IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 21, 2019, accepted October 28, 2019, date of publication November 4, 2019, date of current version November 27, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2951393

Similarity Learning-Induced Symmetric Nonnegative Matrix Factorization for Image Clustering

WEI YAN[1], BOB ZHANG[1], (Member, IEEE), ZUYUAN YANG[2], (Member, IEEE), AND SHENGLI XIE[2], (Fellow, IEEE)

[1]Department of Computer and Information Science, University of Macau, Macau 999078, China
[2]Guangdong Key Laboratory of IoT Information Technology, School of Automation, Guangdong University of Technology, Guangzhou 510006, China

Corresponding author: Bob Zhang (bobzhang@umac.mo)

This work was supported by the National Natural Science Foundation of China under Grant 61602540 and Grant 61722304.

ABSTRACT As a typical variation of nonnegative matrix factorization (NMF), symmetric NMF (SNMF) is capable of exploiting information of the cluster embedded in the matrix of similarity. The traditional SNMF-based methods for clustering first performs the techniques of the similarity learning on input data to learn a matrix of similarity, which is subsequently factorized by SNMF or one of its variants to learn information from the cluster. While these methods have led to satisfactory clustering results, it is suboptimal, since they do not explicitly exploit the fact that processes of the similarity learning and the clustering are depend on each other. In this paper, we describe a new SNMF model, termed similarity learning-induced SNMF (SLSNMF). SLSNMF can be considered as a unified framework that jointly considers these two processes. SLSNMF improves the clustering performance of SNMF by thoroughly exploring the mutual reinforcement between the process of similarity learning and the process of clustering until convergence. We incorporate a constraint into the standard SNMF model to learn the matrices of similarity and cluster simultaneously. Meanwhile, for solving this new model, we use the strategy of alternating iterative and derive an efficient algorithm, whose convergence is theoretically guaranteed. Experimental results over three benchmark image data sets demonstrate that SLSNMF outperforms the state-of-the-art methods for clustering.

INDEX TERMS Symmetric nonnegative matrix factorization (SNMF), nonnegative matrix factorization (NMF), clustering, unsupervised learning.

I. INTRODUCTION

Clustering is a fundamental task in machine learning and data mining. This task aims to group data into a number of partitions. Recently, Non-negative Matrix Factorization (NMF) [1]–[6] and Symmetric Non-negative Matrix Factorization (SNMF) [7]–[9] have been applied to data clustering with impressive outcomes. Specifically, NMF with an orthogonality constraint is equivalent to the traditional K-means clustering method, which enables it to efficiently group linearly separable data [10]. SNMF is closely related to spectral clustering (both of them solve the same problem with different constraints) [11], making it effectively cluster nonlinearly separable data.

The associate editor coordinating the review of this manuscript and approving it for publication was Shiqing Zhang.

SNMF aims to seek two of the same non-negative matrix factors. Mathematically, given a symmetric non-negetive matrix $\mathbf{X}$, SNMF finds the solution $\mathbf{V}$ such that $\mathbf{X} \approx \mathbf{V}\mathbf{V}^{\mathrm{T}}$. In graph clustering setting, $\mathbf{X}$ is matrix of the similarity which encodes the relationships between samples. The result $\mathbf{V}$ serves as the matrix of the clusters. To improve performance of SNMF for clustering, one often adds some constraints to the basic SNMF. For example, an effective pairwise constrained SNMF model was proposed in [12]. Also, by considering both information of the constraint and information of the geometrical structure, another SNMF variant was proposed in [13].

The existing SNMF-based clustering algorithms perform clustering in two steps. In the first step, a matrix of similarity is obtained from the input data by using the techniques of learning. In the second step, matrix of the cluster is learned by performing SNMF on the matrix of the similarity. Although this step-wise method has achieved satisfactory results, it is suboptimal, since this step-wise method totally ignores the dependence between the process of similarity learning and the process of clustering. Information of the cluster may not transfer well to the matrix of the indicator, which will compromise performance of clustering. Considering this disadvantage, it is natural to unify the SNMF-based clustering and similarity learning optimization into a single framework, enhancing performance of clustering. In fact, researchers have proved that a unified framework is effective in dealing with two dependent tasks [13]–[23]. A similar framework of joint learning has been proposed in [13]. This framework performs clustering and similarity learning simultaneously. However, this method is semi-supervised, and its performance highly depends on the amount of label information available. In [23], Kang *et al*. proposed a framework LKGr. LKGr unified the processes of graph construction and kernel learning by the strategy of iteration, where the graph and consensus kernel can be enhanced by each other.

In this paper, a novel regularized SNMF model, termed similarity learning-induced SNMF (SLSNMF), is proposed to enhance performance of SNMF in clustering. The main contributions of this paper are twofold.

1) A new unsuperivised SNMF-based clustering method is proposed, where the similarity learning and the clustering are unified into a joint framework. By formulating a constrained problem of optimization, matrices of the similarity and indicator are learned simultaneously, where both of them make mutual enforcement until convergence.
2) Based on the strategy of iterative updating, an efficient algorithm is presented to solve the new problem of optimization. This algorithm is proved to monotonically decrease the cost of the new objective function.

The remainder of this paper will be organized as follows. We briefly present the background of NMF and SNMF in Section 2. Section 3 provides our proposed similarity learning-induced symmetric non-negative matrix factorization, the updating rules, and the convergence analysis. We describe its performance in Section 4. Finally, Section 5 concludes this paper and provides suggestions for future work.

II. NMF AND SNMF

Given a nonnegative matrix $\mathbf{X} \in \mathcal{R}_+^{m\times n}$, NMF provides two low-rank nonnegative factor matrices $\mathbf{U} \in \mathcal{R}_+^{m\times k}$ and $\mathbf{V} \in \mathcal{R}_+^{n\times k}$ with $k < \min\{m, n\}$, whose product is a well approximation to $\mathbf{X}$. Mathematically, it is formulated as

$$\mathbf{X} \approx \mathbf{U}\mathbf{V}^{\mathrm{T}} \tag{1}$$

The quality of the above approximation can be measured by a cost function based on the Frobenius norm. Then, the objective function of NMF is formulated as:

$$\min ||\mathbf{X} - \mathbf{U}\mathbf{V}^{\mathrm{T}}||_F^2, \quad \text{s.t. } \mathbf{U} \geq 0, \quad \mathbf{V} \geq 0. \tag{2}$$

where $||\cdot||_F$ represents Frobenius norm.

NMF and its variants offer superior results for clustering on linearly separable data compared to others such as the traditional clustering method K-means [24]. In a clustering setting, the columns of $\mathbf{U}$ are the cluster centroids, and the columns of $\mathbf{V}^{\mathrm{T}}$ represents the clustering assignments. Note that the index of the largest value in each column of $\mathbf{V}^{\mathrm{T}}$ is the clustering assignment of the corresponding sample [11].

In contrast to NMF, SNMF is effective in clustering nonlinear data. In practice, the input of SNMF is a matrix of the similarity $\mathbf{S} \in \mathcal{R}^{n\times n}$. Mathematically, SNMF conducts a symmetric non-negative matrix factorization. Choosing the Frobenius norm to quantify the approximation quality, the objective function of SNMF is:

$$\min ||\mathbf{S} - \mathbf{V}\mathbf{V}^{\mathrm{T}}||_F^2, \quad \text{s.t. } \mathbf{V} \geq 0. \tag{3}$$

where $||\cdot||_F$ denotes Frobenius norm. SNMF can obtain similar or better performance than most of the nonlinear clustering methods, including the popular spectral clustering methods [25].

III. PROPOSED SLSNMF

A. PROBLEM FORMULATION

Typical SNMF-based clustering methods are based on two steps. In the first step, the similarity matrix is obtained from the input data through reconstruction-based methods. As for the second step, the SNMF-based model is utilized to the learned similarity matrix, resulting in a cluster assignment. In such a stepwise manner, the dependance between these two processes is totally ignored, and the resultant cluster indicator matrix may not well exploit the cluster information from the input data, resulting in a suboptimal clustering performance. To this end, we aim to simultaneously learn the similarity and cluster indicator matrices in a unified framework, in which the dependency between them can be well explored.

