AIMS Mathematics, 10(5): 11676–11707.
DOI:10.3934/math.2025529
Received: 19 March 2025
Revised: 11 May 2025
Accepted: 16 May 2025
Published: 21 May 2025

https://www.aimspress.com/journal/Math

Research article

Clustering research on graph regularization nonnegative matrix factorization based on auxiliary variables under orthogonal conditions

Caiping Wang[1], Wen Li[2], Junjian Zhao[2,*] and Yasong Chen[2,*]

[1] Information Center, Zhangjiakou Cigarette Factory Co., Ltd., Zhangjiakou 075000, China
[2] School of Mathematical Sciences, Tiangong University, Tianjin 300387, China

* **Correspondence:** Email: zhaojunjian@tiangong.edu.cn, chenyasong@tiangong.edu.cn.

Abstract: This paper addressed the challenge of image clustering by integrating graph and orthogonality mechanisms into the nonnegative matrix factorization algorithm. It presented a novel model named graph regularized nonnegative matrix factorization with auxiliary variable orthogonal subspace (GNMFOSV). This innovative approach not only provided a rigorous proof of algorithm convergence using auxiliary variables, thereby filling a significant gap in the theoretical validation of similar algorithms under orthogonal conditions, but also effectively captured the nonlinear relationships in the reconstructed data. Additionally, it enhanced the sparsity of the decomposition results, leading to a notable improvement in clustering performance. To verify the effectiveness of the proposed method, comprehensive clustering tests were conducted on diverse datasets. The experimental results clearly demonstrated that the GNMFOSV algorithm outperformed existing methods in terms of clustering performance, indicating its great potential for practical applications.

Keywords: NMF; graph Laplacian; orthogonality; clusterting; auxiliary variables
Mathematics Subject Classification: 62H30, 62H35, 68Q32, 68T10, 90C26

1. Introduction

In pattern recognition, cluster analysis is a formal study of methods and algorithms for dividing data into meaningful and useful clusters (groups) based on some intrinsic similarity measures [1]. It is a crucial intermediate step in exploring the potential structure of massive data for subsequent analysis. In the field of internet applications, clustering is used to identify users with different interest characteristics, thus realizing the personalized service recommendation process [2, 3]. In the field of biology, clustering can be applied to discover patterns in genes, helping to identify subtypes of certain diseases (e.g., cancers) [4–6].

The nonnegative matrix factorization (NMF) models (algorithms) have been widely studied as one

of the important tools for achieving clustering. The NMF model finds wide applications in various domains, owing to its part-based representation and physical interpretability. In [7, 8], NMF was used to decompose the original nonnegative high-dimensional matrix X into the product of two low-dimensional nonnegative matrices, WH, ensuring that X and WH are sufficiently close under some metric. Usually, W is referred to as the basis matrix and H is referred to as the coefficient matrix. Each element in the matrices W and H obtained above is nonnegative, and their rank is much lower than X itself, so this decomposition is called dimension reduction decomposition. This form of decomposition effectively saves storage and computational space, improving computational efficiency. In clustering analysis, the coefficient matrix H is used as a substitute matrix for the high-dimensional original matrix X for clustering tasks. The matrix H here is much lower in dimension than the original matrix X, and it will save a lot of computation and space.

Lee and Seung [9] extended NMF by introducing a simple algorithm based on multiplication update rules (MUR) and demonstrating its effectiveness in learning parts of objects and discovering semantic features in text. Additionally, NMF has been successfully extended and applied in various fields, including clustering and classification [10–12], image analysis [13–15], spectral analysis [16, 17], and blind source separation [18, 19], and so on. Moreover, NMF has brought a series of new theories and algorithms for clustering problems. Numerous studies have shown that by introducing different constraints, such as sparsity [12, 20, 21], smoothness [16, 22, 23], and orthogonality [24–26], the performance of standard NMF can usually be further improved. This is because the addition of these constraints introduces other useful features beyond non negativity to the standard NMF model. In fact, many datasets not only do not follow linear distributions, but also exist in certain nonlinear spaces, such as manifold spaces. This requires the introduction of nonlinear strategies for effective processing, such as constructing a Laplace matrix [27, 28] to maintain this nonlinear geometric structure between data points. This approach is known as manifold learning, which assumes that high-dimensional data has a nonlinear mapping in a low-dimensional space.

Cai et al. effectively combined manifold learning with NMF by introducing a graph regularization term, which is the graph regularized nonnegative matrix factorization (GNMF [29]) method. This algorithm aims to establish smooth connections between data representations, encode geometric information between data in linear and nonlinear spaces, and find the representation of data on manifolds. The advantage of this method is that it not only preserves the inherent geometric structure, but also obtains a compact representation of hidden semantic information. Based on this, Li et al. [30] presents a graph-based local concept coordinate factorization (GLCF) method, which respects the intrinsic structure of the data through manifold kernel learning in the warped eproducing kernal Hilbert space. Hu et al. [31] combined the GNMF algorithm with the Convex Non-Negative Matrix Factorization (CNMF) algorithm and proposed the graph convex nonnegative matrix factorization (GCNMF) algorithm with manifold regularization terms; Meng et al. [32] introduced graph regularization terms in the sample space and feature space, and proposed the dual graph sparse nonnegative matrix factorization (DSNMF) algorithm. In addition, algorithms such as graph regularized discriminative non-negative matrix factorization (GDNMF) [33], dual regularization nonnegative matrix factorization(DNMF) [34], Incomplete multi-view clustering (IMVC) [35], collaborative topological graph learning (CTGL) [36], multi-view Deep nonnegative matrix factorization with multi-kernel learning (MvMKDNMF) [37], nonnegtive mtrix factorization with graph regularization for multi-view data representation (GMultiNMF) [38], and Pairwise

constrained Graph Regularized Convex Nonnegative Matrix Factorization (PGCNMF) [39] applied constraints to the data through manifold learning, constructed different Laplacian regularization terms, and provided many basic methods for data analysis through the idea of manifold learning.

Recently, orthogonal nonnegative matrix factorization (ONMF [24]) has received widespread attention by imposing orthogonal constraints on NMF, making it more suitable for clustering tasks as the constraint matrix composed of factors can be viewed as an indicator matrix. This is because local representation requires the fundamental vectors of the factor matrices to be different from each other, which means that local representation is closely related to orthogonality.

The orthogonality of a given matrix A refers to $A^TA = I$ or $AA^T = I$ (I is an appropriate identity matrix), which means that any two different columns of A are orthogonal or rows of A are orthogonal.

In [20] Li et al. found that for some datasets, part-based representation involves clustering the rows of the input data matrix, and the clustering result is described by the basis matrix W. The orthogonal constraint on W is column orthogonality, $W^TW = I$. In document clustering [40, 41], clustering is performed on the columns of the input data matrix, represented by the coefficient matrix H, and the orthogonality on each row of H can improve clustering performance. The orthogonality constraint here is now on coefficient matrix H satisfying $HH^T = I$. The orthogonality models of the above two categories have greatly improved clustering performance [42–44]. Due to the excellent performance of ONMF in clustering, many update rules and variants of ONMF have been proposed [45–47].

In fact, ONMF was initially proposed by Ding [24], and empirical evidence demonstrated that introducing a certain degree of orthogonality constraint into the algorithm can enhance sparsity, reduce redundancy in feature vectors, and ensure the uniqueness of the solution and interpretability of clustering results. In their method, orthogonality is achieved by solving an optimization problem with orthogonality constraints.

There, ONMF was considered as a continuous relaxation (without discrete constraints) under the K-means method. Research on various data mining tasks has shown that compared to clustering methods based on K-means and NMF, the ONMF model performs better. However, in order to achieve the orthogonality of the basis vectors, this requires intensive computation, especially for word clustering in the document matrix, which results in high computational costs. Next, in 2010, Li et al. proposed nonnegative matrix factorization on orthogonal subspace (NMFOS) [48], introducing the concept of penalty functions to incorporate orthogonality constraints into the objective function. This orthogonality, which is achieved by adding it as a penalty term to the objective function, is established to reduce computational burden and to some extent improve the sparsity of the decomposition matrices (W matrix or H matrix). However, this model has two issues. First, when applied to certain document datasets, the algorithm may exhibit zero locking phenomenon; Second, the convergence proof proposed in it is not rigorous enough (due to insufficient correctness of matrix differentiation), that is, it fails to effectively prove that the iterative algorithm does indeed converge.

Overall, ONMF models have shown good performance, but still face the following two challenges: significant computational costs when used as constraints, and difficulty or absence in proving the convergence of algorithms when orthogonality $W^TW = I$ or $HH^T = I$ is introduced as a penalty term. Based on the above observations, this paper will study a new ONMF-based model graph-regularized nonnegative matrix factorization orthogonal subspace with auxiliary variable(GNMFOSV). By adding an orthogonal penalty term and a graph Laplacian term, we achieve the orthogonality of the factor matrix through factorization, enhancing the sparsity of the coefficient matrix to some extent.

Leveraging the geometric structure properties of the data, we enhance the algorithm's capability to recover the underlying nonlinear relationships in the data. More importantly, we will introduce an intermediate variable V as an alternating update in the auxiliary function, and prove the convergence of our algorithm using the idea of alternating updates of three variables. Meanwhile, experiments on a large number of datasets have also shown that this algorithm greatly improves clustering performance and has a certain degree of robustness to the dataset. In other words, experimental results on many datasets have shown that our algorithm maintains overall optimality.

The main contributions of this paper are as follows:

• The ONMF method obtains sparser results through orthogonal constraints. However, in some complex practical application scenarios, it is still difficult to interpret the physical meanings or practical implications of the basis vectors and coefficient matrices obtained after decomposition, which is not conducive to a deep understanding of the internal structure and features of the data. At the same time, it is also very difficult to prove the convergence of the algorithm. The GNMF method preserves the inherent geometric structure of the data and also acquires a compact representation of the hidden semantic information. Nevertheless, this method overemphasizes local information and relatively ignores the global structure and dependency relationships of the data. In this paper, in order to improve the interpretability of ONMF and the theoretical convergence of the algorithm, and also to retain the global structure and dependency relationships of the data in the GNMF model, the two methods are integrated, and an auxiliary function is ingeniously added, resulting in a rigorous convergence proof. The ablation experiments also fully demonstrate that integrating the graph regularization method, orthogonal constraints, and the auxiliary variable mechanism can significantly improve the clustering accuracy and stability, outperforming single methods on multiple datasets, which fully verifies the rationality and innovation of the model design.

• In our algorithm, different levels of orthogonality can be achieved by adjusting the parameter α_1 and α_2, and various regularization strengths can be obtained by tuning the parameter λ, making the algorithm more flexible and applicable to a broader range of scenarios.

