GRAPH REGULARIZED NONNEGATIVE TUCKER DECOMPOSITION FOR TENSOR DATA REPRESENTATION

Yuning Qiu[1], *Guoxu Zhou*[1⋆], *Yu Zhang* [2], *Shengli Xie* [1]

[1] School of Automation, Guangdong University of Technology, Guangzhou 510006, China
[2] Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford CA, USA

ABSTRACT

Nonnegative Tucker Decomposition (NTD) is one of the most popular technique for feature extraction and representation from nonnegative tensor data with preserving internal structure information. From the perspective of geometry, high-dimensional data are usually drawn in low-dimensional submanifold of the ambient space. In this paper, we propose a novel Graph reguralized Nonnegative Tucker Decomposition (GNTD) method which is able to extract the low-dimensional parts-based representation and preserve the geometrical information simultaneously from high-dimensional tensor data. We also present an effictive algorithm to solve the proposed GNTD model. Experimental results demonstrate the effectiveness and high efficiency of the proposed GNTD method.

Index Terms— Mannifold learning, nonnegative tensor, Tucker decomposition, dimensionality reduction, clustering.

1. INTRODUCTION

A tensor can be viewed as a multi-index numerical array, where the order of a tensor denotes the number of its dimension [1]. With the rapid development of data acquisition technologies, the data are always representated by tensors, e.g., multichannel electroencephalography (EEG) tensor data, video volume tensor data, Functional Magnetic Resonance Imaging (fMRI) tensor data. Very often, tensor data are first vectorized, and then low rank approximation methods, such as, singular value decomposition (SVD), principle component analysis (PCA), nonnegative matrix factorization (NMF) etc., are implemented for extracting their low-dimensional representation.

However, vectorization of tensor data often destroys internal structure of data. For this reason, many tensor decompoistion methods have been developed to study the low-dimensional representation in tensor domain. Tucker decomposition is one of the most widely used methods for high-order tensor analysis [2]. It is interesing to note that most of the tensor data in applications are nonnegative because they are recorded from a variety of physical signals [3]. Thus, Nonnegative Tucker Decomposition (NTD) has been a popular method to analyze the nonnegative tensor data. Particularly, with nonnegativity constraints, NTD could not only extract the parts-based reprentation like NMF, but also improve the uniqueness of Tucker decomposition [4]. However, the underlying manifold structure of tensor data have not been exploited by existing NTD methods. Previous works have demonstrated that this geometrical information can significantly improve the performance in data representation and clustering tasks [5] [6] [7] [8].

Recently, several works have been introduced to incoporate the geometrical information into tensor decomposition [9] [10] [11] [3]. Wang *et al*. [9] studied a neighborhood preserving nonnegative tensor factorization method for image representation by incoporating locally linear embedding regularization into Nonnegative Parallel Factor Analysis (NPARAFAC). Wang *et al*. [10] proposed a Laplacian Regularized Nonnegative Tensor Decomposition (LRNTD) method for image clustering and representation by considering the manifold structure of image space while performing NPAPAFAC. However, NPAPAFAC is not as efficient as NTD in data representation because it usually requires more component vectors for each mode than NTD [12]. Jiang *et al*. [11] developed a Graph-Laplacian Tucker Decomposition (GLTD) to explore attributes and similarity information without considering parts-based representation. Li *et al*. [3] introduced a Manifold Regularization Nonnegative Tucker Decompostition (MR-NTD) by incorporating the manifold regularization into core tensor while performing Population Nonnegative Tucker Decomposition (PNTD) [4]. However, the number of entries in core tensor will increase exponentially as tensor dimension increase, which fails to obtain low-dimensional representation of tensor data.

Therefore, in order to simultaneously preserve the internal multilinear structure and geometrical information while acquiring the low-dimensional parts-based representation, Graph regularized Nonnegative Tucker Decomposition (GNTD) is proposed by incoporating the graph regularization and nonnegative constraints into Tucker decomposition. Also, we develop an effective algorithm to solve the proposed model. Our experimental results demonstrate that the proposed GNTD could significantly improve the performance in

⋆ Correspondence author. E-mail address: guoxu.zhou@qq.com

clustering analysis.

