SIAM J. IMAGING SCIENCES
Vol. 18, No. 1, pp. 20–59

Bilateral Tensor Low-Rank Representation for Insufficient Observed Samples in Multidimensional Image Clustering and Recovery*

Meng Ding†, Xi-Le Zhao‡, Jing-Hua Yang§, Zhengchun Zhou¶, and Michael K. Ng‖

Abstract. In this work, we study the subspace clustering and recovery of multidimensional images. Existing matrix-based/tensor-based subspace clustering methods successfully consider unilateral information (i.e., the similarity between image samples) to cluster samples into subspaces by using low-rank representation. The key issue of the unilateral representation-based methods is that the number of samples in each subspace should be sufficient for subspace representation. In practice, the clustering performance can be degraded when there is only a small number of observed samples in each subspace. To address the problem of insufficient observed samples, we propose to introduce hidden tensor data to supplement an insufficient number of observed samples. We employ both observed samples and hidden tensor data under low-rank constraints so that a new bilateral tensor low-rank representation (BTLRR) in subspace clustering is formulated. We show that a closed-form solution of block-diagonal tensor structure is obtained in subspace clustering of observed samples and hidden tensor data. Also the proposed BTLRR optimization problem can be solved by using the convex relaxation technique and augmented Lagrangian multiplier algorithm. The proposed BTLRR can fully explore the bilateral information of observations, including not only the similarity between samples but also the relationship among features. Extensive numerical results on multidimensional image data clustering and recovery illustrate that the effectiveness and robustness of the proposed bilateral representation are better than those of state-of-the-art methods (e.g., the popular LRR and TLRR methods).

Key words. bilateral tensor low-rank representation, multidimensional image, subspace clustering, data recovery, exact recovery

MSC codes. 62H30, 62H35, 68U10

DOI. 10.1137/24M1655093

*Received by the editors April 17, 2024; accepted for publication (in revised form) September 16, 2024; published electronically January 3, 2025.
https://doi.org/10.1137/24M1655093

Funding: This work was supported by NSFC (12201522, 12371456, 12171072, 62131005, 12401605), Natural Science Foundation of Sichuan Province of China (2024NSFTD0015, 2024NSFSC1389, 2024NSFSC1467), Fundamental Research Funds for the Central Universities (2682023CX069, 2682024CX017), Sichuan Science and Technology Program (2024NSFJQ0038, 2023ZYD0007, 2024NSFSC0038), National Key Research and Development Program of China (2020YFA0714001), Postdoctoral Fellowship Program of CPSF (GZC20232198), China Postdoctoral Science Foundation (2024M752661), HKRGC GRF (17201020, 17300021), HKRGC CRF (C7004-21GF), and Joint NSFC and RGC (NHKU769/21).

†School of Mathematics, Southwest Jiaotong University, Chengdu, China (dingmeng56@163.com).
‡Corresponding author. School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, China (xlzhao122003@163.com).
§School of Information Science and Technology, Southwest Jiaotong University, Chengdu, China (yangjinghua110@126.com).
¶School of Information Science and Technology, Southwest Jiaotong University, Chengdu, China (zzc@swjtu.edu.cn).
‖Corresponding author. Department of Mathematics, Hong Kong Baptist University, Hong Kong (michael-ng@hkbu.edu.hk).

1. Introduction. The problem of simultaneous data recovery and clustering has been developed in many applications of image processing [4, 6, 17, 18, 23, 33, 45, 48, 52, 56, 61], e.g., image/video denoising [15, 30, 34, 53, 54, 62, 59], face recognition [51, 43], and data segmentation [24, 35, 38, 39, 47, 64]. Matrix-based methods are well studied in image data recovery and subspace clustering. Robust principal component analysis (RPCA) [3] recovered the original low-rank matrix component from the sparsely corrupted observation. Low-rank representation (LRR) [29] and sparse subspace clustering (SSC) [9] were studied by vectorizing multidimensional image data and clustering sample vectors into subspaces under predefined priors like low rankness [29] and sparsity [9]. In the literature, their variants are developed and studied; see, for example, [1, 30, 55]. On the other hand, deep learning has been successful in many image processing applications [10, 16, 37, 42].

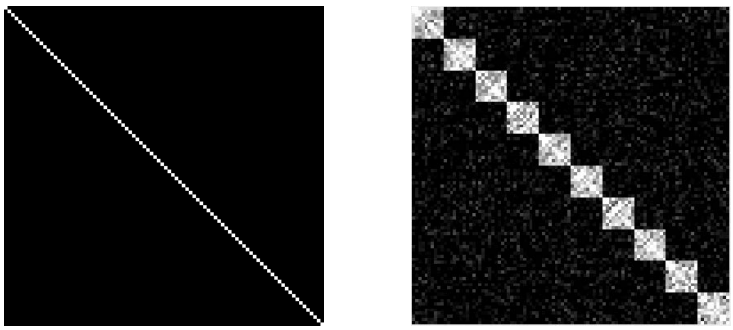
Besides the vectorization operator being employed to reshape multidimensional samples into a list of vectors, the work [11] integrated Tucker decomposition [44] and sparse coding to simultaneously seek the low-rank spatial (two-dimensional) structures over samples and the sparse representation in the feature space for data clustering. To characterize the intrinsic structure of multidimensional data, a tensor LRR (TLRR) [62] employed a tensor linear representation for capturing the multilinear relations among image samples. More precisely, TLRR aims at finding a low-rank tensor representation in a given tensor dictionary as follows:

$$\min_{\mathcal{Z}} \|\mathcal{Z}\|_* \text{ s.t. } \mathcal{X}_O = \mathcal{A} * \mathcal{Z}, \tag{1.1}$$

where $\mathcal{X}_O \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ are the observed data (each lateral slice, i.e., $\mathcal{X}_O(:, j, :)$, is an observed sample), $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ denotes the representation base, $\mathcal{Z} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ denotes the representation coefficient, $\|\mathcal{Z}\|_*$ represents the tensor nuclear norm (see Definitiion 2.10), and $*$ denotes the tensor-tensor product (see Definition 2.1). Zhou et al. [62] showed that the minimizer to (1.1) has the tensor block-diagonal structure formed by nonzero tubes (see the structure of $\mathcal{Z}$ in Figure 2(b)), which indicates the clustering membership of the samples. Usually, TLRR methods [8, 62] set the dictionary $\mathcal{A}$ as the observed tensor data (i.e., $\mathcal{A} = \mathcal{X}_O$) or the denoised result.

In the above matrix-based/tensor-based methods for subspace clustering, one fundamental issue is that the number of samples from each subspace should be sufficient for subspace representation. For example, let $\mathcal{X}_O$ be a set of samples drawn from a union of P tensor subspaces (see Definition 2.11) $\{\mathcal{S}_p\}_{p=1}^P$, each $\mathcal{S}_p$ has a tensor tubal rank (see Definition 2.9) of r_p. Suppose that $\mathcal{X}_O = [\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_P]$, and $\mathcal{X}_p$ contains n_p samples drawn from the pth tensor subspace $\mathcal{S}_p$. Then the sampling of $\mathcal{X}_O$ is sufficient if and only if $\text{rank}(\mathcal{X}_p) = r_p$, $p = 1, \ldots, P$. It is clear that problem (1.1) admits a trivial solution $\mathcal{Z}^\star = \mathcal{I}$ ($\mathcal{I}$ is the identity tensor (see Definition 2.3)) if $\text{rank}(\mathcal{X}_p) < r_p$, $p = 1, \ldots, P$. To illustrate this issue, we show an example in Figure 1. We demonstrate that when the observed tensor data are insufficient, $\mathcal{Z} = \mathcal{I}$ may only be a feasible solution to problem (1.1); see Figure 1(a). In the figure, we construct all frontal slices $(\mathcal{Z}^\star)^{(k)}$ from $\mathcal{Z}$ to construct the affinity matrix $\boldsymbol{Z}$ as follows,

$$\boldsymbol{Z} = \frac{1}{2n_3} \sum_{k=1}^{n_3} \left(\left| (\mathcal{Z}^\star)^{(k)} \right| + \left| (\mathcal{Z}^\star)^{(k)} \right|^\top \right), \tag{1.2}$$



(a) TLRR (b) BTLRR

Figure 1. *Illustrated example under insufficient samples:* (a) *and* (b) *are the affinity matrices obtained by performing* (1.2) *on solutions to unilateral representation* (1.1) *and our bilateral representation* (1.4)*, respectively. Here we generate an observed tensor as* $\mathcal{X}_O = [\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_{10}]$ *with* $\mathcal{X}_p = \mathcal{D}_p * \mathcal{T}_p$ $(p = 1, 2, \ldots, 10)$*, where the entries of* $\mathcal{D}_p \in \mathbb{R}^{100\times 15\times 30}$ *and* $\mathcal{T}_p \in \mathbb{R}^{15\times 10\times 30}$ *are randomly sampled from independent and identically distributed (i.i.d.)* $\mathcal{N}(0,1)$*. We randomly add the sparse noise with noise ratio* 0.05*. The number of samples* $(s = 10)$ *at each subspace is less than the rank* $(r = 15)$ *of each subspace, i.e., the data samples are insufficient.*

where $|(\mathcal{Z}^\star)^{(k)}|$ is a matrix where its entries take the absolute value of the entry of $(\mathcal{Z}^\star)^{(k)}$ and $\cdot^\top$ denotes the matrix transpose operation.

Note that TLRR only considers the unilateral similarity between samples, thereby its clustering performance can be degraded when the observed samples are insufficient and/or severely corrupted. This problem has been studied in the work [30], latent LRR (LatLRR), for the insufficient matrix data. However, LatLRR is limited to the matrix data. When handling the multidimensional image data or tensor (e.g., face data, videos, and multispectral images), LatLRR has to flatten the tensor data into a matrix. Such a preprocessing unavoidably destroys the multidimensional structure and subsequently leads to the performance degradation. Moreover, LatLRR represents the data entries using few other entries, which cannot fully exploit the complex relationships between tensor samples; please see the details in section 3.3.

To address the problem of insufficient sampling in the tensor case, we introduce the hidden tensor data to supplement the insufficient number of observed samples. We employ both observed and hidden tensor data and consider the following problem:

$$\begin{aligned} &\min_{\mathcal{Z}_{O,H}} \|\mathcal{Z}_{O,H}\|_* \\ &\text{s.t. } \mathcal{X}_O = [\mathcal{X}_O, \mathcal{X}_H] * \mathcal{Z}_{O,H}, \end{aligned} \tag{1.3}$$

where $\mathcal{X}_H \in \mathbb{R}^{n_1\times n_4\times n_3}$ denotes the hidden (unobserved) tensor data, and the concatenation (along the column dimension) of $\mathcal{X}_O$ and $\mathcal{X}_H$, denoted as $[\mathcal{X}_O, \mathcal{X}_H]$ (see section 2.1), is used as the dictionary. We note that $[\mathcal{X}_O, \mathcal{X}_H]$ can be sufficient to represent the underlying subspaces. Let $\mathcal{Z}^\star_{O,H} = [\mathcal{Z}^\star_O; \mathcal{Z}^\star_H]$ be the optimal solution to (1.3) such that $\mathcal{Z}^\star_O$ and $\mathcal{Z}^\star_H$, respectively, correspond to $\mathcal{X}_O$ and $\mathcal{X}_H$, where $[\mathcal{Z}^\star_O; \mathcal{Z}^\star_H]$ denotes the concatenation of $\mathcal{Z}^\star_O$ and $\mathcal{Z}^\star_H$ along the first dimension (see section 2.1), then $\mathcal{Z}^\star_O$ has a nontrivial tensor block-diagonal structure that exactly indicates the tensor subspace membership of each observed sample. Given the observed tensor data $\mathcal{X}_O$, recovering $\mathcal{Z}^\star_O$ in problem (1.3) is generally not practical due to the

absence of $\mathcal{X}_H$. Based on Theorem 3.1 (see section 3), we can equivalently reformulate the constraint in (1.3) as the following bilateral tensor representation relationship

$$\mathcal{X}_O = \mathcal{X}_O * \mathcal{Z}_O^\star + \mathcal{L}_H^\star * \mathcal{X}_O,$$

where $\mathcal{Z}_O^\star$ and $\mathcal{L}_H^\star$ are unknown variables related to $\mathcal{X}_O$ and $\mathcal{X}_H$. Assume $\mathcal{X}_O$ and $\mathcal{X}_H$ are sampled from the same collection of low-rank tensor subspaces, one can derive that both $\mathcal{Z}_O^\star$ and $\mathcal{L}_H^\star$ should be low rank (the detailed proof is shown in section 3). Therefore, we propose a new convex bilateral TLRR (BTLRR) for subspace clustering as follows:

$$\begin{aligned} \min_{\mathcal{Z}_O, \mathcal{L}_H} \ & \|\mathcal{Z}_O\|_* + \|\mathcal{L}_H\|_* \\ \text{s.t. } & \mathcal{X}_O = \mathcal{X}_O * \mathcal{Z}_O + \mathcal{L}_H * \mathcal{X}_O. \end{aligned} \tag{1.4}$$

In our new BTLRR, we compute $\mathcal{Z}_O$ and $\mathcal{L}_H$ from the observed $\mathcal{X}_O$ without using the hidden $\mathcal{X}_H$. The detailed explanation can be found in section 3. The proposed BTLRR method can explore the bilateral information within the observations, i.e., the subspace membership of observed samples and the relationship among sample features. Our proposed BTLRR approach provides the estimated block-diagonal affinity matrices, as shown in Figure 1(b), that are capable of learning the block-diagonal structure, which is close to that identified by the sufficient sampling.

In this paper, we study the problem of recovering the representation tensor $\mathcal{Z}_O^\star$ by using the observed tensor data $\mathcal{X}_O$ only. The main contributions of this paper are given as follows.

- We utilize both observed samples and hidden tensor data and propose the bilateral tensor low-rank representation to resolve the problem of insufficient observed samples of existing unilateral representation-based methods. The proposed BTLRR can fully explore the subspace membership of samples and the relationship among features.
- We provide the theoretical guarantee of the proposed BTLRR. For the noise-free observed tensor data, we prove that the noiseless BTLRR admits the closed-form solutions; meanwhile, for the tensor data with sparse corruptions, we can find globally optimal solutions of our BTLRR by establishing the connection between BTLRR and tensor RPCA [31] under certain conditions.
- We give an analysis of exploring the effects of the hidden tensor data. Under the low-rank assumption, we show that $\mathcal{X}_H$ and the optimal minimizer to (1.3) can be approximately recovered by solving a convex tensor nuclear norm minimization problem.
- We employ an efficient algorithm based on the augmented Lagrangian multiplier method to solve the resulting optimization model. Extensive experiments including subspace clustering and data recovery based on various datasets demonstrate the outstanding performance of our method compared to state-of-the-art methods.

The rest of this paper is organized as follows. In section 2, we present the notations and preliminaries used in this work. Also we review the related works, including matrix- and tensor-based subspace clustering methods. In section 3, we study the effect of hidden tensor data and formulate the proposed BTLRR model. We also provide the exact recoverability analysis of the proposed model. In section 4, we present the proposed optimization algorithm.

In section 5, we perform various experiments to demonstrate the effectiveness of the proposed method. In section 6, some concluding remarks are given.

2. Notations and preliminaries.

2.1. Notations. We denote the scalar, the vector, the matrix, and the tensor by the symbols x (or X), $\boldsymbol{x}$, $\boldsymbol{X}$, and $\mathcal{X}$, respectively. Given a third-order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, its (i,j,k)th entry is denoted as $[\mathcal{X}]_{i,j,k}$, and its ith horizontal, jth lateral, and kth frontal slice are denoted as $\mathcal{X}(i,:,:)$, $\mathcal{X}(:,j,:)$, and $\mathcal{X}(:,:,k)$ (or $\mathcal{X}^{(k)}$), respectively. The (i,j)th entry of a matrix $\boldsymbol{X} \in \mathbb{R}^{n_1 \times n_2}$ is denoted as $[\boldsymbol{X}]_{i,j}$. The inner product of $\mathcal{X}$ and $\mathcal{Y}$ in $\mathbb{R}^{n_1 \times n_2 \times n_3}$ is defined as $\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{k=1}^{n_3} \langle \mathcal{X}^{(k)}, \mathcal{Y}^{(k)} \rangle$. The trace of a matrix $\boldsymbol{X}$ is denoted as $\mathrm{tr}(\boldsymbol{X})$. For tensors $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{B} \in \mathbb{R}^{n_4 \times n_2 \times n_3}$, the concatenation of $\mathcal{A}$ and $\mathcal{B}$ along the first dimension is a tensor $\mathcal{C} = [\mathcal{A}; \mathcal{B}] = [\begin{smallmatrix} \mathcal{A} \\ \mathcal{B} \end{smallmatrix}] \in \mathbb{R}^{(n_1+n_4) \times n_2 \times n_3}$, where $\mathcal{C}(1:n_1,:,:) = \mathcal{A}$ and $\mathcal{C}(n_1+1:n_1+n_4,:,:) = \mathcal{B}$. Given two tensors $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_4 \times n_3}$, the concatenation of $\mathcal{X}$ and $\mathcal{Y}$ along the second dimension is a tensor $\mathcal{Z} = [\mathcal{X}, \mathcal{Y}]$ of size $n_1 \times (n_2 + n_4) \times n_3$, where $\mathcal{Z}(:, 1:n_2, :) = \mathcal{X}$ and $\mathcal{Z}(:, n_2+1:n_2+n_4, :) = \mathcal{Y}$. We denote the Frobenius norm as $\|\mathcal{X}\|_F = \sqrt{\sum_{i,j,k} [\mathcal{X}]_{i,j,k}^2}$, the ℓ_1 norm as $\|\mathcal{X}\|_1 = \sum_{i,j,k} |[\mathcal{X}]_{i,j,k}|$, and the ℓ_∞ norm as $\|\mathcal{X}\|_\infty = \max_{i,j,k} |[\mathcal{X}]_{i,j,k}|$. The matrix nuclear norm is defined as the sum of the singular values, i.e., $\|\boldsymbol{X}\|_* = \sum \sigma(\boldsymbol{X})$.

2.2. Preliminaries. For $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $\bar{\mathcal{X}}$ denotes the discrete Fourier transform (DFT) along the third dimension of $\mathcal{X}$. Defining the DFT matrix $\boldsymbol{F}_{n_3} = [\boldsymbol{f}_1, \ldots, \boldsymbol{f}_k, \ldots, \boldsymbol{f}_{n_3}] \in \mathbb{C}^{n_3 \times n_3}$, where $\boldsymbol{f}_k = [\omega^{0 \times (k-1)}; \omega^{1 \times (k-1)}; \ldots; \omega^{(n_3-1) \times (k-1)}] \in \mathbb{C}^{n_3 \times 1}$ with $\omega = e^{-(2\pi\sqrt{-1}/n_3)}$, we obtain each tube $\bar{\mathcal{X}}(i,j,:) = \boldsymbol{F}_{n_3} \mathcal{X}(i,j,:)$. The block circular matrix $\mathtt{bcirc}(\mathcal{X})$ with size $n_1 n_3 \times n_2 n_3$ of $\mathcal{X}$ is defined as

$$\mathtt{bcirc}(\mathcal{X}) = \begin{bmatrix} \mathcal{X}^{(1)} & \mathcal{X}^{(n_3)} & \ldots & \mathcal{X}^{(2)} \\ \mathcal{X}^{(2)} & \mathcal{X}^{(1)} & \ldots & \mathcal{X}^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{X}^{(n_3)} & \mathcal{X}^{(n_3-1)} & \ldots & \mathcal{X}^{(1)} \end{bmatrix}.$$

Two operators $\mathtt{unfold}$ and $\mathtt{fold}$ are denoted as

$$\mathtt{unfold}(\mathcal{X}) = \begin{bmatrix} \mathcal{X}^{(1)} \\ \mathcal{X}^{(2)} \\ \vdots \\ \mathcal{X}^{(n_3)} \end{bmatrix} \in \mathbb{R}^{n_1 n_3 \times n_2}, \ \mathtt{fold}(\mathtt{unfold}(\mathcal{X})) = \mathcal{X}.$$

We also give the definition of $\bar{\boldsymbol{X}} \in \mathbb{C}^{n_1 n_3 \times n_2 n_3}$ as follows,

$$\bar{\boldsymbol{X}} = \mathtt{bdiag}(\bar{\mathcal{X}}) = \begin{bmatrix} \bar{\mathcal{X}}^{(1)} & & & \\ & \bar{\mathcal{X}}^{(2)} & & \\ & & \ddots & \\ & & & \bar{\mathcal{X}}^{(n_3)} \end{bmatrix},$$

where $\mathtt{bdiag}(\cdot)$ reshapes the tensor $\bar{\mathcal{X}}$ to a block-diagonal matrix.

Based on the above definitions, we define the tensor-tensor product, tensor singular value decomposition (TSVD), and tensor nuclear norm (TNN) as follows.

Definition 2.1 (tensor-tensor product [21]). *The* tensor-tensor product *of* $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ *and* $\mathcal{B} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$ *is defined as a tensor of size* $n_1 \times n_4 \times n_3$*:*

$$\mathcal{A} * \mathcal{B} = \texttt{fold}(\texttt{bcirc}(\mathcal{A}) \times \texttt{unfold}(\mathcal{B})).$$

Definition 2.2 (conjugate transpose [21, 31]). *For a tensor* $\mathcal{X}$ *of size* $n_1 \times n_2 \times n_3$*, its conjugate transpose* $\mathcal{X}^\top$ *of size* $n_2 \times n_1 \times n_3$ *is obtained by conjugate transposing each frontal slice and then reversing the order of transposed frontal slices* 2 *to* n_3*.*

Definition 2.3 (identity tensor [21]). *An identity tensor* $\mathcal{I} \in \mathbb{R}^{n \times n \times n_3}$ *is given by setting the first frontal slice to be the* $n \times n$ *identify matrix and other frontal slices to be zeros.*

Definition 2.4 (tensor inverse [20, 60]). *The inverse of a tensor* $\mathcal{A} \in \mathbb{R}^{n \times n \times n_3}$ *is denoted as* $\mathcal{A}^{-1}$ *satisfying*

$$\mathcal{A} * \mathcal{A}^{-1} = \mathcal{A}^{-1} * \mathcal{A} = \mathcal{I},$$

where $\mathcal{I}$ *is the identity tensor of size* $n \times n \times n_3$*.*

Definition 2.5 (tensor pseudoinverse [62]). *For a tensor* $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$*, its pseudoinverse is defined as a tensor* $\mathcal{X}^\dagger \in \mathbb{R}^{n_2 \times n_1 \times n_3}$*, which satisfies* $\mathcal{X} * \mathcal{X}^\dagger * \mathcal{X} = \mathcal{X}$*,* $\mathcal{X}^\dagger * \mathcal{X} * \mathcal{X}^\dagger = \mathcal{X}^\dagger$*,* $\mathcal{X} * \mathcal{X}^\dagger = (\mathcal{X} * \mathcal{X}^\dagger)^\top$*, and* $\mathcal{X}^\dagger * \mathcal{X} = (\mathcal{X}^\dagger * \mathcal{X})^\top$*.*

Definition 2.6 (orthogonal tensor [21, 31]). *A tensor* $\mathcal{Q} \in \mathbb{R}^{n \times n \times n_3}$ *is orthogonal if* $\mathcal{Q}^\top * \mathcal{Q} = \mathcal{Q} * \mathcal{Q}^\top = \mathcal{I}$*.*

Definition 2.7 (F-diagonal tensor [21]). *A tensor is F-diagonal if each of its frontal slices is a diagonal matrix.*

Based on the above definitions, we can define TSVD for a tensor $\mathcal{X}$ as follows.

Definition 2.8 (TSVD and skinny TSVD [21, 31]). *Given a third-order tensor* $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$*, its TSVD is*

$$\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top,$$

where $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ *and* $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ *are orthogonal tensors, and* $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ *is an F-diagonal tensor. Then the skinny TSVD of* $\mathcal{X}$ *is* $\mathcal{X} = \mathcal{U}_\mathcal{X} * \mathcal{S}_\mathcal{X} * \mathcal{V}_\mathcal{X}^\top$*, where* $\mathcal{U}_\mathcal{X} = \mathcal{U}(:, 1:r, :)$*,* $\mathcal{S}_\mathcal{X} = \mathcal{S}(1:r, 1:r, :)$*, and* $\mathcal{V}_\mathcal{X} = \mathcal{V}(:, 1:r, :)$ *in which* r *denotes the tensor tubal rank of* $\mathcal{X}$ *(see Definition 2.9).*

Note that there is one definition of tensor rank based on TSVD similar to matrix rank.