• Extensive experiments were conducted on 13 datasets, demonstrating the algorithm's effectiveness in terms of clustering performance. The results not only highlight its efficiency in learning and enhanced discriminative power but also indicate its ability to acquire improved partial linear representations.

The remaining content of this paper is as follows: Section 2 reviews related work, including NMF, NMFOS, and GNMF. Section 3 provides a detailed introduction to our method, including the derivation of update rules and the corresponding algorithm, along with a proof of the algorithm's convergence. Section 4 presents experiments and analysis. Section 5 offers conclusions and future directions.

2. Symbols and classic models

In this section, some symbols and notations used throughout the paper will be provided in Tables 1 below, followed by classic models of NMF, NFMOS, GNMF and their multiplicative update algorithms.

Table 1. The symbols used throughout the paper.

Symbols	Annotations
$\mathbb{N}$	The set of all natural numbers
$R_+^{\upsilon\times\omega}$	The set of all nonnegative matrices with $\upsilon, \omega \in \mathbb{N}$
X	Input Nonnegative data matrix $X = [X_1, \cdots, X_N] \in R_+^{M\times N}$
W	The basis matrix $W = [W_{ik}] \in R_+^{M\times K}$
H	The coefficient matrix $H = [H_{kj}] \in R_+^{K\times N}$
V	The auxiliary matrix $V = [V_{kj}] \in R_+^{K\times N}$
M	The number of data features(or denoted as data dimension)
N	The number of data instances(or denoted as data points)
K	The number of data clusters
i	Elements in the sets $\{1, 2, \cdots, M\}$
j	Elements in the sets $\{1, 2, \cdots, N\}$
k	Elements in the sets $\{1, 2, \cdots, K\}$
$\|\cdot\|_F$	Frobenius norm
$Tr(\cdot)$	The trace operator
L	The graph Laplacian operators of the data manifold
S	The weighted matrix
D	The diagonal matrices with $D_{ii} = \sum_j S_{ij}$
p	The size of nearestneighbors
I	The identity matrix
α_1, α_2	The penalty parameter for controlling orthogonality $\alpha_1 \geq 0, \alpha_2 \geq 0$
λ	The regularization parameter $\lambda \geq 0$
ψ, ϕ	Lagrange multiplier

2.1. NMF

NMF [9] is a classic matrix factorization algorithm that requires all elements in the decomposed matrix to be nonnegative. Consider the objective matrix $X = [x_1, \cdots, x_N] \in R_+^{M\times N}$ and treat each x_j of X as a column vector. Nonnegative matrix factorization of X refers to finding two matrices W and H ($W = [w_{ik}] \in R_+^{M\times K}$, $H = [h_{kj}] \in R_+^{K\times N}$, $i = 1, 2, \cdots, M$, $j = 1, 2, \cdots, N$ and $k = 1, 2, \cdots, K$), such that the matrix-product WH is as equal as possible to X under a certain metric:

$$X \approx WH. \tag{2.1}$$

Generally speaking, W is called the basis matrix and H is called the coefficient matrix. This is because (2.1) tells us that each column vector x_j of X can be linearly represented using the column vectors of W in this factorization, where the coefficients come from matrix H. The specific representation is as follows:

$$x_j \approx \sum_{k=1}^{K} w_k h_{kj}, \tag{2.2}$$

where w_k represents the column vectors of W, and K in the (2.2) is much smaller than M, N, or satisfies $(M + N)K < MN$. It is precisely because of (2.2) that the decomposed matrix W is called the basis

matrix and matrix H is called the coefficient matrix. Afterward, matrix H will be used in clustering analysis applications, similar to the indicator set matrix in K-means. Meanwhile, the fact that K is much smaller than M and N makes this decomposition a dimensionality reduction decomposition, so clustering with H (not X) will greatly save time and space, and the mathematical model of NMF is

$$\begin{aligned} &\min_{W,H} \|X - WH\|_F^2, \\ &s.t. W \geq 0, H \geq 0, \end{aligned} \tag{2.3}$$

where $W = [w_1, \cdots, w_M] \in R_+^{M\times K}$, $H = [h_1, \cdots, h_N] \in R_+^{K\times N}$. In addition, we will import an iterative multiplicative update algorithm from Lee and Seung [50].

$$W_{ik} \leftarrow W_{ik}\frac{(XH^T)_{ik}}{(WHH^T)_{ik}},\ H_{kj} \leftarrow H_{kj}\frac{(W^TX)_{kj}}{(W^TWH)_{kj}}.$$

2.2. NMFOS

Chris Ding et al. [24] extended the traditional NMF by incorporating an orthogonal constraint on the basis or coefficient matrix into the objective function's constraints, and this model is the important ONMF model. More importantly, the equivalence between ONMF and kernel K-means methods was also proved. Li et al. [48] found that the Lagrangian multiplier used in the nonlinear constraint optimization in ONMF is a symmetric matrix that needs to be updated at each iteration, resulting in a large computational workload. So, to overcome this limitation, they proposed a novel approach called NMFOS (using penalty terms). Then, an improved optimization model was obtained in the following form:

$$\min_{W,H} \|X - WH\|_F^2 + \alpha\|HH^T - I\|_F^2,\ s.t.\ W \geq 0, H \geq 0, \tag{2.4}$$

where $X = [x_1, \cdots, x_N] \in R_+^{M\times N}$, $W = [w_1, \cdots, w_M] \in R_+^{M\times K}$, $H = [h_1, \cdots, h_N] \in R_+^{K\times N}$, I is the identity matrix with K rows and K columns.The nonnegative parameter $\alpha \geq 0$ controls the degree of orthogonality of the vectors in H. When $\alpha = 0$, NMFOS reduces to NMF. Through model (2.4), the following update rules were derived:

$$W_{ik} \leftarrow W_{ik}\frac{(XH^T)_{ik}}{(WHH^T)_{ik}},\ H_{kj} \leftarrow H_{kj}\frac{(W^TX + 2\alpha H)_{kj}}{(W^TWH + 2\alpha HH^TH)_{kj}}. \tag{2.5}$$

The advantage of this algorithm is that by adjusting the penalty parameter α, orthogonality can be achieved during the decomposition process without the need for additional constraints, while reducing computational complexity. By the way, it should be noted that Mirzal [51] has proved the convergence of (2.5) from [24].

2.3. GNMF

The extraction of the internal structure of data is an important application extension based on NMF models. Graph regularization is a manifold-based data processing method widely used to analyze the intrinsic geometric structure of manifold data [49, 52, 53]. The starting point of this idea is that manifolds can enhance the smoothness of data in both linear and nonlinear spaces by minimizing the formula later in (2.6), and by extracting the geometric structure information from data points and

features, manifolds can also enhance the clustering performance of NMF (see details in [34, 54, 55]). The study of NMF models based on such minimization forms is the classic GNMF model [29].

Next, let's briefly introduce the specific construction process of GNMF. To effectively capture the geometric structure of both data and feature manifolds, the p-nearest neighbor graph is constructed based on the input data matrix $X = [x_1, \cdots, x_N] \in R_+^{M\times N}$. Considering a graph with N vertices, which make up the current X, each vertex corresponds to a data point x_j ($j = 1, 2, \cdots, N$). For each data point x_j, we identify its p-nearest neighbors and connect them with edges (theoretically, p can take any value from $\{1, 2, \cdots, N\}$, but in practice, it is usually between 3 and 5). At the same time, a weight matrix S is defined, which is used to measure the degree to which adjacent points are not only similar in distance but also similar in geometric structure, thereby strengthening the influence of geometric structure on the model. There are many options for defining the weight matrix $S = \{S_{jl}\} \in R_+^{N\times N}$ on the graph, with the commonly used ones being the following three options:

• $0-1$ Weight Matrix: If $N(x_j)$ represents the set of p-nearest neighbors for data point x_j, then

$$S_{jl} = \begin{cases} 1, & x_j \in N(x_l) \text{ or } x_l \in N(x_j), \\ 0, & \text{otherwise}, \end{cases} \quad j, l \in \{1, \cdots, N\}.$$

• Heat Kernal Weight Matrix: If point x_j is connected to point x_l, then,

$$S_{jl} = e^{-\frac{\|x_j - x_l\|^2}{\sigma}}.$$

• Dot-Product Weight Matrix: If point x_j is connected to point x_l, then,

$$S_{jl} = x_j^T x_l.$$

It is worth noting that if the data point x is normalized to 1, the above dot-product of two vectors is equal to their cosine similarity.

With the help of the weight matrix S, the Laplacian matrix $L \in R_+^{N\times N}$ is defined to characterize the degree of similarity between points in geometric structure, and the Laplacian matrix L is defined as $L = D - S$, where $D \in R_+^{N\times N}$ is the degree matrix with $D_{ii} = \sum_j S_{ij}$. Note that, D is a diagonal matrix. Inspired by matrix factorization and manifold learning, Cai et al. [29] proposed the GNMF algorithm. The core graph regularization formula is as follows:

$$\frac{1}{2}\sum_{l=1}^{n}\sum_{j=1}^{n}\|h_l - h_j\|_2^2 S_{ij} = \sum_{j=1}^{n} D_{jj} h_j^T h_j - \sum_{l=1}^{n}\sum_{j=1}^{n} S_{lj} h_i^T h_j = Tr(HDH^T) - Tr(HSH^T) = Tr(HLH^T), \tag{2.6}$$

where $Tr(\cdot)$ denotes the trace operator and measures the pairwise similarity of the original data points. By minimizing (2.6), an intuitive result can be obtained: If two data points are close in terms of the intrinsic geometric structure of the original data distribution, they will also be close with respect to the new basis vectors h_l and h_j. Then, the optimization model of GNMF is given by:

$$\min \|X - WH\|_F^2 + \lambda \operatorname{Tr}\left(HLH^T\right), \ s.t.\ W \geq 0, H \geq 0, \tag{2.7}$$

where $W = [w_1, \cdots, w_M] \in R_+^{M\times K}$, $H = [h_1, \cdots, h_N] \in R_+^{K\times N}$. The regularization parameter $\lambda \geq 0$ controls the smoothness of the model. By utilizing the Lagrange method, the model (2.7) can be

solved and the iterative update rules are obtained below:

$$W_{ik} \leftarrow W_{ik}\frac{(XH^T)_{ik}}{(WHH^T)_{ik}},\ H_{kj} \leftarrow H_{kj}\frac{\left(W^TX + \lambda HS\right)_{kj}}{(W^TWH + \lambda HD)_{kj}}. \tag{2.8}$$

It is evident that when $\lambda = 0$, the update rules in (2.8) go back to the form of NMF update rules. In this paper, such manifold techniques are combined with NMFOS to enhance the capability of capturing underlying nonlinear relationships in data recovery.

