectral Unmixing in the Presence of Spectral Library Mismatches,” IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 9, 2016.
Table IV
The average PSNR, SSIM, FSIM and ERGAS values for 7 CVs tested by observed and the nine utilized LRTC methods.

| Method     | SR | PSNR  | SSIM | FSIM | ERGAS | PSNR  | SSIM | FSIM | ERGAS | PSNR  | SSIM | FSIM | ERGAS | Time(s) |
|------------|----|-------|------|------|-------|-------|------|------|-------|-------|------|------|-------|---------|
| Observed   | 5% | 5.793 | 0.011| 0.420| 1794.940| 6.028 | 0.019| 0.423| 1163.038| 6.540 | 0.034| 0.429| 1096.477| 10.000 |
| HilRTC     | 5% | 27.336| 0.548| 0.695| 297.173 | 21.141| 0.622| 0.774| 214.642 | 24.981| 0.772 | 0.862| 137.634 | 13.780  |
| TNN        | 5% | 23.033| 0.771| 0.886| 113.454 | 30.453| 0.855| 0.928 | 79.511 | 33.697| 0.910 | 0.955| 56.639 | 44.979  |
| LRTCtv-1   | 5% | 19.497| 0.579| 0.692| 272.767 | 21.205| 0.655| 0.771| 228.491 | 25.812| 0.817 | 0.881| 126.740| 281.989 |
| PSTNN      | 5% | 16.151| 0.312| 0.664| 364.981 | 27.897| 0.778| 0.890| 102.835 | 33.258| 0.906 | 0.952| 58.772 | 44.823  |
| FTNN       | 5% | 25.286| 0.766| 0.871| 113.494 | 28.544| 0.858| 0.917| 93.393  | 32.214| 0.924 | 0.954| 61.414  | 373.005 |
| WSTNN      | 5% | 29.257| 0.872| 0.920| 88.184  | 32.635| 0.924| 0.952| 62.072  | 36.557| 0.960 | 0.975| 40.820  | 197.168 |
| NMCP       | 5% | 30.432| 0.885| 0.933| 77.007  | 33.934| 0.933| 0.961| 52.983  | 37.359| 0.963 | 0.980| 35.450  | 211.281 |
| EMCP       | 5% | 30.768| 0.887| 0.937| 74.465  | 34.512| 0.934| 0.964| 50.102  | 38.204| 0.964 | 0.982| 32.399  | 226.812 |
| BEMCP      | 5% | 31.284| 0.893| 0.942| 70.978  | 34.683| 0.935| 0.965| 49.164  | 38.244| 0.964 | 0.982| 32.285  | 247.195 |

Fig. 4. The convergence behaviours of LRTC algorithm, with respect to MSI, MRI, CV data.
solutions of linear matrix equations via nuclear norm minimization,” *SIAM Review*, vol. 52, no. 3, pp. 471–501, 2010.

[30] D. Qiu, M. Bai, M. K. Ng, and X. Zhang, “Nonlocal robust tensor recovery with nonconvex regularization,” *Inverse Problems*, vol. 37, no. 3, p. 035001, 2021.

[31] C. Lu, C. Zhu, C. Xu, S. Yan, and Z. Lin, “Generalized singular value thresholding,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 29, no. 1, 2015.

[32] C.-H. Zhang, “Nearly unbiased variable selection under minimax concave penalty,” *The Annals of statistics*, vol. 38, no. 2, pp. 894–942, 2010.

[33] P. K. Pokala, R. V. Hemadri, and C. S. Seelamantula, “Iteratively Reweighted Minimax-Concave Penalty Minimization for Accurate Low-rank Plus Sparse Matrix Decomposition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[34] S. Boyd, N. Parikh, and E. Chu, *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011.

[35] Z. Lin, R. Liu, and Z. Su, “Linearized Alternating Direction Method with Adaptive Penalty for Low-Rank Representation,” *Advances in Neural Information Processing Systems*, vol. 24, pp. 612–620, 2011.

[36] M. E. Kilmer and C. D. Martin, “Factorization strategies for third-order tensors,” *Linear Algebra and its Applications*, vol. 435, no. 3, pp. 641–658, 2011.

[37] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, “Tensor Robust Principal Component Analysis with a New Tensor Nuclear Norm,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 4, pp. 925–938, 2020.

[38] L. Mirsky, “A trace inequality of John von Neumann,” *Monatshefte für mathematik*, vol. 79, no. 4, pp. 303–306, 1975.

[39] Z. Lin, M. Chen, and Y. Ma, “The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices,” *arXiv preprint arXiv:1009.5055*, 2010.

[40] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[41] L. Wald, *Data fusion: definitions and architectures: fusion of images of different spatial resolutions*. Presses des MINES, 2002.

[42] J. Liu, P. Musialski, P. Wonka, and J. Ye, “Tensor Completion for Estimating Missing Values in Visual Data,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 208–220, 2013.

[43] X. Li, Y. Ye, and X. Xu, “Low-rank tensor completion with total variation for visual data inpainting,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, 2017.

[44] Z. Zhang and S. Aeron, “Exact Tensor Completion Using t-SVD,” *IEEE Transactions on Signal Processing*, vol. 65, no. 6, pp. 1511–1526, 2017.

[45] T.-X. Jiang, T.-Z. Huang, X.-L. Zhao, and L.-J. Deng, “Multi-dimensional imaging data recovery via minimizing the partial sum of tubal nuclear norm,” *Journal of Computational and Applied Mathematics*, vol. 372, p. 112680, 2020.

[46] T.-X. Jiang, M. K. Ng, X.-L. Zhao, and T.-Z. Huang, “Framelet Representation of Tensor Nuclear Norm for Third-Order Tensor Completion,” *IEEE Transactions on Image Processing*, vol. 29, pp. 7233–7244, 2020.