Definition 2.9 (tensor tubal rank [20, 31]). *For any* $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$*, the* tensor tubal rank *is defined as the number of nonzero singular tubes of* $\mathcal{S}$ *in which the TSVD of* $\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$*, i.e.,* $\operatorname{rank}(\mathcal{X}) = \#\{i : \mathcal{S}(i,i,:) \neq 0\}$.

Given two tensors $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{B} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$, $\operatorname{rank}(\mathcal{A} * \mathcal{B}) \leq \min\{\operatorname{rank}(\mathcal{A}), \operatorname{rank}(\mathcal{B})\}$ holds [63].

Then we give the definition of TNN used in our work.

Definition 2.10 (TNN [31]). *Given a tensor* $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, *its* tensor nuclear norm *is defined as*

$$\|\mathcal{X}\|_* = \frac{1}{n_3} \sum_{k=1}^{n_3} \|\bar{\mathcal{X}}^{(k)}\|_*,$$

where $\bar{\mathcal{X}}^{(k)}$ *is the* k*th frontal slice of* $\bar{\mathcal{X}}$.

A *tensor space* is a set of tensors that are closed under tensor addition and scalar multiplication. We refer the tensor space as the set $\mathbb{K} = \{\forall \mathcal{K} \in \mathbb{R}^{n_1 \times 1 \times n_3}\}$. Given a set of tensors $\{\mathcal{D}_{(1)}, \ldots, \mathcal{D}_{(P)}\}$, each of $\mathcal{D}_{(p)}$ is the pth lateral slice of $\mathcal{D} \in \mathbb{R}^{n_1 \times P \times n_3}$. Then $\{\mathcal{D}_{(p)}\}_{p=1}^{P}$ is *linearly independent* if there is not a nonzero $\mathcal{T} \in \mathbb{R}^{P \times 1 \times n_3}$ satisfying $\mathcal{D} * \mathcal{T} = \mathbf{0}$.

Definition 2.11 (tensor subspace [62]). *Given a set* $\{\mathcal{D}_{(1)}, \ldots, \mathcal{D}_{(P)}\} \subseteq \mathbb{K}$, *the elements* $\mathcal{D}_{(p)}$ *are linearly independent. Then the set* $\Xi = \{\mathcal{M} | \mathcal{M} = \mathcal{D} * \mathcal{T}, \forall \mathcal{T} \in \mathbb{R}^{P \times 1 \times n_3}\}$ *is called a tensor subspace of dimension* $\dim(\mathcal{D}) = P$. *Here* $\{\mathcal{D}_{(1)}, \ldots, \mathcal{D}_{(P)}\}$ *is the basis of* Ξ.

2.3. Related works.

2.3.1. Matrix-based methods. The matrix-type method aims at finding the linear representation of all vector-valued samples and then clusters samples into corresponding subspaces. To be specific, given a two-dimensional (2D) dataset $\{\boldsymbol{X}_j\}_{j=1}^{n_2}$, where $\boldsymbol{X}_j \in \mathbb{R}^{n_1 \times n_3}$ and n_2 is the total number of samples, the matrix methods usually reshape the 2D sample $\boldsymbol{X}_j$ into a vector $\boldsymbol{x}_j \in \mathbb{R}^{n_1 n_3}$ and then forms a matrix $\boldsymbol{X} \in \mathbb{R}^{n_1 n_3 \times n_2}$.

Candès et al. [3] proposed RPCA to estimate the clean 2D data from the observed $\boldsymbol{X}$. However, RPCA assumed that the data are collected from a single subspace. The works in [9, 29] sought the linear representation (i.e., $\boldsymbol{Z}$) w.r.t a given dictionary $\boldsymbol{A}$ by solving the following formulation:

$$\begin{aligned} \min_{\boldsymbol{Z},\boldsymbol{E}} \ & R_1(\boldsymbol{Z}) + \lambda R_2(\boldsymbol{E}) \\ \text{s.t.} \ & \boldsymbol{X} = \boldsymbol{A}\boldsymbol{Z} + \boldsymbol{E}, \end{aligned}$$

where each column of $\boldsymbol{X}$ is obtained by $\boldsymbol{X}(:, j) = \boldsymbol{x}_j$, $\boldsymbol{E}$ is the noise, $R_1(\boldsymbol{Z})$ and $R_2(\boldsymbol{E})$ denote the regularization terms of the representation matrix $\boldsymbol{Z}$ and the noise $\boldsymbol{E}$, respectively (see Figure 2(a) for an illustration). A variety of regularization terms have been used for characterizing the priors of $\boldsymbol{Z}$ and $\boldsymbol{E}$. For example, SSC [9] used the ℓ_1-norm $\|\boldsymbol{Z}\|_1$ and $\|\boldsymbol{E}\|_1$ in favor of the sparse representation and noise. LRR [29] employed the nuclear norm $\|\boldsymbol{Z}\|_*$ to find the low-rank representation matrix, and the $\ell_{2,1}$-norm is used on the noise to improve the robustness to outliers. In order to resolve the insufficient samples problem of LRR, LatLRR [30] introduced the hidden data and proposed the following optimization problem:

$$\begin{aligned} \min_{\boldsymbol{Z},\boldsymbol{L},\boldsymbol{E}} \ & R_1(\boldsymbol{Z}) + R_2(\boldsymbol{L}) + \lambda R_3(\boldsymbol{E}) \\ \text{s.t.} \ & \boldsymbol{X} = \boldsymbol{X}\boldsymbol{Z} + \boldsymbol{L}\boldsymbol{X} + \boldsymbol{E}. \end{aligned}$$

LatLRR employed nuclear norm regularizations $\|\boldsymbol{Z}\|_*$ and $\|\boldsymbol{L}\|_*$ and the sparse regularization $\|\boldsymbol{E}\|_1$.

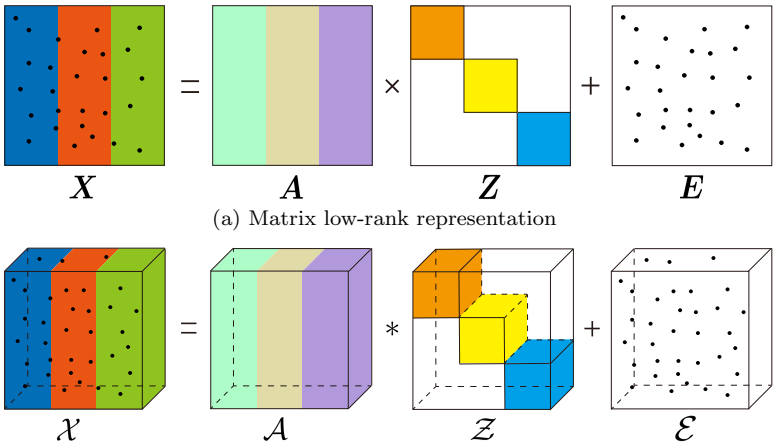


Figure 2. *Illustration of matrix LRR (top) and TLRRs (bottom).*

In fact, SSC, LRR, and their variants [30, 55, 58] are matrix-type methods, which can only process the matrix data. When handling the high-dimensional data (i.e., tensor data), these methods have to first reshape the tensor data into a matrix and destroy the high-dimensional structure of tensor data, resulting in the limited clustering performance. Therefore, the tensor-based subspace clustering methods are proposed to keep the multidimensional structure.

2.3.2. Tensor-based methods. Differently from the matrix-based methods, the tensor-type methods arrange all samples into a tensor without destroying the tensor data structure. In particular, considering the 2D samples $\{\boldsymbol{X}_j\}_{j=1}^{n_2}$, the tensor methods formulate a third-order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ whose jth lateral slice is such that $\mathcal{X}(:,j,:) = \boldsymbol{X}_j$.

Existing tensor-based methods employ tensor decomposition to find the low-rank correlation hidden in the tensor data. For instance, Fu et al. [11] used Tucker decomposition on the tensor data to explore the low-rank spatial correlations among samples for data clustering. However, the work [11] still reshaped the 2D samples into vectors. Recently, based on TSVD, Lu et al. [31] studied the tensor RPCA (TRPCA) problem that extends RPCA from the 2D matrix to the three-dimensional tensor data and proposed the following optimization problem,

$$\begin{aligned} &\min_{\hat{\mathcal{X}},\mathcal{E}} \operatorname{rank}(\hat{\mathcal{X}}) + \lambda\|\mathcal{E}\|_1 \\ &\text{s.t. } \mathcal{X} = \hat{\mathcal{X}} + \mathcal{E}, \end{aligned} \tag{2.1}$$

where $\hat{\mathcal{X}}$ represents the underlying clean tensor data, and $\|\mathcal{E}\|_1$ is used to characterize the sparsity of the tensor corruption $\mathcal{E}$. Lu et al. [31] proved that the underlying low-rank and sparse tensors can be exactly recovered by the convex version of (2.1) with an overwhelming probability under certain conditions. To better explore the tensor subspace information, TLRR [62] represented the tensor data as a predefined dictionary by a low-rank coefficient under the tensor-tensor product framework. Generally, TLRR can be modeled as

$$\begin{aligned} &\min_{\mathcal{Z},\mathcal{E}} \mathcal{R}_1(\mathcal{Z}) + \lambda\mathcal{R}_2(\mathcal{E}) \\ &\text{s.t. } \mathcal{X} = \mathcal{A} * \mathcal{Z} + \mathcal{E}, \end{aligned}$$

where $\mathcal{R}_1(\mathcal{Z})$ and $\mathcal{R}_2(\mathcal{E})$ denote the regularizations of the coefficient $\mathcal{Z}$ and the noise $\mathcal{E}$, respectively (see Figure 2(b) for illustration). Zhou et al. [62] utilized the regularization's TNN (i.e., $\|\mathcal{Z}\|_*$) and the ℓ_1-norm (i.e., $\|\mathcal{E}\|_1$) that can recover the underlying tensor data and cluster them as well. Johnson et al. [19] used the weighted TNN to enhance the low-tensor rankness of the representation tensor. TLRR not only characterizes the linear relations among tensor samples, but also reveals the cluster structure. Nonetheless, the performance of TLRR would drop when the observations are insufficient and/or corrupted seriously.

3. BTLRR. In this section, we investigate the proposed BTLRR method for addressing the problem of insufficient observed data samples.

In order to study the effect of hidden tensor data, we need to explore the minimizer of (1.3).

Theorem 3.1. *Given any third-order tensors $\mathcal{X}_O$ ($\mathcal{X}_O \neq 0$) and $\mathcal{X}_H$, assume that $\mathcal{U}*\mathcal{S}*\mathcal{V}^\top$ is the skinny TSVD of $[\mathcal{X}_O, \mathcal{X}_H]$, then Problem (1.3) has the following unique and closed-form solution $\mathcal{Z}^\star_{O,H} = [\mathcal{Z}^\star_O; \mathcal{Z}^\star_H]$,*

$$\mathcal{Z}^\star_O = \mathcal{V}_O * \mathcal{V}_O^\top \text{ and } \mathcal{Z}^\star_H = \mathcal{V}_H * \mathcal{V}_O^\top, \tag{3.1}$$

where $\mathcal{V} = [\mathcal{V}_O; \mathcal{V}_H]$ and $\mathcal{X}_O = \mathcal{U} * \mathcal{S} * \mathcal{V}_O^\top$.

The proof of Theorem 3.1 is given in Appendix A. With Theorem 3.1, we will show that the recovery of hidden tensor data can be obtained when there is no noise in the observed data samples:

$$\begin{aligned}
\mathcal{X}_O &= [\mathcal{X}_O, \mathcal{X}_H] * \mathcal{Z}^\star_{O,H} \\
&= \mathcal{X}_O * \mathcal{Z}^\star_O + \mathcal{X}_H * \mathcal{Z}^\star_H \\
&= \mathcal{X}_O * \mathcal{Z}^\star_O + \mathcal{X}_H * \mathcal{V}_H * \mathcal{V}_O^\top \\
&= \mathcal{X}_O * \mathcal{Z}^\star_O + \mathcal{U} * \mathcal{S} * \mathcal{V}_H^\top * \mathcal{V}_H * \mathcal{V}_O^\top \\
&= \mathcal{X}_O * \mathcal{Z}^\star_O + \mathcal{U} * \mathcal{S} * \mathcal{V}_H^\top * \mathcal{V}_H * \mathcal{S}^\dagger * \mathcal{U}^\top * \mathcal{X}_O.
\end{aligned}$$

Let

$$\mathcal{L}^\star_H = \mathcal{U} * \mathcal{S} * \mathcal{V}_H^\top * \mathcal{V}_H * \mathcal{S}^\dagger * \mathcal{U}^\top. \tag{3.2}$$

Then we equivalently reformulate the observed $\mathcal{X}_O$ as the following bilateral tensor representation:

$$\mathcal{X}_O = \mathcal{X}_O * \mathcal{Z}^\star_O + \mathcal{L}^\star_H * \mathcal{X}_O. \tag{3.3}$$

Even we have insufficiently observed samples, we can still make use of (3.3) and observed tensor samples to study the effect of hidden tensor data.

Suppose that $\mathcal{X}_O$ and $\mathcal{X}_H$ are two sets of samples drawn from a union of k tensor subspaces $\{\mathcal{S}_i\}_{i=1}^k$, each $\mathcal{S}_i$ having a tensor tubal rank of r_i. Then with (3.1), we have

$$\operatorname{rank}(\mathcal{Z}^\star_O) \leq \operatorname{rank}(\mathcal{V}_O) \leq \operatorname{rank}([\mathcal{X}_O, \mathcal{X}_H]) \leq \sum_{i=1}^k r_i = r.$$

According to (3.2), we also obtain

$$\operatorname{rank}(\mathcal{L}_H^\star) \le \operatorname{rank}(\mathcal{V}_H) \le \operatorname{rank}([\mathcal{X}_O, \mathcal{X}_H]) \le r.$$

Both $\mathcal{Z}_O^\star$ and $\mathcal{L}_H^\star$ should be low tubal rank tensors. We can recover $\mathcal{Z}_O^\star$ and $\mathcal{L}_H^\star$ from the observed data $\mathcal{X}_O$ by the following optimization problem:

$$\begin{aligned} &\min_{\mathcal{Z}_O, \mathcal{L}_H} \operatorname{rank}(\mathcal{Z}_O) + \operatorname{rank}(\mathcal{L}_H) \\ &\text{s.t. } \mathcal{X}_O = \mathcal{X}_O * \mathcal{Z}_O + \mathcal{L}_H * \mathcal{X}_O. \end{aligned}$$

The two low tubal-rank tensors $\mathcal{Z}_O$ and $\mathcal{L}_H$ play different important roles. The representation tensor $\mathcal{Z}_O$ can capture the similarity between samples, and $\mathcal{L}_H * \mathcal{X}_O$ can be interpreted as extracting the relationship among features. These two terms boost each other to achieve a satisfactory clustering and recovery performance. For simplicitly, we omit the subscripts of $\mathcal{Z}_O$, $\mathcal{L}_H$, and $\mathcal{X}_O$ in the following discussions.

3.1. BTLRR for noiseless data. Since minimizing the tubal rank is generally NP-hard, we employ its convex relaxation and its associated TNN [31] in the optimization model. Therefore, when the observed tensor data are noiseless, the following optimization problem is studied:

$$\begin{aligned} &\min_{\mathcal{Z}, \mathcal{L}} \|\mathcal{Z}\|_* + \|\mathcal{L}\|_* \\ &\text{s.t. } \mathcal{X} = \mathcal{X} * \mathcal{Z} + \mathcal{L} * \mathcal{X}. \end{aligned} \tag{3.4}$$

Next we give solutions to (3.4) without noise as the following theorem.

Theorem 3.2. *Suppose that $\mathcal{U}_\mathcal{X} * \mathcal{S}_\mathcal{X} * \mathcal{V}_\mathcal{X}^\top$ is the skinny TSVD of $\mathcal{X}$, and $\bar{\boldsymbol{S}}_\mathcal{X} = \texttt{bdiag}(\bar{\mathcal{S}}_\mathcal{X})$ has full rank. The solutions to (3.4) are given as follows,*

$$\mathcal{Z}^\star = \mathcal{V}_\mathcal{X} * \mathcal{W} * \mathcal{V}_\mathcal{X}^\top \text{ and } \mathcal{L}^\star = \mathcal{U}_\mathcal{X} * (\mathcal{I} - \mathcal{W}) * \mathcal{U}_\mathcal{X}^\top, \tag{3.5}$$

where $\mathcal{W}$ is a tensor, which satisfies (i) *$\mathcal{W}$ is compatible with $\mathcal{S}_\mathcal{X}$, i.e., $[\bar{\boldsymbol{W}}]_{i,j} = 0$ if $[\bar{\boldsymbol{S}}_\mathcal{X}]_{i,i} \neq [\bar{\boldsymbol{S}}_\mathcal{X}]_{j,j}$, where $\bar{\boldsymbol{W}} = \texttt{bdiag}(\bar{\mathcal{W}})$ and* (ii) *both $\bar{\boldsymbol{W}}$ and $\bar{\boldsymbol{I}} - \bar{\boldsymbol{W}}$ are positive semidefinite.*

The proof can be found in Appendix B.

Here we compare our proposed method with current methods. Liu and Yan [30] proposed LatLRR for clustering the vector-valued samples. Compared with LatLRR, the proposed BTLRR uses tensors to model the multidimensional samples, which can keep the intrinsic multidimensional structure within collected data. Also we employ the tensor representation to explore the complex membership between tensor samples and the relationship among features, better than the vector representation used in LatLRR. Zhou et al. [62] proposed TLRR for subspace clustering and data recovery. Compared with TLRR, we introduce the hidden tensor data and recover the corresponding effects by a TNN-based convex optimization problem, which can resolve the problem of insufficiently observed data samples. Also the proposed BTLRR takes into account the similarity between samples and the relationship among features of the tensor data.

3.2. BTLRR for noisy data. To address the problem of insufficiently observed tensor samples corrupted by sparse noise, we propose the following BTLRR model by using $\mathcal{X}-\mathcal{E}$ (i.e., the underlying clean tensor data) as the dictionary:

$$\begin{aligned} &\min_{\mathcal{Z},\mathcal{L},\mathcal{E}} \|\mathcal{Z}\|_* + \|\mathcal{L}\|_* + \lambda\|\mathcal{E}\|_1 \\ &\text{s.t. } \mathcal{X} = (\mathcal{X}-\mathcal{E}) * \mathcal{Z} + \mathcal{L} * (\mathcal{X}-\mathcal{E}) + \mathcal{E}. \end{aligned} \tag{3.6}$$

Next, we show that problem (3.6) admits the closed-form solutions in the following theorem.

Theorem 3.3. *Suppose that* $(\hat{\mathcal{X}}^\star, \mathcal{E}^\star)$ *is an optimal solution to the TRPCA problem in* (2.1), $\mathcal{U}_{\hat{\mathcal{X}}^\star} * \mathcal{S}_{\hat{\mathcal{X}}^\star} * \mathcal{V}_{\hat{\mathcal{X}}^\star}^\top$ *is the skinny TSVD of* $\hat{\mathcal{X}}^\star$, *and* $\bar{\boldsymbol{S}}_{\hat{\mathcal{X}}^\star} = \texttt{bdiag}(\bar{\mathcal{S}}_{\hat{\mathcal{X}}^\star})$ *has full rank. Then the relaxation problem in* (3.6) *has a minimizer* $(\mathcal{Z}^\star, \mathcal{L}^\star, \mathcal{E}^\star)$, *where*

$$\mathcal{Z}^\star = \mathcal{V}_{\hat{\mathcal{X}}^\star} * \mathcal{W} * \mathcal{V}_{\hat{\mathcal{X}}^\star}^\top \text{ and } \mathcal{L}^\star = \mathcal{U}_{\hat{\mathcal{X}}^\star} * (\mathcal{I}-\mathcal{W}) * \mathcal{U}_{\hat{\mathcal{X}}^\star}^\top, \tag{3.7}$$

where $\mathcal{W}$ *is a tensor, which satisfies* (i) $\mathcal{W}$ *is compatible with* $\mathcal{S}_{\hat{\mathcal{X}}^\star}$, *i.e.,* $[\bar{\boldsymbol{W}}]_{i,j} = 0$ *if* $[\bar{\boldsymbol{S}}_{\hat{\mathcal{X}}^\star}]_{i,i} \neq [\bar{\boldsymbol{S}}_{\hat{\mathcal{X}}^\star}]_{j,j}$, *where* $\bar{\boldsymbol{W}} = \texttt{bdiag}(\bar{\mathcal{W}})$; *and* (ii) *both* $\bar{\boldsymbol{W}}$ *and* $\bar{\boldsymbol{I}} - \bar{\boldsymbol{W}}$ *are positive semidefinite. Conversely, let* $(\mathcal{Z}^\star, \mathcal{L}^\star, \mathcal{E}^\star)$ *be any optimal solution to the problem in* (3.6). *Then* $(\mathcal{X}-\mathcal{E}^\star, \mathcal{E}^\star)$ *is a minimizer of the TRPCA problem in* (2.1).

The proof of Theorem 3.3 can be found in Appendix C.

Remark 3.4. We first establish the theoretical performance guarantees for the proposed tensor BTLRR (both noiseless and noisy cases), which are not studied in the matrix work LatLRR. The main challenge is how to tackle the tensor algebra instead of the standard matrix algebra. From the TSVD of the observed tensor data, Theorem 3.2 theoretically provides the solution of the proposed noiseless BTLRR. Furthermore, Theorem 3.3 builds the close connection between the solutions of our noisy BTLRR and TRPCA [31] and gives the closed-form solution to problem (3.6).

In practice, we use the predenoised data as an estimate of $\mathcal{X}-\mathcal{E}$ and propose the following robust BTLRR optimization model,

$$\begin{aligned} &\min_{\mathcal{Z},\mathcal{L},\mathcal{E}} \|\mathcal{Z}\|_* + \|\mathcal{L}\|_* + \lambda\|\mathcal{E}\|_1 \\ &\text{s.t. } \mathcal{X} = \widetilde{\mathcal{X}} * \mathcal{Z} + \mathcal{L} * \widetilde{\mathcal{X}} + \mathcal{E}, \end{aligned} \tag{3.8}$$

where $\widetilde{\mathcal{X}}$ denotes the predenoised tensor data obtained by a denoising method, e.g., TRPCA [31].

3.3. Interpretation of BTLRR. Compared with LatLRR flattening the tensor data into a matrix, the proposed BTLRR is capable of directly handling the tensor data, which can preserve the multidimensional structures of tensor samples. More importantly, the proposed bilateral tensor representation can fully exploit the complex relationships between tensor samples. Therefore, the proposed BTLRR is capable of effectively handling the insufficient tensor samples problem. To show the advantage of our method in characterizing the relationship, we give the expressions and involved entries for each tensor data entry in LatLRR and BTLRR for comparison.

LatLRR represents noiseless $\boldsymbol{X}$ as $\boldsymbol{X} = \boldsymbol{X}\boldsymbol{Z} + \boldsymbol{L}\boldsymbol{X}$, where $\boldsymbol{Z} \in \mathbb{R}^{n_2 \times n_2}$ and $\boldsymbol{L} \in \mathbb{R}^{n_1 n_3 \times n_1 n_3}$. In LatLRR, each entry $[\mathcal{X}]_{i,j,k}$ is expressed as

$$[\mathcal{X}]_{i,j,k} = \sum_{j_1=1}^{n_2} [\mathcal{X}]_{i,j_1,k} \times [\boldsymbol{Z}]_{j_1,j} + \sum_{i_1=1}^{n_1} \sum_{k_1=1}^{n_3} [\boldsymbol{L}]_{(k-1)n_1+i,(i_1-1)n_3+k_1} \times [\mathcal{X}]_{i_1,j,k_1},$$

where $[\mathcal{X}]_{i,j,k}$ involves $n_2 + n_1 n_3$ entries (i.e., $\mathcal{X}(i,:,k)$ and $\mathcal{X}(:,j,:)$) and requires $n_2^2 + n_1^2 n_3^2$ parameters.