In [26], Kong et al. proposed an effective algorithm, termed iterative locally linear embedding (ILLE), to learn the similarity matrix $\mathbf{S}$. The objective function of ILLE is formulated as:

$$\begin{aligned}\mathcal{O}(\mathbf{S}) = \|\mathbf{X} - \mathbf{X}\mathbf{S}\|_F^2 + \alpha \mathrm{Tr}(\mathbf{S}^{\mathrm{T}}\mathbf{S}) + \beta\, \|\mathbf{S}\|_1 \\ \text{s.t. } \mathbf{S} \geq 0, \quad \alpha \geq 0,\ \beta \geq 0.\end{aligned} \tag{4}$$

where α and β are regularization parameters. $||\cdot||_1$ is the L1-norm. The first term in (4) is the reconstruction term. The second term penalizes the complexity of $\mathbf{S}$. The third term is applied to obtain a sparse solution [27]. Considering the similarity matrix learnt by (4) is not symmetric, we set the matrix $\widetilde{\mathbf{S}}$ to be symmetric by:

$$\widetilde{\mathbf{S}} = \frac{(\mathbf{S} + \mathbf{S}^T)}{2}$$

Combining the traditional SNMF model with the above-mentioned similarity learning model, a new SNMF model is formulated as:

$$\begin{aligned}\mathcal{O}(\mathbf{V}, \mathbf{S}) = \left\|\widetilde{\mathbf{S}} - \mathbf{V}\mathbf{V}^{\mathrm{T}}\right\|_F^2 + \|\mathbf{X} - \mathbf{X}\mathbf{S}\|_F^2 \\ +\alpha \mathrm{Tr}(\mathbf{S}^{\mathrm{T}}\mathbf{S}) + \beta\, \|\mathbf{S}\|_1\end{aligned} \tag{5}$$

In this model, the similarity matrix $\widetilde{\mathbf{S}}$ and clustering indicator matrix $\mathbf{V}$ communicate with each other until convergence. This shows that the dependence between the similarity learning and the clustering processes are explicitly explored.

B. ALGORITHM FOR SOLVING SLSNMF WITH CONVERGENCE ANALYSIS

For the new model in (5), we optimize it based on the widely used alternating update strategy, i.e., updating one factor while keeping the other fixed. Specifically, with initialized $\mathbf{S}$ and $\mathbf{V}$, optimizing $\mathbf{S}$ while keeping $\mathbf{V}$ fixed, and then optimizing $\mathbf{V}$ by keeping $\mathbf{S}$ fixed until convergence. Based on the majorization-minimization technique [28], we obtained the updating rules under which we can theoretically guarantee the objective function to be monotonically non-increasing.

Regarding the updating rules, we introduce an essential definition, an important lemma [29], and five propositions for the following analysis.

Definition 1: $\mathcal{J}(h, \hat{h})$ is an auxiliary function for $\mathcal{O}(h)$ if it satisfies the following two conditions

$$\mathcal{J}(h, \hat{h}) \geq \mathcal{O}(h), \quad \mathcal{J}(h, h) = \mathcal{O}(h).$$

Lemma 1: If $\mathcal{J}$ is an auxiliary function of $\mathcal{O}(h)$, then $\mathcal{O}(h)$ is nonincreasing under the update

$$h^{t+1} = \arg\min_{h} \mathcal{J}(h, h^t)$$

where t denotes the iteration number.

Please see [29] for the proof of the above lemma.

Proposition 1: (Quadratic Upper Bound) [30]: For any matrices $\mathbf{S} \in \mathbb{R}_+^{n\times n}$, $\mathbf{V} \in \mathbb{R}_+^{m\times n}$, and $\widehat{\mathbf{V}} \in \mathbb{R}_+^{m\times n}$, if $\mathbf{S} = \mathbf{S}^\mathsf{T}$, then it holds

$$\mathrm{Tr}(\widehat{\mathbf{V}}^\mathsf{T}\widehat{\mathbf{V}}\mathbf{S}) \leq \sum_{ik} \frac{(\mathbf{V}\mathbf{S})_{ik}\widehat{\mathbf{V}}_{ik}^2}{\mathbf{V}_{ik}} \tag{6}$$

Proposition 2: (Linear Upper Bound) [30]: For any matrices $\mathbf{S} \in \mathbb{R}_+^{n\times n}$, $\mathbf{V} \in \mathbb{R}_+^{m\times n}$, and $\widehat{\mathbf{V}} \in \mathbb{R}_+^{m\times n}$, it holds

$$\mathrm{Tr}(\mathbf{S}^\mathsf{T}\widehat{\mathbf{V}}) \leq \sum_{ik} \mathbf{S}_{ik}\left(\frac{\widehat{\mathbf{V}}_{ik}^2 + \mathbf{V}_{ik}^2}{2\mathbf{V}_{ik}}\right) \tag{7}$$

Proposition 3: (Quadratic Lower Bound) [30]: For any matrices $\mathbf{S} \in \mathbb{R}_+^{n\times n}$, $\mathbf{V} \in \mathbb{R}_+^{m\times n}$, and $\widehat{\mathbf{V}} \in \mathbb{R}_+^{m\times n}$, it holds

$$-\mathrm{Tr}(\widehat{\mathbf{V}}^\mathsf{T}\widehat{\mathbf{V}}\mathbf{S}) \leq -\sum_{ikl} \mathbf{S}_{kl}\mathbf{V}_{ik}\mathbf{V}_{il}\left(1 + \log\frac{\widehat{\mathbf{V}}_{ik}\widehat{\mathbf{V}}_{il}}{\mathbf{V}_{ik}\mathbf{V}_{il}}\right) \tag{8}$$

Proposition 4: For any matrices $\mathbf{V} \in \mathbb{R}_+^{m\times n}$, $\widehat{\mathbf{V}} \in \mathbb{R}_+^{m\times n}$, it holds [31]

$$\mathrm{Tr}(\widehat{\mathbf{V}}\widehat{\mathbf{V}}^\mathsf{T}\widehat{\mathbf{V}}\widehat{\mathbf{V}}^\mathsf{T}) \leq \sum_{ik} \frac{(\mathbf{V}\mathbf{V}\mathbf{V}^\mathsf{T})_{ik}\widehat{\mathbf{V}}_{ik}^4}{\mathbf{V}_{ik}^3} \tag{9}$$

Proposition 5: For any matrices $\mathbf{S} \in \mathbb{R}_+^{n\times n}$, $\mathbf{V} \in \mathbb{R}_+^{m\times n}$, and $\widehat{\mathbf{V}} \in \mathbb{R}_+^{m\times n}$, if $\mathbf{S} = \mathbf{S}^\mathsf{T}$, then it holds [32]

$$\mathrm{Tr}(\widehat{\mathbf{V}}\mathbf{S}\widehat{\mathbf{V}}) \leq \frac{1}{2}\sum_{ik} \frac{(\mathbf{V}^\mathsf{T}\mathbf{S})_{ik}\widehat{\mathbf{V}}_{ik}^2}{\mathbf{V}_{ik}} + \frac{1}{2}\sum_{ik} \frac{(\mathbf{S}\mathbf{V}^\mathsf{T})_{ik}\widehat{\mathbf{V}}_{ik}^2}{\mathbf{V}_{ik}} \tag{10}$$