3. A new class of NMF-type model and algorithm

In this section, we will introduce a new method called the GNMFOSV. More importantly, we also provide a convergence proof for this type of model to compensate for the lack or inadequacy of algorithm convergence proof. The specific operation is as follows: First, using the idea of NMFOS, the orthogonal constraint conditions on the coefficient matrix H are added to the objective function by using the penalty function method. Second, inspired by GNMF, our new model utilizes the geometric structures inherent in the data manifold and feature manifold to evolve the manifold term into additional regularization terms and add them to the objective function for further clustering research. From the experimental results, it can be seen that combining orthogonality and graph regularization (manifold) and adding them separately to the objective function not only solves the problem of considering both the nonnegativity and orthogonality of the coefficient matrix, but also improves the clustering performance via GNMFOSV.

3.1. Methodology

As mentioned above, we incorporate both the orthogonality constraint and the Laplacian constraint into the objective function to construct GNMFOSV:

$$\min J(W,H) = \min_{W,H}\frac{1}{2}\|X - WH\|_F^2 + \frac{\alpha}{4}\|HH^T - I\|_F^2 + \frac{\lambda}{2}Tr(HLH^T),\ s.t.\ W \geq 0, H \geq 0, \tag{3.1}$$

where $X = [x_1, \cdots, x_N] \in R_+^{M\times N}$, $W = [w_1, \cdots, w_M] \in R_+^{M\times K}$, $H = [h_1, \cdots, h_N] \in R_+^{K\times N}$, $L \in R_+^{N\times N}$, I is the identity matrix with K rows and K columns. Due to the property that a nonnegative matrix with pairwise orthogonal rows may have at most one nonzero entry in each column, the generation of strictly uncorrelated feature vectors is possible. Thus, we incorporate an orthogonal penalty term, $\alpha\|HH^T - I\|_F^2$, into the foundational model. However, directly imposing orthogonality constraints on the coefficient matrix poses challenges for convergence proof. Therefore, we introduce an auxiliary variable, denoted as $V \in R_+^{K\times N}$, with the matrix dimension equivalent to that of H. Subsequently, two new parameters, α_1 and α_2, are introduced to split the orthogonal constraint penalty term, $\alpha\|HH^T - I\|_F^2$, into two second-order terms. This is equivalently replaced by $\frac{\alpha_1}{2}\|I - HV^T\|_F^2 + \frac{\alpha_2}{2}\|V - H\|_F^2$, i.e.,

$$\alpha\|HH^T - I\|_F^2 = \frac{\alpha_1}{2}\|I - HV^T\|_F^2 + \frac{\alpha_2}{2}\|V - H\|_F^2. \tag{3.2}$$

Therefore, we equivalently replace model (3.1) with the following new model:

$$\begin{aligned}&\min J(W,V,H) = \min_{W,H,V}\frac{1}{2}\|X - WH\|_F^2 + \frac{\lambda}{2}Tr(HLH^T) + \frac{\alpha_1}{2}\|I - HV^T\|_F^2 + \frac{\alpha_2}{2}\|V - H\|_F^2,\\ &s.t. W \geq 0, H \geq 0, V \geq 0.\end{aligned} \tag{3.3}$$

This formulation incorporates the orthogonality constraint with an auxiliary variable as a penalty term in the objective function(see Figure 1). The advantage of introducing the auxiliary variable lies in its ability to deconstruct the complex orthogonality constraints found in NMFOS, thereby addressing the common challenges of convergence proof encountered in traditional orthogonal subspace NMF methods, and facilitating rapid convergence and favorable comparative performance in our experiments.

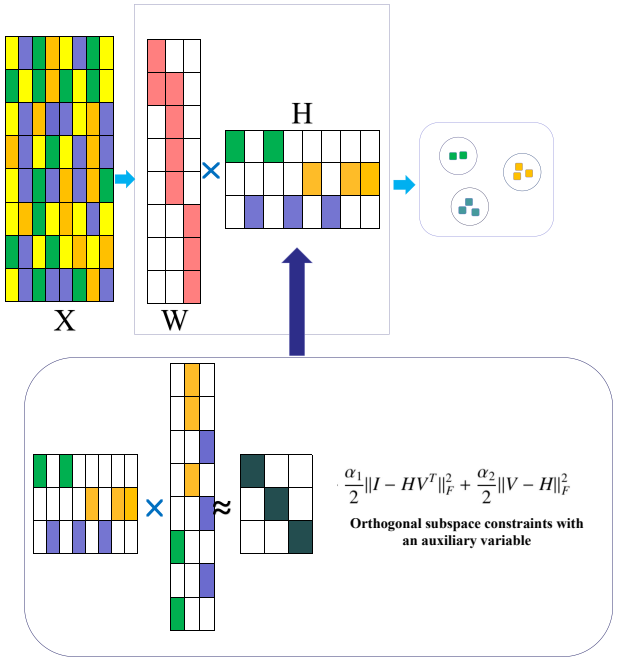


Figure 1. The framework of GNMFOSV.

Here, $\alpha_1 \geq 0$, $\alpha_2 \geq 0$ are the two orthogonal penalty parameters, and $\lambda \geq 0$ is the regularization parameter, respectively. The first two parameters adjust the level of orthogonality in H and ensure the validity of Eq (3.2), while the latter balances the reconstruction error in the first term and the graph regularization in the third term of the objective function. For convenience, we employ a $0-1$ weight scheme to construct the p-nearest neighbor graph. Clearly, when $\alpha_1 = \alpha_2 = \lambda = 0$, model (3.3) degenerates into model (2.3). Since the objective function in model (3.3) is non-convex with respect to both W and H, obtaining closed-form solutions is challenging. Now, we need to explore how to develop an effective algorithm to solve this model. We adopt an iterative optimization algorithm and utilize multiplicative update rules to obtain local optima (a similar idea can be found in [56–58], but not the same). We can rewrite the objective function of $J(W, V, H)$ in trace form as follows:

$$\begin{aligned} J(W,V,H) =& \frac{1}{2}Tr((X-WH)^T(X-WH)) + \frac{\lambda}{2}Tr(HLH^T) \\ &+ \frac{\alpha_1}{2}Tr((I-HV^T)^T(I-HV^T)) + \frac{\alpha_2}{2}Tr((V-H)^T(V-H)) \\ =& \frac{1}{2}Tr(X^TX - 2X^TWH + H^TW^TWH) + \frac{\lambda}{2}Tr(HLH^T) \\ &+ \frac{\alpha_1}{2}Tr(I - 2HV^T + VH^THV^T) + \frac{\alpha_2}{2}Tr(V^TV - 2V^TH + H^TH). \end{aligned} \tag{3.4}$$

Let ψ, ϕ, and θ be the Lagrange multipliers associated with the constraints $W \geq 0$, $H \geq 0$, and $V \geq 0$ in the constrained model (3.3). Define matrices $\Psi = [\psi_{ik}]$, $\Phi = [\phi_{kj}]$, and $\Theta = [\theta_{kj}]$. The Lagrangian function for the constrained optimization problem can then be expressed as:

$$\begin{aligned} L =& J(W, V, H) - Tr(\Psi^T H) - Tr(\Phi^T W) - Tr(\Theta^T V) \\ =& \frac{1}{2} Tr(X^T X - 2X^T WH + H^T W^T WH) + \frac{\lambda}{2} Tr(HLH^T) \\ &+ \frac{\alpha_1}{2} Tr(I - 2HV^T + VH^T HV^T) + \frac{\alpha_2}{2} Tr(V^T V - 2V^T H + H^T H) \\ &- Tr(\Psi^T H) - Tr(\Phi^T W) - Tr(\Theta^T V). \end{aligned}$$

Now, we will have the following equations by taking the derivatives of the Lagrangian function L with respect to W, V and H.

$$\frac{\partial L}{\partial W} = WHH^T - XH^T - \Psi, \tag{3.5}$$

$$\frac{\partial L}{\partial V} = \alpha_1(VH^T H - H) + \alpha_2(V - H) - \Theta, \tag{3.6}$$

and

$$\frac{\partial L}{\partial H} = W^T WH - W^T X + \lambda HD - \lambda HS + \alpha_1(HV^T V - V) + \alpha_2(H - V) - \Phi. \tag{3.7}$$

By setting Eqs (3.5)–(3.7) to zero, we can derive the expressions for the Lagrange multipliers:

$$\Psi = WHH^T - XH^T, \tag{3.8}$$

$$\Theta = \alpha_1(VH^T H - H) + \alpha_2(V - H), \tag{3.9}$$

and

$$\Phi = W^T WH - W^T X + \lambda HD - \lambda HS + \alpha_1(HV^T V - V) + \alpha_2(H - V). \tag{3.10}$$

Furthermore, at both ends of Eqs (3.8)–(3.10), we multiply them by factors W_{ik}, V_{kj}, and H_{kj}, respectively. With the help of Karush-Kuhn-Tucker (KKT) conditions $\psi_{ik}W_{ik} = 0$, $\theta_{kj}V_{kj} = 0$, and $\phi_{kj}H_{kj} = 0$, we will obtain the following three equations:

$$\psi_{ik}W_{ik} = \left[WHH^T - XH^T\right]_{ik} W_{ik} = 0, \tag{3.11}$$

$$\theta_{kj}V_{kj} = \left[\alpha_1(VH^T H - H) + \alpha_2(V - H)\right]_{kj} V_{kj} = 0, \tag{3.12}$$

$$\phi_{kj}H_{kj} = \left[W^T WH - W^T X + \lambda HD - \lambda HS + \alpha_1(HV^T V - V) + \alpha_2(H - V)\right]_{kj} H_{kj} = 0. \tag{3.13}$$

Therefore, based on Eqs (3.11)–(3.13), we derive the iterative formulas for updating W, V, and H in the optimization of the model (3.3):

$$W_{ik} \longleftarrow W_{ik} \frac{\left[XH^T\right]_{ik}}{[WHH^T]_{ik}}, \tag{3.14}$$

$$V_{kj} \longleftarrow V_{kj} \frac{[(\alpha_1 + \alpha_2)H]_{kj}}{[\alpha_1 VH^T H + \alpha_2 V]_{kj}}, \tag{3.15}$$

$$H_{kj} \longleftarrow H_{kj} \frac{\left[W^T X + \lambda HS + (\alpha_1 + \alpha_2)V\right]_{kj}}{\left[W^T WH + \lambda HD + \alpha_1 HV^T V + \alpha_2 H\right]_{kj}}. \tag{3.16}$$

We develop an efficient algorithm to solve model (3.3) in the Algorithm 1. The proof of the convergence of the Algorithm 1 can be found in the Supplementary.