The rest of this paper is organized as follows. In Section 2, we review NMF and NTD briefly. In Section 3, the GNTD method is proposed. Finally, experiments for clustering tasks are presented in Section 4, and conclusions are drawn in Section 5.

2. NONNEGATIVE MATRIX FACTORIZATION AND NONNEGATIVE TUCKER DECOMPOSITION

Given $\mathcal{Y} \in \mathbb{R}_+^{I_1 \times I_2 \times \cdots \times I_N}$ be an Nth-order nonnegative tensor data set containing the collection of $(N-1)$th-order tensor data $\mathcal{X} \in \mathbb{R}_+^{I_1 \times I_2 \times \cdots \times I_{N-1}}$. Nonnegative Tucker Decomposition (NTD) could be formulated as a nonnegative core tensor $\mathcal{G} \in \mathbb{R}_+^{r_1 \times r_2 \times \cdots \times r_N}$ multiplies N nonnegative factors $\mathbf{A}^{(n)} \in \mathbb{R}_+^{I_n \times r_n}, (n = 1, 2, \cdots, N)$ in each modality [12]. The aim of NTD is to find N nonnegative factor matrices and the nonnegative core tensor by minimizing

$$\begin{aligned} \min_{\mathcal{G}, \mathbf{A}^{(n)}, n \in \mathcal{I}_n} & \left\| \mathcal{Y} - \mathcal{G} \times_{n \in \mathcal{I}_N} \mathbf{A}^{(n)} \right\|_F^2, \\ & \text{s.t. } \mathcal{G} \geq 0, \mathbf{A}^{(n)} \geq 0, n \in \mathcal{I}_n, \end{aligned} \tag{1}$$

where $\mathcal{I}_n = \{1, 2, \cdots, N\}$ is a set of positive integers no larger than N. Equivalently, NTD could be represented as a matrix form:

$$\begin{aligned} & \min_{\mathbf{G}_{(N)}, \mathbf{A}^{(N)}, \mathbf{H}} \left\| \mathbf{Y}_{(N)} - \mathbf{A}^{(N)} \mathbf{G}_{(N)} \mathbf{H}_N^T \right\|_F^2, \\ & \text{s.t. } \mathbf{G}_{(N)} \geq 0, \mathbf{A}^{(N)} \geq 0, \mathbf{H} \geq 0, \end{aligned} \tag{2}$$

where $\mathbf{Y}_{(N)} \in \mathbb{R}_+^{I_N \times I_1 I_2 \cdots I_{N-1}}$ and $\mathbf{G}_{(N)}$ are the mode-N unfolding matrices of tensor $\mathcal{Y}$ and tensor $\mathcal{G}$ [13] respectively, and $\mathbf{H}_N = \mathbf{A}^{(N-1)} \otimes \mathbf{A}^{(N-2)} \cdots \otimes \mathbf{A}^{(1)}$, $\otimes$ denotes Kronecker product. In mode-N unfolding case, $\mathbf{A}^{(N)}$ is the coefficient matrix corresponding to the basic factor matrix $\mathbf{G}_{(N)} \mathbf{H}_N^T$.

Alternatively, NTD could be viewed as a special case of NMF. An intuitive difference between (2) and NMF is that the basic factor matrix in the former will further perform NTD from mode one to mode $(N-1)$, which not only yield multilinear representation but also improve the sparsity of basic vectors [4].

3. GRAPH REGULARIZED NONNEGATIVE TUCKER DECOMPOSITION

In the following, we present the way to construct a graph of tensor data, and then incoporate the graph regularization into regular NTD. Finally we give an effective algorithm to solve GNTD problem.