Substituting $\widetilde{\mathbf{S}} = (\mathbf{S} + \mathbf{S}^\mathsf{T})/2$ into the objective function in (5), we have

$$\begin{aligned}\mathcal{O}(\mathbf{V}, \mathbf{S}) &= \left\|\widetilde{\mathbf{S}} - \mathbf{V}\mathbf{V}^\mathsf{T}\right\|_F^2 + \|\mathbf{X} - \mathbf{X}\mathbf{S}\|_F^2 \\ &\quad + \alpha\mathrm{Tr}(\mathbf{S}^\mathsf{T}\mathbf{S}) + \beta\|\mathbf{S}\|_1 \\ &= \frac{1}{2}\mathrm{Tr}(\mathbf{S}\mathbf{S} + \mathbf{S}^\mathsf{T}\mathbf{S}) - \mathrm{Tr}(\mathbf{S}\mathbf{V}\mathbf{V}^\mathsf{T} + \mathbf{S}^\mathsf{T}\mathbf{V}\mathbf{V}^\mathsf{T}) \\ &\quad + \mathrm{Tr}(\mathbf{V}\mathbf{V}^\mathsf{T}\mathbf{V}\mathbf{V}^\mathsf{T}) + \mathrm{Tr}(\mathcal{K} - 2\mathcal{K}\mathbf{S} + \mathbf{S}^\mathsf{T}\mathcal{K}\mathbf{S}) \\ &\quad + \alpha\mathrm{Tr}(\mathbf{S}^\mathsf{T}\mathbf{S}) + \beta\mathrm{Tr}(\mathbf{E}\mathbf{S})\end{aligned} \tag{11}$$

where $\mathcal{K} = \mathbf{X}^\mathsf{T}\mathbf{X}$.

Initially, we obtain the updating rule for $\mathbf{S}$ while keeping $\mathbf{V}$ fixed. By discarding the constant terms non-related to $\mathbf{S}$, the objective function in (11) is rewritten as

$$\begin{aligned}\mathcal{O}_1(\mathbf{S}) &= \frac{1}{2}\mathrm{Tr}(\mathbf{S}\mathbf{S} + \mathbf{S}^\mathsf{T}\mathbf{S}) - \mathrm{Tr}(\mathbf{S}\mathbf{V}\mathbf{V}^\mathsf{T} + \mathbf{S}^\mathsf{T}\mathbf{V}\mathbf{V}^\mathsf{T}) \\ &\quad + \mathrm{Tr}(\mathbf{S}^\mathsf{T}\mathcal{K}\mathbf{S}) + \alpha\mathrm{Tr}(\mathbf{S}^\mathsf{T}\mathbf{S}) \\ &\quad + \beta\mathrm{Tr}(\mathbf{E}\mathbf{S}) - \mathrm{Tr}(2\mathcal{K}\mathbf{S})\end{aligned} \tag{12}$$

Then, we construct the following function:

$$\begin{aligned}\mathcal{J}_1(\mathbf{S}, \widehat{\mathbf{S}}) &= \sum_{ik} \frac{(0.5\mathbf{S}^\mathsf{T})_{ik}\widehat{\mathbf{S}}_{ik}^2}{\mathbf{S}_{ik}} + \sum_{ik} \frac{(0.5\mathbf{S})_{ik}\widehat{\mathbf{S}}_{ik}^2}{\mathbf{S}_{ik}} \\ &\quad + \alpha\sum_{ik} \frac{\mathbf{S}_{ik}\widehat{\mathbf{S}}_{ik}^2}{\mathbf{S}_{ik}} + \beta\sum_{ik}\left(\frac{\mathbf{S}_{ik}^2 + \widehat{\mathbf{S}}_{ik}^2}{2\mathbf{S}_{ik}}\right) \\ &\quad + \frac{(\mathcal{K}\mathbf{S})_{ik}\widehat{\mathbf{S}}_{ik}^2}{\mathbf{S}_{ik}} - \mathrm{Tr}(\mathbf{V}\mathbf{V}^\mathsf{T}\widehat{\mathbf{S}}) \\ &\quad - \mathrm{Tr}(\mathbf{V}\mathbf{V}^\mathsf{T}\widehat{\mathbf{S}}^\mathsf{T}) - 2\mathrm{Tr}(\mathcal{K}\widehat{\mathbf{S}})\end{aligned} \tag{13}$$

where $\widehat{\mathbf{S}} \in \mathbb{R}^{n\times n}$. With Propositions 1, 2, 3 and 5, we obtain:

$$\mathcal{J}_1(\mathbf{S}, \widehat{\mathbf{S}}) \geq \mathcal{O}_1(\widehat{\mathbf{S}}), \quad \textit{and } \mathcal{J}_1(\widehat{\mathbf{S}}, \widehat{\mathbf{S}}) = \mathcal{O}_1(\widehat{\mathbf{S}})$$

According to Definition 1, we can verify that $\mathcal{J}_1(\mathbf{S}, \widehat{\mathbf{S}})$ is the auxiliary function $\mathcal{O}_1(\widehat{\mathbf{S}})$. Also, $\mathcal{J}_1(\mathbf{S}, \widehat{\mathbf{S}})$ is convex with respect to $\widehat{\mathbf{S}}$, because each term in $\mathcal{J}_1(\mathbf{S}, \widehat{\mathbf{S}})$ is convex. Thus, through setting the partial derivative of $\mathcal{J}_1(\mathbf{S}, \widehat{\mathbf{S}})$ with respect to factor $\widehat{\mathbf{S}}$ to zero, that is,

$$\begin{aligned}\frac{\partial\mathcal{J}_1(\mathbf{S}, \widehat{\mathbf{S}})}{\partial\widehat{\mathbf{S}}} &= (-2\mathbf{V}\mathbf{V}^\mathsf{T} - 2\mathcal{K})_{ik} + \frac{(\mathbf{S}^\mathsf{T})_{ik}\widehat{\mathbf{S}}_{ik}}{\mathbf{S}_{ik}} \\ &\quad + \frac{\mathbf{S}_{ik}\widehat{\mathbf{S}}_{ik}}{\mathbf{S}_{ik}} + \beta\left(\frac{\widehat{\mathbf{S}}_{ik}}{\mathbf{S}_{ik}}\right) \\ &\quad + 2\alpha\frac{\mathbf{S}_{ik}\widehat{\mathbf{S}}_{ik}}{\mathbf{S}_{ik}} + 2\frac{(\mathcal{K}\mathbf{S})_{ik}\widehat{\mathbf{S}}_{ik}}{\mathbf{S}_{ik}} \\ &= 0\end{aligned} \tag{14}$$

we can achieve the optimal solution of $\widehat{\mathbf{S}}$ for $\min_{\widehat{\mathbf{S}}} \mathcal{J}_1(\mathbf{S}, \widehat{\mathbf{S}})$ as

$$\widehat{\mathbf{S}}_{ik} = \mathbf{S}_{ik}\frac{(2\mathbf{V}\mathbf{V}^T + 2\mathcal{K})_{ik}}{(\mathbf{S}^\mathsf{T} + (\mathbf{I} + 2\alpha + 2\mathcal{K})\mathbf{S} + \beta)_{ik}} \tag{15}$$

Here, at the tth iteration, the solution for $\mathbf{S}$ is

$$\mathbf{S}_{ik}^{t+1} = \mathbf{S}_{ik}^t\frac{(2\mathbf{V}^t(\mathbf{V}^t)^\mathsf{T} + 2\mathcal{K})_{ik}}{((\mathbf{S}^t)^\mathsf{T} + (\mathbf{I} + 2\alpha + 2\mathcal{K})\mathbf{S}^t + \beta)_{ik}} \tag{16}$$

To obtain the solution for $\mathbf{V}$ by keeping $\mathbf{S}$ fixed, the objective function in (11) is rewritten through discarding the constant terms nonrelated to $\mathbf{V}$, which results in

$$\mathcal{O}_2(\widehat{\mathbf{S}}) = \mathrm{Tr}(\mathbf{V}\mathbf{V}^{\mathsf{T}}\mathbf{V}\mathbf{V}^{\mathsf{T}} - \mathbf{S}\mathbf{V}\mathbf{V}^{\mathsf{T}} - \mathbf{S}^{\mathsf{T}}\mathbf{V}\mathbf{V}^{\mathsf{T}})$$