BTLRR represents $\mathcal{X}$ as $\mathcal{X} = \mathcal{X} * \mathcal{Z} + \mathcal{L} * \mathcal{X}$, where $\mathcal{Z} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ and $\mathcal{L} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$. In BTLRR, each entry $[\mathcal{X}]_{i,j,k}$ is expressed as

$$\begin{aligned}[\mathcal{X}]_{i,j,k} = &\sum_{k_1=1}^{k} \sum_{j_1=1}^{n_2} [\mathcal{X}]_{i,j_1,k+1-k_1} \times [\mathcal{Z}]_{j_1,j,k_1} + \sum_{k_1=k+1}^{n_3} \sum_{j_1=1}^{n_2} [\mathcal{X}]_{i,j_1,k+1+n_3-k_1} \times [\mathcal{Z}]_{j_1,j,k_1} \\ &+ \sum_{k_1=1}^{k} \sum_{i_1=1}^{n_1} [\mathcal{L}]_{i,i_1,k+1-k_1} \times [\mathcal{X}]_{i_1,j,k_1} + \sum_{k_1=k+1}^{n_3} \sum_{i_1=1}^{n_1} [\mathcal{L}]_{i,i_1,k+1+n_3-k_1} \times [\mathcal{X}]_{i_1,j,k_1},\end{aligned}$$

where $[\mathcal{X}]_{i,j,k}$ involves $n_2 n_3 + n_1 n_3$ entries (i.e., $\mathcal{X}(i,:,:)$ and $\mathcal{X}(:,j,:)$) and requires $(n_1^2 + n_2^2) n_3$ parameters.

According to the above entries' expressions, LatLRR only involves the entries $\mathcal{X}(i,:,k)$, which means that LatLRR just considers the interrelationship between samples. In contrast, the proposed BTLRR involves $\mathcal{X}(i,:,:)$, i.e., BTLRR not only explores the interrelationship between samples but also the internal relationship embedded in each sample. Exploring the internal relationship within the sample can help reveal the similarities and difference between samples belonging to one cluster, which is beneficial for finding representative samples and improving the clustering performance. Moreover, the proposed BTLRR adopts the tensor-tensor product to transform the data samples into the frequency domain, which can help reduce the noise and make the proposed method more robust to noise.

To numerically illustrate the advantage of BTLRR in characterizing the relationships between tensor samples, we, respectively, perform BTLRR and LatLRR on the insufficient tensor data and build the affinity matrices, which are shown in Figure 3. We also list the relative errors between the affinity matrices estimated by BTLRR and LatLRR and the ground truth (i.e., the block-diagonal matrix, where each block is the identity matrix with the proper size). From Figure 3, one can see that the proposed BTLRR achieves the lower relative error in estimating the affinity matrix for different numbers of samples. Meanwhile, the affinity matrices obtained by our method have much clearer block-diagonal structures.

4. Proposed algorithm. In this work, we adapt the alternating direction method of multipliers (ADMM) [2, 5, 14, 25, 26, 27] to solve problem (3.8). However, directly applying ADMM will be computationally expensive when the size n_1 and sample number n_2 are large, as they require to compute the SVD of n_3 matrices of sizes $n_1 \times n_1$ and $n_2 \times n_2$. Hence, to reduce the computational cost, we provide an equivalent reformulation based on the subspace projection for problem (3.8). Specifically, assume the skinny TSVD of $\widetilde{\mathcal{X}}$ is $\mathcal{U}_{\widetilde{\mathcal{X}}} * \mathcal{S}_{\widetilde{\mathcal{X}}} * \mathcal{V}_{\widetilde{\mathcal{X}}}^{\top}$. Defining $\mathcal{A} = \mathcal{U}_{\widetilde{\mathcal{X}}} * \mathcal{S}_{\widetilde{\mathcal{X}}} \in \mathbb{R}^{n_1 \times r_{\widetilde{\mathcal{X}}} \times n_3}$, where $r_{\widetilde{\mathcal{X}}} = \operatorname{rank}(\widetilde{\mathcal{X}})$, then we replace $\widetilde{\mathcal{X}}$ and $\mathcal{Z}$ of $\widetilde{\mathcal{X}} * \mathcal{Z}$

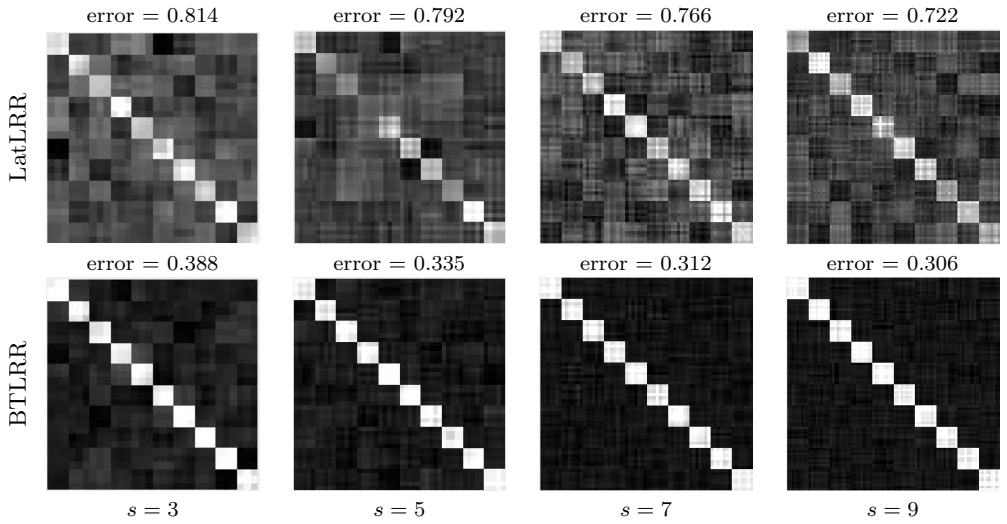


Figure 3. *Illustrations of the effects of LatLRR (first row) and BTLRR (second row) in handling the insufficient tensor data. Here we generate an observed tensor as* $\mathcal{X}_O = [\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_{10}]$ *with* $\mathcal{X}_p = \mathcal{D}_p * \mathcal{T}_p$ $(p = 1, 2, \ldots, 10)$*, where the entries of* $\mathcal{D}_p \in \mathbb{R}^{100\times 10\times 10}$ *and* $\mathcal{T}_p \in \mathbb{R}^{10\times s\times 10}$ *are randomly sampled from i.i.d.* $\mathcal{N}(0,1)$*. We randomly add the sparse noise with noise ratio* 0.05*. The number of samples (i.e., s) at each subspace is less than the rank* $(r = 10)$ *of each subspace, hence, the data samples are insufficient.*

in (3.8) with $\mathcal{A}$ and $\mathcal{Z}_\flat$, respectively, where $\mathcal{Z}_\flat \in \mathbb{R}^{r_{\widetilde{\mathcal{X}}}\times n_2\times n_3}$ is a variable which will be optimized. In a similar way, $\mathcal{L}$ and $\widetilde{\mathcal{X}}$ of $\mathcal{L} * \widetilde{\mathcal{X}}$ in (3.8) can be, respectively, replaced with $\mathcal{L}_\flat$ and $\mathcal{B} = \mathcal{S}_{\widetilde{\mathcal{X}}} * \mathcal{V}_{\widetilde{\mathcal{X}}}^\top \in \mathbb{R}^{r_{\widetilde{\mathcal{X}}}\times n_2\times n_3}$, where $\mathcal{L}_\flat \in \mathbb{R}^{n_1\times r_{\widetilde{\mathcal{X}}}\times n_3}$ is a variable needing to be optimized. Then problem (3.8) can be reformulated into the following problem:

$$\begin{aligned} &\min_{\mathcal{Z}_\flat, \mathcal{L}_\flat, \mathcal{E}} \|\mathcal{Z}_\flat\|_* + \|\mathcal{L}_\flat\|_* + \lambda\|\mathcal{E}\|_1 \\ &\text{s.t. } \mathcal{X} = \mathcal{A} * \mathcal{Z}_\flat + \mathcal{L}_\flat * \mathcal{B} + \mathcal{E}. \end{aligned} \tag{4.1}$$

Using the above reformulation, we only need to compute the SVD of matrices with smaller sizes of $n_1 \times r_{\widetilde{\mathcal{X}}}$ and $n_2 \times r_{\widetilde{\mathcal{X}}}$. The following Theorem 4.1 builds the relationship of solutions to problems (3.8) and (4.1) to show the feasibility of such reformulation. The proof is performed in Appendix D.

Theorem 4.1. *Assume* $(\mathcal{Z}_\flat^\star, \mathcal{L}_\flat^\star, \mathcal{E}^\star)$ *is a minimizer to* (4.1)*, then* $(\mathcal{V}_{\widetilde{\mathcal{X}}} * \mathcal{Z}_\flat^\star, \mathcal{L}_\flat^\star * \mathcal{U}_{\widetilde{\mathcal{X}}}^\top, \mathcal{E}^\star)$ *is a solution to Problem* (3.8)*.*

To solve (4.1), we introduce two auxiliary variables $\mathcal{Z}_\flat = \mathcal{F}$ and $\mathcal{L}_\flat = \mathcal{P}$ and minimize the following augmented Lagrange function,

$$\begin{aligned} &\|\mathcal{F}\|_* + \|\mathcal{P}\|_* + \lambda\|\mathcal{E}\|_1 + \langle \mathcal{G}, \mathcal{Z}_\flat - \mathcal{F}\rangle + \frac{\rho}{2}\|\mathcal{Z}_\flat - \mathcal{F}\|_F^2 + \langle \mathcal{Q}, \mathcal{L}_\flat - \mathcal{P}\rangle + \frac{\theta}{2}\|\mathcal{L}_\flat - \mathcal{P}\|_F^2 \\ &\quad + \langle \mathcal{Y}, \mathcal{X} - \mathcal{A} * \mathcal{Z}_\flat - \mathcal{L}_\flat * \mathcal{B} - \mathcal{E}\rangle + \frac{\mu}{2}\|\mathcal{X} - \mathcal{A} * \mathcal{Z}_\flat - \mathcal{L}_\flat * \mathcal{B} - \mathcal{E}\|_F^2, \end{aligned} \tag{4.2}$$

where $\mathcal{Y}$, $\mathcal{G}$, and $\mathcal{Q}$ are the Lagrange multipliers, and μ, ρ, and θ are the nonnegative penalty parameters. Finally, we solve the above unconstrained problem through alternately updating $\mathcal{Z}_\flat$, $\mathcal{L}_\flat$, $\mathcal{F}$, $\mathcal{P}$, and $\mathcal{E}$ by fixing other variables. Details of updating each variable are as follows.

Updating $\mathcal{Z}_\flat$*:* We minimize the following problem,

$$\begin{aligned}\mathcal{Z}_{\flat,t+1} &= \underset{\mathcal{Z}_\flat}{\operatorname{argmin}} \frac{\rho_t}{2}\left\|\mathcal{Z}_\flat - \mathcal{F}_t + \frac{\mathcal{G}_t}{\rho_t}\right\|_F^2 + \frac{\mu_t}{2}\left\|\mathcal{X} - \mathcal{A} * \mathcal{Z}_\flat - \mathcal{L}_{\flat,t} * \mathcal{B} - \mathcal{E}_t + \frac{\mathcal{Y}_t}{\mu_t}\right\|_F^2 \\ &= \left(\rho_t \mathcal{I} + \mu_t \mathcal{A}^\top * \mathcal{A}\right)^{-1} * \left(\mu_t \mathcal{A}^\top * \mathcal{A}_1 + \rho_t \mathcal{F}_t - \mathcal{G}_t\right),\end{aligned} \tag{4.3}$$

where $\mathcal{A}_1 = \mathcal{X} - \mathcal{L}_{\flat,t} * \mathcal{B} + \mathcal{Y}_t/\mu_t - \mathcal{E}_t$. Based on the property $\|\mathcal{X}\|_F^2 = \frac{1}{n_3}\|\bar{\boldsymbol{X}}\|_F^2$ and the fact that $\bar{\boldsymbol{X}}$ is block-diagonal, $\mathcal{Z}_{\flat,t+1}$ can be updated by computing $\bar{\mathcal{Z}}_{\flat,t+1}$ first via

$$\bar{\mathcal{Z}}_{\flat,t+1}^{(k)} = \bar{\mathcal{Z}}_1^{(k)}\left(\mu_t (\bar{\mathcal{A}}^{(k)})^\top \bar{\mathcal{A}}_1^{(k)} + \rho_t \bar{\mathcal{F}}_t^{(k)} - \bar{\mathcal{G}}_t^{(k)}\right),$$

where $\mathcal{Z}_1 = \left(\rho_t \mathcal{I} + \mu_t \mathcal{A}^\top * \mathcal{A}\right)^{-1}$. Then we can obtain $\mathcal{Z}_{\flat,t+1} = \texttt{ifft}(\bar{\mathcal{Z}}_{\flat,t+1}, [\,], 3)$.

Updating $\mathcal{L}_\flat$*:* We minimize the following problem,

$$\begin{aligned}\mathcal{L}_{\flat,t+1} &= \underset{\mathcal{L}_\flat}{\operatorname{argmin}} \frac{\theta_t}{2}\left\|\mathcal{L}_\flat - \mathcal{P}_t + \frac{\mathcal{Q}_t}{\theta_t}\right\|_F^2 + \frac{\mu_t}{2}\left\|\mathcal{X} - \mathcal{A} * \mathcal{Z}_{\flat,t+1} - \mathcal{L}_\flat * \mathcal{B} - \mathcal{E}_t + \frac{\mathcal{Y}_t}{\mu_t}\right\|_F^2 \\ &= \left(\mu_t \mathcal{B}_1 * \mathcal{B}^\top + \theta_t \mathcal{P}_t - \mathcal{Q}_t\right) * \left(\theta_t \mathcal{I} + \mu_t \mathcal{B} * \mathcal{B}^\top\right)^{-1},\end{aligned} \tag{4.4}$$

where $\mathcal{B}_1 = \mathcal{X} - \mathcal{A} * \mathcal{Z}_{\flat,t+1} + \mathcal{Y}_t/\mu_t - \mathcal{E}_t$.

Updating $\mathcal{F}$*:* We minimize the following problem,

$$\mathcal{F}_{t+1} = \underset{\mathcal{F}}{\operatorname{argmin}} \|\mathcal{F}\|_* + \frac{\rho_t}{2}\left\|\mathcal{Z}_{\flat,t+1} - \mathcal{F} + \frac{\mathcal{G}_t}{\rho_t}\right\|_F^2, \tag{4.5}$$

which can be solved via the tensor singular value thresholding operator [31], i.e.,

$$\mathcal{F}_{t+1} = \mathcal{U}_1 * \Phi_{1/\rho_t}(\mathcal{S}_1) * \mathcal{V}_1^\top, \tag{4.6}$$

where $\mathcal{U}_1 * \mathcal{S}_1 * \mathcal{V}_1^\top$ is the TSVD of $\mathcal{Z}_{\flat,t+1} + \mathcal{G}_t/\rho_t$, and $\Phi_\tau(\mathcal{S}) = \texttt{ifft}((\bar{\mathcal{S}} - \tau)_+, [\,], 3)$, in which $s_+ = \max\{s, 0\}$.

Updating $\mathcal{P}$*:* We minimize the following problem,

$$\begin{aligned}\mathcal{P}_{t+1} &= \underset{\mathcal{P}}{\operatorname{argmin}} \|\mathcal{P}\|_* + \frac{\theta_t}{2}\left\|\mathcal{L}_{\flat,t+1} - \mathcal{P} + \frac{\mathcal{Q}_t}{\theta_t}\right\|_F^2 \\ &= \mathcal{U}_2 * \Phi_{1/\theta_t}(\mathcal{S}_2) * \mathcal{V}_2^\top,\end{aligned} \tag{4.7}$$

where $\mathcal{U}_2 * \mathcal{S}_2 * \mathcal{V}_2^\top$ is the TSVD of $\mathcal{L}_{\flat,t+1} + \mathcal{Q}_t/\theta_t$.

Updating $\mathcal{E}$*:* We minimize the following problem,

$$\begin{aligned}\mathcal{E}_{t+1} &= \underset{\mathcal{E}}{\operatorname{argmin}} \lambda\|\mathcal{E}\|_1 + \frac{\mu_t}{2}\left\|\mathcal{E} - \mathcal{X} + \mathcal{A} * \mathcal{Z}_{\flat,t+1} + \mathcal{L}_{\flat,t+1} * \mathcal{B} - \frac{\mathcal{Y}_t}{\mu_t}\right\|_F^2 \\ &= \Psi_{\lambda/\mu_t}\left(\mathcal{X} - \mathcal{A} * \mathcal{Z}_{\flat,t+1} - \mathcal{L}_{\flat,t+1} * \mathcal{B} + \mathcal{Y}_t/\mu_t\right),\end{aligned} \tag{4.8}$$

where $\Psi_{\lambda/\mu_t}(\cdot)$ denotes the soft-thresholding operator [7].

Updating $\mathcal{Y}$*,* $\mathcal{G}$*, and* $\mathcal{Q}$*:* Update the multipliers by

Algorithm 4.1. ADMM-based algorithm for solving (4.1).

Input: Tensor data $\mathcal{X}$, $\mathcal{A}=\mathcal{U}_{\widetilde{\mathcal{X}}}*\mathcal{S}_{\widetilde{\mathcal{X}}}$ and $\mathcal{B}=\mathcal{S}_{\widetilde{\mathcal{X}}}*\mathcal{V}_{\widetilde{\mathcal{X}}}^{\top}$ with skinny TSVD $\mathcal{U}_{\widetilde{\mathcal{X}}}*\mathcal{S}_{\widetilde{\mathcal{X}}}*\mathcal{V}_{\widetilde{\mathcal{X}}}^{\top}$ of $\widetilde{\mathcal{X}}$.
Initialize: $\mathcal{Z}_\flat=\mathcal{F}=0$, $\mathcal{L}_\flat=\mathcal{P}=0$, $\mathcal{E}=0$, $\mathcal{Y}=0$, $\mathcal{G}=0$, $\mathcal{Q}=0$, $\mu_0=\rho_0=\theta_0>0$, $\eta=1.1$, and $\varepsilon=10^{-8}$.
While not converged **do**

1. Update $\mathcal{Z}_\flat$ by (4.3).
2. Update $\mathcal{L}_\flat$ by (4.4).
3. Update $\mathcal{F}$ by (4.6).
4. Update $\mathcal{P}$ by (4.7).
5. Update $\mathcal{E}$ by (4.8).
6. Update the multipliers by (4.9), (4.10), and (4.11).
7. Update the parameters $(\mu_{t+1},\rho_{t+1},\theta_{t+1})=\eta(\mu_t,\rho_t,\theta_t)$.
8. Check the convergence conditions:

$$\max\left(\|\mathcal{X}-\mathcal{A}*\mathcal{Z}_{\flat,t}-\mathcal{L}_{\flat,t}*\mathcal{B}-\mathcal{E}_t\|_\infty,\|\mathcal{Z}_{\flat,t}-\mathcal{F}_t\|_\infty,\|\mathcal{L}_{\flat,t}-\mathcal{P}_t\|_\infty\right)<\varepsilon,$$

end while
Output: $\widehat{\mathcal{X}}=\mathcal{A}*\mathcal{Z}_{\flat,t+1}+\mathcal{L}_{\flat,t+1}*\mathcal{B}$, $\widehat{\mathcal{Z}}=\mathcal{V}_{\widetilde{\mathcal{X}}}*\mathcal{Z}_{\flat,t+1}$, $\widehat{\mathcal{L}}=\mathcal{L}_{\flat,t+1}*\mathcal{U}_{\widetilde{\mathcal{X}}}^{\top}$, $\widehat{\mathcal{E}}=\mathcal{E}_{t+1}$.

$$\mathcal{Y}_{t+1}=\mathcal{Y}_t+\mu_t(\mathcal{X}-\mathcal{A}*\mathcal{Z}_{\flat,t+1}-\mathcal{L}_{\flat,t+1}*\mathcal{B}-\mathcal{E}_{t+1}), \tag{4.9}$$

$$\mathcal{G}_{t+1}=\mathcal{G}_t+\rho_t(\mathcal{Z}_{\flat,t+1}-\mathcal{F}_{t+1}), \tag{4.10}$$

$$\mathcal{Q}_{t+1}=\mathcal{Q}_t+\theta_t(\mathcal{L}_{\flat,t+1}-\mathcal{P}_{t+1}). \tag{4.11}$$

The proposed algorithm for handling (4.1) is outlined in Algorithm 4.1. Therefore, we first obtain the optimal minimizer $(\mathcal{Z}_\flat^\star,\mathcal{L}_\flat^\star,\mathcal{E}^\star)$ to problem (4.1), and then get the solution $(\mathcal{V}_{\widetilde{\mathcal{X}}}*\mathcal{Z}_\flat^\star,\mathcal{L}_\flat^\star*\mathcal{U}_{\widetilde{\mathcal{X}}}^{\top},\mathcal{E}^\star)$ to problem (3.8).

4.1. Complexity analysis. Computing $\mathcal{Z}_{\flat,t+1}$ and $\mathcal{L}_{\flat,t+1}$ both cost $\mathcal{O}(r_{\widetilde{\mathcal{X}}}n_1n_2n_3+r_{\widetilde{\mathcal{X}}}(n_1+n_2)n_3\log(n_3))$. The major cost of computing $\mathcal{F}_{t+1}$ involves an n_3 SVD on $r_{\widetilde{\mathcal{X}}}\times n_2$ matrices at the cost of $\mathcal{O}(r_{\widetilde{\mathcal{X}}}^2n_2n_3)$ and the inverse DFT at the cost of $\mathcal{O}(r_{\mathcal{X}}n_2n_3\log(n_3))$. Similarly, computing $\mathcal{P}_{t+1}$ costs $\mathcal{O}(r_{\widetilde{\mathcal{X}}}n_1n_3(r_{\widetilde{\mathcal{X}}}+\log(n_3)))$. When updating $\mathcal{E}_{t+1}$, the main two steps of computing tensor-tensor product cost $\mathcal{O}((r_{\widetilde{\mathcal{X}}}+\log(n_3))n_1n_2n_3)$. So Algorithm 4.1 takes $\mathcal{O}((r_{\widetilde{\mathcal{X}}}+\log(n_3))n_1n_2n_3)$ at each iteration. However, directly solving the original (3.8) costs $\mathcal{O}((n_1+n_2)(n_1^2+n_2^2)n_3+(n_1^2+n_1n_2n_2^2)n_3\log(n_3))$. Therefore, the reformulation (4.1) reduces the cost complexity significantly—usually $r_{\widetilde{\mathcal{X}}}\ll\min(n_1,n_2)$ holds.

4.2. Convergence guarantee. Now we give the convergence analysis of the proposed algorithm in the following theorem, which is proved in Appendix E.

Theorem 4.2. *Let* $\{\mathcal{H}_t\}=\{\mathcal{Z}_{\flat,t},\mathcal{L}_{\flat,t},\mathcal{E}_t,\mathcal{F}_t,\mathcal{G}_t,\mathcal{P}_t,\mathcal{Q}_t,\mathcal{Y}_t\}$ *be the sequences generated by Algorithm* 4.1, *then* (i) *there exists a limit point* $\mathcal{H}^\star=\{\mathcal{Z}_\flat^\star,\mathcal{L}_\flat^\star,\mathcal{E}^\star,\mathcal{F}^\star,\mathcal{G}^\star,\mathcal{P}^\star,\mathcal{Q}^\star,\mathcal{Y}^\star\}$ *satisfying the Karush-Kuhn-Tucker conditions of* (4.2), *i.e.,*

$$\mathcal{Z}_\flat^\star = \mathcal{F}^\star,\ \mathcal{L}_\flat^\star = \mathcal{P}^\star,\ \mathcal{A}^\top * \mathcal{Y}^\star = \mathcal{G}^\star,\ \mathcal{Y}^\star * \mathcal{B}^\top = \mathcal{Q}^\star,$$
$$\mathcal{X} = \mathcal{A} * \mathcal{Z}_\flat^\star + \mathcal{L}_\flat^\star * \mathcal{B} + \mathcal{E}^\star,\ \mathcal{Y}^\star \in \lambda\partial\|\mathcal{E}^\star\|_1, \mathcal{G}^\star \in \partial\left\|\mathcal{F}^\star\right\|_*,\ \mathcal{Q}^\star \in \partial\left\|\mathcal{P}^\star\right\|_*$$

and (ii) *the sequences* $\{\mathcal{Z}_{\flat,t}, \mathcal{L}_{\flat,t}, \mathcal{E}_t, \mathcal{F}_t, \mathcal{P}_t\}$ *are Cauchy sequences and thus converge to the critical point of* (4.2).