3.1. Graph Construction

Many research works have demonstrated that high-dimensional data are usually located in low-dimensional submanifold of the ambient space, and this underlying geometrical information of data could be obtained by modeling a neighbor graph [3] [5] [6] [7] [8]. Given a total of I_N tensors $\mathcal{X}_i \in \mathbb{R}_+^{I_1 \times I_2 \times \cdots \times I_{N-1}} (i = 1, \cdots, I_N)$ of order $(N-1)$, we encode their geometrical information by connecting each tensor subjcet with its p-nearest neighbors. We construct a relationship matrix $\mathbf{W} \in \mathbb{R}^{I_N \times I_N}$ to denote their connections in the graph:

$$w_{ij} = \begin{cases} 1, \text{ if } \mathcal{X}_i \in \mathcal{N}_p(\mathcal{X}_j), \text{and } \mathcal{X}_j \in \mathcal{N}_p(\mathcal{X}_i) \\ 0, \text{ otherwise} \end{cases} \tag{3}$$

where $\mathcal{N}_p(\mathcal{X}_i)$ represents the set of p samples closest to $\mathcal{X}_i$ in the graph. There are many techniques to measure the distance between two tensors. To simplfy the problem, we use the Frobenius Norm Distance in this case.

3.2. GNTD Model

As discussed in the Section 2, NTD could be considered as a special case of NMF with sparser and multilinear basics vectors. It is resonable to incoperate the graph regularization into mode-N low-dimensional representation $\mathbf{A}^{(N)}$. Therefore, GNTD is obtained by minimizing the following objective function:

$$\begin{aligned} \min_{\mathcal{G}, \mathbf{A}^{(n)}, n \in \mathcal{I}_n} \mathcal{O}_{GNTD} = & \frac{1}{2} \left\| \mathcal{Y} - \mathcal{G} \times_{n \in \mathcal{I}_N} \mathbf{A}^{(n)} \right\|^2 \\ & + \frac{\lambda}{2} \text{Tr}(\mathbf{A}^{(N)^T} \mathbf{L} \mathbf{A}^{(N)}), \\ & \text{s.t. } \mathcal{G} \geq 0, \mathbf{A}^{(n)} \geq 0, n \in \mathcal{I}_n. \end{aligned} \tag{4}$$

where $\text{Tr}(\cdot)$ represents the trace of a matrix, λ is a nonnegative parameter for balancing the importance of a graph regularization term and reconstruction error term, $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is Laplacian matrix and $\mathbf{D}_{jj} = \sum_k \mathbf{W}_{jk}$.

We use the Lagrange multiplier method, and consider the mode-n unfolding form, and then define the following Lagrange function inherit from (4):

$$\begin{aligned} \mathcal{L}_{GNTD} = & \frac{1}{2} \left\| \mathbf{Y}_{(n)} - \mathbf{A}^{(n)} \mathbf{G}_{(n)} (\bigotimes_{p \neq n} \mathbf{A}^{(p)^T}) \right\|^2 \\ & + \frac{\lambda}{2} \text{Tr}(\mathbf{A}^{(N)^T} \mathbf{L} \mathbf{A}^{(N)}) + \text{Tr}(\mathbf{\Phi}_n \mathbf{G}_{(n)}^T) + \sum_{l=1}^{N} \text{Tr}(\mathbf{\Psi}_l \mathbf{A}^{(l)^T}), \end{aligned} \tag{5}$$

where $\bigotimes_{p \neq n} \mathbf{A}^{(p)^T} = \mathbf{A}^{(N)^T} \otimes \cdots \otimes \mathbf{A}^{(p+1)^T} \otimes \mathbf{A}^{(p-1)^T} \otimes \cdots \otimes \mathbf{A}^{(1)^T}$. $\mathbf{\Psi}_l$ and $\mathbf{\Phi}_n$ represent the Lagrange multipliers matrices of $\mathbf{A}^{(l)}$ and $\mathbf{G}_{(n)}$, respectively.

3.3. Optimization Algorithm

To solve (5), we adopt the block coordinate descent framework, i.e., update core tensor or one factor matrix each time while fixing the others.