Then, we construct the following function:

$$\mathcal{J}_2(\mathbf{V}, \widehat{\mathbf{V}}) = \sum_{ik} \frac{(\mathbf{V}\mathbf{V}^{\mathsf{T}}\mathbf{V})_{ik}\widehat{\mathbf{V}}_{ik}^4}{\mathbf{V}_{ik}^3} - \sum_{ik} (\mathbf{S}+\mathbf{S}^{\mathsf{T}})_{kl}\mathbf{V}_{ik}\mathbf{V}_{il}(1+\log\frac{\widehat{\mathbf{V}}_{ik}\widehat{\mathbf{V}}_{il}}{\mathbf{V}_{ik}\mathbf{V}_{il}}) \tag{17}$$

where $\widehat{\mathbf{V}} \in \mathbb{R}^{m\times c}$. With Propositions 3 and 4, we conclude that

$$\mathcal{J}_2(\mathbf{V}, \widehat{\mathbf{V}}) \geq \mathcal{O}_2(\widehat{\mathbf{V}}), \quad \text{and } \mathcal{J}_2(\widehat{\mathbf{V}}, \widehat{\mathbf{V}}) = \mathcal{O}_2(\widehat{\mathbf{V}})$$

According to Definition 1, $\mathcal{J}_2(\mathbf{V}, \widehat{\mathbf{V}})$ is the auxiliary function for $\mathcal{O}_2(\widehat{\mathbf{V}})$, and it is convex with respect to $\widehat{\mathbf{V}}$. To find the minimum of $\mathcal{J}_2(\mathbf{V}, \widehat{\mathbf{V}})$, we have

$$\frac{\partial \mathcal{J}_2(\mathbf{V}, \widehat{\mathbf{V}})}{\partial \widehat{\mathbf{V}}_{ik}} = 4\frac{(\mathbf{V}\mathbf{V}^{\mathsf{T}}\mathbf{V})_{ik}\widehat{\mathbf{V}}_{ik}^3}{\mathbf{V}_{ik}^3} - 2\frac{(\mathbf{S}\mathbf{V}+\mathbf{S}^{\mathsf{T}}\mathbf{V})_{ik}\mathbf{V}_{ik}}{\widehat{\mathbf{V}}_{ik}} = 0 \tag{18}$$

and the optimal $\widehat{\mathbf{V}}$ in elementwise for $\min_{\widehat{\mathbf{V}}} \mathcal{J}_2(\mathbf{V}, \widehat{\mathbf{V}})$

$$\widehat{\mathbf{V}}_{ik} = \mathbf{V}_{ik}\sqrt[4]{\frac{(\mathbf{S}\mathbf{V}+\mathbf{S}^{\mathsf{T}}\mathbf{V})_{ik}}{2(\mathbf{V}\mathbf{V}^{\mathsf{T}}\mathbf{V})_{ik}}} \tag{19}$$

Therefore, at the tth iteration, the update formula for $\mathbf{V}$ is

$$\mathbf{V}_{ik}^{t+1} = \mathbf{V}_{ik}^{t}\sqrt[4]{\frac{(\mathbf{S}^{t+1}\mathbf{V}^{t}+(\mathbf{S}^{t+1})^{\mathsf{T}}\mathbf{V}^{t})_{ik}}{2(\mathbf{V}^{t}(\mathbf{V}^{t+1})^{\mathsf{T}}\mathbf{V}^{t})_{ik}}} \tag{20}$$

According to Lemma 1, under the update formulas in (16) and (20), we obtain

$$\mathcal{O}(\mathbf{V}^{t}, \mathbf{S}^{t+1}) \leq \mathcal{O}_1(\mathbf{V}^{t}, \mathbf{S}^{t}) \tag{21}$$

$$\mathcal{O}(\mathbf{V}^{t+1}, \mathbf{S}^{t+1}) \leq \mathcal{O}_1(\mathbf{V}^{t}, \mathbf{S}^{t+1}) \tag{22}$$

Hence, the value of objective function $\mathcal{O}(\mathbf{V}, \mathbf{S})$ is non-increasing under the update rules in (16) and (20). Meanwhile, it is not difficult to verify that the value of the objective function is larger than 0. Finally, we conclude that the convergence of the proposed algorithm is theoretically guaranteed. We summarize the algorithm in Algorithm 1.

Algorithm 1 Algorithm of the Proposed SLSNMF

Input:
input data $\mathbf{X} \in \mathcal{R}^{m\times n}$, cluster number k, parameters α and β, and *maxIter*

Output:
cluster indicator matrix $\mathbf{V}^{\mathsf{T}}$;

Initialize matrices $\mathbf{V}$, and $\mathbf{S}$;
$t = 1$;
while ($t < maxIter$) **do**
    Update $\mathbf{V}$ with rule (20);
    Update $\mathbf{S}$ with rule (16);
    $t = t + 1$;
end while
return the index of the largest value in each column of $\mathbf{V}^{\mathsf{T}}$ as the clustering indicator;

IV. EXPERIMENTAL RESULTS

For this section, we evaluated the performances of SLSNMF on data clustering using three benchmark image data sets (PIE, MNIST and ORL). To show the advantages of the proposed SLSNMF, we compared it with several algorithms, including three recent SNMF algorithms, i.e., NS-SymNMF [33], SymNMF [11], $\alpha-$SNMF [8], and classical K-means. Regarding the effectiveness of our joint optimization algorithm SLSNMF, we also designed a comparison method, termed "disjoint SL/SNMF". In disjoint SL/SNMF, the similarity matrix is first obtained by solving the objective function in (4), followed by performing SNMF on the learned similarity matrix to obtain the cluster assignment.

To evaluate the clustering performance, two standard metrics: the accuracy (AC) and the normalized mutual information (NMI) [34] are applied. For these two metrics, a value close to 1 implies a good clustering result.

The value of accuracy represents the percentage of correctly predicted labels, and is calculated by:

$$\mathrm{AC} = \frac{\sum_{i=1}^{n}\delta(r_i, map(l_i))}{n} \tag{23}$$

where $\delta(a, b)$ takes 1 if $a = b$ and 0 otherwise, and $map(l_i)$ is the permutation mapping function that maps each cluster label l_i to the corresponding label from the data set. In our experiments, the best map is obtained via the Kuhn-Munkres algorithm [35].

The NMI matrix is used for measuring the similarity of two clusters. Given two clusters C and C', the NMI(C, C') is defined as:

$$\mathrm{NMI(C, C')} = \frac{\mathrm{MI}(C, C')}{max(\mathrm{H}(C), \mathrm{H}(C'))}, \tag{24}$$

where H(C) and H(C') denote the entropies of C and C'. The corresponding mutual information matrix (MI) is given as:

$$\mathrm{MI}(C, C') = \sum_{c_i\in C, c_j'\in C'} p(c_i, c_j') \cdot log\frac{p(c_i, c_j')}{p(c_i)\cdot p(c_j')}, \tag{25}$$

where $p(c_i)$ and $p(c_j')$ represent the probabilities that a randomly selected data belongs to the clusters c_i and c_j, respectively, and $p(c_i, c_j')$ is the joint probability that the selected point belongs to both clusters.

For a fair comparison, we use the technique of Gaussian kernel to obtain the similarity matrix for all methods. Given the input data $\mathbf{X}$, the kernel matrix $\mathcal{K}(x_i, x_j)$ is defined as $\mathcal{K}(x_i, x_j) = \exp(-\|x_i - x_j\|^2/t)$, where t denotes the bandwidth parameter. For the three data sets (PIE, MNIST,

TABLE 1. AC value compariosn on PIE database.