Algorithm 1 GNMFOSV

Require: Original data matrix $X = [x_1, \cdots, x_N] \in R_+^{M\times N}$, the number of clusters K, orthogonality parameters $\alpha_1 \geq 0, \alpha_2 \geq 0$, regularization parameter $\lambda \geq 0$, the maximum number of iterations T;

Ensure: Locally optimal basis matrix W and solution coefficient matrix H

1: Initialization randomly choose initial matrices $W^0 \in R_+^{M\times K}$, $V^0 \in R_+^{K\times N}$ and $H^0 \in R_+^{K\times N}$;
2: Start updating $t = 1, \cdots, T$;
3: Updating W according to Eq (3.14);
4: Updating V according to Eq (3.15);
5: Updating H according to Eq (3.16);
6: $t = t + 1$;
7: If the termination conditions are not satisfied or $t \leq T$, return step 3; otherwise terminates.

3.2. Complexity analysis of algorithm

In this subsection, we analyze the computational complexity of Algorithm 1, which is the amount of space requirements needed to run the algorithm. The notation $O(n)$ is used to represent the computational complexity of the algorithm and is of the same order as n. Recall that M denotes the number of features, N denotes the number of samples, K denotes the number of clusters, and p denotes the number of nearest neighbors. The computational complexity of XH^T, WHH^T, $W^T X$, and $W^T WH$ are all MNK. Moreover, the computational complexity of HD and HS are both N^2K, while the computational complexity of $VH^T H$ and $HV^T V$ is NK^2. Besides, we need to provide an additional complexity of $O(N^2M)$ to construct the p-nearest neighbor graph. In conclusion, the computational complexity of each iteration in Algorithm 1 is $O(MNK + N^2M + N^2K + NK^2)$. During the experiment processing, we found that when conducting algorithm comparison experiments, the differences in time consumption among various algorithms on the author's computer were not significant. Therefore, data on time consumption was not included.

4. Experimental results and analysis

In this section, via lots of experiments, we will verify and study the performance of our algorithm in many aspects, including algorithm parameter selection analysis, clustering effect comparison experiment, and convergence experiment. In order to show these, we have compared the results based on the following datasets.

4.1. Datasets

To verify the performance superiority of the proposed model, we conduct comprehensive experiments on 13 real-world datasets involving face images, digit images, and documents.

•ORL[*]: Olivetti research laboratory dataset(ORL). This dataset contains 400 face images collected from 40 individuals and each subject has 10 grayscale images with size 32*32.

•AR [59]: Augmented reality dataset(AR). This dataset contains images of 65 men and 55 women, with each person contributing 14 images, resulting in a total of 1680 images. These images are resized to 32*32 pixels.

•Yale[†]: This dataset contains 165 grayscale images collected from 15 individuals. Each individual has 11 images with different facial expressions or configurations: with or without glasses, left-light, center-light, right-light, normal, happy, sad, sleepy and surprised. All image sizes are 32*32.

•Handwritten digit[‡]: The dataset consists of 2000 samples collected from 10 numeric characters (0–9).

•UCI digit[§]: This dataset consists of features of handwritten numerals (0–9) from University of California, Irvine(UCI) machine learning repository. They contain 2000 instances belonging to 10 classes.

•Diag-Bcw[¶]: Breast Cancer Wisconsin (denoted as Diag-Bcw) dataset, whereas the features are computed from a digitized image of a fine needle aspirate of a breast mass. It contains 569 documents in 2 categories.

•Caltech101 [60]: Caltech101 contains 9144 pictures of objects belonging to 101 categories, including motorbike, airplane, cannon, etc. There are about 40 to 800 images per category but most categories have about 50 images.

•Isolet[‖]: This dataset was generated as follows: 150 subjects spoke the name of each letter of the alphabet twice. The speakers are grouped into sets of 30 speakers each, and are referred to as isolet1, isolet2, isolet3, isolet4, and isolet5. We select the subset isolet3 in this experiment. It contains 1560 samples in 26 categories.

•Vote[**]: Congressional Voting Records (denoted as Vote) are derived from vote collections for each of the U.S. House of Representatives Congressmen on the 16 key votes identified by the congressional quarterly almanac(CQA). This dataset contains 435 documents in 2 categories.

•Abalone[††]: This dataset corresponds to the physical measurements of abalones and was collected from Marine Resources Division. It contains 2282 documents in 4 categories.

A summary of the above datasets is provided in Table 2 for their statistical characteristics.

[*] https://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
[†] http://cvc.cs.yale.edu/cvc/projects/yalefaces/yalefaces.html
[‡] https://lig-membres.imag.fr/grimal/data.html
[§] http://archive.ics.uci.edu/dataset/72/multiple+features
[¶] https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic
[‖] http://archive.ics.uci.edu/dataset/54/isolet
[**] https://archive.ics.uci.edu/dataset/105/congressional+voting+records
[††] https://archive.ics.uci.edu/dataset/1/abalone

Table 2. Statistics characteristics of the datasets.

Datasets	Samples (N)	Features (M)	Classes (C)	Data Types
ORL	400	1024	40	Face image
AR	1680	1024	120	Face image
Yale	165	1024	15	Face image
Handwritten digit	2000	649	10	Digit image
UCI-fac	2000	216	10	Digit image
UCI-fou	2000	76	10	Digit image
UCI-kar	2000	64	10	Digit image
Diag-Bcw	569	30	2	Digit image
Caltech7	1474	3766	7	Object image
Caltech20	2386	3766	20	Object image
isolet3	1560	617	26	Spoken document
Vote	435	16	2	Document
Abalone	2282	8	4	Document

4.2. Algorithms used for comparison

The proposed method is compared with the well-known K-means method and six other popular NMF based algorithms: NMF, ONMF, GNMF, sparse constrained nonnegative matrix factorization with matrix H (SNMF-H), sparse constrained nonnegative matrix factorization with matrix W (SNMF-W), NMFOS. As to the parameter setup of all compared methods, we adopt their default settings if feasible. For a fair comparison, we only consider algorithms based on the Frobenius norm. The description of each algorithm is as follows:

• K-means [61]: This algorithm operates on the original matrix without extracting or utilizing information contained in the original matrix.

• NMF [9]: The goal of this algorithm is to find two nonnegative matrices with significantly lower dimensions than the original matrix, while achieving strategies for features extraction and dimensionality reduction.

• ONMF [24]: Building upon NMF, an orthogonal constraint is introduced to the basis matrix (or coefficient matrix).

• GNMF [29]: By integrating graph theory and manifold assumptions with the original NMF, this algorithm preserves the intrinsic geometric structure of the data.

• SNMF-W and SNMF-H [62]: These two algorithms can generate sparse representations and represent the data as a linear combination of a small number of basis vectors.

•RMNMF [63]: The Robust manifold nonnegative matrix factorization (RMNMF) algorithm combines the $\ell_{2,1}$-norm with spectral clustering to enhance robustness to outliers and preserve data geometry, while an orthonormal constraint addresses the uniqueness of the solution.

•KOGNMF [64]: Kernel-based orthogonal graph regularized NMF (KOGNMF). By integrating graph constraints into the nonlinear NMF framework, this algorithm develops a kernel-based graph-regularized NMF.

•min-vol NMF [65]: The min-vol NMF algorithm incorporates volume regularization to address rank-deficient basis matrices, making it applicable to real-world problems such as multispectral image

unmixing. It utilizes an alternating fast projected gradient method for optimization.

• NMFOS [48]: Building upon ONMF, this algorithm incorporates an orthogonal constraint on the coefficient matrix as a penalty term in the objective function.

4.3. Settings

To assess clustering performance, we will adopt purity, normalized mutual information(NMI), accuracy(ACC), and rand index(RI) for comparing clustering results, which are widely used in nonnegative learning literature.

In this subsection, we will provide descriptions and definitions of these four indicators, as well as the selection of initial matrices and the stopping criteria for algorithms.

Purity is used for measuring the extent to which each cluster contained data instances from primary one class, and it is computed by

$$Purity = \sum_{i=1}^{K} \frac{\max_j \left(n_i^j\right)}{N}. \tag{4.1}$$

The notation n_i^j represents the count of occurrences where the j-th input class is assigned to the i-th cluster, with N denoting the total number of samples. The described process involves assigning a class to each cluster, and the samples for this class that appear most frequently within the cluster. Therefore, its range is [0, 1] and a higher value implies better clustering performance.

NMI is commonly employed in clustering to measure the similarity between two clustering results. Its range is [0, 1], with higher values indicating greater similarity between the two clustering results [63]. NMI is defined by the following formula:

$$NMI\left(C, C'\right) = \frac{MI\left(C, C'\right)}{\max\left(H(C), H\left(C'\right)\right)}. \tag{4.2}$$

Here, C and C' represent a set of true labels and a set of clustering labels, respectively, and $H(\cdot)$ denotes the entropy function.

ACC is used for discovering the one-to-one relationship between classes and clusters, along with measuring the extent to which each cluster's contained data instances from the corresponding class. The range of ACC values is from 0 to 1, where a higher value indicates better clustering results [66]. The expression for ACC is defined as follows:

$$ACC = \frac{\sum_{i=1}^{N} \delta\left(c_i, \operatorname{map}\left(c_j'\right)\right)}{N}, \tag{4.3}$$

where N represents the total number of samples, and map$(\cdot)$ denotes the mapping of the computed clustering label c_j' to the true clustering label c_i. The definition of δ is as follows:

$$\delta(i, j) = \begin{cases} 1, & i = j, \\ 0, & \text{others}. \end{cases} \tag{4.4}$$

RI is a clustering evaluation indicator that measures the similarity between two clusters. It determines the agreement between the pairwise relationships of data points in the true cluster and the obtained cluster. RI is calculated by using the following formula:

$$RI = \frac{a+b}{\binom{n}{2}}, \tag{4.5}$$

where a represents the number of pairwise agreements between data points assigned to the same cluster in both the true and obtained clusterings, b represents the number of pairwise agreements between data points assigned to different clusters in both the true and obtained clusterings, and n denotes the total number of data points, with $\binom{n}{2}$ being the combinatorial number.

By unifying (4.1)–(4.3) and (4.5) together, the definitions of all considered clustering indicators are given in Table 3.

Table 3. Definitions of all considered clustering metrics.

Metric	Definition	Range
Purity	$\sum_{i=1}^{K} \frac{\max_j\left(n_i^j\right)}{N}$	[0, 1]
Normalized Mutual Information (NMI)	$\frac{MI(C,C')}{\max(H(C),H(C'))}$	[0, 1]
Accuracy (ACC)	$\frac{\sum_{i=1}^{N}\delta\left(c_i,\mathrm{map}\left(c_j'\right)\right)}{N}$	[0, 1]
Rand Index(RI)	$\frac{a+b}{\binom{n}{2}}$	[0, 1]

In the clustering experiment, we use the number of clusters K in the dataset to be equal to the maximum number of categories C in the target dataset. Due to the sensitivity of most methods to initial values, for experimental accuracy, we conducted 10 separate runs for all algorithms and calculate their mean and standard deviation. As for the stopping criterion, we only needed to set a maximum number $T = 100$ of iterations.