Remark 4.3. To solve the proposed tensor optimization problem, we develop an ADMM-based algorithm and establish its theoretical convergence by exploiting the structure of the proposed model and designing the updating rule of penalty parameters (see line 7 of Algorithm 1). The factor η is larger than 1 so that the boundedness requirement can be established in the convergence proof.

5. Experimental results.

In this section, we showcase the effectiveness of the proposed BTLRR using extensive experiments on both subspace clustering and data recovery of multidimensional images. All simulations are performed using MATLAB R2018b on a desktop with a 3.70 GHz Intel Core i7-8700 M CPU and 32 GB RAM.

5.1. Experimental settings.

5.1.1. Baselines.

We compare BTLRR with some relevant baselines. For the subspace clustering task, baselines include LRR [29], SSC [9], LatLRR [30], block diagonal representation (BDR) [32], and TLRR [62].

For the data recovery task, baselines include RPCA [3], LatLRR [30], high-order RPCA (HoRPCA) [13], TRPCA [31], and TLRR [62].

5.1.2. Algorithm settings.

The proposed BTLRR involves two parameters, i.e., the regularization parameter λ and the penalty parameter μ. Empirically, the parameter λ is selected from one of the values in $\{a/\sqrt{\max(n_1, n_2)n_3}\}$, where $a = [0.6 : 0.2 : 3]$. The parameter μ is selected from one value of $\{0.001, 0.005, 0.01, 0.05\}$. For the parameter settings of baselines, we mainly follow the corresponding papers' suggestions and take proper parameters to obtain their best performance under our experimental settings.

For the dictionary-based methods, e.g., the works in [29, 62], a high-quality dictionary is very important to the methods' performance. Indeed, we can use the raw observed data as the dictionary—that is similar to the scheme in LRR [29]. However, when the data are grossly corrupted, choosing the observed data as the dictionary would drop the performance. Therefore, in this work, we select the result $\widetilde{\mathcal{X}}$ by applying TRPCA onto the observed $\mathcal{X}$ as the dictionary for TLRR and our BTLRR. Similarly, for LRR and LatLRR, the dictionary is set as the estimation by RPCA.

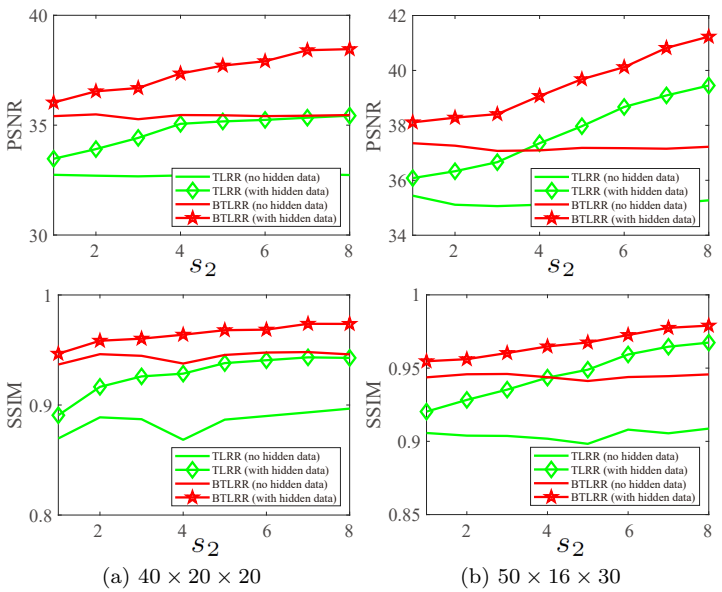
5.1.3. Metrics.

In the clustering experiments, we use three metrics, including accuracy (ACC) [9], normalized mutual information (NMI) [46], and purity (PUR) [36]. ACC computes the accuracy of the best matching permutation between the clustered and the ground-truth labels. NMI measures the mutual dependence between the clustered and true labels. PUR is calculated by the percentage of the number of correctly matched labels over the total number of data points. Higher ACC, NMI, and PUR values demonstrate higher clustering accuracy.

In the data recovery, we use the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) [49] as the denoising performance metrics. PSNR generally measures the error between the corresponding pixels of original and estimated images, SSIM is used for measuring the similarity between images. Higher PSNR and SSIM values indicate better recovered quality.

5.2. Synthetic data experiments. Here, we present the synthetic data simulations, including clustering and recovery, to demonstrate the effectiveness of the proposed method. The procedures of generating $\mathcal{X}_O$ and $\mathcal{X}_H$ are detailed as follows: (1) we first generate three kinds of tensors $\mathcal{D}_i \in \mathbb{R}^{n_1 \times r \times n_3}$, $\mathcal{T}_i \in \mathbb{R}^{r \times s_1 \times n_3}$, and $\mathcal{K}_i \in \mathbb{R}^{r \times s_2 \times n_3}$ $(i = 1, \ldots, I)$ following the i.i.d. Gaussian distribution with unit variance and zero mean; (2) we then generate the observed tensor data by $\mathcal{X}_O = [\mathcal{D}_1 * \mathcal{T}_1, \ldots, \mathcal{D}_I * \mathcal{T}_I] \in \mathbb{R}^{n_1 \times I s_1 \times n_3}$, and the hidden tensor data by $\mathcal{X}_H = [\mathcal{D}_1 * \mathcal{K}_1, \ldots, \mathcal{D}_I * \mathcal{K}_I] \in \mathbb{R}^{n_1 \times I s_2 \times n_3}$.

5.2.1. Data recovery. In this simulation, we test the effect of the hidden tensor data. We consider two observed tensors: one is of size $40 \times 20 \times 20$ by setting $n_1 = 40$, $n_3 = 20$, $I = 5$, $r = 5$, and $s_1 = 4$; one is of size $50 \times 16 \times 30$ by setting $n_1 = 50$, $n_3 = 30$, $I = 4$, $r = 6$, and $s_1 = 4$. Since the number of samples of each subspace is less than the corresponding subspace rank, each observed datum is insufficient. For each subspace, we also generate the hidden tensor data as $\mathcal{D}_i * \mathcal{K}_i \in \mathbb{R}^{n_1 \times s_2 \times n_3}$ for $i = 1, \ldots, I$. We randomly add the sparse noise with noise ratio (NR) $= 0.1$. Figure 4 shows the recovery performance of TLRR and BTLRR under the cases with/without hidden tensor data. One can get that (1) with the hidden tensor data, the denoising performance of both TLRR and the proposed BTLRR can be improved; (2) our BTLRR can achieve the better restored results than TLRR with the hidden tensor data.



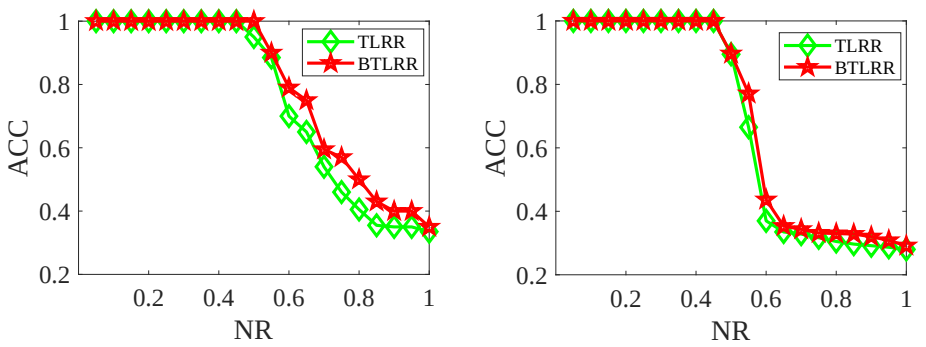
(a) $40 \times 20 \times 20$ (b) $50 \times 16 \times 30$

Figure 4. *Illustration of the improved recovery performance of the hidden effects.*

One can conclude that taking the hidden tensor data into account, the proposed method can resolve the insufficient data problem.

5.2.2. Subspace clustering. We test the clustering performance of TLRR and the proposed method on two data size cases. One size is that we set $n_1 = 20$, $n_3 = 10$, the total number of subspaces $I = 5$, the rank of each subspace $r = 5$, and the number of sample $s_1 = 4$, then we get a tensor of data with a size of $20 \times 20 \times 10$. The another one is of size $30 \times 60 \times 20$ by setting $n_1 = 30$, $n_3 = 20$, $I = 10$, $r = 10$, and $s_1 = 6$. We randomly set the pixels to random values in $[0, 1]$ with NR from 0.05 to 1. Figure 5 presents the clustering accuracy of TLRR and BTLRR under different noise levels. One can see that BTLRR is more robust than TLRR under sparse corruption, i.e., the robustness of TLRR can be improved by incorporating the hidden tensor data.

5.3. Subspace clustering experiments. In this subsection, we test the clustering performance of the proposed method. The subspace clustering aims at grouping the observed data into different clusters by exploring the correlation among each sample belonging to one group. For the matrix-based methods, we form a matrix where each column is the vectorization of each image and then perform the matrix clustering methods to get the affinity matrix $\boldsymbol{Z}$. For the tensor-based methods, we arrange each image as the lateral slice to get a third-order tensor, perform the tensor clustering methods to get the learned coefficient tensor $\mathcal{Z}^\star$, and get $\boldsymbol{Z}$ constructed by (1.2). Then, we use NCut [41] to obtain the clustering results. The used multidimensional image datasets in this section are described in Table 1, and the image samples are shown in Figure 6.



(a) $20 \times 20 \times 10$ (b) $30 \times 60 \times 20$

Figure 5. *Illustration of the robustness of the proposed method under different noises.*

Table 1
Descriptions of three clustering datasets.

Datasets	# Total image	Size	# Class	# Per Class
UCSD	1200	32×32	18	≈ 66
YaleB	840	80×60	28	30
MIT	3240	32×32	10	324

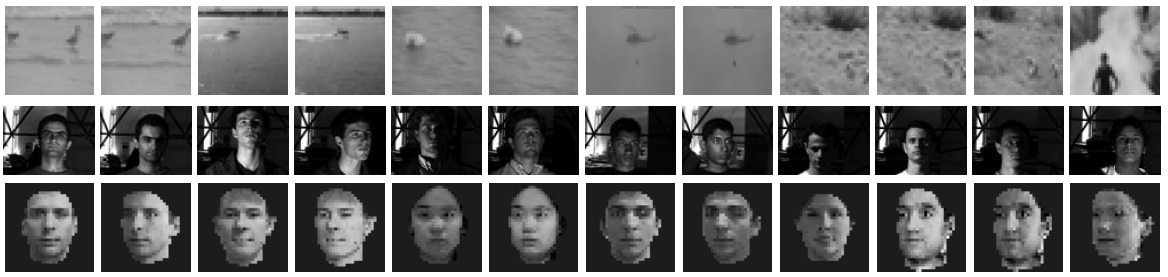
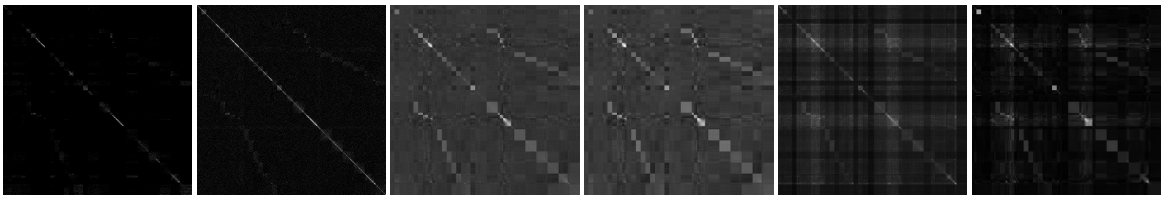


Figure 6. *Image samples of clustering datasets. From top to bottom: UCSD, YaleB, and MIT, respectively.*

Table 2
Clustering results (ACC, NMI, PUR) on the UCSD dataset under different NRs.

NRs	Metrics	SSC	BDR	LRR	LatLRR	TLRR	BTLRR
0.1	ACC	0.7314	0.6220	0.8323	0.8400	0.8111	**0.8910**
	NMI	0.8430	0.5884	0.8946	0.9132	0.8722	**0.9195**
	PUR	0.8105	0.6663	0.8772	0.8842	0.8647	**0.9076**
0.2	ACC	0.5840	0.4960	0.7687	0.8138	0.7443	**0.8624**
	NMI	0.6548	0.4877	0.8082	0.8602	0.7619	**0.9052**
	PUR	0.6782	0.5433	0.8095	0.8578	0.7993	**0.8888**



(a) SSC (b) BDR (c) LRR (d) LatLRR (e) TLRR (f) BTLRR

Figure 7. *The comparison of block-diagonal structures learned by different methods for the UCSD dataset under NR* $= 0.1$.

5.3.1. UCSD dataset. The first experiment that we use is the University of California San Diego (UCSD)[1] scene dataset. This database contains 1200 different environment images with size 32×32 for 18 video sequences. We set the random sparse noise as NR $= 0.1$ and 0.2.

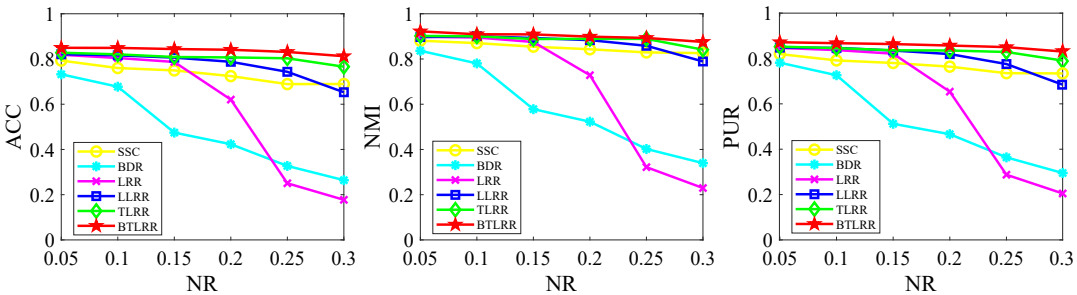
From Table 2, one can see that the proposed algorithm achieves the most promising performance over various metrics under different noise levels. Figure 7 plots the block-diagonal structure of the affinity matrix obtained by different methods. One can see that the proposed BTLRR admits the clearer block-diagonal structure than other methods, which indicates the superiority of learning the correlation along samples of our method.

5.3.2. YaleB dataset. The second experiment uses a subscene of the YaleB dataset [12]. The used YaleB dataset has 840 face images over 28 subjects under different lighting conditions

[1]Statistical Visual Computing Lab UC San Diego, http://www.svcl.ucsd.edu/projects/background_subtraction/ ucsdbgsub_dataset.htm.

Table 3
Clustering results (ACC, NMI, PUR) on the YaleB dataset under different NRs.

NRs	Metrics	SSC	BDR	LRR	LatLRR	TLRR	BTLRR
0.1	ACC	0.7642	0.6774	0.8044	0.8162	0.8205	**0.8490**
	NMI	0.8682	0.7805	0.8950	0.9011	0.9002	**0.9093**
	PUR	0.7961	0.7262	0.8371	0.8468	0.8457	**0.8675**
0.2	ACC	0.7243	0.4232	0.6204	0.7876	0.8070	**0.8408**
	NMI	0.8430	0.5227	0.7283	0.8833	0.8893	**0.8988**
	PUR	0.7640	0.4662	0.6537	0.8195	0.8352	**0.8571**



(a) ACC (b) NMI (c) PUR

Figure 8. *Clustering results on the YaleB data under different noises.*

and various facial expressions, where each subject contains about 30 expressions with size 80×60. We randomly set the pixels to values in $[0, 1]$ with the NR from 0.05 to 0.3.

Table 3 lists the detailed ACC, NMI, and PUR metrics of all algorithms under NR $= 0.1$ and 0.2. One can see that our BTLRR achieves the highest metrics in all cases. We also test the performance of all methods on different NRs that range from 0.05 to 0.3. Figure 8 displays the clustering results under different noises. One can see that the proposed method outperforms the baselines in all cases, showing the promising performance.

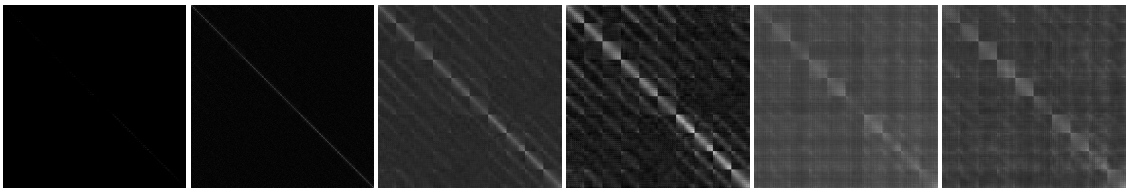
5.3.3. MIT dataset. For the third experiment, we use the MIT face dataset, which contains a total of 3240 images for 10 subjects. Each one has 324 images with size 32×32. The sparse noise with NR $= 0.1$ or 0.2 is randomly added to the observed data.

Table 4 shows the reconstruction performance of different methods under NR $= 0.1$ and 0.2. Similarly to the previous experiment, the proposed BTLRR achieves the highest values over different metrics. This superiority leads to the better visual block-diagonal effect, as shown in Figure 9. One can see that the proposed method admits the superior ability to explore the subspace membership of samples and the relationship among features.

The superior performance of the proposed method lies in two aspects. (1) BTLRR can preserve the multidimensional structure of the tensor data, while matrix-type methods unfold the tensor data along one certain mode, resulting in structural damage and limited performance. (2) By introducing the hidden data, our method can fully explore the samples similarity and features relationship, while TLRR only uses the similarity between samples.

Table 4
Clustering results (ACC, NMI, PUR) on the MIT dataset under different NRs.

NRs	Metrics	SSC	BDR	LRR	LatLRR	TLRR	BTLRR
0.1	ACC	0.5123	0.1735	0.5769	0.5815	0.8466	**0.9031**
	NMI	0.5456	0.0375	0.7103	0.7059	0.8311	**0.9366**
	PUR	0.5564	0.1927	0.6664	0.6707	0.8466	**0.9031**
0.2	ACC	0.2374	0.1429	0.5535	0.5764	0.8127	**0.8929**
	NMI	0.1583	0.0155	0.6806	0.6941	0.8158	**0.9200**
	PUR	0.2501	0.1524	0.6475	0.6662	0.8241	**0.8929**



(a) SSC (b) BDR (c) LRR (d) LatLRR (e) TLRR (f) BTLRR

Figure 9. *The comparison of block-diagonal structures learned by different methods for the MIT dataset under NR* $= 0.1$.

5.4. Data recovery experiments. We test our method on the multidimensional image recovery. The testing datasets include the multispectral images (MSIs) and videos.

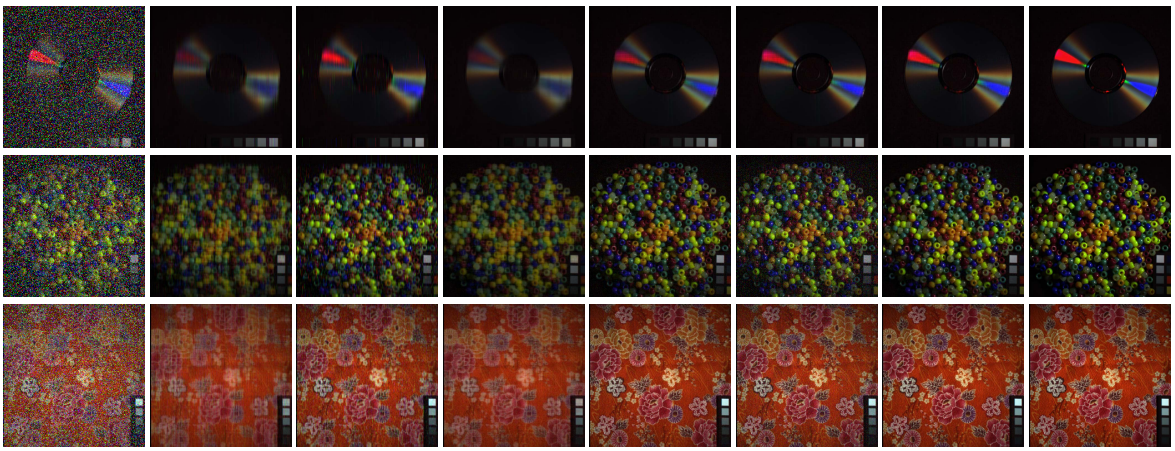
5.4.1. MSI dataset. In the first data recovery experiment, the MSI database CAVE,[2] consisting of 32 scenes, is used. Each MSI datum is with size of $256 \times 256 \times 31$. We set the values of each image in $[0, 1]$, and the sparse noise with NR $= 0.2$ and 0.3 is also added.

Figure 10 shows the images recovered by all methods. One can see that RPCA, LatLRR, and HoRPCA both have undesired stripe noise in the recovered images. The methods TRPCA and TLRR work reasonably well. However, the images obtained by the proposed method are visually closer to the original ones, and our BTLRR outperforms the baselines in terms of keeping the edges and details of recovered images. Figure 11 plots the denoising performance for all images. One can see that the proposed method achieves the highest PSNR and SSIM values in most scenarios. More detailed numerical comparisons under different noise levels for some images can be found in Table 5. One can see that the proposed method stands out under most metrics and cases.

5.4.2. Video dataset. For the second data recovery experiment, we test the performance of our method on videos, including *hall, salesman, suzie, foreman,* and *news.* The size of each video is $144 \times 176 \times 150$. We also randomly add the sparse noise with NR $= 0.2$.

Table 6 lists the numerical results for recovering the videos by all methods. The recovered values are averaged by all the frames. One can see that the proposed method obtains the highest PSNR and SSIM values for all videos. We also plot the PSNR and SSIM curves w.r.t. each frame of recovered results by all methods in Figure 12. Obviously, the proposed method achieves the highest PSNR and SSIM values in most frames. Figure 13 shows the estimated

[2] https://cave.cs.columbia.edu/repository/Multispectral.