Class	K-means	α-SNMF	SymNMF	NS-SymNMF	Disjoint SL/SNMF	SLSNMF
2	60.88	93.57	97.86	97.26	90.12	**98.93**
3	53.17	79.13	89.13	76.19	87.70	**94.37**
4	42.38	66.90	75.24	69.94	72.80	**85.24**
5	40.33	65.00	83.90	71.14	72.24	**95.05**
6	39.63	62.66	79.09	63.10	66.83	**87.94**
7	34.90	60.82	76.63	64.18	67.01	**82.24**
8	31.88	58.10	76.82	62.41	65.51	**86.13**
9	31.14	56.69	73.10	60.77	60.37	**83.94**
10	39.26	47.95	62.83	55.48	57.88	**78.81**
Av.	41.51	65.65	79.40	68.94	71.16	**88.07**

TABLE 2. NMI value compariosn on PIE database.

Class	K-means	α-SNMF	SymNMF	NS-SymNMF	Disjoint SL/SNMF	SLSNMF
2	10.32	81.23	93.71	90.88	65.64	**94.38**
3	24.72	66.04	80.58	62.94	74.01	**84.23**
4	27.76	57.56	78.52	62.04	67.69	**81.63**
5	26.00	58.70	87.47	71.22	71.53	**90.05**
6	32.32	61.39	84.38	66.07	67.08	**85.61**
7	32.55	63.34	81.44	66.95	70.73	**82.99**
8	33.84	61.98	85.33	69.07	71.63	**86.51**
9	39.53	62.56	82.92	67.48	69.17	**86.57**
10	34.92	58.19	75.90	66.27	69.41	**83.78**
Av.	28.99	63.44	83.36	69.21	69.65	**86.19**

and ORL), the parameter t is empirically set 0.7, 0.7, and 0.25, respectively.

A. PIE

The PIE database is composed of 2856 grayscale images of 68 individuals. There exist 42 images per subject, one per different light or illumination conditions.

First, we verified the convergence of the proposed SLSNMF algorithm. We conducted the experiments under different class numbers and various parameters. It is observed that the results are similar and all of them obtain good convergence performances. Here, we give the result for the case $k = 6$ in Fig. 1.

Next, we evaluated the clustering performance of SLSNMF, and compared it to those of the classical K-means, α-SNMF [8], SymSNMF [11], NS-SymNMF [33], and disjoint SL/SNMF. Fig. 2 shows the clustering performance of these algorithms in terms of AC, with more detailed results of AC and NMI shown in Table 1 and 2. It can be seen that the proposed SLSNMF is superior to other clustering methods in terms of AC and NMI. This demonstrates the effectiveness of the proposed method. Notably, the clustering performance of SLSNMF is higher than that of disjoint SL/SNMF method, implying that the effectiveness of the joint learning framework.

We can observe that our proposed method SLSNMF consistently outperforms others. This indicates it is useful at exploiting the mutual enhancement of similarity learning and clustering. When compared to the second best results, i.e., average results in terms of AC and NMI for SymSNMF, SLSNMF obtains 8.67 % and 2.83 % improvements, respectively.

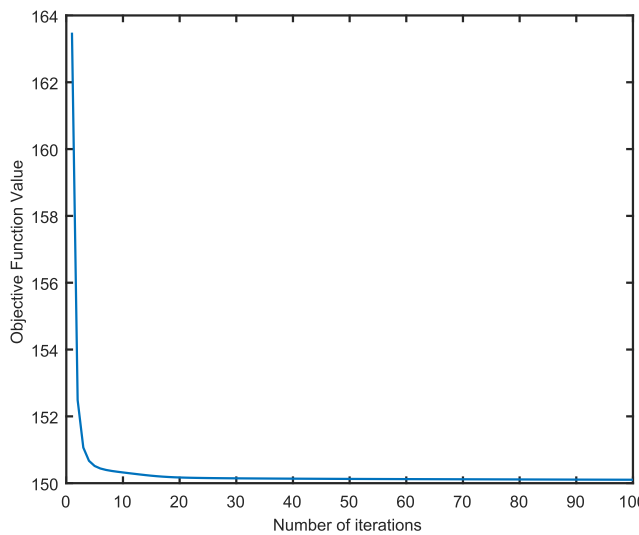


FIGURE 1. Convergence curve of SLSNMF on the PIE database when $k = 6$.

B. MNIST

The MNIST database contains normalized; center cropped images of handwritten digits ranging from zero to nine of

TABLE 3. AC value compariosn on MNIST database.

Class	K-means	α-SNMF	SymNMF	NS-SymNMF	Disjoint SL/SNMF	SLSNMF
2	55.67	85.33	90.60	87.00	88.67	**90.67**
3	54.78	80.44	83.33	83.33	86.22	**88.22**
4	55.33	73.17	81.50	82.50	84.67	**85.67**
5	55.93	78.13	80.53	84.40	78.93	**85.47**
6	51.33	78.78	83.22	81.78	83.33	**85.22**
7	50.52	66.10	85.03	83.62	81.52	**85.24**
8	41.47	66.08	83.17	79.17	80.00	**84.58**
9	40.30	67.26	80.96	78.52	81.15	**83.85**
10	40.93	66.73	79.07	75.40	80.47	**82.67**
Av.	49.59	73.56	83.05	81.75	82.77	**85.73**

TABLE 4. NMI value compariosn on MNIST database.

Class	K-means	αSNMF	SymNMF	NS-SymNMF	Disjoint SL/SNMF	SLSNMF
2	55.53	74.12	80.80	77.50	78.80	**81.62**
3	54.14	68.56	72.45	71.24	73.34	**78.05**
4	50.68	66.55	65.85	69.22	71.94	**76.78**
5	53.45	67.84	71.11	72.61	69.94	**75.44**
6	52.85	68.88	74.53	73.39	74.98	**75.05**
7	47.09	66.06	74.81	75.42	74.76	**75.97**
8	42.28	62.69	74.07	71.75	69.18	**74.18**
9	44.30	64.43	69.91	67.88	70.23	**72.45**
10	43.50	59.14	68.94	61.27	68.46	**70.63**
Av.	49.31	66.47	72.50	71.14	72.40	**75.57**

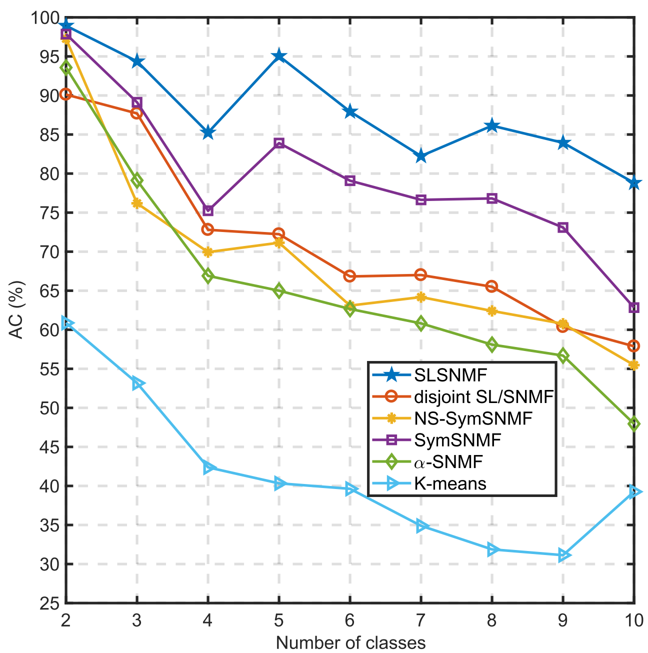


FIGURE 2. Clustering performance comparison of various methods on the PIE data set in terms of AC.