3.3.1. Solutions of Factor Matrices $\mathbf{A}^{(n)}, n \neq N$

The gradient of $\mathcal{L}_{GNTD}$ with respect to $\mathbf{A}^{(n)}, n \neq N$ is

$$\begin{aligned}\frac{\partial \mathcal{L}_{GNTD}}{\partial \mathbf{A}^{(n)}} =& \mathbf{A}^{(n)}\mathbf{G}_{(n)}(\bigotimes_{p\neq n}\mathbf{A}^{(p)^T}\mathbf{A}^{(p)})\mathbf{G}_{(n)}^T \\ &- \mathbf{Y}_{(n)}(\bigotimes_{p\neq n}\mathbf{A}^{(p)})\mathbf{G}_{(n)}^T + \mathbf{\Psi}_n.\end{aligned} \tag{6}$$

By considering the Karush-Kuhn-Tucker (KKT) condition, we have $\frac{\partial \mathcal{L}_{GNTD}}{\partial \mathbf{A}^{(n)}} = 0$ and $\psi_{n(j,k)}a_{(j,k)}^{(n)} = 0$, where $\psi_{n(j,k)}$ and $a_{(j,k)}^{(n)}$ denote the (j,k)-th entry in $\mathbf{\Psi}_n$ and $\mathbf{A}^{(n)}$, thereby leading to the following equation:

$$\left(\mathbf{A}^{(n)}\mathbf{G}_{(n)}(\bigotimes_{p\neq n}\mathbf{A}^{(p)^T}\mathbf{A}^{(p)})\mathbf{G}_{(n)}^T - \mathbf{Y}_{(n)}(\bigotimes_{p\neq n}\mathbf{A}^{(p)})\mathbf{G}_{(n)}^T\right)_{j,k} \cdot a_{(j,k)}^{(n)} = 0. \tag{7}$$

Therefore, we obtain the following update rule:

$$\mathbf{A}^{(n)} \leftarrow \mathbf{A}^{(n)} \odot \frac{\mathcal{P}_+(\mathbf{Y}_{(n)}(\bigotimes_{p\neq n}\mathbf{A}^{(p)})\mathbf{G}_{(n)}^T)}{\mathbf{A}^{(n)}\mathbf{G}_{(n)}(\bigotimes_{p\neq n}\mathbf{A}^{(p)^T}\mathbf{A}^{(p)})\mathbf{G}_{(n)}^T}, \tag{8}$$

where $\mathcal{P}_+(\xi) = \max(0,\xi)$, $\odot$ is the element-wise product.

3.3.2. Solutions of Factor Matrix $\mathbf{A}^{(N)}$

The gradient of $\mathcal{L}_{GNTD}$ with respect to $\mathbf{A}^{(N)}$ is

$$\begin{aligned}\frac{\partial \mathcal{L}_{GNTD}}{\partial \mathbf{A}^{(N)}} =& \mathbf{A}^{(N)}\mathbf{G}_{(N)}(\bigotimes_{p\neq N}\mathbf{A}^{(p)^T}\mathbf{A}^{(p)})\mathbf{G}_{(N)}^T \\ &- \mathbf{Y}_{(N)}(\bigotimes_{p\neq N}\mathbf{A}^{(p)})\mathbf{G}_{N}^T + \lambda\mathbf{L}\mathbf{A}^{(N)} + \mathbf{\Psi}_N.\end{aligned} \tag{9}$$

Similarly to the solutions of $\mathbf{A}^{(n)}$, we consider the KKT condition and obtain the following update rule:

$$\begin{aligned}&\mathbf{A}^{(N)} \leftarrow \mathbf{A}^{(N)}\odot \\ &\frac{\mathcal{P}_+(\mathbf{Y}_{(N)}(\bigotimes_{p\neq N}\mathbf{A}^{(p)})\mathbf{G}_{(N)}^T + \lambda\mathbf{W}\mathbf{A}^{(N)})}{\mathbf{A}^{(N)}\mathbf{G}_{(N)}(\bigotimes_{p\neq N}\mathbf{A}^{(p)^T}\mathbf{A}^{(p)})\mathbf{G}_{(N)}^T + \lambda\mathbf{D}\mathbf{A}^{(N)}}.\end{aligned} \tag{10}$$