(a) Noisy (b) RPCA (c) LatLRR (d) HoRPCA (e) TRPCA (f) TLRR (g) BTLRR (h) Original

Figure 10. *Denoising results (R:*30*th band, G:*20*th band, B:*10*th band) on three MSIs under NR* = 0.3. *From top to bottom:* cd, beads, *and* cloth, *respectively.*

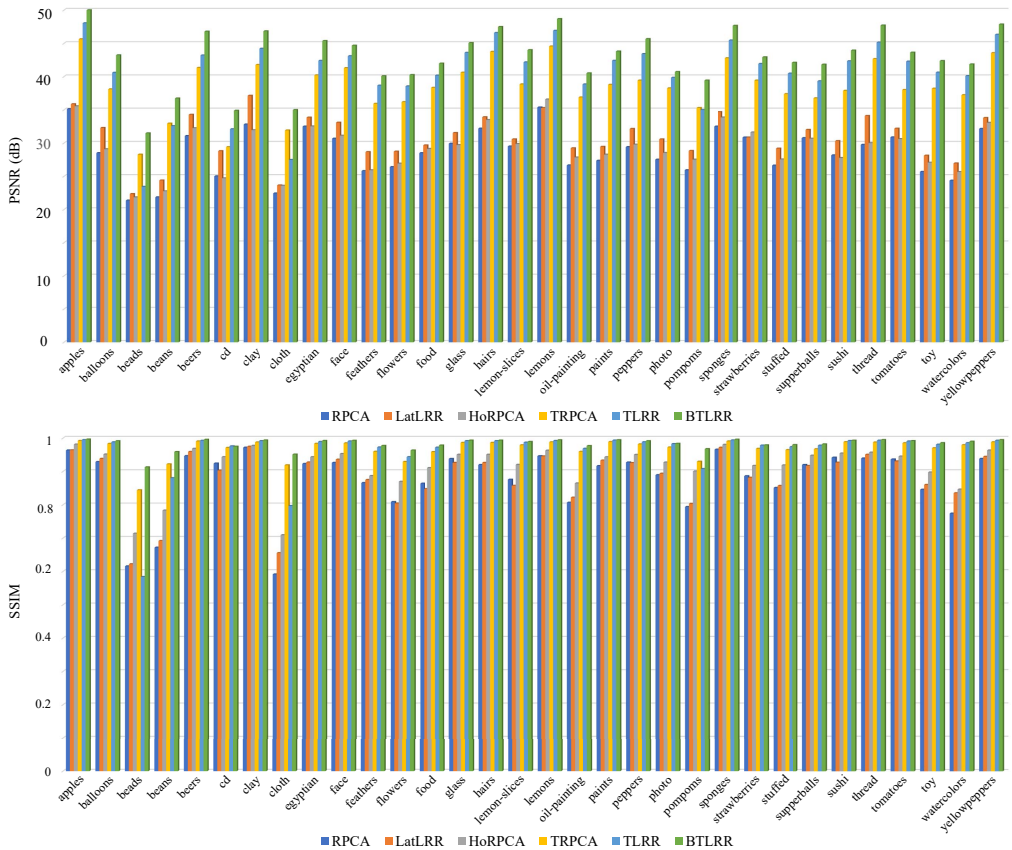


Figure 11. *Denoising performance on all MSIs under NR* = 0.3. *Top: PSNR (dB); bottom: SSIM.*

Table 5
Recovery performance (PSNR, SSIM) on MSIs under different NRs.

Images	Metrics	RPCA	LatLRR	HoRPCA	TRPCA	TLRR	BTLRR
Toy (NR = 0.2)	PSNR	27.17	29.71	28.79	41.71	43.91	**47.68**
	SSIM	0.8909	0.9162	0.9283	0.9906	0.9912	**0.9940**
Feather (NR = 0.2)	PSNR	26.72	30.16	26.91	38.90	41.63	**44.72**
	SSIM	0.8999	0.9256	0.9127	0.9858	0.9878	**0.9915**
Flowers (NR = 0.2)	PSNR	27.60	30.94	28.13	39.73	42.86	**45.60**
	SSIM	0.8676	0.8943	0.9056	0.9846	0.9860	**0.9893**
Beads (NR = 0.3)	PSNR	21.36	22.34	21.84	28.27	23.43	**31.48**
	SSIM	0.6156	0.6219	0.7137	0.8439	0.5836	**0.9127**
Flowers (NR = 0.3)	PSNR	26.40	28.73	26.92	36.18	38.55	**40.25**
	SSIM	0.8086	0.8042	0.8694	0.9289	0.9434	**0.9633**

Table 6
Recovery performance (PSNR, SSIM) on videos under NR = 0.2.

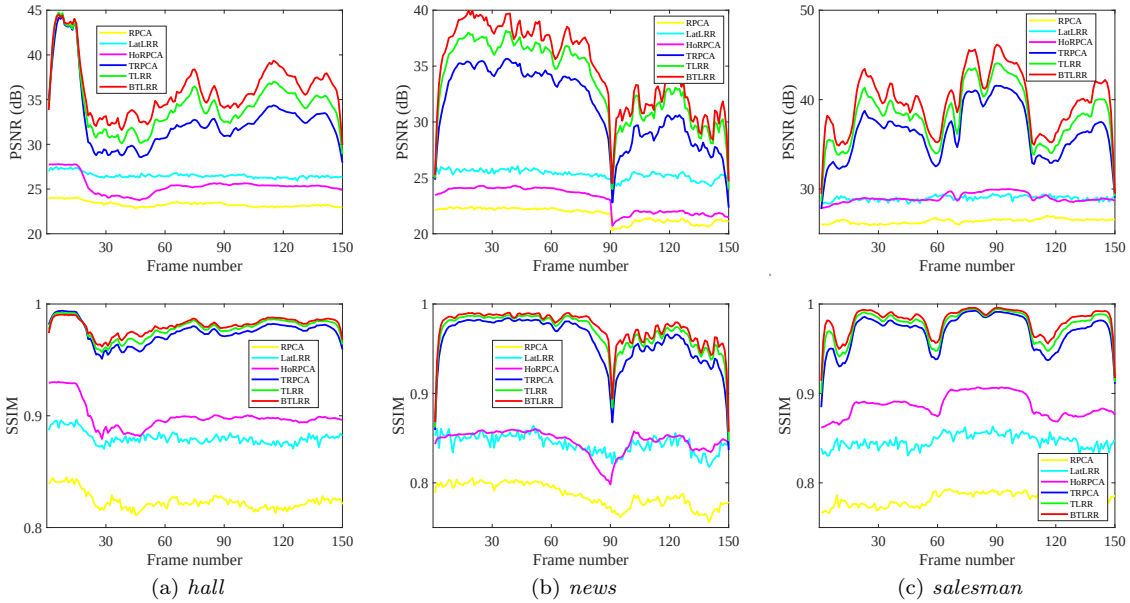
Videos	Metrics	RPCA	LatLRR	HoRPCA	TRPCA	TLRR	BTLRR
Hall	PSNR	23.31	26.53	25.37	32.73	34.55	**36.12**
	SSIM	0.8236	0.8803	0.8991	0.9744	0.9784	**0.9808**
Salesman	PSNR	26.42	28.97	29.03	36.37	38.20	**39.74**
	SSIM	0.7822	0.8468	0.8868	0.9692	0.9763	**0.9816**
Suzie	PSNR	27.02	30.89	27.62	32.95	34.91	**35.59**
	SSIM	0.8263	0.8799	0.8587	0.9287	0.9385	**0.9414**
Foreman	PSNR	23.72	25.66	25.65	30.27	32.23	**33.10**
	SSIM	0.7156	0.7787	0.8355	0.9184	0.9157	**0.9220**
News	PSNR	21.70	25.33	23.04	31.39	33.56	**34.83**
	SSIM	0.7864	0.8438	0.8463	0.9584	0.9679	**0.9732**

videos. One can see that all methods work to a certain extent, outputting less noise images compared to the observed ones. Nevertheless, one can see that the method LatLRR has noise in the recovered videos. The methods RPCA and HoRPCA oversmooth the recovered images. The methods TRPCA and TLRR obtain the smooth regions. In comparison, the proposed BTLRR seems to perform better in preserving the sharp edges and smoothness of the recovered videos.

5.5. Discussions.

5.5.1. Parameter analysis. In Figure 14, we show the sensitivity of the proposed method w.r.t. two parameters λ and μ for data recovery on two sets of data *cd* (NR = 0.3) and *hall* (NR = 0.2). One can see that our method can achieve similar high metrics in a relatively wide range of two parameters. Hence, the proposed method is robust to the change of parameters λ and μ.

5.5.2. Numerical convergence. In Theorem 4.2, we analysis the convergence behavior of the proposed algorithm. In Figure 15, we present the error values of Error $= \max(\|\mathcal{X} - \mathcal{A} * \mathcal{Z}_{\flat,t} - \mathcal{L}_{\flat,t} * \mathcal{B} - \mathcal{E}_t\|_\infty, \|\mathcal{Z}_{\flat,t} - \mathcal{F}_t\|_\infty, \|\mathcal{L}_{\flat,t} - \mathcal{P}_t\|_\infty)$ on tensor data clustering (NR = 0.1)



(a) *hall* (b) *news* (c) *salesman*

Figure 12. *PSNR and SSIM values of each frame of recovered videos by different methods under* $NR = 0.2$.

(a) Noisy (b) RPCA (c) LatLRR (d) HoRPCA (e) TRPCA (f) TLRR (g) BTLRR (h) Original

Figure 13. *Denoising results on three videos under* $NR = 0.2$. *From top to bottom:* hall *(the* $115th$ *frame),* news *(the* $77th$ *frame), and* salesman *(the* $4th$ *frame), respectively.*

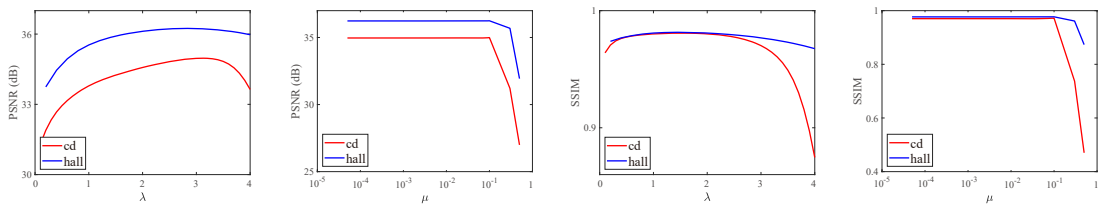
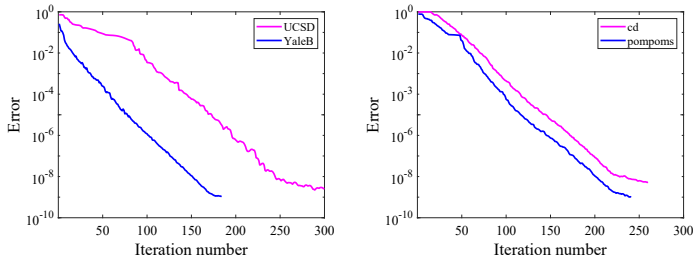


Figure 14. *PSNR (dB) and SSIM values of recovered images under different data,* λ, *and* μ.



(a) data clustering (b) data recovery

Figure 15. *Numerical convergence performance of our algorithm.*

and recovery (NR = 0.3). One can see that the error curves rapidly reach a low value, which demonstrates the numerical convergence performance of the proposed method.

6. Conclusion. In this work, by introducing the hidden tensor data, we propose the BTLRR method for handling the problem of insufficient data. Under mild conditions, we provide the closed-form solutions of the proposed model under both noisy/noiseless cases. Meanwhile, we recover the hidden tensor data effects by optimizing a convex TNN problem, which can be solved efficiently. The proposed BTLRR takes into account the bilateral information, i.e., the similarity between samples and the relationship among features. Experiments on two important multidimensional image analysis tasks (including clustering and recovery) and various datasets demonstrate the superior performance improvements of our BTLRR.

Appendix A. Proof of Theorem 3.1. We first give the following lemma introduced in [62].

Lemma A.1 (see [62]). *Assume that $\mathcal{X}_O = \mathcal{A} * \mathcal{Z}$ with $\mathcal{A} \neq 0$ has feasible solutions, then problem (1.1) admits the unique minimizer as follows,*

$$\mathcal{Z}^\star = \mathcal{A}^\dagger * \mathcal{X}_O,$$

where $\mathcal{A}^\dagger$ is the pseudoinverse of $\mathcal{A}$.

Based on Lemma A.1, we detail the proof of Theorem 3.1.

Proof. Based on the definition of the skinny TSVD, one can calculate that the equation $\mathcal{X}_O = [\mathcal{X}_O, \mathcal{X}_H] * \mathcal{Z}_{O,H}$ is equal to $\mathcal{U} * \mathcal{S} * \mathcal{V}_O^\top = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top * \mathcal{Z}_{O,H}$, which can also be calculated as $\mathcal{V}_O^\top = \mathcal{V}^\top * \mathcal{Z}_{O,H}$, where $\mathcal{V} = [\mathcal{V}_O; \mathcal{V}_H]$ and $\mathcal{V}_O$ and $\mathcal{V}_H$, respectively, correspond to $\mathcal{X}_O$ and $\mathcal{X}_H$. So problem (1.3) can be equally reformulated as the following optimization problem:

$$\min_{\mathcal{Z}_{O,H}} \|\mathcal{Z}_{O,H}\|_* \text{ s.t. } \mathcal{V}_O^\top = \mathcal{V}^\top * \mathcal{Z}_{O,H}.$$

By Lemma A.1, problem (1.3) admits a unique minimizer

$$\mathcal{Z}_{O,H}^\star = \mathcal{V} * \mathcal{V}_O^\top = [\mathcal{V}_O * \mathcal{V}_O^\top; \mathcal{V}_H * \mathcal{V}_O^\top].$$

According to $\mathcal{Z}_{O,H}^\star = [\mathcal{Z}_O^\star; \mathcal{Z}_H^\star]$, we have the conclusions. ∎

Appendix B. Proof of Theorem 3.2. We first present one lemma from [62].

Lemma B.1. *For any four tensors $\mathcal{G}_i$, $i=1,\ldots,4$ of compatible dimensions, the following inequality holds:*

$$\left\|\begin{bmatrix} \mathcal{G}_1 & \mathcal{G}_2 \\ \mathcal{G}_3 & \mathcal{G}_4 \end{bmatrix}\right\|_* \geq \|\mathcal{G}_1\|_* + \|\mathcal{G}_4\|_*.$$

We further obtain the following property.

Proposition B.2. *For any four tensors $\mathcal{G}_i$, $i=1,\ldots,4$ of compatible dimensions, the following inequality holds,*

$$\left\|\begin{bmatrix} \mathcal{G}_1 & \mathcal{G}_2 \\ \mathcal{G}_3 & \mathcal{G}_4 \end{bmatrix}\right\|_* \geq \|\mathcal{G}_1\|_*,$$

where the equality holds if and only if $\mathcal{G}_i = 0$, $i=2,3,4$.

Proof. According to the definition of TNN for any two tensors $\mathcal{H}_1$ and $\mathcal{H}_2$, we have

$$\begin{aligned}\|[\mathcal{H}_1,\mathcal{H}_2]\|_* &= \frac{1}{n_3}\sum_{k=1}^{n_3}\|[\bar{\mathcal{H}}_1^{(k)},\bar{\mathcal{H}}_2^{(k)}]\|_* \\ &\geq \frac{1}{n_3}\sum_{k=1}^{n_3}\|\bar{\mathcal{H}}_1^{(k)}\|_* \\ &= \|\mathcal{H}_1\|_*\end{aligned}$$

and/or

$$\begin{aligned}\|[\mathcal{H}_1;\mathcal{H}_2]\|_* &= \frac{1}{n_3}\sum_{k=1}^{n_3}\|[\bar{\mathcal{H}}_1^{(k)};\bar{\mathcal{H}}_2^{(k)}]\|_* \\ &\geq \frac{1}{n_3}\sum_{k=1}^{n_3}\|\bar{\mathcal{H}}_1^{(k)}\|_* \\ &= \|\mathcal{H}_1\|_*,\end{aligned}$$

where the equality holds if and only if $\mathcal{H}_2 = 0$. Then we can simply complete the proof. ∎

Then we provide the following lemma.

Lemma B.3. *For any tensor $\mathcal{Y}\in\mathbb{R}^{n\times n\times n_3}$, we have $\|\mathcal{Y}\|_* \geq \frac{1}{n_3}\mathrm{tr}(\bar{\boldsymbol{Y}})$, where the equality holds if and only if $\bar{\boldsymbol{Y}}$ is positive semidefinite and symmetric.*

Proof. The proof is simply based on the definition of TNN, Lemma B.1, Proposition B.2, and [57, Lemma 4], so we omit it. ∎

Transforming from (3.4) by eliminating $\mathcal{L}$ therein, according to Lemma A.1, we may consider the following unconstrained problem:

$$\min_{\mathcal{Z}} f(\mathcal{Z}) \triangleq \|\mathcal{Z}\|_* + \|\mathcal{X} * (\mathcal{I} - \mathcal{Z}) * \mathcal{X}^\dagger\|_*. \tag{B.1}$$

Then we have the following result.

Lemma B.4. *Assume that* $\mathcal{U}_\mathcal{X} * \mathcal{S}_\mathcal{X} * \mathcal{V}_\mathcal{X}^\top$ *is the skinny TSVD of* $\mathcal{X}$*, and* $\bar{\boldsymbol{S}}_\mathcal{X} = \texttt{bdiag}(\bar{\mathcal{S}}_\mathcal{X})$ *has full rank. The optimization problem* (B.1) *has a minimum objective function value rank*$(\mathcal{X})$*.*

Proof. From [31], the set of subgradients of the TNN of a tensor $\mathcal{Z}$ is

$$\partial_\mathcal{Z}\|\mathcal{Z}\|_* = \{\mathcal{U}_\mathcal{Z} * \mathcal{V}_\mathcal{Z}^\top + \mathcal{M} | \mathcal{U}_\mathcal{Z}^\top * \mathcal{M} = \mathbf{0}, \mathcal{M} * \mathcal{V}_\mathcal{Z} = \mathbf{0}, \|\mathcal{M}\| \le 1\},$$

where $\mathbf{0}$ denotes the zero tensor, $\|\mathcal{M}\|$ is the tensor spectral norm [31], and $\mathcal{U}_\mathcal{Z} * \mathcal{S}_\mathcal{Z} * \mathcal{V}_\mathcal{Z}^\top$ is the skinny TSVD of the tensor $\mathcal{Z}$. Next, we prove that $\mathcal{Z}^\star = 1/2\mathcal{X}^\dagger * \mathcal{X}$ is a minimizer to (B.1). One needs to show that

$$\begin{aligned}
&\mathbf{0} \in \partial_\mathcal{Z} f(\mathcal{Z}^\star) \\
&\quad = \partial_\mathcal{Z}\|\mathcal{Z}^\star\|_* + \partial_\mathcal{Z}\|\mathcal{X} * (\mathcal{I} - \mathcal{Z}^\star) * \mathcal{X}^\dagger\|_* \\
&\quad = \partial_\mathcal{Z}\|\mathcal{Z}^\star\|_* - \mathcal{X}^\top * \partial_{\mathcal{X} * (\mathcal{I} - \mathcal{Z}) * \mathcal{X}^\dagger}\|\mathcal{X} * (\mathcal{I} - \mathcal{Z}^\star) * \mathcal{X}^\dagger\|_* * (\mathcal{X}^\dagger)^\top.
\end{aligned}$$

Note that $\mathcal{U}_\mathcal{X} * (1/2\mathcal{I}) * \mathcal{U}_\mathcal{X}^\top = \mathcal{X} * (\mathcal{I} - \mathcal{Z}^\star) * \mathcal{X}^\dagger$ is the skinny TSVD of $\mathcal{X} * (\mathcal{I} - \mathcal{Z}^\star) * \mathcal{X}^\dagger$, and $\mathcal{V}_\mathcal{X} * (1/2\mathcal{I}) * \mathcal{V}_\mathcal{X}^\top = \mathcal{Z}^\star$ is the skinny TSVD of $\mathcal{Z}^\star$, where $\mathcal{U}_\mathcal{X} * \mathcal{S}_\mathcal{X} * \mathcal{V}_\mathcal{X}^\top$ is the skinny TSVD of $\mathcal{X}$. Therefore, $\partial_\mathcal{Z} f(\mathcal{Z}^\star)$ contains

$$\begin{aligned}
&\mathcal{V}_\mathcal{X} * \mathcal{V}_\mathcal{X}^\top - \mathcal{X}^\top * \mathcal{U}_\mathcal{X} * \mathcal{U}_\mathcal{X}^\top * (\mathcal{X}^\dagger)^\top \\
&= \mathcal{V}_\mathcal{X} * \mathcal{V}_\mathcal{X}^\top - \mathcal{V}_\mathcal{X} * \mathcal{S}_\mathcal{X} * \mathcal{U}_\mathcal{X}^\top * \mathcal{U}_\mathcal{X} * \mathcal{U}_\mathcal{X}^\top * \mathcal{U}_\mathcal{X} * \mathcal{S}_\mathcal{X}^\dagger * \mathcal{V}_\mathcal{X}^\top \\
&= \mathbf{0}.
\end{aligned}$$

Substituting $\mathcal{Z}^\star = 1/2\mathcal{X}^\dagger * \mathcal{X}$ into (B.1), we can get the minimum objective function value $\mathrm{rank}(\mathcal{X})$. ∎

Next, we provide the form of the optimal solution to (B.1).

Lemma B.5. *Assume that* $\mathcal{U}_\mathcal{X} * \mathcal{S}_\mathcal{X} * \mathcal{V}_\mathcal{X}^\top$ *is the skinny TSVD of* $\mathcal{X}$*, and* $\bar{\boldsymbol{S}}_\mathcal{X} = \texttt{bdiag}(\bar{\mathcal{S}}_\mathcal{X})$ *has full rank. The optimal solutions to the unconstrained optimization problem* (B.1) *can be written as* $\mathcal{Z}^\star = \mathcal{V}_\mathcal{X} * \mathcal{W} * \mathcal{V}_\mathcal{X}^\top$*.*

Proof. Let $(\mathcal{V}_\mathcal{X})_\perp$ be the orthogonal complement of $\mathcal{V}_\mathcal{X}$. According to Lemma B.4, $\mathrm{rank}(\mathcal{X})$ is the minimum objective function value of (B.1). Thus we get

$$\begin{aligned}
&\mathrm{rank}(\mathcal{X}) \\
&\quad = \|\mathcal{Z}^\star\|_* + \|\mathcal{X} * (\mathcal{I} - \mathcal{Z}^\star) * \mathcal{X}^\dagger\|_* \\
&\quad = \left\| \begin{bmatrix} \mathcal{V}_\mathcal{X}^\top \\ (\mathcal{V}_\mathcal{X})_\perp^\top \end{bmatrix} * \mathcal{Z}^\star * \begin{bmatrix} \mathcal{V}_\mathcal{X}, (\mathcal{V}_\mathcal{X})_\perp \end{bmatrix} \right\|_* + \|\mathcal{X} * (\mathcal{I} - \mathcal{Z}^\star) * \mathcal{X}^\dagger\|_* \\
&\quad = \left\| \begin{bmatrix} \mathcal{V}_\mathcal{X}^\top * \mathcal{Z}^\star * \mathcal{V}_\mathcal{X} & \mathcal{V}_\mathcal{X}^\top * \mathcal{Z}^\star * (\mathcal{V}_\mathcal{X})_\perp \\ (\mathcal{V}_\mathcal{X})_\perp^\top * \mathcal{Z}^\star * \mathcal{V}_\mathcal{X} & (\mathcal{V}_\mathcal{X})_\perp^\top * \mathcal{Z}^\star * (\mathcal{V}_\mathcal{X})_\perp \end{bmatrix} \right\|_* + \|\mathcal{X} * (\mathcal{I} - \mathcal{Z}^\star) * \mathcal{X}^\dagger\|_* \\
&\quad \ge \|\mathcal{V}_\mathcal{X}^\top * \mathcal{Z}^\star * \mathcal{V}_\mathcal{X}\|_* + \|\mathcal{U}_\mathcal{X} * \mathcal{S}_\mathcal{X} * \mathcal{V}_\mathcal{X}^\top * (\mathcal{I} - \mathcal{Z}^\star) * \mathcal{V}_\mathcal{X} * \mathcal{S}_\mathcal{X}^\dagger * \mathcal{U}_\mathcal{X}^\top\|_* \\
&\quad = \|\mathcal{V}_\mathcal{X} * \mathcal{V}_\mathcal{X}^\top * \mathcal{Z}^\star * \mathcal{V}_\mathcal{X} * \mathcal{V}_\mathcal{X}^\top\|_* \\
&\qquad + \|\mathcal{U}_\mathcal{X} * \mathcal{S}_\mathcal{X} * \mathcal{V}_\mathcal{X}^\top * (\mathcal{I} - \mathcal{V}_\mathcal{X} * \mathcal{V}_\mathcal{X}^\top * \mathcal{Z}^\star * \mathcal{V}_\mathcal{X} * \mathcal{V}_\mathcal{X}^\top) * \mathcal{V}_\mathcal{X} * \mathcal{S}_\mathcal{X}^\dagger * \mathcal{U}_\mathcal{X}^\top\|_* \\
&\quad = \|\mathcal{V}_\mathcal{X} * \mathcal{V}_\mathcal{X}^\top * \mathcal{Z}^\star * \mathcal{V}_\mathcal{X} * \mathcal{V}_\mathcal{X}^\top\|_* + \|\mathcal{X} * (\mathcal{I} - \mathcal{V}_\mathcal{X} * \mathcal{V}_\mathcal{X}^\top * \mathcal{Z}^\star * \mathcal{V}_\mathcal{X} * \mathcal{V}_\mathcal{X}^\top) * \mathcal{X}^\dagger\|_* \\
&\quad \ge \mathrm{rank}(\mathcal{X}),
\end{aligned}$$

where the last inequality holds by viewing $\mathcal{Z} = \mathcal{V}_\mathcal{X} * \mathcal{V}_\mathcal{X}^\top * \mathcal{Z}^\star * \mathcal{V}_\mathcal{X} * \mathcal{V}_\mathcal{X}^\top$ as a feasible solution to (B.1). Then all inequalities must be equalities. By Proposition B.2, we have

$$\mathcal{V}_\mathcal{X}^\top * \mathcal{Z}^\star * (\mathcal{V}_\mathcal{X})_\perp = (\mathcal{V}_\mathcal{X})_\perp^\top * \mathcal{Z}^\star * \mathcal{V}_\mathcal{X} = (\mathcal{V}_\mathcal{X})_\perp^\top * \mathcal{Z}^\star * (\mathcal{V}_\mathcal{X})_\perp = \mathbf{0}.$$

That is

$$\begin{bmatrix} \mathcal{V}_\mathcal{X}^\top \\ (\mathcal{V}_\mathcal{X})_\perp^\top \end{bmatrix} * \mathcal{Z}^\star * \begin{bmatrix} \mathcal{V}_\mathcal{X}, (\mathcal{V}_\mathcal{X})_\perp \end{bmatrix} = \begin{bmatrix} \mathcal{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

where $\mathcal{W} = \mathcal{V}_\mathcal{X}^\top * \mathcal{Z}^\star * \mathcal{V}_\mathcal{X}$. Hence the following equality,

$$\mathcal{Z}^\star = \begin{bmatrix} \mathcal{V}_\mathcal{X}, (\mathcal{V}_\mathcal{X})_\perp \end{bmatrix} * \begin{bmatrix} \mathcal{W} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} * \begin{bmatrix} \mathcal{V}_\mathcal{X}^\top \\ (\mathcal{V}_\mathcal{X})_\perp^\top \end{bmatrix} = \mathcal{V}_\mathcal{X} * \mathcal{W} * \mathcal{V}_\mathcal{X}^\top,$$

holds. ∎

Based on the above lemmas and proposition, we present the following lemma that gives the whole closed-form solutions to the unconstrained optimization problem (B.1).