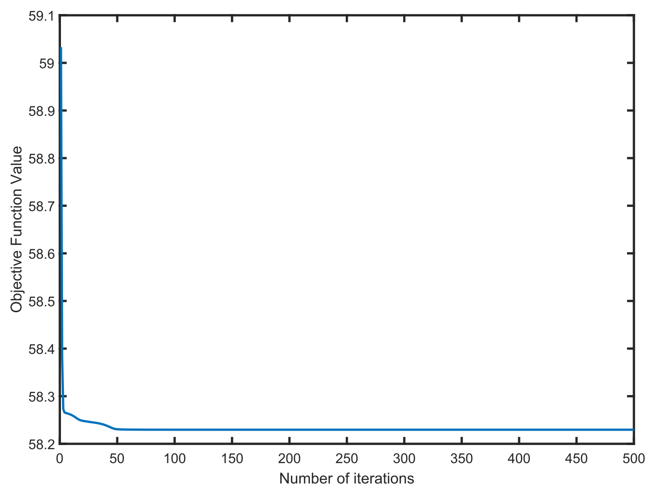


FIGURE 3. Convergence curve of SLSNMF on the MNIST database when $k = 4$.

size 28 × 28 [36]. The convergence curve of SLSNMF on the MNIST data set is shown in Fig. 3. Fig. 4 shows the AC of the comparison methods along with SLSNMF versus the class number, with more detailed results given in Table 3. Furthermore, Table 4 presents the corresponding NMI of all methods. We can observe that the average AC and NMI results of SLSNMF are among the best when compared to the other algorithms.

SLSNMF achieves a 2.68 % improvement in AC and 3.07 % improvement in NMI on average, compared to the next best method (i.e., SymNMF).

TABLE 5. AC value compariosn on ORL database.

Class	K-means	α-SNMF	SymNMF	NS-SymNMF	Disjoint SL/SNMF	SLSNMF
2	55.50	76.00	87.50	88.50	85.50	**89.50**
3	59.67	81.67	80.67	82.67	82.00	**84.00**
4	51.25	68.25	64.00	66.75	68.25	**70.75**
5	57.80	61.00	65.00	67.20	62.00	**70.60**
6	53.00	56.00	55.67	60.50	55.83	**64.50**
7	47.43	51.14	54.43	54.43	52.14	**58.57**
8	41.50	49.25	51.50	53.25	52.75	**54.63**
9	39.67	42.00	51.11	49.11	49.33	**55.89**
10	43.90	47.00	50.30	48.60	47.30	**52.70**
Av.	49.97	59.15	62.24	63.45	61.68	**66.79**

TABLE 6. NMI value compariosn on ORL database.

Class	K-means	αSNMF	SymNMF	NS-SymNMF	Disjoint SL/SNMF	SLSNMF
2	48.64	64.74	68.06	71.06	70.70	**73.52**
3	47.74	60.15	68.18	68.79	68.54	**70.65**
4	46.49	57.56	64.66	63.37	63.37	**65.22**
5	40.83	58.74	59.40	61.57	59.36	**64.29**
6	41.50	54.57	57.56	59.28	59.85	**60.38**
7	36.61	55.34	60.19	58.43	58.77	**61.16**
8	37.11	54.25	58.57	56.67	56.85	**60.39**
9	36.02	56.71	59.66	60.41	57.70	**61.82**
10	36.50	53.27	59.53	58.93	55.68	**60.54**
Av.	41.27	57.26	61.76	62.06	61.21	**64.22**

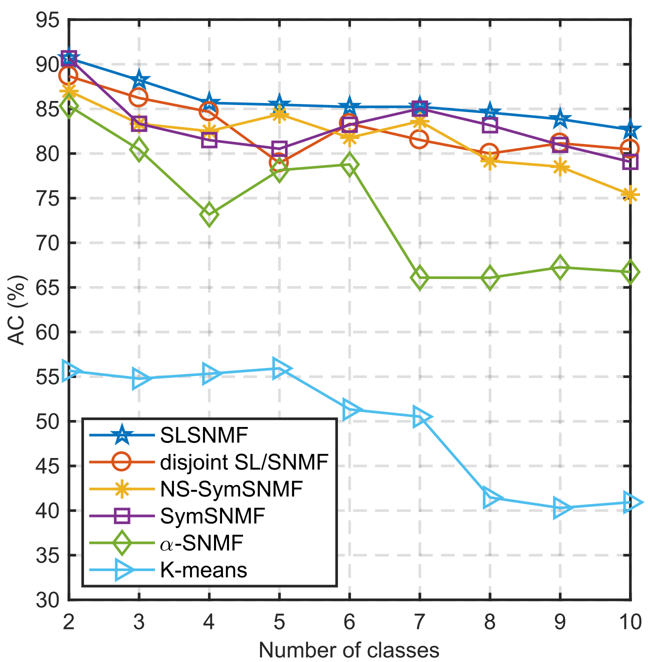


FIGURE 4. Clustering performance comparison of various methods on the MNIST data set in terms of AC.

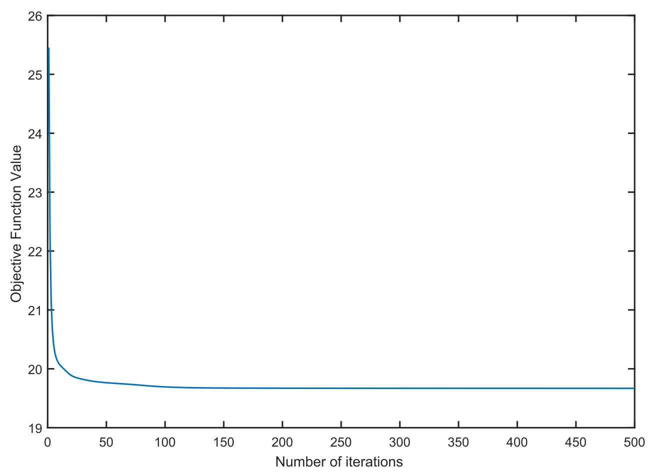


FIGURE 5. Convergence curve of SLSNMF on the ORL database when $k = 5$.

C. ORL

The ORL database includes 40 subjects with ten gray-scale images for each subject. All the images were resized to 32×32. We provide the performance of convergence with respect to the proposed method SLSNMF in Fig 5. Next, we show the AC of the comparison methods along with SLSNMF versus the class number in Fig. 6. In addition, we present the comprehensive results of AC and NMI in Table 5 and Table 6, respectively. One can observe from the figures and tables that the proposed method outperforms the comparison methods in terms of AC and NMI.

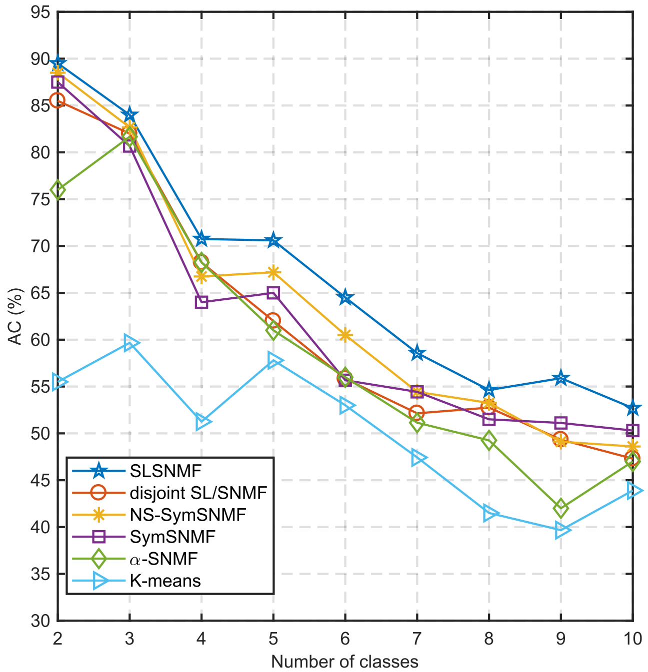


FIGURE 6. **Clustering performance comparison of various methods on the ORL data set in terms of AC.**

When matched with the algorithm that performed second best (NS-SymNMF), SLSNMF achieves a 3.34 % improvement in AC, and 2.16 % improvement in NMI on average.