3.3.3. Solution of Core Tensor $\mathcal{G}$

We consider the vectorization form of (5), and rewrite the objective function as

$$\begin{aligned}\mathcal{C}_{GNTD} =& \frac{1}{2}\left\|\text{vec}(\mathcal{Y}) - \mathbf{F}\text{vec}(\mathcal{G})\right\|^2 + \sum_{l=1}^{N}\text{Tr}(\Psi_n\mathbf{A}^{(l)^T}) \\ &+ \text{vec}(\mathcal{G})^T\text{vec}(\Omega) + \frac{\lambda}{2}\text{Tr}(\mathbf{A}^{(N)^T}\mathbf{L}\mathbf{A}^{(N)}),\end{aligned} \tag{11}$$

where $\mathbf{F} = \mathbf{A}^{(1)} \otimes \mathbf{A}^{(2)} \cdots \otimes \mathbf{A}^{(N)} \in \mathbb{R}_+^{I_1I_2\cdots I_N \times r_1r_2\cdots r_N}$, $\text{vec}(\Omega)$ represents the Lagrange multipliers of $\text{vec}(\mathcal{G})$. The gradient of $\mathcal{C}_{GNTD}$ with respect to $\text{vec}(\mathcal{G})$ is

$$\frac{\partial \mathcal{C}_{GNTD}}{\text{vec}(\mathcal{G})} = \mathbf{F}^T\mathbf{F}\text{vec}(\mathcal{G}) - \mathbf{F}\text{vec}(\mathcal{Y}) + \text{vec}(\Omega). \tag{12}$$

Similarly to the solutions of $\mathbf{A}^{(n)}$, we consider the KKT condition, and obtain the following update rule:

$$\text{vec}(\mathcal{G}) \leftarrow \text{vec}(\mathcal{G}) \odot \frac{\mathcal{P}_+(\mathbf{F}\text{vec}(\mathcal{G}))}{\mathbf{F}^T\mathbf{F}\text{vec}(\mathcal{G})} \tag{13}$$

4. EXPERIMENTS ANALYSIS

In this section, we compare the performance of GNTD with that of other methods, on the task of clustering images in two publicly available real-world data sets.

4.1. Data sets Description

- *COIL-100* [1]: The first data set we used is the Columbia Object Image Library (COIL-100) data set. COIL-100 consists of 7200 color images of size 128×128 for 100 objects, each of which has 72 images taken from varying poses. In our experiments, each color image was resized into $64 \times 64 \times 3$.
- *AT&T ORL* [2]: The AT&T ORL data set consists of 400 grayscale 112×92 face images of 40 distinct subjects. Each person has 10 different images under different time with varying lighting, facial expressions, and facial details. In our experiment, all of the images in AT&T ORL were resized into 32×27.

The images of COIL100 and AT&T ORL form two tensor of $64 \times 64 \times 3 \times 72k$ and $32 \times 27 \times 10k$ (where k is the number of clusters).

4.2. Compared Methods

In order to verify the performance of our GNTD method, we have maken comparisions against the Canonical K-means clustering (K-means), NMF [2], GNMF [8] and NTD [4].

[1] Available at www.cs.columbia.edu/CAVE/software/softlib/coil-100.php
[2] Available at www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