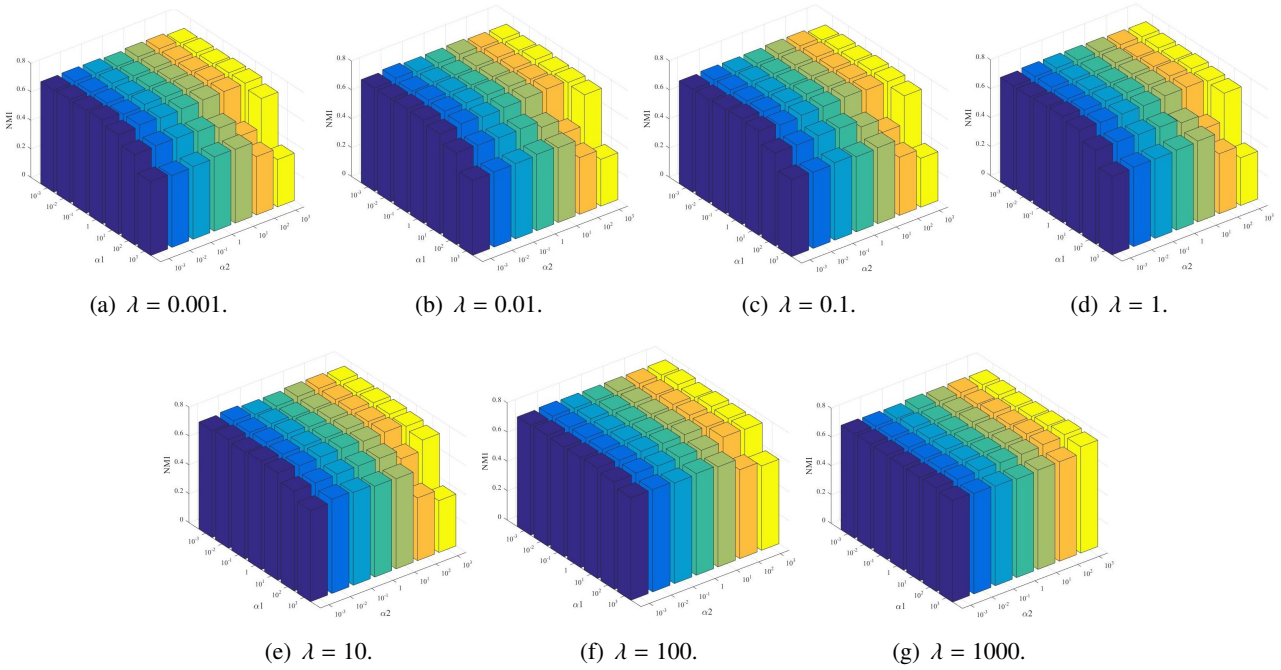
The experiments were conducted on a computing system operating under Windows 10 Home Edition, 64-bit (version 10.0, Build 19045). The hardware configuration comprises an Intel(R) Core(TM) i5-7300HQ CPU, functioning at a clock speed of 2.50 GHz with four cores, and is supplemented by 8 GB of RAM. The programming environment employed for the implementation of the algorithms is MATLAB R2015b.

4.4. Parameter sensitivity

In this subsection, we will discuss the sensitivity of the parameters.

There are four parameters in our model: the regularization parameter λ, the orthogonal penalty parameters α_1, the orthogonal penalty parameters α_2, and the nearest neighbor point p. To reveal the impact of these terms on the performance of the model, we select a batch of parameters such that λ, α_1, and α_2 are all in the range of $\left\{10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}\right\}$. For the parameters λ, α_1 and α_2, we employed NMI as the evaluation metric to conduct a parameter sensitivity experiment on the ORL dataset. The number of iterations is empirically fixed at 100 times. As stated above, the 0–1 weighting is used for computing the weighting graph in all experiments if not stated otherwise. Figure 2 plots the NMI of the ORL dataset with different parameters. The $x-axis$ represents α_2, the $y-axis$ represents α_1, and the $z-axis$ represents NMI. As shown in Figure 2, algorithm performance changes with variations in three parameters. For example, we have the following observations that the proposed GNMFOSV algorithm obtains superior performance when $\lambda = 100$, α_1 varies in $\left\{10^{-2}, 10^{-1}, 10^{0}, 10^{1}\right\}$,

and achieves consistent good performance regardless of the range of α_2. The specific analysis of Figure 2 is presented in the following paragraphs.



(a) $\lambda = 0.001$. (b) $\lambda = 0.01$. (c) $\lambda = 0.1$. (d) $\lambda = 1$.

(e) $\lambda = 10$. (f) $\lambda = 100$. (g) $\lambda = 1000$.

Figure 2. Parameter sensitivity on the ORL dataset, where NMI of the proposed method are reported while fixing the nearest neighbor point p and clustering number K.

First, a small λ and a large α_1 may deteriorate the clustering performance. For example, from Figure 2(c), GNMFOSV performs poorly when α_1 is greater than 100. However, when the value of α_1 is in a smaller range, the performance of GNMFOSV is very stable and has significant advantages. So, the value of α_1 will be set to $\left\{10^{-2}, 10^{-1}, 10^{0}, 10^{1}\right\}$ in the following experiments.

Moreover, when $\lambda = 100$, the best results appear for $\alpha_1 = 0.01$, and $\alpha_2 = 1000$ and this reflects that a small λ and a large α_1 are needed. Then, by experiments, we will use $\alpha_1 = 0.01$, $\alpha_2 = 1000$, and $\lambda = 100$.

Second, to choose nearest neighbor point p when $\alpha_1 = 0.01$, $\alpha_2 = 1000$, and $\lambda = 100$. The clustering results of NMI with respect to the neighborhood size p varies from 1 to 10 are shown in Figure 3. Based on the ORL dataset, GNMFOSV algorithm arrives at the good NMI for the ORL dataset when p varies from 2 to 5. Other indicators and datasets are similar.

In summary, our experiments (all datasets are analyzed) informed us that the parameter selection results are $p = 3$, $\alpha_1 = 0.01$, $\alpha_2 = 1000$, and $\lambda = 100$. These selected parameter values will be used in numerical experiments for clustering throughout the entire paper. This means that the robustness of the proposed method can be guaranteed when the parameters are selected in an appropriate range. These observations also indicate that the GNMFOSV algorithm is sensitive to the values of parameters, which provides corresponding moderation for practical applications.

At the same time, we also conducted parameter sensitivity tests on 12 other datasets. We found that the model proposed in this paper demonstrates excellent robustness. Regrettably, constrained by the page limit, only the experimental results of the ORL dataset are presented.

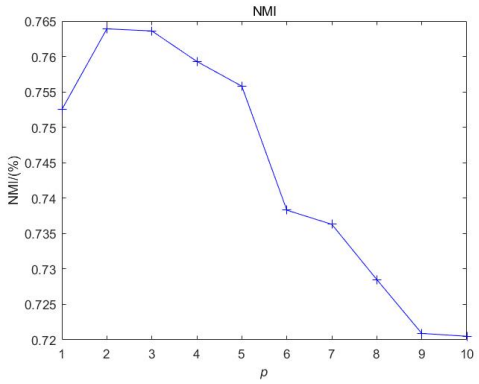


Figure 3. The clustering results of NMI on ORL dataset with respect to the neighborhood size p.

4.5. Ablation study

In the ablation study of this subsection, we choose $p = 3$, $\alpha_1 = 0.01$, $\alpha_2 = 1000$, and $\lambda = 100$ as parameter selection of GNMFOSV. The ablation study means that the lost functions of the proposed GNMFOSV are studied by experiments of removing different parts. The combinations of each part in (3.3) are denoted as follows.

•Baseline(NMF): The loss function composed of only the NMF initial item(the first item) in (3.3) is set as the baseline, which is

$$\min J(W, H) = \min_{W,H} \frac{1}{2}\|X - WH\|_F^2 \ s.t.\ W \geq 0, H \geq 0.$$

•GNMF: The loss function composed of the NMF term and the second term in Eq (3.3), which is

$$\min J(W, H) = \min_{W,H} \frac{1}{2}\|X - WH\|_F^2 + \frac{\lambda}{2} Tr(HLH^T) \ s.t.\ W \geq 0, H \geq 0.$$

•NMFOSV: The loss function, consisting of the NMF term along with the third and fourth terms in Eq (3.3),

$$\min J(W, H, V) = \min_{W,H,V} \frac{1}{2}\|X - WH\|_F^2 + \frac{\alpha_1}{2}\|I - HV^T\|_F^2 + \frac{\alpha_2}{2}\|V - H\|_F^2 \ s.t.\ W \geq 0, H \geq 0, V \geq 0.$$

•GNMFOSV: The loss function of Eq (3.3), which is indeed the proposed algorithm in this paper.

Here $X = [x_1, \cdots, x_N] \in R_+^{M\times N}$, $W = [w_1, \cdots, w_M] \in R_+^{M\times K}$, $H = [h_1, \cdots, h_N] \in R_+^{K\times N}$, $L \in R_+^{N\times N}$, I is the identity matrix with K rows and K columns. The clustering performance values presented in Tables 4–6 demonstrate that, all core modules of the proposed image clustering model are indispensable. The experimental results indicate that the integration of graph regularization methods, orthogonal constraints, and the auxiliary variable mechanism can significantly enhance clustering accuracy and stability. Outperforming comparative methods on multiple datasets, this validates the rationality and innovativeness of the proposed model.

Table 4. Results (mean ± standard deviation) on AR.

Metric	NMF	GNMF	NMFOSV	GNMFOSV
Purity	0.5262±0.0166	0.4824±0.0161	0.5103±0.0131	**0.5718±0.0147**
NMI	0.7636±0.0090	0.7268±0.0098	0.7563±0.0101	**0.8011±0.0085**
ACC	0.4904±0.0154	0.4464±0.0187	0.4562±0.0092	**0.5438±0.0191**

Table 5. Results (mean ± standard deviation) on Yale.

Metric	NMF	GNMF	NMFOSV	GNMFOSV
Purity	0.4024±0.0294	0.3909±0.0103	0.3972±0.0083	**0.4661±0.0153**
NMI	0.4381±0.0161	0.4329±0.0146	0.4412±0.0261	**0.5024±0.0175**
ACC	0.3872±0.0355	0.3757±0.0151	0.3810±0.0113	**0.6459±0.0163**

Table 6. Results (mean ± standard deviation) on Caltech20.