Lemma B.6. *Assuming that $\mathcal{U}_\mathcal{X} * \mathcal{S}_\mathcal{X} * \mathcal{V}_\mathcal{X}^\top$ is the skinny TSVD of the tensor $\mathcal{X}$ and $\bar{\boldsymbol{S}}_\mathcal{X} = \texttt{bdiag}(\bar{\mathcal{S}}_\mathcal{X})$ has full rank, the solutions to the unconstrained optimization problem (B.1) are $\mathcal{Z}^\star = \mathcal{V}_\mathcal{X} * \mathcal{W} * \mathcal{V}_\mathcal{X}^\top$, where the tensor $\mathcal{W}$ satisfies (1) $\mathcal{W}$ is compatible with $\mathcal{S}_\mathcal{X}$, i.e., $[\bar{\boldsymbol{W}}]_{i,j} = 0$ if $[\bar{\boldsymbol{S}}_\mathcal{X}]_{i,i} \neq [\bar{\boldsymbol{S}}_\mathcal{X}]_{j,j}$, where $\bar{\boldsymbol{W}} = \texttt{bdiag}(\bar{\mathcal{W}})$; (2) both $\bar{\boldsymbol{W}}$ and $\bar{\boldsymbol{I}} - \bar{\boldsymbol{W}}$ are positive semidefinite.*

Proof. First, we prove the necessity. Suppose $\mathcal{Z}^\star$ is a minimizer. By Lemma B.5, $\mathcal{Z}^\star$ can be written as $\mathcal{Z}^\star = \mathcal{V}_\mathcal{X} * \mathcal{W} * \mathcal{V}_\mathcal{X}^\top$. We show that $\bar{\boldsymbol{W}}$ satisfies the stated conditions. According to Lemma B.3, we have

$$\begin{aligned}
&\operatorname{rank}(\mathcal{X}) \\
&\quad = \|\mathcal{Z}^\star\|_* + \|\mathcal{X} * (\mathcal{I} - \mathcal{Z}^\star) * \mathcal{X}^\dagger\|_* \\
&\quad = \|\mathcal{W}\|_* + \|\mathcal{S}_\mathcal{X} * (\mathcal{I} - \mathcal{W}) * \mathcal{S}_\mathcal{X}^\dagger\|_* \\
&\quad \geq \|\mathcal{W}\|_* + 1/n_3 \operatorname{tr}(\bar{\boldsymbol{S}}_\mathcal{X}(\bar{\boldsymbol{I}} - \bar{\boldsymbol{W}})\bar{\boldsymbol{S}}_\mathcal{X}^\dagger) \\
&\quad = \|\mathcal{W}\|_* + 1/n_3 \operatorname{tr}(\bar{\boldsymbol{I}} - \bar{\boldsymbol{W}}) \\
&\quad = \|\mathcal{W}\|_* + \operatorname{rank}(\mathcal{X}) - 1/n_3 \operatorname{tr}(\bar{\boldsymbol{W}}) \\
&\quad \geq \operatorname{rank}(\mathcal{X}).
\end{aligned}$$

Then the all inequalities must be equalities. From $\|\mathcal{W}\|_* = 1/n_3 \operatorname{tr}(\bar{\boldsymbol{W}})$ and Lemma B.3, we get that $\bar{\boldsymbol{W}}$ is positive semidefinite. By the first inequality and Lemma B.3 and [57, Lemma 4], we get that $\bar{\boldsymbol{S}}_\mathcal{X}(\bar{\boldsymbol{I}} - \bar{\boldsymbol{W}})\bar{\boldsymbol{S}}_\mathcal{X}^\dagger$ is positive semidefinite and symmetric, i.e.,

$$\frac{[\bar{\boldsymbol{S}}_\mathcal{X}]_{i,i}}{[\bar{\boldsymbol{S}}_\mathcal{X}]_{j,j}}[\bar{\boldsymbol{I}} - \bar{\boldsymbol{W}}]_{i,j} = \frac{[\bar{\boldsymbol{S}}_\mathcal{X}]_{j,j}}{[\bar{\boldsymbol{S}}_\mathcal{X}]_{i,i}}[\bar{\boldsymbol{I}} - \bar{\boldsymbol{W}}]_{i,j},$$

then $\bar{\boldsymbol{W}}_{i,j} = 0$ if $[\bar{\boldsymbol{S}}_\mathcal{X}]_{i,i} \neq [\bar{\boldsymbol{S}}_\mathcal{X}]_{j,j}$. Notice that $\bar{\boldsymbol{I}} - \bar{\boldsymbol{W}} = \bar{\boldsymbol{S}}_\mathcal{X}(\bar{\boldsymbol{I}} - \bar{\boldsymbol{W}})\bar{\boldsymbol{S}}_\mathcal{X}^\dagger$, then $\bar{\boldsymbol{I}} - \bar{\boldsymbol{W}}$ is also positive semidefinite.

Next, we prove the sufficiency. Suppose that $\mathcal{Z}^\star = \mathcal{V}_\mathcal{X} * \mathcal{W} * \mathcal{V}_\mathcal{X}^\top$ satisfies all conditions in the theorem. Substituting $\mathcal{Z}^\star$ into the objective function, we have

$$\begin{aligned}
&\|\mathcal{Z}^\star\|_* + \|\mathcal{X} * (\mathcal{I} - \mathcal{Z}^\star) * \mathcal{X}^\dagger\|_* \\
&= \|\mathcal{W}\|_* + \|\mathcal{S}_\mathcal{X} * (\mathcal{I} - \mathcal{W}) * \mathcal{S}_\mathcal{X}^\dagger\|_* \\
&= \|\mathcal{W}\|_* + 1/n_3 \mathrm{tr}(\bar{\boldsymbol{S}}_\mathcal{X}(\bar{\boldsymbol{I}} - \bar{\boldsymbol{W}})\bar{\boldsymbol{S}}_\mathcal{X}^\dagger) \\
&= \|\mathcal{W}\|_* + 1/n_3 \mathrm{tr}(\bar{\boldsymbol{I}} - \bar{\boldsymbol{W}}) \\
&= \|\mathcal{W}\|_* + \mathrm{rank}(\mathcal{X}) - 1/n_3 \mathrm{tr}(\bar{\boldsymbol{W}}) \\
&= \mathrm{rank}(\mathcal{X}) \\
&= \min_\mathcal{Z} \|\mathcal{Z}\|_* + \|\mathcal{X} * (\mathcal{I} - \mathcal{Z}) * \mathcal{X}^\dagger\|_*.
\end{aligned}$$

Based on Lemma B.3, since $\bar{\boldsymbol{I}} - \bar{\boldsymbol{W}} = \bar{\boldsymbol{S}}_\mathcal{X}(\bar{\boldsymbol{I}} - \bar{\boldsymbol{W}})\bar{\boldsymbol{S}}_\mathcal{X}^\dagger$, and both $\bar{\boldsymbol{W}}$ and $\bar{\boldsymbol{I}} - \bar{\boldsymbol{W}}$ are positive semidefinite, the second and fifth equalities hold. This completes the proof. ∎

Next, we prove Theorem 3.2.

Proof. Let $\mathcal{W}$ satisfy all the conditions in the theorem. Since the row space of $\mathcal{Z}^\star = \mathcal{V}_\mathcal{X} * \mathcal{W} * \mathcal{V}_\mathcal{X}^\top$ belongs to that of $\mathcal{X}$, one can get that $(\mathcal{Z}^\star, \mathcal{X} * (\mathcal{I} - \mathcal{Z}^\star) * \mathcal{X}^\dagger)$ is feasible to problem (3.4) according to Lemma B.5. Assume that there exists another solution $(\mathcal{Z}', \mathcal{L}')$ which satisfies

$$\mathcal{X} = \mathcal{X} * \mathcal{Z}' + \mathcal{L}' * \mathcal{X},$$

$$\|\mathcal{Z}'\|_* + \|\mathcal{L}'\|_* < \|\mathcal{Z}^\star\|_* + \|\mathcal{L}^\star\|_*.$$

Fixing $\mathcal{Z}$ in problem (3.4) and by Lemma A.1, we get

$$\|\mathcal{Z}'\|_* + \|(\mathcal{X} - \mathcal{X} * \mathcal{Z}') * \mathcal{X}^\dagger\|_* \le \|\mathcal{Z}'\|_* + \|\mathcal{L}'\|_*.$$

Thus

$$\|\mathcal{Z}'\|_* + \|(\mathcal{X} - \mathcal{X} * \mathcal{Z}') * \mathcal{X}^\dagger\|_* < \|\mathcal{Z}^\star\|_* + \|(\mathcal{X} - \mathcal{X} * \mathcal{Z}^\star) * \mathcal{X}^\dagger\|_*,$$

which is a contradiction w.r.t. the optimality of $\mathcal{Z}^\star$ in Lemma B.6. Hence the proof is completed. ∎

Appendix C. Proof of Theorem 3.3.

Equipped with Theorem 3.2, we prove Theorem 3.3.

Proof. We first prove the first part of the theorem. Suppose that $(\mathcal{Z}', \mathcal{L}', \mathcal{E}')$ is a better solution than $(\mathcal{Z}^\star, \mathcal{L}^\star, \mathcal{E}^\star)$ to the relaxed problem (3.6), i.e.,

$$\|\mathcal{Z}'\|_* + \|\mathcal{L}'\|_* + \lambda\|\mathcal{E}'\|_1 < \|\mathcal{Z}^\star\|_* + \|\mathcal{L}^\star\|_* + \lambda\|\mathcal{E}^\star\|_1. \tag{C.1}$$

Without loss of generality, we assume that $(\mathcal{Z}', \mathcal{L}', \mathcal{E}')$ is the optimal solution to problem (3.6). According to the form (3.5) in Theorem 3.2, $(\mathcal{Z}', \mathcal{L}', \mathcal{E}')$ can be reformulated as follows,

$$\mathcal{Z}' = \mathcal{V}_{\hat{\mathcal{X}}} * \mathcal{W} * \mathcal{V}_{\hat{\mathcal{X}}}^\top \text{ and } \mathcal{L}' = \mathcal{U}_{\hat{\mathcal{X}}} * (\mathcal{I} - \mathcal{W}) * \mathcal{U}_{\hat{\mathcal{X}}}^\top, \tag{C.2}$$

where $\hat{\mathcal{X}} = \mathcal{X} - \mathcal{E}'$, and $\mathcal{W}$ satisfies the conditions in Theorem 3.2. Substituting (C.2) into the objective function of problem (3.6), we have

$$\|\mathcal{Z}'\|_* + \|\mathcal{L}'\|_* + \lambda\|\mathcal{E}'\|_1 = \operatorname{rank}(\mathcal{X} - \mathcal{E}') + \lambda\|\mathcal{E}'\|_1, \tag{C.3}$$

where the equality $\|\mathcal{Z}'\|_* + \|\mathcal{L}'\|_* = \operatorname{rank}(\hat{\mathcal{X}}) = \operatorname{rank}(\mathcal{X} - \mathcal{E}')$ is guaranteed by the conditions in Theorem 3.2. On the other hand, taking the equation in (3.7) into the objective function of problem (3.6) and using the conditions in the theorem, we have

$$\|\mathcal{Z}^\star\|_* + \|\mathcal{L}^\star\|_* + \lambda\|\mathcal{E}^\star\|_1 = \operatorname{rank}(\mathcal{X} - \mathcal{E}^\star) + \lambda\|\mathcal{E}^\star\|_1. \tag{C.4}$$

From (C.1), (C.3), and (C.4), we obtain

$$\operatorname{rank}(\mathcal{X} - \mathcal{E}') + \lambda\|\mathcal{E}'\|_1 < \operatorname{rank}(\mathcal{X} - \mathcal{E}^\star) + \lambda\|\mathcal{E}^\star\|_1, \tag{C.5}$$

which results in a contradiction to the assumption that $(\hat{\mathcal{X}}^\star, \mathcal{E}^\star)$ is optimal to TRPCA.

We then prove the second part. Suppose the TRPCA problem (2.1) has a better solution $(\hat{\mathcal{X}}', \mathcal{E}')$ than $(\mathcal{X} - \mathcal{E}^\star, \mathcal{E}^\star)$:

$$\operatorname{rank}(\hat{\mathcal{X}}') + \lambda\|\mathcal{E}'\|_1 < \operatorname{rank}(\mathcal{X} - \mathcal{E}^\star) + \lambda\|\mathcal{E}^\star\|_1.$$

On one hand, we have

$$\operatorname{rank}(\hat{\mathcal{X}}') + \lambda\|\mathcal{E}'\|_1 = \|(\hat{\mathcal{X}}')^\dagger * \hat{\mathcal{X}}'\|_* + \|\mathbf{0}\|_* + \lambda\|\mathcal{E}^\star\|_1.$$

On the other hand, since $(\mathcal{Z}^\star, \mathcal{L}^\star, \mathcal{E}^\star)$ is optimal to the relaxed BTLRR (3.6), with conditions in Theorem 3.2 satisfied, it can be written as

$$\mathcal{Z}^\star = \mathcal{V}_{\hat{\mathcal{X}}} * \mathcal{W} * \mathcal{V}_{\hat{\mathcal{X}}}^\top \text{ and } \mathcal{L}^\star = \mathcal{U}_{\hat{\mathcal{X}}} * (\mathcal{I} - \mathcal{W}) * \mathcal{U}_{\hat{\mathcal{X}}}^\top, \tag{C.6}$$

where $\hat{\mathcal{X}} = \mathcal{X} - \mathcal{E}^\star$. Taking (C.6) into the objective function of problem (3.6), we have

$$\operatorname{rank}(\mathcal{X} - \mathcal{E}^\star) + \lambda\|\mathcal{E}^\star\|_1 = \|\mathcal{Z}^\star\|_* + \|\mathcal{L}^\star\|_* + \lambda\|\mathcal{E}^\star\|_1. \tag{C.7}$$

The above equality is guaranteed by the conditions in Theorem 3.2. So we get the following inequality:

$$\|(\hat{\mathcal{X}}')^\dagger * \hat{\mathcal{X}}'\|_* + \|\mathbf{0}\|_* + \lambda\|\mathcal{E}^\star\|_1 < \|\mathcal{Z}^\star\|_* + \|\mathcal{L}^\star\|_* + \lambda\|\mathcal{E}^\star\|_1.$$

Since $((\hat{\mathcal{X}}')^\dagger * \hat{\mathcal{X}}', \mathbf{0}, \mathcal{E}^\star)$ is feasible for (3.6), the above inequality is contradictory to the fact that $(\mathcal{Z}^\star, \mathcal{L}^\star, \mathcal{E}^\star)$ is any optimal solution to the problem (3.6). This completes the proof. ∎

Appendix D. Proof of Theorem 4.1.

Proof. Assume that $(\mathcal{Z}^\star, \mathcal{L}^\star, \mathcal{E}^\star)$ is a minimizer to problem (3.8). Then we have $\mathcal{Z}^\star = \widetilde{\mathcal{X}}^\dagger * (\mathcal{X} - \mathcal{L}^\star * \widetilde{\mathcal{X}} - \mathcal{E}^\star)$ and $\mathcal{L}^\star = (\mathcal{X} - \widetilde{\mathcal{X}} * \mathcal{Z}^\star - \mathcal{E}^\star) * \widetilde{\mathcal{X}}^\dagger$. So there exists $\mathcal{Z}_\flat^\star$ such that $\mathcal{Z}^\star = \mathcal{V}_{\widetilde{\mathcal{X}}} * \mathcal{Z}_\flat^\star$ and $\mathcal{L}_\flat^\star$ such that $\mathcal{L}^\star = \mathcal{L}_\flat^\star * \mathcal{U}_{\widetilde{\mathcal{X}}}^\top$, where $\mathcal{U}_{\widetilde{\mathcal{X}}} * \mathcal{S}_{\widetilde{\mathcal{X}}} * \mathcal{V}_{\widetilde{\mathcal{X}}}^\top$ is the skinny TSVD of $\widetilde{\mathcal{X}}$. In addition, we have

$$\begin{aligned}
\|\mathcal{Z}^\star\|_* &= \|\mathcal{V}_{\widetilde{\mathcal{X}}} * \mathcal{Z}_\flat^\star\|_* \\
&= \frac{1}{n_3}\sum_{k=1}^{n_3} \|\bar{\mathcal{V}}_{\widetilde{\mathcal{X}}}^{(k)} (\bar{\mathcal{Z}}_\flat^\star)^{(k)}\|_* \\
&= \frac{1}{n_3}\sum_{k=1}^{n_3} \operatorname{tr}\left(\sqrt{((\bar{\mathcal{Z}}_\flat^\star)^{(k)})^\top (\bar{\mathcal{V}}_{\widetilde{\mathcal{X}}}^{(k)})^\top \bar{\mathcal{V}}_{\widetilde{\mathcal{X}}}^{(k)} (\bar{\mathcal{Z}}_\flat^\star)^{(k)}}\right) \\
&\overset{(a)}{=} \frac{1}{n_3}\sum_{k=1}^{n_3} \operatorname{tr}\left(\sqrt{((\bar{\mathcal{Z}}_\flat^\star)^{(k)})^\top (\bar{\mathcal{Z}}_\flat^\star)^{(k)}}\right) \\
&= \frac{1}{n_3}\sum_{k=1}^{n_3} \|(\bar{\mathcal{Z}}_\flat^\star)^{(k)}\|_* \\
&= \|\mathcal{Z}_\flat^\star\|_*
\end{aligned}$$

and

$$\begin{aligned}
\|\mathcal{L}^\star\|_* &= \|\mathcal{L}_\flat^\star * \mathcal{U}_{\widetilde{\mathcal{X}}}^\top\|_* \\
&= \frac{1}{n_3}\sum_{k=1}^{n_3} \|(\bar{\mathcal{L}}_\flat^\star)^{(k)} (\bar{\mathcal{U}}_{\widetilde{\mathcal{X}}}^\top)^{(k)}\|_* \\
&= \frac{1}{n_3}\sum_{k=1}^{n_3} \operatorname{tr}\left(\sqrt{(\bar{\mathcal{L}}_\flat^\star)^{(k)} (\bar{\mathcal{U}}_{\widetilde{\mathcal{X}}}^\top)^{(k)} (\bar{\mathcal{U}}_{\widetilde{\mathcal{X}}})^{(k)} ((\bar{\mathcal{L}}_\flat^\star)^{(k)})^\top}\right) \\
&\overset{(b)}{=} \frac{1}{n_3}\sum_{k=1}^{n_3} \operatorname{tr}\left(\sqrt{(\bar{\mathcal{L}}_\flat^\star)^{(k)} ((\bar{\mathcal{L}}_\flat^\star)^{(k)})^\top}\right) \\
&= \frac{1}{n_3}\sum_{k=1}^{n_3} \|(\bar{\mathcal{L}}_\flat^\star)^{(k)}\|_* \\
&= \|\mathcal{L}_\flat^\star\|_*,
\end{aligned}$$

where (a) and (b) hold since $(\bar{\mathcal{V}}_{\widetilde{\mathcal{X}}}^{(k)})^\top \bar{\mathcal{V}}_{\widetilde{\mathcal{X}}}^{(k)} = \boldsymbol{I}$ and $(\bar{\mathcal{U}}_{\widetilde{\mathcal{X}}}^{(k)})^\top \bar{\mathcal{U}}_{\widetilde{\mathcal{X}}}^{(k)} = \boldsymbol{I}$. Then, we substitute $\mathcal{Z}^\star = \mathcal{V}_{\widetilde{\mathcal{X}}} * \mathcal{Z}_\flat^\star$ and $\mathcal{L}^\star = \mathcal{L}_\flat^\star * \mathcal{U}_{\widetilde{\mathcal{X}}}^\top$ into problem (3.8) and transform it into the equivalent problem (4.1). Therefore, we can get that if $(\mathcal{Z}_\flat^\star, \mathcal{L}_\flat^\star, \mathcal{E}^\star)$ is a minimizer to (4.1), problem (3.8) admits a solution, i.e., $(\mathcal{V}_{\widetilde{\mathcal{X}}} * \mathcal{Z}_\flat^\star, \mathcal{L}_\flat^\star * \mathcal{U}_{\widetilde{\mathcal{X}}}^\top, \mathcal{E}^\star)$. ∎

Appendix E. Proof of Theorem 4.2.

Before proving Theorem 4.2, we present some important lemmas.