D. PARAMETERS ANALYSIS

Our SLSNMF has two regularization parameters (i.e., α and β). Their impact on SLSNMF were investigated on the clustering performance.

Both α and β were varied in the range of {0.01, 100} on the three data sets. The average performance is shown in Fig. 7. In each of the experiments, 10 repetitions were conducted. From Fig. 7, we can observe that the proposed method performs well when α is selected in the range of [10, 100], where the optimal performance is obtained over PIE and MNIST when α is around 100 and β is around 10^{-2}. As for the ORL data set, SLSNMF achieved good performance when α was 100 and β was 0.1.

E. COMPUTATIONAL COMPLEXITY ANALYSIS

This subsection will conduct a computational complexity analysis of SLSNMF. The updating rules for SLSNMF have four arithmetic operations: (1) addition; (2) multiplication; (3) division; (4) square root operation. Based on the updating rules in (16) and (20), the arithmetic operations of each iteration are $\mathcal{O}(n^2k)$, $\mathcal{O}(n^2k)$, $\mathcal{O}(n^2)$, and $\mathcal{O}(nk)$, respectively. Thus, the overall computational complexity for SLSNMF is $\mathcal{O}(tn^2k)$, where t is the iteration number.

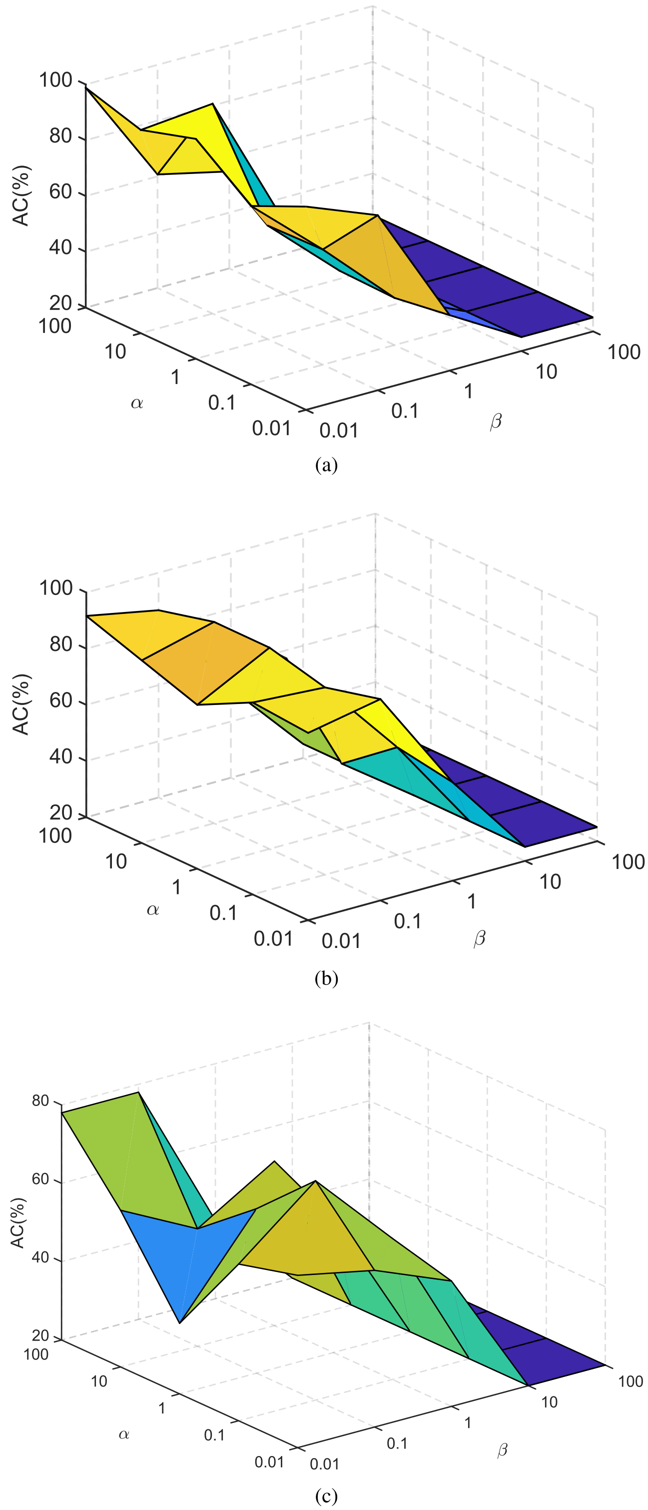


FIGURE 7. **Clustering performance of the proposed SLSNMF versus parameters (i.e., α and β) on three data sets. (a) PIE. (b) MNIST. (c) ORL.**

V. CONCLUSION AND FUTURE WORK

This paper proposed a novel constrained SNMF method, termed SLSNMF. In contrast to classic SNMF-based clustering methods, which are stepwise, SLSNMF is able to simultaneously perform similarity learning and clustering. To this end, SLSNMF unified the similarity learning and clustering processes into a single-constrained optimization problem. The dependency between the similarity learning and clustering processes is thoroughly explored so that the resultant cluster indicator matrix can well exploit the cluster information from the input data. To effectively solve this problem, an alternative iterative algorithm was derived with theoretical convergence. Experimentation using three benchmark image data sets showed the superiority of our method.

The proposed SLSNMF is currently a single-view learning method. Due to the availability of multi-view data, it is natural and capable of extending the proposed SLSNMF method to a multi-view learning one [37]–[40]. Therefore, we will consider this as part of our future work. It is also of great interest to extend the proposed SLSNMF method to a robust one [41]. This would offer researchers a better understanding of the effectiveness of SLSNMF for the problem of robust image clustering.

REFERENCES

[1] N. Guan, T. Liu, Y. Zhang, D. Tao, and L. S. Davis, "Truncated cauchy non-negative matrix factorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 1, pp. 246–259, Jan. 2017.

[2] Z. Yang, Y. Zhang, Y. Xiang, W. Yan, and S. Xie, "Non-negative matrix factorization with dual constraints for image clustering," *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published.

[3] D. Wang, X. Gao, and X. Wang, "Semi-supervised nonnegative matrix factorization via constraint propagation," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 233–244, Jan. 2016.

[4] Z. Li, J. Tang, and X. He, "Robust structured nonnegative matrix factorization for image representation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1947–1960, May 2018.

[5] Z. Yang, Y. Zhang, W. Yan, Y. Xiang, and S. Xie, "A fast non-smooth nonnegative matrix factorization for learning sparse representation," *IEEE Access*, vol. 4, pp. 5161–5168, 2016.

[6] W. Yan, B. Zhang, S. Ma, and Z. Yang, "A novel regularized concept factorization for document clustering," *Knowl.-Based Syst.*, vol. 135, pp. 147–158, Nov. 2017.

[7] Z. Yang, J. Corander, and E. Oja, "Low-rank doubly stochastic matrix decomposition for cluster analysis," *J. Mach. Learn. Res.*, vol. 17, no. 1, pp. 6454–6478, 2016.

[8] Z. He, S. Xie, R. Zdunek, G. Zhou, and A. Cichocki, "Symmetric non-negative matrix factorization: Algorithms and applications to probabilistic clustering," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 2117–2131, Dec. 2011.

[9] Z. Zhu, X. Li, K. Liu, and Q. Li, "Dropping symmetry for fast symmetric nonnegative matrix factorization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 5154–5164.

[10] C. Ding, T. Li, W. Peng, and H. Park, "Orthogonal nonnegative matrix t-factorizations for clustering," in *Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2006, pp. 126–135.

[11] D. Kuang, S. Yun, and H. Park, "SymNMF: Nonnegative low-rank approximation of a similarity matrix for graph clustering," *J. Global Optim.*, vol. 62, no. 3, pp. 545–574, Jul. 2015.