Metric	NMF	GNMF	NMFOSV	GNMFOSV
Purity	0.5547±0.0133	0.5428±0.0028	0.5582±0.0121	**0.6114±0.0185**
NMI	0.2910±0.0110	0.2731±0.0044	0.2841±0.0076	**0.3816±0.0130**
ACC	0.2767±0.0191	0.2403±0.0095	0.2811±0.0126	**0.3465±0.0170**

4.6. Clustering results and analysis

In this subsection, extensive experimental results are displayed to validate the effectiveness of the proposed model. The experimental results of GNMFOSV with other comparison algorithms are given in Tables 7–19, and each result is the average of 10 runs after turning the parameters. The clustering results of all comparison algorithms are evaluated according to the former 4 indicators. The best result for each indicator is marked in bold. "-" indicates that clustering results could not be obtained or there was insufficient memory.

Table 7. Results (mean ± standard deviation) on ORL.

Metric	K-means	NMF	ONMF	GNMF	SNMF-H	SNMF-W	RNMF	KOGNMF	min-vol NMF	NMFOS	GNMFOSV
Purity	0.6140¡Å0.0350	0.6350±0.0302	0.6180±0.0399	0.6540±0.0022	0.6431±0.0292	0.6020±0.0205	0.5432±0.0211	0.1407±0.0039	0.5199±0.0325	0.5730±0.0309	**0.7185±0.0328**
NMI	0.6082±0.0214	0.5967±0.0146	0.5927±0.0233	0.6171±0.0015	0.5937±0.0213	0.5930±0.0139	0.4951±0.0207	0.0130±0.0017	0.5417±0.0179	0.5597±0.0179	**0.6572±0.0152**
ACC	0.5875±0.0518	0.6080±0.0338	0.5785±0.0444	0.6065±0.0022	0.6415±0.0368	0.5805±0.0270	0.5119±0.0146	0.1357±0.0027	0.5058±0.0406	0.5375±0.0359	**0.7185±0.0411**
RI	0.9008±0.0081	0.9020±0.0041	0.8919±0.0068	0.9097±0.0005	0.9002±0.0050	0.8969±0.0039	0.4001±0.0110	0.1014±0.0008	0.8801±0.0065	0.8827±0.0055	**0.9165±0.0058**

Table 8. Results (mean ± standard deviation) on AR.

Metric	K-means	NMF	ONMF	GNMF	SNMF-H	SNMF-W	RNMF	KOGNMF	min-vol NMF	NMFOS	GNMFOSV
Purity	0.4222±0.0135	0.5262±0.0166	0.4876±0.0171	0.4824±0.0161	0.5258±0.0199	0.5387±0.0130	0.5593±0.0191	0.4190±0.0153	0.2874±0.0092	0.5095±0.0100	**0.5718±0.0147**
NMI	0.6992±0.0059	0.7636±0.0090	0.7289±0.0131	0.7268±0.0098	0.7536±0.0125	0.7707±0.0077	0.7557±0.0127	0.6722±0.0078	0.6084±0.0081	0.7633±0.0074	**0.8011±0.0085**
ACC	0.3883±0.0154	0.4904±0.0154	0.4511±0.0193	0.4464±0.0187	0.4873±0.0201	0.5011±0.0136	0.4973±0.0206	0.3860±0.0156	0.2629±0.0118	0.4666±0.0138	**0.5438±0.0191**
RI	0.9867±0.0003	0.9885±0.0004	0.9874±0.0005	0.9865±0.0005	0.9879±0.0005	0.9885±0.0003	0.9865±0.0009	0.9836±0.0013	0.9843±0.0005	0.9878±0.0004	**0.9896±0.0006**

Table 9. Results (mean ± standard deviation) on Yale.

Metric	K-means	NMF	ONMF	GNMF	SNMF-H	SNMF-W	RNMF	KOGNMF	min-vol NMF	NMFOS	GNMFOSV
Purity	0.4079±0.0346	0.4024±0.0294	0.4055±0.0181	0.3909±0.0103	0.4018±0.0277	0.4042±0.0228	0.3285±0.0251	0.4067±0.0194	0.4430±0.0283	0.3964±0.0333	**0.4661±0.0153**
NMI	0.4408±0.0392	0.4381±0.0161	0.4502±0.0206	0.4329±0.0146	0.4361±0.0188	0.4393±0.0227	0.3477±0.0259	0.4566±0.0231	0.4573±0.0213	0.4353±0.0306	**0.5024±0.0175**
ACC	0.3836±0.0333	0.3872±0.0355	0.3860±0.0217	0.3757±0.0151	0.3793±0.0325	0.3800±0.0336	0.3072±0.0221	0.3903±0.0162	0.4200±0.0281	0.3793±0.0377	**0.4593±0.0163**
RI	0.8785±0.0164	0.8975±0.0062	0.8980±0.0043	0.8974±0.0037	0.8956±0.0045	0.8958±0.0051	0.8751±0.0061	0.8810±0.0083	0.8963±0.0050	0.8973±0.0045	**0.9014±0.0025**

Table 10. Results (mean ± standard deviation) on Handwritten digit.

Metric	K-means	NMF	ONMF	GNMF	SNMF-H	SNMF-W	RNMF	KOGNMF	min-vol NMF	NMFOS	GNMFOSV
Purity	0.5461±0.0065	0.5382±0.0299	0.5358±0.0160	0.5521±0.0060	0.5030±0.0271	0.5537±0.0156	0.4924±0.0249	0.3105±0.0102	0.4701±0.0048	0.5583±0.0445	**0.6010±0.0216**
NMI	0.5276±0.0098	0.5162±0.0260	0.5063±0.0118	0.5563±0.0052	0.5061±0.0298	0.5070±0.0133	0.4629±0.0158	0.2847±0.0106	0.4663±0.0041	0.5210±0.0256	**0.6128±0.0205**
ACC	0.4865±0.0650	0.5450±0.0449	0.5220±0.0520	0.5900±0.0476	0.4510±0.0340	0.5125±0.0527	0.4531±0.0258	0.2970±0.0165	0.4189±0.0162	0.5015±0.0549	**0.6075±0.0748**
RI	0.8679±0.0086	0.8688±0.0068	0.8786±0.0017	0.8737±0.0005	0.8578±0.0115	0.8665±0.0053	-	0.8432±0.0043	-	0.8529±0.0110	**0.8913±0.0030**

Table 11. Results (mean ± standard deviation) on UCI-fac.

Metric	K-means	NMF	ONMF	GNMF	SNMF-H	SNMF-W	RNMF	KOGNMF	min-vol NMF	NMFOS	GNMFOSV
Purity	0.6095±0.0233	0.5316±0.0365	0.5812±0.0235	0.5934±0.0057	0.5284±0.0204	0.5320±0.0185	0.5342±0.0256	0.1362±0.0038	0.3986±0.0692	0.5427±0.0247	**0.6355±0.0232**
NMI	0.5809±0.0165	0.4674±0.0282	0.5283±0.0185	0.5349±0.0043	0.4658±0.0136	0.4682±0.0155	0.4491±0.0186	0.0097±0.0016	0.3376±0.0605	0.4695±0.0186	**0.6028±0.0136**
ACC	0.5855±0.0462	0.5119±0.0432	0.5478±0.0285	0.5597±0.0057	0.4965±0.0340	0.5107±0.0196	0.5040±0.0345	0.1324±0.0045	0.3730±0.0656	0.5232±0.0267	**0.6118±0.0276**
RI	0.8940±0.0057	0.8760±0.0074	0.8856±0.0059	0.8869±0.0008	0.8741±0.0036	0.8754±0.0032	0.8682±0.0076	0.8202±0.0003	0.0484±0.8464	0.8768±0.0045	**0.8966±0.0052**

Table 12. Results (mean ± standard deviation) on UCI-fou.

Metric	K-means	NMF	ONMF	GNMF	SNMF-H	SNMF-W	RNMF	KOGNMF	min-vol NMF	NMFOS	GNMFOSV
Purity	0.6140±0.0350	0.6350±0.0302	0.6180±0.0399	0.6540±0.0022	0.6431±0.0292	0.6020±0.0205	0.5432±0.0211	0.1407±0.0039	0.5199±0.0325	0.5730±0.0309	**0.7185±0.0328**
NMI	0.6082±0.0214	0.5967±0.0146	0.5927±0.0233	0.6171±0.0015	0.5937±0.0213	0.5930±0.0139	0.4951±0.0207	0.0130±0.0017	0.5417±0.0179	0.5597±0.0179	**0.6572±0.0152**
ACC	0.5875±0.0518	0.6080±0.0338	0.5785±0.0444	0.6065±0.0022	0.6415±0.0368	0.5805±0.0270	0.5119±0.0146	0.1357±0.0027	0.5058±0.0406	0.5375±0.0359	**0.7185±0.0411**
RI	0.9008±0.0081	0.9020±0.0041	0.8919±0.0068	0.9097±0.0005	0.9002±0.0050	0.8969±0.0039	0.4001±0.0110	0.1014±0.0008	0.8801±0.0065	0.8827±0.0055	**0.9165±0.0058**

Table 13. Results (mean ± standard deviation) on UCI-kar.

Metric	K-means	NMF	ONMF	GNMF	SNMF-H	SNMF-W	RNMF	KOGNMF	min-vol NMF	NMFOS	GNMFOSV
Purity	0.6170±0.0347	0.6990±0.0240	0.6085±0.0268	0.6085±0.0172	0.6650±0.0352	0.6910±0.0474	0.5809±0.0384	0.2394±0.0123	0.6721±0.0408	0.6145±0.0507	**0.7759±0.0247**
NMI	0.5258±0.0267	0.6286±0.0208	0.5186±0.0215	0.5553±0.0104	0.6438±0.0311	0.6590±0.0286	0.5110±0.0238	0.1495±0.0128	0.6244±0.0328	0.6339±0.0392	**0.7212±0.0245**
ACC	0.5890±0.0469	0.6675±0.0310	0.5760±0.0318	0.5350±0.0237	0.6330±0.0436	0.6910±0.0564	0.5567±0.0443	0.2243±0.0098	0.6402±0.0609	0.5410±0.0591	**0.7480±0.0317**
RI	0.8877±0.0088	0.9125±0.0054	0.8962±0.0056	0.8994±0.0029	0.8597±0.0088	0.9084±0.0087	0.8814±0.0090	0.8288±0.0023	0.9028±0.0131	0.9142±0.0098	**0.9266±0.0070**

Table 14. Results (mean ± standard deviation) on Diag-Bcw.