Lemma E.1 (Theorem 4 of [28]). *Assume that $\boldsymbol{P} \in \mathbb{R}^{N\times N}$, and $\|\cdot\|$ is a unitary invariant matrix norm. Let $\boldsymbol{Q} \in \mathbb{R}^{N\times N}$ satisfying $\boldsymbol{Q} \in \partial\|\boldsymbol{P}\|$, where $\partial\|\boldsymbol{P}\|$ denotes the set of subdifferentials of $\|\cdot\|$ at $\boldsymbol{P}$. Then $\|\boldsymbol{Q}\|^* \le 1$, where $\|\cdot\|^*$ is the dual norm of $\|\cdot\|$.*

Lemma E.2. *The sequences* $\{\mathcal{G}_t\}$, $\{\mathcal{Q}_t\}$, *and* $\{\mathcal{Y}_t\}$ *generated by Algorithm* 4.1 *are bounded.*

Proof. To minimize $\mathcal{F}$ at the $(t+1)$th iteration in (4.5), the optimal $\mathcal{F}_{t+1}$ should satisfy the following first-order optimality condition:

$$\mathbf{0} \in \partial\|\mathcal{F}_{t+1}\|_* + \rho_t\left(\mathcal{F}_{t+1} - \mathcal{Z}_{\flat,t+1} - \mathcal{G}_t/\rho_t\right).$$

From (4.10), i.e., $\mathcal{G}_{t+1} = \mathcal{G}_t + \rho_t(\mathcal{Z}_{\flat,t+1} - \mathcal{F}_{t+1})$, we get $\mathcal{G}_{t+1} \in \partial\|\mathcal{F}_{t+1}\|_*$. According to $\|\mathcal{F}_{t+1}\|_* = \frac{1}{n_3}\sum_{k=1}^{n_3}\|\bar{\mathcal{F}}_{t+1}^{(k)}\|_*$, we have $(\frac{\partial\|\mathcal{F}_{t+1}\|_*}{\partial\bar{\mathcal{F}}_{t+1}})^{(k)} = \frac{1}{n_3}\frac{\partial\|\bar{\mathcal{F}}_{t+1}^{(k)}\|_*}{\partial\bar{\mathcal{F}}_{t+1}^{(k)}}$ $(k = 1,\ldots,n_3)$, where $\frac{\partial\|\bar{\mathcal{F}}_{t+1}^{(k)}\|_*}{\partial\bar{\mathcal{F}}_{t+1}^{(k)}} = \{\boldsymbol{U}_{k,t+1}\boldsymbol{V}_{k,t+1}^\top + \boldsymbol{M}_{k,t+1} | \boldsymbol{U}_{k,t+1}^\top\boldsymbol{M}_{k,t+1} = 0, \boldsymbol{M}_{k,t+1}\boldsymbol{V}_{k,t+1} = 0, \|\boldsymbol{M}_{k,t+1}\| \le 1\}$ [50], $\|\boldsymbol{M}_{k,t+1}\|$ is the spectral norm of $\boldsymbol{M}_{k,t+1}$, and $\boldsymbol{U}_{k,t+1}\boldsymbol{S}_{k,t+1}\boldsymbol{V}_{k,t+1}^\top$ is the skinny SVD of $\bar{\mathcal{F}}_{t+1}^{(k)}$. Noting that the dual norm of the matrix nuclear norm is the spectral norm [40], by using Lemma E.1 and setting $\boldsymbol{G}_{k,t+1} \in \frac{\partial\|\bar{\mathcal{F}}_{t+1}^{(k)}\|_*}{\partial\bar{\mathcal{F}}_{t+1}^{(k)}}$, we have $\|\boldsymbol{G}_{k,t+1}\| \le 1$, then $\|\boldsymbol{G}_{k,t+1}\|_F^2 \le r_{\mathcal{F}_{t+1}}$, where $r_{\mathcal{F}_{t+1}} = \operatorname{rank}(\mathcal{F}_{t+1})$. From $\mathcal{G}_{t+1} \in \frac{\partial\|\mathcal{F}_{t+1}\|_*}{\partial\mathcal{F}_{t+1}}$ and $\bar{\mathcal{F}}_{t+1} = \mathcal{F}_{t+1} \times_3 \boldsymbol{F}_{n_3}$,[3] where $\boldsymbol{F}_{n_3} \in \mathbb{R}^{n_3\times n_3}$ is the DFT matrix, and using the chain rule of matrix calculus, then $\mathcal{G}_{t+1} \times_3 \boldsymbol{F}_{n_3} \in \frac{\partial\|\mathcal{F}_{t+1}\|_*}{\partial\mathcal{F}_{t+1}} \times_3 \boldsymbol{F}_{n_3}$, i.e., $\bar{\mathcal{G}}_{t+1} \in \frac{\partial\|\mathcal{F}_{t+1}\|_*}{\partial\bar{\mathcal{F}}_{t+1}}$. From $\|\boldsymbol{G}_{k,t+1}\|_F^2 \le r_{\mathcal{F}_{t+1}}$, we have $\|\bar{\mathcal{G}}_{t+1}\|_F^2 \le \frac{r_{\mathcal{F}_{t+1}}}{n_3}$. Then $\|\mathcal{G}_{t+1}\|_F^2 = \frac{1}{n_3}\|\bar{\mathcal{G}}_{t+1}\|_F^2 \le \frac{r_{\mathcal{F}_{t+1}}}{n_3^2}$ is bounded.

Similarly, to minimize $\mathcal{P}$ at the $(t+1)$th iteration in (4.7), the optimal $\mathcal{P}_{t+1}$ satisfies the following optimal condition:

$$\mathbf{0} \in \partial\|\mathcal{P}_{t+1}\|_* + \theta_t(\mathcal{P}_{t+1} - \mathcal{L}_{\flat,t+1} - \mathcal{Q}_t/\theta_t).$$

Combining with (4.11), i.e., $\mathcal{Q}_{t+1} = \mathcal{Q}_t + \theta_t\left(\mathcal{L}_{\flat,t+1} - \mathcal{P}_{t+1}\right)$, we can get that $\mathcal{Q}_{t+1} \in \partial\|\mathcal{P}_{t+1}\|_*$. Similarly to the boundedness of $\{\mathcal{G}_{t+1}\}$, we have $\|\frac{\partial\|\mathcal{P}_{t+1}\|_*}{\partial\mathcal{P}}\|_F^2 = \|\frac{\partial\|\mathcal{P}_{t+1}\|_*}{\partial\bar{\mathcal{P}}} \times_3 \boldsymbol{F}_{n_3}^{-1}\|_F^2 \le \frac{r_{\mathcal{P}_{t+1}}}{n_3^2}$ is bounded, where $r_{\mathcal{P}_{t+1}} = \operatorname{rank}(\mathcal{P}_{t+1})$. Then from $\mathcal{Q}_{t+1} \in \partial\|\mathcal{P}_{t+1}\|_*$, the sequence $\{\mathcal{Q}_{t+1}\}$ is also bounded.

Last, to minimize $\mathcal{E}$ at the $(t+1)$th iteration in (4.8), the optimal $\mathcal{E}_{t+1}$ should satisfy the first-order optimization condition

$$\mathbf{0} \in \frac{\lambda}{\mu_t}\partial\|\mathcal{E}_{t+1}\|_1 + \mathcal{E}_{t+1} - \mathcal{D}_{t+1},$$

where $\mathcal{D}_{t+1} = \mathcal{X} - \mathcal{A} * \mathcal{Z}_{\flat,t+1} - \mathcal{L}_{\flat,t+1} * \mathcal{B} + \mathcal{Y}_t/\mu_t$. Using (4.9), we can get $\mathcal{Y}_{t+1} = \mu_t(\mathcal{D}_{t+1} - \mathcal{E}_{t+1})$. Thus

$$\mathbf{0} \in \lambda\partial\|\mathcal{E}_{t+1}\|_1 - \mathcal{Y}_{t+1} \Rightarrow \mathcal{Y}_{t+1} \in \lambda\partial\|\mathcal{E}_{t+1}\|_1.$$

According to Lemma E.1, and the fact that the dual of the ℓ_1-norm is the ℓ_∞-norm, we have $\|\mathcal{Y}_{t+1}\|_\infty \le \lambda$, which is bounded. Then we get that the sequence $\{\mathcal{Y}_{t+1}\}$ is bounded. ∎

[3]The notation $\times_3$ denotes the mode-3 product of a tensor and a matrix [22]. To be specific, the mode-n product of $\mathcal{Y} \in \mathbb{R}^{m_1\times m_2\times\cdots\times m_p}$ and $\boldsymbol{A} \in \mathbb{R}^{l\times m_n}$ is denoted as $\mathcal{Y}\times_n \boldsymbol{A}$ of size $m_1\times\cdots\times m_{n-1}\times l\times m_{n+1}\times\cdots\times m_p$, each element is $(\mathcal{Y}\times_n \boldsymbol{A})_{i_1,\ldots,i_{n-1},j,i_{n+1},\ldots,i_p} = \sum_{i_n=1}^{m_n}[\mathcal{Y}]_{i_1,i_2,\ldots,i_p}[\boldsymbol{A}]_{j,i_n}$, where $j = 1,\ldots,l$, $[\mathcal{Y}]_{i_1,i_2,\ldots,i_p}$ is the $(i_1,i_2,\ldots,i_p)$th element of $\mathcal{Y}$, and $[\boldsymbol{A}]_{j,i_n}$ is the (j,i_n)th element of $\boldsymbol{A}$.

Lemma E.3. *Let* $\{\mathcal{L}(\mathcal{Z}_{\flat,t};\mathcal{L}_{\flat,t};\mathcal{E}_t;\mathcal{F}_t;\mathcal{G}_t;\mathcal{P}_t;\mathcal{Q}_t;\mathcal{Y}_t;\rho_t;\mu_t;\theta_t)\}$ $(t\geq 1)$ *be the sequence generated by Algorithm* 4.1*, then the sequence*

$$\{\mathcal{L}(\mathcal{Z}_{\flat,t+1};\mathcal{L}_{\flat,t+1};\mathcal{E}_{t+1};\mathcal{F}_{t+1};\mathcal{G}_{t+1};\mathcal{P}_{t+1};\mathcal{Q}_{t+1};\mathcal{Y}_{t+1};\rho_t;\mu_t;\theta_t)\}$$

is bounded.

Proof. From the updating rules of the Lagrange multipliers $\mathcal{G}_t=\mathcal{G}_{t-1}+\rho_{t-1}(\mathcal{Z}_{\flat,t}-\mathcal{F}_t)$, $\mathcal{Q}_t=\mathcal{Q}_{t-1}+\theta_{t-1}\left(\mathcal{L}_{\flat,t}-\mathcal{P}_t\right)$, and $\mathcal{Y}_t=\mathcal{Y}_{t-1}+\mu_{t-1}\left(\mathcal{X}-\mathcal{A}*\mathcal{Z}_{\flat,t}-\mathcal{L}_{\flat,t}*\mathcal{B}-\mathcal{E}_t\right)$, we get that

$$\begin{aligned}
&\mathcal{L}(\mathcal{Z}_{\flat,t};\mathcal{L}_{\flat,t};\mathcal{E}_t;\mathcal{F}_t;\mathcal{G}_t;\mathcal{P}_t;\mathcal{Q}_t;\mathcal{Y}_t;\rho_t;\mu_t;\theta_t)\\
&\quad=\mathcal{L}(\mathcal{Z}_{\flat,t};\mathcal{L}_{\flat,t};\mathcal{E}_t;\mathcal{F}_t;\mathcal{G}_{t-1};\mathcal{P}_t;\mathcal{Q}_{t-1};\mathcal{Y}_{t-1};\rho_{t-1};\mu_{t-1};\theta_{t-1})\\
&\qquad+\frac{\rho_t+\rho_{t-1}}{2}\left\|\mathcal{Z}_{\flat,t}-\mathcal{F}_t\right\|_F^2+\frac{\theta_t+\theta_{t-1}}{2}\left\|\mathcal{L}_{\flat,t}-\mathcal{P}_t\right\|_F^2\\
&\qquad+\frac{\mu_t+\mu_{t-1}}{2}\left\|\mathcal{X}-\mathcal{A}*\mathcal{Z}_{\flat,t}-\mathcal{L}_{\flat,t}*\mathcal{B}-\mathcal{E}_t\right\|_F^2\\
&\quad=\mathcal{L}(\mathcal{Z}_{\flat,t};\mathcal{L}_{\flat,t};\mathcal{E}_t;\mathcal{F}_t;\mathcal{G}_{t-1};\mathcal{P}_t;\mathcal{Q}_{t-1};\mathcal{Y}_{t-1};\rho_{t-1};\mu_{t-1};\theta_{t-1})\\
&\qquad+\frac{\rho_t+\rho_{t-1}}{2\rho_{t-1}^2}\left\|\mathcal{G}_t-\mathcal{G}_{t-1}\right\|_F^2+\frac{\theta_t+\theta_{t-1}}{2\theta_{t-1}^2}\left\|\mathcal{Q}_t-\mathcal{Q}_{t-1}\right\|_F^2+\frac{\mu_t+\mu_{t-1}}{2\mu_{t-1}^2}\left\|\mathcal{Y}_t-\mathcal{Y}_{t-1}\right\|_F^2.
\end{aligned}$$

Thus from Algorithm 4.1, we have

$$\begin{aligned}
&\mathcal{L}(\mathcal{Z}_{\flat,t+1};\mathcal{L}_{\flat,t+1};\mathcal{E}_{t+1};\mathcal{F}_{t+1};\mathcal{G}_t;\mathcal{P}_{t+1};\mathcal{Q}_t;\mathcal{Y}_t;\rho_t;\mu_t;\theta_t)\\
&\quad\leq\mathcal{L}(\mathcal{Z}_{\flat,t+1};\mathcal{L}_{\flat,t+1};\mathcal{E}_{t+1};\mathcal{F}_{t+1};\mathcal{G}_t;\mathcal{P}_t;\mathcal{Q}_t;\mathcal{Y}_t;\rho_t;\mu_t;\theta_t)\\
&\quad\leq\mathcal{L}(\mathcal{Z}_{\flat,t};\mathcal{L}_{\flat,t};\mathcal{E}_t;\mathcal{F}_t;\mathcal{G}_t;\mathcal{P}_t;\mathcal{Q}_t;\mathcal{Y}_t;\rho_t;\mu_t;\theta_t)\\
&\quad\leq\mathcal{L}(\mathcal{Z}_{\flat,t};\mathcal{L}_{\flat,t};\mathcal{E}_t;\mathcal{F}_t;\mathcal{G}_{t-1};\mathcal{P}_t;\mathcal{Q}_{t-1};\mathcal{Y}_{t-1};\rho_{t-1};\mu_{t-1};\theta_{t-1})\\
&\qquad+\frac{\rho_t+\rho_{t-1}}{2\rho_{t-1}^2}\left\|\mathcal{G}_t-\mathcal{G}_{t-1}\right\|_F^2+\frac{\theta_t+\theta_{t-1}}{2\theta_{t-1}^2}\left\|\mathcal{Q}_t-\mathcal{Q}_{t-1}\right\|_F^2+\frac{\mu_t+\mu_{t-1}}{2\mu_{t-1}^2}\left\|\mathcal{Y}_t-\mathcal{Y}_{t-1}\right\|_F^2\\
&\quad\leq\cdots\\
&\quad\leq\mathcal{L}(\mathcal{Z}_{\flat,1};\mathcal{L}_{\flat,1};\mathcal{E}_1;\mathcal{F}_1;\mathcal{G}_0;\mathcal{P}_1;\mathcal{Q}_0;\mathcal{Y}_0;\rho_0;\mu_0;\theta_0)\\
&\qquad+\sum_{p=1}^{t}\frac{\rho_p+\rho_{p-1}}{2\rho_{p-1}^2}\left\|\mathcal{G}_p-\mathcal{G}_{p-1}\right\|_F^2+\sum_{p=1}^{t}\frac{\theta_p+\theta_{p-1}}{2\theta_{p-1}^2}\left\|\mathcal{Q}_p-\mathcal{Q}_{p-1}\right\|_F^2\\
&\qquad+\sum_{p=1}^{t}\frac{\mu_p+\mu_{p-1}}{2\mu_{p-1}^2}\left\|\mathcal{Y}_p-\mathcal{Y}_{p-1}\right\|_F^2.
\end{aligned}$$

Since $\{\mathcal{G}_t\}$ is bounded, it holds that $\|\mathcal{G}_p-\mathcal{G}_{p-1}\|_F^2$ is also bounded. Similarly, $\|\mathcal{Q}_p-\mathcal{Q}_{p-1}\|_F^2$ and $\|\mathcal{Y}_p-\mathcal{Y}_{p-1}\|_F^2$ are bounded due to the boundedness of $\{\mathcal{Q}_t\}$ and $\{\mathcal{Y}_t\}$. Note that the following,

$$\sum_{p=1}^{\infty}\frac{\rho_p+\rho_{p-1}}{2\rho_{p-1}^2}=\sum_{k=1}^{\infty}\frac{\eta+1}{2\rho_0\eta^{p-1}}=\frac{\eta(\eta+1)}{2\rho_0(\eta-1)},\eta>1,$$

is bounded, which can be applied for θ_p and μ_p. And $\mathcal{L}(\mathcal{Z}_{\flat,1};\mathcal{L}_{\flat,1};\mathcal{E}_1;\mathcal{F}_1;\mathcal{G}_0;\mathcal{P}_1;\mathcal{Q}_0;\mathcal{Y}_0;\rho_0;\mu_0;\theta_0)$ is finite, hence

$$\{\mathcal{L}(\mathcal{Z}_{\flat,t+1};\mathcal{L}_{\flat,t+1};\mathcal{E}_{t+1};\mathcal{F}_{t+1};\mathcal{G}_{t+1};\mathcal{P}_{t+1};\mathcal{Q}_{t+1};\mathcal{Y}_{t+1};\rho_t;\mu_t;\theta_t)\}$$

is bounded. ∎

Lemma E.4. *The sequences* $\{\mathcal{Z}_{\flat,t}\}$, $\{\mathcal{L}_{\flat,t}\}$, $\{\mathcal{E}_t\}$, $\{\mathcal{F}_t\}$, *and* $\{\mathcal{P}_t\}$ *are bounded.*

Proof. We have the following reformulation

$$\begin{aligned}
&\mathcal{L}(\mathcal{Z}_{\flat,t+1};\mathcal{L}_{\flat,t+1};\mathcal{E}_{t+1};\mathcal{F}_{t+1};\mathcal{G}_t;\mathcal{P}_{t+1};\mathcal{Q}_t;\mathcal{Y}_t;\rho_t;\mu_t;\theta_t)+\frac{1}{2\rho_t}\|\mathcal{G}_t\|_F^2+\frac{1}{2\theta_t}\|\mathcal{Q}_t\|_F^2+\frac{1}{2\mu_t}\|\mathcal{Y}_t\|_F^2\\
&=\|\mathcal{F}_{t+1}\|_*+\|\mathcal{P}_{t+1}\|_*+\lambda\|\mathcal{E}_{t+1}\|_1+\frac{\rho_t}{2}\left\|\mathcal{Z}_{t+1}-\mathcal{F}_{t+1}+\frac{\mathcal{G}_t}{\rho_t}\right\|_F^2+\frac{\theta_t}{2}\left\|\mathcal{L}_{t+1}-\mathcal{P}_{t+1}+\frac{\mathcal{Q}_t}{\theta_t}\right\|_F^2\\
&\quad+\frac{\mu_t}{2}\left\|\mathcal{X}-\mathcal{A}*\mathcal{Z}_{\flat,t+1}-\mathcal{L}_{\flat,t+1}*\mathcal{B}-\mathcal{E}_t+\frac{\mathcal{Y}_t}{\mu_t}\right\|_F^2\\
&\geq\|\mathcal{F}_{t+1}\|_*+\|\mathcal{P}_{t+1}\|_*+\lambda\|\mathcal{E}_{t+1}\|_1.
\end{aligned}$$

From Lemmas E.2 and E.3, we thus get that the sequences $\{\mathcal{F}_t\}$, $\{\mathcal{P}_t\}$, and $\{\mathcal{E}_t\}$ are bounded. Since $\mathcal{Z}_{\flat,t+1}=\mathcal{F}_{t+1}+1/\rho_t(\mathcal{G}_{t+1}-\mathcal{G}_t)$, $\mathcal{L}_{\flat,t+1}=\mathcal{P}_{t+1}+1/\theta_t\,(\mathcal{Q}_{t+1}-\mathcal{Q}_t)$ (i.e., (4.10) and (4.11)) and from Lemma E.2, we have $\{\mathcal{Z}_{\flat,t}\}$ and $\{\mathcal{L}_{\flat,t}\}$ are bounded. ∎

Next, we give the proof of Theorem 4.2.

Proof.

1. *Proof of the first part.* From Lemmas E.2 and E.4, the sequence $\{\mathcal{H}_t\}_{t=1}^{\infty}$ generated by Algorithm 4.1 is bounded. According to the Bolzano–Weierstrass theorem, there exists at least one accumulation point of the sequence $\{\mathcal{H}_t\}_{t=1}^{\infty}$. We denote one of the points as

$$\mathcal{H}^\star=\{\mathcal{Z}_\flat^\star,\mathcal{L}_\flat^\star,\mathcal{E}^\star,\mathcal{F}^\star,\mathcal{G}^\star,\mathcal{P}^\star,\mathcal{Q}^\star,\mathcal{Y}^\star\}.$$

Simply, we assume $\{\mathcal{H}_t\}_{t=1}^{\infty}$ converges to $\mathcal{H}^\star$.

By (4.10), taking the limit, we have

$$\lim_{t\to\infty}(\mathcal{Z}_{\flat,t+1}-\mathcal{F}_{t+1})=\lim_{t\to\infty}(\mathcal{G}_{t+1}-\mathcal{G}_t)/\rho_t=0,$$

thus we have $\mathcal{Z}_\flat^\star=\mathcal{F}^\star$.

Similarly, according to (4.11) and (4.9), the following formulations hold:

$$\mathcal{L}_\flat^\star=\mathcal{P}^\star \text{ and } \mathcal{X}=\mathcal{A}*\mathcal{Z}_\flat^\star+\mathcal{L}_\flat^\star*\mathcal{B}+\mathcal{E}^\star.$$

Meanwhile, (4.3) in the $\mathcal{Z}_\flat$-subproblem can be rewritten as

$$\begin{aligned}
&\mu_t\mathcal{A}^\top*(\mathcal{A}*\mathcal{Z}_{\flat,t+1}+\mathcal{L}_{\flat,t}*\mathcal{B}+\mathcal{E}_t-\mathcal{X}-\mathcal{Y}_t/\mu_t)+\rho_t(\mathcal{Z}_{\flat,t+1}-\mathcal{F}_t+\mathcal{G}_t/\rho_t)=0\\
&\quad\Rightarrow\mathcal{A}^\top*\mathcal{Y}^\star-\mathcal{G}^\star=0.
\end{aligned}$$

Similarly, according to (4.4), we have

$$\mu_t(\mathcal{A} * \mathcal{Z}_{\flat,t+1} + \mathcal{L}_{\flat,t+1} * \mathcal{B} + \mathcal{E}_t - \mathcal{X} - \mathcal{Y}_t/\mu_t) * \mathcal{B}^\top + \theta_t(\mathcal{L}_{\flat,t+1} - \mathcal{P}_t + \mathcal{Q}_t/\theta_t) = 0$$
$$\Rightarrow \mathcal{Y}^\star * \mathcal{B}^\top - \mathcal{Q}^\star = 0.$$

Next, in the $\mathcal{E}$-subproblem (4.8), we have

$$\mathbf{0} \in \lambda \partial \|\mathcal{E}_{t+1}\|_1 - \mathcal{Y}_{t+1} \Rightarrow \mathcal{Y}^\star \in \lambda \partial \|\mathcal{E}^\star\|_1.$$

As shown in the $\mathcal{F}$ and $\mathcal{P}$-subproblems, we have

$$\mathbf{0} \in \partial \left\|\mathcal{F}_{t+1}\right\|_* - \mathcal{G}_{t+1} \Rightarrow \mathcal{G}^\star \in \partial \left\|\mathcal{F}^\star\right\|_*,$$
$$\mathbf{0} \in \partial \left\|\mathcal{P}_{t+1}\right\|_* - \mathcal{Q}_{t+1} \Rightarrow \mathcal{Q}^\star \in \partial \left\|\mathcal{P}^\star\right\|_*.$$

2. *Proof of the second part.* First, we show that $\{\mathcal{F}_t\}$ is a Cauchy sequence. From (4.10), we have

$$\mathcal{F}_t = \mathcal{Z}_{\flat,t} - \frac{\mathcal{G}_t - \mathcal{G}_{t-1}}{\rho_{t-1}}.$$

Thus

$$\begin{aligned}
&\|\mathcal{F}_{t+1} - \mathcal{F}_t\|_F^2 \\
&= \left\|\mathcal{F}_{t+1} - \left(\mathcal{Z}_{\flat,t} + \frac{\mathcal{G}_t}{\rho_t}\right) + \frac{\mathcal{G}_t}{\rho_t} + \frac{\mathcal{G}_t - \mathcal{G}_{t-1}}{\rho_{t-1}}\right\|_F^2 \\
&\le \left\|\mathcal{F}_{t+1} - \tilde{\mathcal{F}}_t\right\|_F^2 + \left\|\frac{\mathcal{G}_t}{\rho_t} + \frac{\mathcal{G}_t - \mathcal{G}_{t-1}}{\rho_{t-1}}\right\|_F^2,
\end{aligned}$$

where $\tilde{\mathcal{F}}_t = \mathcal{Z}_{\flat,t} + \mathcal{G}_t/\rho_t$. By (4.5), we have

$$\mathcal{F}_{t+1} = \operatorname*{argmin}_{\mathcal{F}} \|\mathcal{F}\|_* + \frac{\rho_t}{2}\left\|\mathcal{F} - \tilde{\mathcal{F}}_t\right\|_F^2.$$

Thus $\mathcal{F}_{t+1}$ obeys the first-order optimality condition

$$-\rho_t\left(\mathcal{F}_{t+1} - \tilde{\mathcal{F}}_t\right) \in \partial \|\mathcal{F}_{t+1}\|_*.$$

Then following the proof of Lemma E.2, we have

$$\left\|\mathcal{F}_{t+1} - \tilde{\mathcal{F}}_t\right\|_F^2 = \frac{1}{\rho_t^2}\|\partial \|\mathcal{F}_{t+1}\|_*\|_F^2 \le \frac{r_{\mathcal{F}_{t+1}}}{n_3^2\rho_t^2}. \tag{E.1}$$

In conclusion, we have

$$\|\mathcal{F}_{t+1} - \mathcal{F}_t\|_F^2 \le \frac{r_{\mathcal{F}_{t+1}}}{n_3^2\rho_t^2} + \frac{\|\mathcal{G}_t\|_F^2}{\rho_t^2} + \frac{\|\mathcal{G}_t\|_F^2}{\rho_{t-1}^2} + \frac{\|\mathcal{G}_{t-1}\|_F^2}{\rho_{t-1}^2}.$$

Noting that $\{\mathcal{G}_t\}$ is bounded, $\rho_t = \eta\rho_{t-1}$, and $\eta = 1.1$, then due to the triangle inequality, for any $n < m$, we have

$$\lim_{m,n\to\infty} \|\mathcal{F}_m - \mathcal{F}_n\|_F^2 \le \lim_{m,n\to\infty} \sum_{t=n}^{m-1} \|\mathcal{F}_{t+1} - \mathcal{F}_t\|_F^2 = 0.$$

We prove thus $\{\mathcal{F}_t\}$ is a Cauchy sequence.