[12] X. Zhang, L. Zong, X. Liu, and J. Luo, "Constrained clustering with nonnegative matrix factorization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 7, pp. 1514–1526, Jul. 2016.

[13] W. Wu, Y. Jia, S. Kwong, and J. Hou, "Pairwise constraint propagation-induced symmetric nonnegative matrix factorization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 12, pp. 6348–6361, Dec. 2018.

[14] J. Wen, X. Fang, Y. Xu, C. Tian, and L. Fei, "Low-rank representation with adaptive graph regularization," *Neural Netw.*, vol. 108, pp. 83–96, Dec. 2018.

[15] X. Fang, S. Teng, Z. Lai, Z. He, S. Xie, and W. K. Wong, "Robust latent subspace learning for image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2502–2515, Jun. 2018.

[16] X. Fang, Y. Xu, X. Li, Z. Lai, and W. K. Wong, "Learning a nonnegative sparse graph for linear regression," *IEEE Trans. Image Process.*, vol. 24, no. 9, pp. 2760–2771, Sep. 2015.

[17] X. Fang, Y. Xu, X. Li, Z. Lai, and W. K. Wong, "Robust semi-supervised subspace clustering via non-negative low-rank representation," *IEEE Trans. Cybern.*, vol. 46, no. 8, pp. 1828–1838, Aug. 2016.

[18] X. Fang, N. Han, J. Wu, Y. Xu, J. Yang, W. K. Wong, and X. Li, "Approximate low-rank projection learning for feature extraction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5228–5241, Nov. 2018.

[19] J. Wen, X. Fang, J. Cui, L. Fei, K. Yan, Y. Chen, and Y. Xu, "Robust sparse linear discriminant analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 2, pp. 390–403, Feb. 2019.

[20] X. Fang, N. Han, W. K. Wong, S. Teng, J. Wu, S. Xie, and X. Li, "Flexible affinity matrix learning for unsupervised and semisupervised classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 4, pp. 1133–1149, Apr. 2019.

[21] J. Wen, N. Han, X. Fang, L. Fei, K. Yan, and S. Zhan, "Low-rank preserving projection via graph regularized reconstruction," *IEEE Trans. Cybern.*, vol. 49, no. 4, pp. 1279–1291, Apr. 2019.

[22] Z. Kang, H. Xu, B. Wang, H. Zhu, and Z. Xu, "Clustering with similarity preserving," *Neurocomputing*, vol. 365, pp. 211–218, Nov. 2019.

[23] Z. Kang, L. Wen, W. Chen, and Z. Xu, "Low-rank kernel learning for graph-based clustering," *Knowledge-Based Syst.*, vol. 163, pp. 510–517, Jan. 2019.

[24] T. Li and C. Ding, "The relationships among various nonnegative matrix factorization methods for clustering," in *Proc. 6th IEEE Int. Conf. Data Mining*, Dec. 2006, pp. 362–371.

[25] U. von Luxburg, "A tutorial on spectral clustering," *Statist. Comput.*, vol. 17, no. 4, pp. 395–416, Dec. 2007.

[26] D. Kong, C. Ding, H. Huang, and F. Nie, "An iterative locally linear embedding algorithm," in *Proc. 29th Int. Conf. Int. Conf. Mach. Learn.*, 2012, pp. 931–938.

[27] Z. Kang, C. Peng, and Q. Cheng, "Kernel-driven similarity learning," *Neurocomputing*, vol. 267, pp. 210–219, Dec. 2017.

[28] Y. Sun, P. Babu, and D. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2016.

[29] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Adv. Neural Inf. Process, Syst.*, 2001, pp. 556–562.

[30] Z. Yang and E. Oja, "Linear and nonlinear projective nonnegative matrix factorization," *IEEE Trans. Neural Netw.*, vol. 21, no. 5, pp. 734–749, May 2010.

[31] H. Wang, H. Huang, and C. Ding, "Simultaneous clustering of multi-type relational data via symmetric nonnegative matrix tri-factorization," in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manage.*, 2011, pp. 279–284.

[32] X. Pei, C. Chen, and Y. Guan, "Joint sparse representation and embedding propagation learning: A framework for graph-based semisupervised learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 12, pp. 2949–2960, Dec. 2017.

[33] S. Lu, M. Hong, and Z. Wang, "A nonconvex splitting method for symmetric nonnegative matrix factorization: Convergence analysis and optimality," *IEEE Trans. Signal Process.*, vol. 65, no. 12, pp. 3120–3135, Jun. 2017.

[34] D. Cai, X. He, and J. Han, "Locally consistent concept factorization for document clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 6, pp. 902–913, Jun. 2011.

[35] L. Lovász and M. D. Plummer, *Matching Theory*, vol. 367. Providence, RI, USA: AMS, 2009.

[36] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[37] J. Wang, F. Tian, H. Yu, C. H. Liu, K. Zhan, and X. Wang, "Diverse non-negative matrix factorization for multiview data representation," *IEEE Trans. Cybern.*, vol. 48, no. 9, pp. 2620–2632, Sep. 2017.

[38] J. Wen, Y. Xu, and H. Liu, "Incomplete multiview spectral clustering with adaptive graph learning," *IEEE Trans. Cybern.*, to be published.

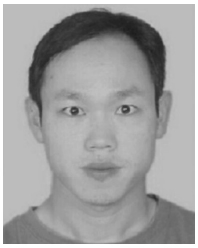
[39] H. Liu and Y. Fu, "Consensus guided multi-view clustering," *ACM Trans. Knowl. Discovery Data*, vol. 12, no. 4, 2018, Art. no. 42.

[40] Z. Kang, G. Shi, S. Huang, W. Chen, X. Pu, J. T. Zhou, and Z. Xu, "Multi-graph fusion for multi-view spectral clustering," 2019, *arXiv:1909.06940*. [Online]. Available: https://arxiv.org/abs/1909.06940

[41] Z. Kang, H. Pan, S. C. H. Hoi, and Z. Xu, "Robust graph learning from noisy data," *IEEE Trans. Cybern.*, to be published.

WEI YAN was born in Jiangxi, China. He received the master's degree from the School of Automation, Guangdong University of Technology, Guangzhou, China, in 2016. He is currently pursuing the Ph.D. degree with the Faculty of Science and Technology, University of Macau. His research interests include machine learning, nonnegative signal processing, and blind signal processing.

BOB ZHANG (M'11) received the B.A. degree in computer science from York University, Toronto, ON, Canada, in 2006, the M.A.Sc. degree in information systems security from Concordia University, Montréal, QC, Canada, in 2007, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2011. He was with the Center for Pattern Recognition and Machine Intelligence, and later was a Postdoctoral Researcher with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA. He is currently an Associate Professor with the Department of Computer and Information Science, University of Macau, Zhuhai, China. His current research interests include medical biometrics, pattern recognition, image processing, and blockchain.



ZUYUAN YANG (M'15) received the B.E. degree from the Hunan University of Science and Technology, Xiangtan, China, in 2003, and the Ph.D. degree from the South China University of Technology, Guangzhou, China, in 2010. He joined the National Program for New Century Excellent Talents in University. He is currently a Researcher with the School of Automation, Guangdong University of Technology, Guangzhou, China. His research interests include machine learning, nonnegative matrix factorization, and image processing. He was a recipient of the excellent Ph.D. Thesis Award Nomination of China and the Guangdong Distinguished Young Scholar.

SHENGLI XIE (M'01–F'19) received the M.S. degree in mathematics from Central China Normal University, Wuhan, China, in 1992, and the Ph.D. degree in control theory and applications from the South China University of Technology, Guangzhou, China, in 1997. He is currently the Director of the Laboratory for Intelligent Information Processing (LIIP) and a Full Professor with the Guangdong University of Technology, Guangzhou. He has authored or coauthored two monographs and more than 100 scientific articles published in journals and conference proceedings. His current research interests include automatic control and signal processing, especially blind signal processing, and image processing.