Metric	K-means	NMF	ONMF	GNMF	SNMF-H	SNMF-W	RNMF	KOGNMF	min-vol NMF	NMFOS	GNMFOSV
Purity	0.8799±0.0008	0.7374±0.0993	0.8660±0.0059	0.8365±0.0370	0.7179±0.1090	0.6769±0.1045	0.7840±0.0195	0.6274±0.0000	0.8746±0.0008	0.7079±0.1047	**0.8922±0.0125**
NMI	0.4652±0.0029	0.1862±0.1602	0.4076±0.0173	0.3666±0.0233	0.1684±0.1837	0.1010±0.1776	0.2319±0.0245	0.0007±0.0007	0.4320±0.0020	0.1434±0.1796	**0.4761±0.0383**
ACC	0.8822±0.0008	0.7165±0.1272	0.8660±0.0059	0.8365±0.0392	0.6869±0.1389	0.6050±0.1476	0.7840±0.0195	0.5158±0.0114	0.8746±0.0008	0.6720±0.1379	**0.8957±0.0125**
RI	0.7918±0.0013	0.6222±0.1087	0.7676±0.0086	0.7260±0.0090	0.6039±0.1200	0.5605±0.1168	0.6614±0.0215	0.4998±0.0009	0.7804±0.0012	0.5927±0.1166	**0.8080±0.0195**

Table 15. Results (mean ± standard deviation) on Caltech7.

Metric	K-means	NMF	ONMF	GNMF	SNMF-H	SNMF-W	RNMF	KOGNMF	min-vol NMF	NMFOS	GNMFOSV
Purity	0.7818±0.0021	0.7898±0.0102	0.7794±0.0046	0.7824±0.0010	0.7875±0.0109	0.7822±0.0178	0.7787±0.0077	0.5467±0.0086	0.5559±0.0025	0.7818±0.0093	**0.8110±0.0066**
NMI	0.2426±0.0062	0.2688±0.0168	0.2415±0.0075	0.2423±0.0007	0.2727±0.0181	0.2685±0.0153	0.2467±0.0146	0.0193±0.0075	0.0563±0.0013	0.2631±0.0177	**0.3482±0.0094**
ACC	0.3390±0.0162	0.3968±0.0441	0.3352±0.0121	0.3215±0.0019	0.4233±0.0490	0.4119±0.0384	0.3616±0.0214	0.1978±0.0107	0.2421±0.0087	0.4216±0.0499	**0.4548±0.0290**
RI	0.6620±0.0034	0.6790±0.0131	0.6618±0.0021	0.6606±0.0004	0.6834±0.0138	0.6803±0.0096	0.6544±0.0052	0.5856±0.0021	0.5890±0.0004	0.6695±0.0142	**0.6812±0.0072**

Table 16. Results (mean ± standard deviation) on Caltech20.

Metric	K-means	NMF	ONMF	GNMF	SNMF-H	SNMF-W	RNMF	KOGNMF	min-vol NMF	NMFOS	GNMFOSV
Purity	0.5539±0.0083	0.5547±0.0133	0.5369±0.0091	0.5428±0.0028	0.5496±0.0123	0.5636±0.0116	0.5439±0.0123	0.3388±0.0037	0.3497±0.0016	0.5579±0.0143	**0.6114±0.0185**
NMI	0.2859±0.0105	0.2910±0.0110	0.2650±0.0051	0.2731±0.0044	0.2837±0.0087	0.2934±0.0107	0.2746±0.0114	0.0467±0.0041	0.0788±0.0012	0.2922±0.0157	**0.3816±0.0130**
ACC	0.2485±0.0140	0.2767±0.0191	0.2481±0.0113	0.2403±0.0095	0.2772±0.0197	0.2739±0.0127	0.2623±0.0189	0.0949±0.0023	0.1099±0.0020	0.2753±0.0131	**0.3465±0.0170**
RI	0.8280±0.0019	0.8300±0.0028	0.8272±0.0012	0.8272±0.0015	0.8299±0.0018	0.8306±0.0022	-	0.8042±0.0002	0.8023±0.0014	0.8302±0.0024	**0.8352±0.0025**

Table 17. Results (mean ± standard deviation) on isolet3.

Metric	K-means	NMF	ONMF	GNMF	SNMF-H	SNMF-W	RNMF	KOGNMF	min-vol NMF	NMFOS	GNMFOSV
Purity	0.5759±0.0210	0.5800±0.0275	0.5807±0.0188	0.5391±0.0219	0.5908±0.0216	0.5665±0.0209	0.5064±0.0260	0.4964±0.0266	0.5435±0.0338	0.5839±0.0240	**0.6227±0.0239**
NMI	0.7161±0.0105	0.6989±0.0156	0.6947±0.0130	0.7129±0.0119	0.7068±0.0131	0.6879±0.0146	0.6209±0.0200	0.6304±0.0142	0.6766±0.0189	0.7008±0.0162	**0.7392±0.0129**
ACC	0.5312±0.0205	0.5505±0.0301	0.5438±0.0212	0.4820±0.0284	0.5562±0.0281	0.5300±0.0236	0.4787±0.0289	0.4631±0.0213	0.5016±0.0317	0.5507±0.0285	**0.5866±0.0257**
RI	0.9575±0.0013	0.9589±0.0020	0.9583±0.0017	0.9448±0.0017	0.9594±0.0028	0.9563±0.0026	0.9520±0.0020	0.9497±0.0020	0.9534±0.0035	0.9589±0.0025	**0.9612±0.0020**

Table 18. Results (mean ± standard deviation) on Vote.

Metric	K-means	NMF	ONMF	GNMF	SNMF-H	SNMF-W	RNMF	KOGNMF	min-vol NMF	NMFOS	GNMFOSV
Purity	0.8712±0.0022	0.8664±0.0020	0.8667±0.0000	0.8667±0.0000	0.8667±0.0000	0.6161±0.0000	0.8791±0.0022	0.7094±0.0540	0.8667±0.0000	0.6161±0.0000	**0.8824±0.0016**
NMI	0.4592±0.0037	0.4393±0.0075	0.4375±0.0000	0.4375±0.0000	0.4440±0.0000	0.0032±0.0000	0.4420±0.0045	0.1426±0.0726	0.4440±0.0000	0.0032±0.0000	**0.4680±0.0045**
ACC	0.8712±0.0022	0.8664±0.0020	0.8666±0.0000	0.8666±0.0000	0.8666±0.0000	0.6160±0.0000	0.8790±0.0022	0.7000±0.0729	0.8666±0.0000	0.6160±0.0000	**0.8824±0.0016**
RI	0.7751±0.0033	0.7680±0.0029	0.7683±0.0000	0.7683±0.0000	0.7683±0.0000	0.5258±0.0000	0.7869±0.0033	0.5886±0.0464	0.7683±0.0000	0.5258±0.0000	**0.7768±0.0024**

Table 19. Results (mean ± standard deviation) on Abalone.

Metric	K-means	NMF	ONMF	GNMF	SNMF-H	SNMF-W	RNMF	KOGNMF	min-vol NMF	NMFOS	GNMFOSV
Purity	0.3624±0.0000	0.3537±0.0132	0.3582±0.0119	0.3501±0.0035	0.3575±0.0099	0.3511±0.0141	0.3849±0.0154	0.3053±0.0030	0.3553±0.0014	0.3588±0.0044	**0.3834±0.0010**
NMI	0.0703±0.0002	0.0628±0.0140	0.0671±0.0144	0.0595±0.0010	0.0664±0.0111	0.0605±0.0141	0.0834±0.0091	0.0017±0.0006	0.0639±0.0015	0.0687±0.0069	**0.0810±0.0009**
ACC	0.3312±0.0011	0.3373±0.0118	0.3413±0.0131	0.3326±0.0066	0.3405±0.0146	0.3385±0.0122	0.3526±0.0118	0.2686±0.0049	0.3242±0.0020	0.3414±0.0062	**0.3745±0.0026**
RI	0.5972±0.0004	0.6099±0.0094	0.6175±0.0088	0.5629±0.0085	0.6055±0.0222	0.6081±0.0160	0.6308±0.0054	0.6198±0.0005	0.5755±0.0068	0.6078±0.0095	**0.6332±0.0011**

From the above tables, we will have the following observations.

First, the proposed method GNMFOSV achieves the best Purity, ACC, NMI, and RI values among all compared algorithms from a total of 13 datasets. It should be pointed that GNMFOSV outperforms other compared methods, with considerable advantages on ORL, AR, UCI-fou, UCI-kar datasets. For example, it makes about 7.7% increases against the second best method SNMF-H on the dataset UCI-fou in terms of ACC. Second, from the data in the above tables, the clustering algorithms of GNMF and GNMFOSV show certain competitive power at Purity, NMI, ACC, and RI comparing to the other methods. It indicates that the graph-based clustering algorithms are good at modeling the intrinsic structure of data. Furthermore, the orthogonal-based learning is promising over original NMF learning, which can be observed from the fact that the orthogonal-based algorithms, i.e., ONMF, NMFOS, and GNMFOSV, substantially boost the clustering performance on document datasets.

Third, compared with GNMF and NMFOS algorithms, neither the graph regularization algorithm nor the orthogonal subspace algorithm can achieve excellent clustering performance. However the GNMFOSV algorithm combines the advantages of graph regularity and orthogonal subspace, and it can significantly exceed for the original single-constrained NMF algorithm. In general, our proposed GNMFOSV algorithm consistently outperforms the other compared algorithms in terms of Purity, NMI, ACC, and RI metrics, which demonstrates the clear advance of our method.