Second, we prove that $\{\mathcal{P}_t\}$ is a Cauchy sequence. From (4.11), we have

$$\mathcal{P}_t = \mathcal{L}_{\flat,t} - (\mathcal{Q}_t - \mathcal{Q}_{t-1})/\theta_{t-1}.$$

Thus

$$\begin{aligned}
&\|\mathcal{P}_{t+1} - \mathcal{P}_t\|_F^2 \\
&= \left\| \mathcal{P}_{t+1} - \left(\mathcal{L}_{\flat,t} + \frac{\mathcal{Q}_t}{\theta_t} \right) + \frac{\mathcal{Q}_t}{\theta_t} + \frac{\mathcal{Q}_t - \mathcal{Q}_{t-1}}{\theta_{t-1}} \right\|_F^2 \\
&\le \left\| \mathcal{P}_{t+1} - \tilde{\mathcal{P}}_t \right\|_F^2 + \left\| \frac{\mathcal{Q}_t}{\theta_t} + \frac{\mathcal{Q}_t - \mathcal{Q}_{t-1}}{\theta_{t-1}} \right\|_F^2,
\end{aligned}$$

where $\tilde{\mathcal{P}}_t = \mathcal{L}_{\flat,t} + \mathcal{Q}_t/\theta_t$. And (4.7) implies that

$$\mathcal{P}_{t+1} = \operatorname*{argmin}_{\mathcal{P}} \|\mathcal{P}\|_* + \frac{\theta_t}{2} \left\| \mathcal{P} - \tilde{\mathcal{P}}_t \right\|_F^2.$$

$\mathcal{P}_{t+1}$ should obey the first-order optimality condition

$$-\theta_t \left(\mathcal{P}_{t+1} - \tilde{\mathcal{P}}_t \right) \in \partial \|\mathcal{P}_{t+1}\|_*.$$

Then we have

$$\left\| \mathcal{P}_{t+1} - \tilde{\mathcal{P}}_t \right\|_F^2 = \frac{1}{\theta_t^2} \|\partial \|\mathcal{P}_{t+1}\|_*\|_F^2 \le \frac{r_{\mathcal{P}_{t+1}}}{n_3^2 \theta_t^2}.$$

In conclusion, we get

$$\|\mathcal{P}_{t+1} - \mathcal{P}_t\|_F^2 \le \frac{r_{\mathcal{P}_{t+1}}}{n_3^2\theta_t^2} + \frac{\|\mathcal{Q}_t\|_F^2}{\theta_t^2} + \frac{\|\mathcal{Q}_t\|_F^2}{\theta_{t-1}^2} + \frac{\|\mathcal{Q}_{t-1}\|_F^2}{\theta_{t-1}^2}.$$

Due to the boundedness of $\{\mathcal{Q}_t\}$, and the triangle inequality, we get

$$\lim_{m,n\to\infty} \|\mathcal{P}_m - \mathcal{P}_n\|_F^2 \le \lim_{m,n\to\infty} \sum_{t=n}^{m-1} \|\mathcal{P}_{t+1} - \mathcal{P}_t\|_F^2 = 0$$

for any $n < m$. Hence $\{\mathcal{P}_t\}$ is a Cauchy sequence.

Third, we show that $\{\mathcal{E}_t\}$ is a Cauchy sequence. From (4.8), let us denote that $\mathcal{E}_{t+1} = \mathcal{T}_{\lambda/\mu_t}(\mathcal{D}_t)$, where $\mathcal{D}_{t+1} = \mathcal{X} - \mathcal{A} * \mathcal{Z}_{\flat,t+1} - \mathcal{L}_{\flat,t} * \mathcal{B} + \mathcal{Y}_t/\mu_t$, then it is easy to see

$$\left\| \mathcal{T}_{\lambda/\mu_t}(\mathcal{D}_t) - \mathcal{D}_t \right\|_F^2 \le \frac{\lambda^2 N}{\mu_t^2}, \tag{E.2}$$

where $N = n_1 n_2 n_3$. Besides, (4.9) implies

$$\mathcal{E}_t = \mathcal{X} - \mathcal{A} * \mathcal{Z}_{\flat,t} - \mathcal{L}_{\flat,t} * \mathcal{B} - \frac{\mathcal{Y}_t - \mathcal{Y}_{t-1}}{\mu_{t-1}}.$$

We have

$$\mathcal{E}_t = \mathcal{D}_t - \frac{2\mathcal{Y}_t - \mathcal{Y}_{t-1}}{\mu_{t-1}}. \tag{E.3}$$

Then for any $n < m$, from (E.2)–(E.3) and the triangle inequality, one can get that

$$\begin{aligned}
&\|\mathcal{E}_m - \mathcal{E}_n\|_F^2 \\
&\quad\le \sum_{t=n}^{m-1} \|\mathcal{E}_{t+1} - \mathcal{E}_t\|_F^2 \\
&\quad= \sum_{t=n}^{m-1} \left\| \mathcal{T}_{\frac{\lambda}{\mu_t}}(\mathcal{D}_t) - \mathcal{D}_t + \frac{2\mathcal{Y}_t - \mathcal{Y}_{t-1}}{\mu_t} \right\|_F^2 \\
&\quad\le \sum_{t=n}^{m-1} \left\| \mathcal{T}_{\frac{\lambda}{\mu_t}}(\mathcal{D}_t) - \mathcal{D}_t \right\|_F^2 + \sum_{t=n}^{m-1} \frac{\|2\mathcal{Y}_t - \mathcal{Y}_{t-1}\|_F^2}{\mu_{t-1}^2} \\
&\quad\le \sum_{t=n}^{m-1} \frac{\lambda_2^2 N + \eta^2 \|2\mathcal{Y}_t - \mathcal{Y}_{t-1}\|_F^2}{\mu_0^2 \eta^{2t}},
\end{aligned}$$

where $\eta = 1.1$ in Algorithm 4.1. We have known $\{1/\eta^{2t}\}_{t=1}^{\infty}$ is a convergent Cauchy sequence. Notice that $\|2\mathcal{Y}_t - \mathcal{Y}_{t-1}\|_F^2$ is bounded, and we can get that $\lim_{m,n\to\infty} \|\mathcal{E}_m - \mathcal{E}_n\|_F^2 = 0$, hence $\mathcal{E}_t$ is a Cauchy sequence, which is convergent.

Forth, (4.10) implies that

$$\mathcal{G}_t = \mathcal{G}_{t-1} + \rho_{t-1}(\mathcal{Z}_{\flat,t} - \mathcal{F}_t) \Rightarrow \mathcal{Z}_{\flat,t} = \mathcal{F}_t + (\mathcal{G}_t - \mathcal{G}_{t-1})/\rho_{t-1}.$$

Due to the boundedness of $\{\mathcal{G}_t - \mathcal{G}_{t-1}\}$, we have that $\{(\mathcal{G}_t - \mathcal{G}_{t-1})/\rho_0\eta^{t-1}\}$ is a Cauchy sequence. Since $\{\mathcal{F}_t\}$ is a Cauchy sequence, thus $\{\mathcal{Z}_{\flat,t}\}$ is a Cauchy sequence. Similarly, it is easy to get that $\{\mathcal{L}_{\flat,t}\}$ is a Cauchy sequence as well according to (4.11). Therefore, the sequences $\{\mathcal{Z}_{\flat,t}\}$, $\{\mathcal{L}_{\flat,t}\}$, $\{\mathcal{E}_t\}$, $\{\mathcal{F}_t\}$, and $\{\mathcal{P}_t\}$ converge to the critical point of (4.2). This completes the proof. ∎

REFERENCES

[1] S. E. Abhadiomhen, Z. Wang, X. Shen, and J. Pan, *Multiview common subspace clustering via coupled low rank representation*, ACM Trans. Intell. Syst. Technol., 12 (2021), 44, https://doi.org/10.1145/3465056.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Found. Trends Mach. Learn., 3 (2011), pp. 1–122, http://dx.doi.org/10.1561/2200000016.

[3] E. J. Candès, X. Li, Y. Ma, and J. Wright, *Robust principal component analysis?*, J. ACM, 58 (2011), pp. 1–37, https://doi.org/10.1145/1970392.1970395.

[4] J. Chang, Y. Chen, L. Qi, and H. Yan, *Hypergraph clustering using a new Laplacian tensor with applications in image processing*, SIAM J. Imaging Sci., 13 (2020), pp. 1157–1178, https://doi.org/10.1137/19M1291601.

[5] C. Chen, B. He, Y. Ye, and X. Yuan, *The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent*, Math. Program., 155 (2016), pp. 57–79, https://doi.org/10.1007/s10107-014-0826-5.
[6] X. Chen and D. Cai, *Large scale spectral clustering with landmark-based representation*, in Proceedings of the AAAI Conference on Artificial Intelligence, AAAI Press, Menlo Park, CA, 2011, pp. 313–318.
[7] D. L. Donoho and I. M. Johnstone, *Adapting to unknown smoothness via wavelet shrinkage*, J. Amer. Statist. Assoc., 90 (1995), pp. 1200–1224, https://doi.org/10.1080/01621459.1995.10476626.
[8] S. Du, Y. Shi, G. Shan, W. Wang, and Y. Ma, *Tensor low-rank sparse representation for tensor subspace learning*, Neurocomputing, 440 (2021), pp. 351–364, https://doi.org/10.1016/j.neucom.2021.02.002.
[9] E. Elhamifar and R. Vidal, *Sparse subspace clustering: Algorithm, theory, and applications*, IEEE Trans. Pattern Anal. Mach. Intell., 35 (2013), pp. 2765–2781, https://doi.org/10.1109/TPAMI.2013.57.
[10] X. Fu, Q. Qi, Y. Zhu, X. Ding, and Z.-J. Zha, *Rain streak removal via dual graph convolutional network*, in Proceedings of the 35th AAAI Conference on Artificial Intelligence, AAAI, Palo Alto, CA, 2021, pp. 1352–1360.
[11] Y. Fu, J. Gao, D. Tien, Z. Lin, and X. Hong, *Tensor LRR and sparse coding-based subspace clustering*, IEEE Trans. Neural Netw. Learn. Syst., 27 (2016), pp. 2120–2133, https://doi.org/10.1109/TNNLS.2016.2553155.
[12] A. Georghiades, P. Belhumeur, and D. Kriegman, *From few to many: Illumination cone models for face recognition under variable lighting and pose*, IEEE Trans. Pattern Anal. Mach. Intell., 23 (2001), pp. 643–660, https://doi.org/10.1109/34.927464.
[13] D. Goldfarb and Z. (T.) Qin, *Robust low-rank tensor recovery: Models and algorithms*, SIAM J. Matrix Anal. Appl., 35 (2014), pp. 225–253, https://doi.org/10.1137/130905010.
[14] A. Goncalves, X. Liu, and A. Banerjee, *Two-Block vs. Multi-block ADMM: An Empirical Evaluation of Convergence*, preprint, arXiv:1907.04524, 2019.
[15] J. Guo, Y. Guo, Q. Jin, M. Kwok-Po Ng, and S. Wang, *Gaussian patch mixture model guided low-rank covariance matrix minimization for image denoising*, SIAM J. Imaging Sci., 15 (2022), pp. 1601–1622, https://doi.org/10.1137/21M1454262.
[16] K. He, X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, Los Alamitos, CA, 2016, pp. 770–778.
[17] J. Hou, F. Zhang, H. Qiu, J. Wang, Y. Wang, and D. Meng, *Robust low-tubal-rank tensor recovery from binary measurements*, IEEE Trans. Pattern Anal. Mach. Intell., 44 (2022), pp. 4355–4373, https://doi.org/10.1109/TPAMI.2021.3063527.
[18] T.-X. Jiang, M. K. Ng, J. Pan, and G.-J. Song, *Nonnegative low rank tensor approximations with multidimensional image applications*, Numer. Math., 153 (2023), pp. 141–170, https://doi.org/10.1007/s00211-022-01328-6.
[19] A. Johnson, J. Francis, B. Madathil, and S. N. George, *A two-way optimization framework for clustering of images using weighted tensor nuclear norm approximation*, in Proceedings of the National Conference on Communications, IEEE, Piscataway, NJ, 2020, pp. 1–5.
[20] M. E. Kilmer, K. Braman, N. Hao, and R. C. Hoover, *Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 148–172, https://doi.org/10.1137/110837711.
[21] M. Kilmer and C. Martin, *Factorization strategies for third-order tensors*, Linear Algebra Appl., 435 (2011), pp. 641–658, https://doi.org/10.1016/j.laa.2010.09.020.
[22] T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, SIAM Rev., 51 (2009), pp. 455–500, https://doi.org/10.1137/07070111X.
[23] C. Li, C. Yang, B. Liu, Y. Yuan, and G. Wang, *LRSC: Learning representations for subspace clustering*, in Proceedings of the 35th AAAI Conference on Artificial Intelligence, AAAI, Palo Alto, CA, 2021, pp. 8340–8348.
[24] Y. Li, P. Hu, Z. Liu, D. Peng, J. T. Zhou, and X. Peng, *Contrastive clustering*, in Proceedings of the 35th AAAI Conference on Artificial Intelligence, AAAI, Palo Alto, CA, 2021, pp. 8547–8555.

[25] D. Lin and X. Yuan, *A note on the alternating direction method of multipliers*, J. Optim. Theory Appl., 155 (2012), pp. 227–238, https://doi.org/10.1007/s10957-012-0003-z.
[26] T. Lin, S. Ma, and S. Zhang, *On the global linear convergence of the ADMM with multiblock variables*, SIAM J. Optim., 25 (2015), pp. 1478–1497, https://doi.org/10.1137/140971178.
[27] T. Lin, S. Ma, and S. Zhang, *Iteration complexity analysis of multi-block ADMM for a family of convex minimization without strong convexity*, J. Sci. Comput., 69 (2016), pp. 52–81, https://doi.org/10.1007/s10915-016-0182-0.
[28] Z. Lin, M. Chen, and Y. Ma, *The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices*, J. Struct. Biol., 181 (2010), pp. 116–127.
[29] G. Liu, Z. Liu, and Y. Yu, *Robust subspace segmentation by low-rank representation*, in Proceedings of the International Conference on Machine Learning, International Machine Learning Society, Madison, WI, 2010, pp. 663–670.
[30] G. Liu and S. Yan, *Latent low-rank representation for subspace segmentation and feature extraction*, in Proceedings of the IEEE 2011 International Conference on Computer Vision, IEEE, Barcelona, Spain, 2011, pp. 1615–1622, https://ieeexplore.ieee.org/document/6126422.
[31] 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 Trans. Pattern Anal. Mach. Intell., 42 (2020), pp. 925–938, https://doi.org/10.1109/TPAMI.2019.2891760.
[32] C. Lu, J. Feng, Z. Lin, T. Mei, and S. Yan, *Subspace clustering by block diagonal representation*, IEEE Trans. Pattern Anal. Mach. Intell., 41 (2019), pp. 487–501, https://doi.org/10.1109/TPAMI.2018.2794348.
[33] L. Lu and R. Vidal, *Combined central and subspace clustering for computer vision applications*, in Proceedings of the International Conference on Machine Learning, International Machine Learning Society, Madison, WI, 2006, pp. 593–600.
[34] Y. Luo, X. Zhao, Z. Li, M. K. Ng, and D. Meng, *Low-rank tensor function representation for multi-dimensional data recovery*, IEEE Trans. Pattern Anal. Mach. Intell., 46 (2024), pp. 3351–3369, https://doi.org/10.1109/TPAMI.2023.3341688.
[35] Y. Ma, H. Derksen, W. Hong, and J. Wright, *Segmentation of multivariate mixed data via lossy data coding and compression*, IEEE Trans. Pattern Anal. Mach. Intell., 29 (2007), pp. 1546–1562, https://doi.org/10.1109/TPAMI.2007.1085.
[36] C. D. Manning, P. Raghavan, and H. Schutze, *Introduction to Information Retrieval*, Cambridge University Press, Cambridge, 2008.
[37] J. Meng, F. Wang, and J. Liu, *Learnable nonlocal self-similarity of deep features for image denoising*, SIAM J. Imaging Sci., 17 (2024), pp. 441–475, https://doi.org/10.1137/22M1536996.
[38] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, *Image segmentation using deep learning: A survey*, IEEE Trans. Pattern Anal. Mach. Intell., 44 (2022), pp. 3523–3542, https://doi.org/10.1109/TPAMI.2021.3059968.
[39] S. Rao, R. Tron, R. Vidal, and Y. Ma, *Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories*, IEEE Trans. Pattern Anal. Mach. Intell., 32 (2010), pp. 1832–1845, https://doi.org/10.1109/TPAMI.2009.191.
[40] B. Recht, M. Fazel, and P. A. Parrilo, *Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization*, SIAM Rev., 52 (2010), pp. 471–501, https://doi.org/10.1137/070697835.
[41] J. Shi and J. Malik, *Normalized cuts and image segmentation*, IEEE Trans. Pattern Anal. Mach. Intell., 22 (2000), pp. 888–905, https://doi.org/10.1109/34.868688.
[42] K. Simonyan and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*, in Proceedings of the International Conference on Learning Representation, 2015, https://iclr.cc/archive/www/doku.php%3Fid=iclr2015:main.html.
[43] M. Tapaswi, M. Law, and S. Fidler, *Video face clustering with unknown number of clusters*, in Proceedings of the IEEE International Conference on Computer Vision, IEEE, Piscataway, NJ, 2019, pp. 5026–5035.
[44] L. R. Tucker, *Some mathematical notes on three-mode factor analysis*, Psychometrika, 31 (1966), pp. 279–311, https://doi.org/10.1007/BF02289464.

[45] R. Vidal, *Subspace clustering*, IEEE Signal Process. Mag., 28 (2011), pp. 52–68, https://doi.org/10.1109/MSP.2010.939739, https://ieeexplore.ieee.org/document/5714408.

[46] N. X. Vinh, J. Epps, and J. Bailey, *Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance*, J. Mach. Learn. Res., 11 (2010), pp. 2837–2854, http://jmlr.org/papers/v11/vinh10a.html.

[47] L. Wang, Z. Ding, and Y. Fu, *Learning transferable subspace for human motion segmentation*, in Proceedings of the 32nd AAAI Conference on Artificial Intelligence, AAAI, Palo Alto, CA, 2018, pp. 4195–4202.

[48] P. Wang, H. Liu, A. M.-C. So, and L. Balzano, *Convergence and recovery guarantees of the k-subspaces method for subspace clustering*, in Proceedings of the International Conference on Machine Learning, International Machine Learning Society, Madison, WI, 162, 2022, pp. 22884–22918.

[49] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, *Image quality assessment: From error visibility to structural similarity*, IEEE Trans. Image Process., 13 (2004), pp. 600–612, https://doi.org/10.1109/TIP.2003.819861.

[50] G. A. Watson, *Characterization of the subdifferential of some matrix norms*, Linear Algebra Appl., 170 (1992), pp. 33–45, https://doi.org/10.1016/0024-3795(92)90407-2.

[51] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, *Robust face recognition via sparse representation*, IEEE Trans. Pattern Anal. Mach. Intell., 31 (2009), pp. 210–227, https://doi.org/10.1109/TPAMI.2008.79.

[52] J. Yang, J. Liang, K. Wang, P. L. Rosin, and M.-H. Yang, *Subspace clustering via good neighbors*, IEEE Trans. Pattern Anal. Mach. Intell., 42 (2020), pp. 1537–1544, https://doi.org/10.1109/TPAMI.2019.2913863.

[53] J.-H. Yang, C. Chen, H.-N. Dai, M. Ding, Z.-B. Wu, and Z. Zheng, *Robust corrupted data recovery and clustering via generalized transformed tensor low-rank representation*, IEEE Trans. Neural Netw. Learn. Syst., 35 (2024), pp. 8839–8853, https://doi.org/10.1109/TNNLS.2022.3215983.

[54] M. Yang, Q. Luo, W. Li, and M. Xiao, *Multiview clustering of images with tensor rank minimization via nonconvex approach*, SIAM J. Imaging Sci., 13 (2020), pp. 2361–2392, https://doi.org/10.1137/20M1318006.

[55] M. Yin, J. Gao, and Z. Lin, *Laplacian regularized low-rank representation and its applications*, IEEE Trans. Pattern Anal. Mach. Intell., 38 (2016), pp. 504–517, https://doi.org/10.1109/TPAMI.2015.2462360.

[56] C. Zhang, Q. Hu, H. Fu, P. Zhu, and X. Cao, *Latent multi-view subspace clustering*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4333–4341.

[57] H. Zhang, Z. Lin, and C. Zhang, *A counterexample for the validity of using nuclear norm as a convex surrogate of rank*, in Proceedings of Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, Heidelberg, 2013, pp. 226–241.

[58] M. Zhang and K. Liu, *Enriched robust multi-view kernel subspace clustering*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition., 2022, pp. 1993–2002.

[59] X. Zhang and M. K. Ng, *Low rank tensor completion with Poisson observations*, IEEE Trans. Pattern Anal. Mach. Intell., 44 (2022), pp. 4239–4251, https://doi.org/10.1109/TPAMI.2021.3059299.

[60] Z. Zhang and S. Aeron, *Exact tensor completion using t-SVD*, IEEE Trans. Signal Process., 65 (2017), pp. 1511–1526, https://doi.org/10.1109/TSP.2016.2639466.

[61] H. Zheng, Y. Lou, G. Tian, and C. Wang, *A scale-invariant relaxation in low-rank tensor recovery with an application to tensor completion*, SIAM J. Imaging Sci., 17 (2024), pp. 756–783, https://doi.org/10.1137/23M1560847.

[62] P. Zhou, C. Lu, J. Feng, Z. Lin, and S. Yan, *Tensor low-rank representation for data recovery and clustering*, IEEE Trans. Pattern Anal. Mach. Intell., 43 (2021), pp. 1718–1732, https://doi.org/10.1109/TPAMI.2019.2954874.

[63] P. Zhou, C. Lu, Z. Lin, and C. Zhang, *Tensor factorization for low-rank tensor completion*, IEEE Trans. Image Process., 27 (2018), pp. 1152–1163, https://doi.org/10.1109/TIP.2017.2762595.

[64] T. Zhou, H. Fu, C. Gong, J. Shen, L. Shao, and F. Porikli, *Multi-mutual consistency induced transfer subspace learning for human motion segmentation*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2020, pp. 10274–10283.