Table 1: Clustering performance on COIL-100 data set

k	Accuracy(%)					Normalized Mutual Information (%)				
	Kmeans	NMF	GNMF	NTD	GNTD	Kmeans	NMF	GNMF	NTD	GNTD
4	92.0±9.7	89.7±11.9	98.4±5.0	89.9±11.5	**99.6±1.2**	88.8±11.7	85.6±15.2	98.0±4.4	85.7±15.1	**99.3±2.4**
6	90.4±8.1	87.3±8.8	96.3±6.8	86.3±9.3	**98.0±4.6**	88.7±7.9	84.8±9.2	96.7±5.7	84.0±9.4	**97.8±4.1**
8	87.7±6.9	83.7±5.7	94.4±7.0	85.0±6.5	**99.3±1.2**	88.8±5.4	86.8±4.8	96.9±3.5	87.1±5.1	**99.0±1.8**
10	82.5±7.5	80.9±6.8	92.0±6.5	81.3±7.4	**95.3±6.6**	87.0±6.1	85.2±6.0	96.9±2.5	85.4±6.0	**97.5±3.1**
12	81.8±8.8	84.4±7.5	88.8±5.0	81.1±6.7	**92.8±5.2**	87.4±4.9	86.9±4.8	95.2±2.3	85.9±4.7	**96.0±2.6**
14	78.7±7.7	76.1±7.1	86.0±5.5	77.3±8.0	**86.5±5.5**	85.3±5.6	83.3±5.8	**94.9±2.4**	83.4±5.9	94.7±2.3
16	80.3±5.6	80.0±5.5	85.9±6.1	78.2±6.1	**90.0±4.8**	87.6±3.8	86.5±3.9	94.9±2.1	85.6±4.0	**96.2±1.9**
18	75.7±7.9	75.0±6.8	85.4±5.0	75.7±6.1	**87.4±5.4**	85.1±4.9	84.2±4.9	94.4±1.9	84.0±5.4	**94.5±2.2**
20	75.1±4.7	74.7±5.5	82.7±4.7	74.5±6.2	**85.8±4.4**	85.0±2.8	83.6±3.9	94.1±1.6	83.0±3.9	**94.5±1.6**
Avg	82.6	81.3	90.0	81.0	**92.7**	87.1	85.2	95.8	84.9	**96.6**

Table 2: Clustering performance on AT&T ORL data set

k	Accuracy(%)					Normalized Mutual Information (%)				
	Kmeans	NMF	GNMF	NTD	GNTD	Kmeans	NMF	GNMF	NTD	GNTD
5	89.4±10.6	89.4±11.0	92.9±8.6	89.4±10.6	**94.6±7.5**	88.3±10.6	89.0±9.1	92.3±6.7	88.3±10.6	**93.5±6.7**
10	84.6±7.2	82.5±8.8	87.5±8.2	84.8±6.6	**88.3±6.3**	90.0±4.2	87.2±5.6	91.3±4.0	88.5±3.6	**91.5±3.6**
15	80.1±6.0	82.5±6.5	84.6±5.1	79.7±7.5	**84.7±5.2**	88.6±2.3	89.0±4.4	90.4±4.0	87.2±4.3	**90.6±3.6**
20	75.3±4.9	74.9±4.3	78.9±4.0	75.2±5.2	**79.9±5.2**	86.8±2.7	86.2±2.6	88.9±2.0	85.9±2.9	**88.8±2.9**
25	71.0±3.8	71.9±3.9	75.4±3.5	69.7±4.7	**76.3±3.5**	85.3±2.2	85.6±2.6	87.5±1.5	83.0±2.6	**88.0±2.3**
30	68.1±3.4	71.4±3.8	75.4±2.5	68.4±4.2	**75.7±3.0**	84.8±1.6	85.4±1.8	**88.3±1.1**	83.4±2.6	88.1±1.1
35	67.7±3.4	68.9±2.9	**74.2±2.8**	68.4±3.7	74.0±2.6	84.7±1.4	85.2±1.7	87.9±1.0	83.3±1.8	**88.0±0.9**
40	64.7	67.8	73.4	65.5	**73.4**	83.9	84.2	87.4	82.5	**87.9**
Avg	75.1	76.2	80.3	75.1	**80.9**	86.6	86.5	89.3	85.3	**89.6**

4.3. Experiments Results

4.3.1. Parameter Selections

For GNMF and GNTD, we construct the graph Laplacian using p-nearest neighbors in which the neighborhood size p is set to 5. The graph regularization parameter λ is searched from the grid $(0.1, 1, 10, 10^2, 10^3, 10^4, 10^5)$. For NTD and GNTD, the size of core tensor are searched from $\frac{1}{4}I_n$ to $\frac{3}{4}I_n$.

4.3.2. Clustering Results

Table 1 and Table 2 demonstrate the clustering results with respect to the COIL-100 and AT&T ORL data sets. The two metrics used in our experiments were the accuracy (AC) and the normalized mutual information metric (NMI), detailed definition could be found in [14]. Implementation details are as follows:

1) For each k, we selected a total of k categories from the data sets randomly.

2) We implemented GNTD and compared methods to the selected k categories samples, and obtained their low-dimensional factor matrices.

3) We applied K-means to the factor matrix and repeated 50 times to mitigate the local convergence issue with different initial points.

We repeat the above three steps 20 times, and recorded the mean and standard error of performance. It could be observed that, with geometrical information and internal structure preserved, GNTD yield better clustering performance compare to NTD and GNMF.

5. CONCLUSIONS

In this paper, we propose a novel Graph regularized Nonnegative Tucker Decomposition (GNTD) method for tensor data representation. GNTD models tensor data in their original domain, thus their internal structure is preserved. Meanwhile, GNTD explicitly takes into account the underlying manifold structure of tensor data. Consequenty, GNTD is able to extract the parts-based representation and preserve the geometrical information from high-dimensional tensor data. Experimental results on the publicly available real-world data sets show that GNTD presents more competent representation and achieves better clustering performance.

6. ACKNOWLEDGEMENTS

This work was partly supported by the National Natural Science Foundation of China 61673124.

7. REFERENCES

[1] Andrzej Cichocki, Danilo Mandic, Lieven De Lathauwer, Guoxu Zhou, Qibin Zhao, Cesar Caiafa, and Huy Anh Phan, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 145–163, 2015.

[2] Guoxu Zhou, Andrzej Cichocki, and Shengli Xie, "Fast nonnegative matrix/tensor factorization based on low-rank approximation," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2928–2940, 2012.

[3] Xutao Li, Michael K Ng, Gao Cong, Yunming Ye, and Qingyao Wu, "Mr-ntd: manifold regularization nonnegative tucker decomposition for tensor data dimension reduction and representation," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 8, pp. 1787–1800, 2017.

[4] Guoxu Zhou, Andrzej Cichocki, Qibin Zhao, and Shengli Xie, "Efficient nonnegative tucker decompositions: Algorithms and uniqueness," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4990–5003, 2015.

[5] Mikhail Belkin and Partha Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Advances in neural information processing systems*, 2002, pp. 585–591.

[6] Joshua B Tenenbaum, Vin De Silva, and John C Langford, "A global geometric framework for nonlinear dimensionality reduction," *science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[7] Mikhail Belkin, Partha Niyogi, and Vikas Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *Journal of machine learning research*, vol. 7, no. Nov, pp. 2399–2434, 2006.

[8] Deng Cai, Xiaofei He, Jiawei Han, and Thomas S Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1548–1560, 2011.

[9] Yu-Xiong Wang, Liang-Yan Gui, and Yu-Jin Zhang, "Neighborhood preserving non-negative tensor factorization for image representation," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 3389–3392.

[10] Can Wang, Xiaofei He, Jiajun Bu, Zhengguang Chen, Chun Chen, and Ziyu Guan, "Image representation using laplacian regularized nonnegative tensor factorization," *Pattern Recognition*, vol. 44, no. 10-11, pp. 2516–2526, 2011.

[11] Bo Jiang, Chris Ding, Jin Tang, and Bin Luo, "Image representation and learning with graph-laplacian tucker tensor decomposition," *IEEE Transactions on Cybernetics*, 2018.

[12] Yong-Deok Kim and Seungjin Choi, "Nonnegative tucker decomposition," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–8.

[13] Tamara G Kolda and Brett W Bader, "Tensor decompositions and applications," *SIAM review*, vol. 51, no. 3, pp. 455–500, 2009.

[14] Wei Xu, Xin Liu, and Yihong Gong, "Document clustering based on non-negative matrix factorization," in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*. ACM, 2003, pp. 267–273.