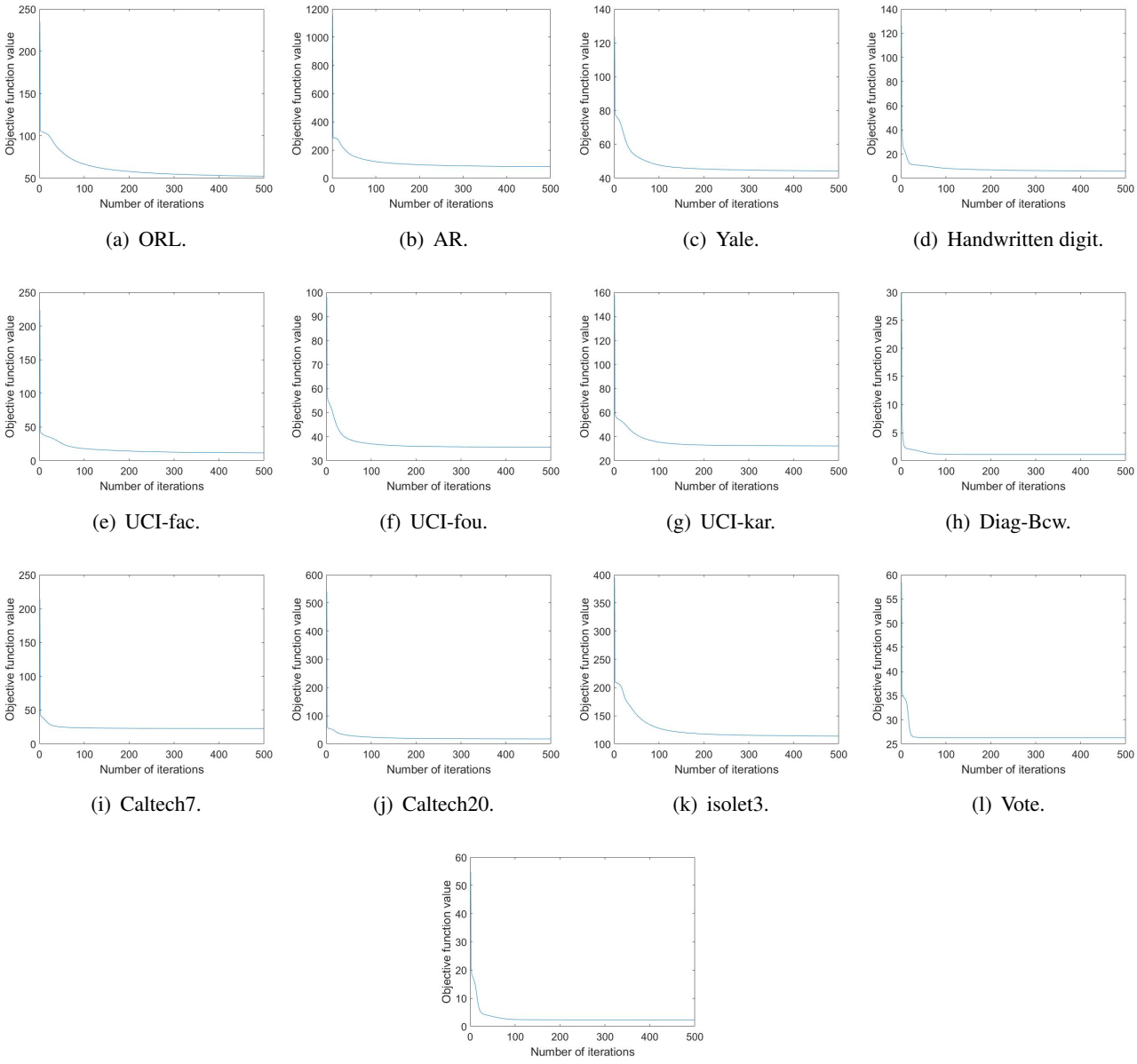
In conclusion, the GNMFOSV method performs acceptably on all datasets, i.e., face image, digit image, object image, and document, which further validates its robustness in terms of steady clustering performance. The quantitative results fully demonstrate the effectiveness of our proposed GNMFOSV approach.

4.7. Convergence analysis

In this subsection, we investigate how fast the convergence speed is. The convergence curves of GNMFOSV on all datasets are demonstrated in Figure 4 based on all 13 datasets. The $x-axis$ denotes the number of iterations, and the $y-axis$ denotes the objective function value.

It can be seen from the convergence curve (tested on each dataset depicted in the figure) that our algorithm can achieve rapid convergence on each data set. In the datasets of AR, Handwritten digit, Diag-Bcw, Caltech7, Caltech20, and Vote, when the number of iterations is less than 20, the GNMFOSV algorithm can achieve stable convergence. On the ORL, Yale, UCI-fac, UCI-fou, UCI-kar and Abalone datasets, the speed of GNMFOSV drops rapidly, and then it becomes a gentle decline, and finally achieves stable convergence at about 100 iterations. On the isolet3 dataset, the convergence curve of GNMFOSV is very smooth, and convergence is achieved around 100 iterations.

So, Figure 4 indicates that the GNMFOSV algorithm has done a very good job in actual convergence experiments, and the choosen of 100 (for iteration) is good enough in actual experiments.



(a) ORL. (b) AR. (c) Yale. (d) Handwritten digit.

(e) UCI-fac. (f) UCI-fou. (g) UCI-kar. (h) Diag-Bcw.

(i) Caltech7. (j) Caltech20. (k) isolet3. (l) Vote.

(m) Abalone.

Figure 4. The convergence curves of the objective function value in GNMFOSV on (a) ORL, (b) AR, (c) Yale, (d) Handwritten digit, (e) UCI-fac, (f) UCI-fou, (g) [UCI-kar, (h) Diag-Bcw, (i) Caltech7, (j) Caltech20, (k) isolet3, (l) Vote, (m) Abalone datasets.

5. Conclusions

In this paper, we have proposed a new NMF cost function for clustering problems. The derived optimization model incorporates the orthogonal constraints, a graph Laplacian term, and auxiliary variable mechanism. The experiments show the superiority of the proposed algorithm (at least on the datasets used in this work). The pro-posed algorithm has the following advantages: (i) It leverages the geometric structure information hidden in both data and feature spaces effectively. Specifically,

this model effectively exploits the geometric structure information hidden in the data and feature spaces. It models the data space as a sub-manifold embedded in a high-dimensional space. After conducting matrix factorization on this sub-manifold, it obtains a low-dimensional representation of data points, which is beneficial for data point clustering. Therefore, this model is more discriminative than algorithms based on orthogonal subspaces that only take into account the Euclidean structure of the data. Also, it can achieve better classification performance than graph-based algorithms because the latter do not consider the orthogonalization of the vectors in the factor matrices. (ii) By introducing an auxiliary variable, it provides a complete proof of convergence, filling the gap in demonstrating convergence under orthogonality conditions for such algorithms. (iii) After conducting 13 groups experiments with data of different sizes and types, the results show that,it is less sensitive to changing the size and type of the data. The performance of the proposed algorithm remains more or less stable even when the size and type of the data changes.

Author contributions

Caiping Wang: Conceptualization; Wen Li: Formal analysis; Caiping Wang, Wen Li, Junjian Zhao and Yasong Chen wrote the main manuscript text. All authors reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The first author is partially supported by the China Tobacco Hebei industrial co., Ltd Technology Project, China (Grant No. HBZY2023A034). The third and fourth authors are partially supported by the Natural Science Foundation of Tianjin City, China (Grant Nos. 24JCYBJC00430 and 18JCYBJC16300), the Natural Science Foundation of China (Grant No. 11401433), and the Ordinary Universities Undergraduate Education Quality Reform Research Project of Tianjin, China (Grant No. A231005802).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. A. Jain, Data clustering: 50 years beyond k-means, *Pattern Recogn. Lett.*, **31** (2010), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011
2. S. Renaud-Deputter, T. Xiong, S. Wang, Combining collaborative filtering and clustering for implicit recommender system, *Proceedings of IEEE 27th International Conference on Advanced Information Networking and Applications (AINA)*, 2013, 748–755. https://doi.org/10.1109/AINA.2013.65

3. J. Das, P. Mukherjee, S. Majumder, P. Gupta, Clustering-based recommender system using principles of voting theory, *Proceedings of International Conference on Contemporary Computing and Informatics (IC3I)*, 2014, 230–235. https://doi.org/10.1109/IC3I.2014.7019655
4. C. Zheng, D. Huang, L. Zhang, X. Kong, Tumor clustering using nonnegative matrix factorization with gene selection, *IEEE Trans. Inf. Technol. Bio.*, **13** (2009), 599–607. https://doi.org/10.1109/TITB.2009.2018115
5. J. Brunet, P. Tamayo, T. Golub, J. Mesirov, Metagenes and molecular pattern discovery using matrix factorization, *Proc. Natl. Acad. Sci. USA*, **101** (2004), 4164–4169. https://doi.org/10.1073/pnas.0308531101
6. S. Wang, P. Wu, M. Zhou, T. Chang, S. Wu, Cell subclass identification in single-cell rna-sequencing data using orthogonal nonnegative matrix factorization, *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, 876–880. https://doi.org/10.1109/ICASSP.2018.8462055
7. P. Paatero, U. Tapper, Positive matrix factorization: a nonnegative factor model with optimal utilization of error estimates of data values, *Environmetrics*, **5** (1994), 111–126. https://doi.org/10.1002/env.3170050203
8. P. Anttila, P. Paatero, U. Tapper, O. Järvinen, Source identification of bulk wet deposition in finland by positive matrix factorization, *Atmos. Environ.*, **29** (1995), 1705–1718. https://doi.org/10.1016/1352-2310(94)00367-T
9. D. D. Lee, H. S. Seung, Learning the parts of objects by nonnegative matrix factorization, *Nature*, **401** (1999), 788–791. https://doi.org/10.1038/44565
10. F. Shahnaz, M. W. Berry, V. P. Pauca, R. J. Plemmons, Document clustering using nonnegative matrix factorization, *Inform. Process. Manag.*, **42** (2006), 373–386. https://doi.org/10.1016/j.ipm.2004.11.005
11. W. Xu, X. Liu, Y. Gong, Document clustering based on nonnegative matrix factorization, *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, 2003, 267–273. https://doi.org/10.1145/860435.860485
12. H. Kim, H. Park, Sparse nonnegative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis, *Bioinformatics*, **23** (2007), 1495–1502. https://doi.org/10.1093/bioinformatics/btm134
13. P. O. Hoyer, Non-negative matrix factorization with sparseness constraints, *J. Mach. Learn. Res.*, **5** (2004), 1457–1469.
14. D. Wang, H. Lu, On-line learning parts-based representation via incremental orthogonal projective nonnegative matrix factorization, *Signal Process.*, **93** (2013), 1608–1623. https://doi.org/10.1016/j.sigpro.2012.07.015
15. N. Gillis, F. Glineur, A multilevel approach for nonnegative matrix factorization, *J. Comput. Appl. Math.*, **236** (2012), 1708–1723. https://doi.org/10.1016/j.cam.2011.10.002
16. V. P. Pauca, J. Piper, R. J. Plemmons, Nonnegative matrix factorization for spectral data analysis, *Linear Algebra Appl.*, **416** (2006), 29–47. https://doi.org/10.1016/j.laa.2005.06.025

17. M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, R. J. Plemmons, Algorithms and applications for approximate nonnegative matrix factorization, *Comput. Stat. Data Anal.*, **52** (2007), 155–173. https://doi.org/10.1016/j.csda.2006.11.006

18. N. Bertin, R. Badeau, E. Vincent, Enforcing harmonicity and smoothness in bayesian nonnegative matrix factorization applied to polyphonic music transcription, *IEEE Trans. Audio Speech*, **18** (2010), 538–549. https://doi.org/10.1109/TASL.2010.2041381

19. G. Zhou, Z. Yang, S. Xie, J. Yang, Online blind source separation using incremental nonnegative matrix factorization with volume constraint, *IEEE Trans. Neural Networ.*, **22** (2011), 550–560. https://doi.org/10.1109/TNN.2011.2109396

20. S. Z. Li, X. W. Hou, H. J. Zhang, Q. S. Cheng, Learning spatially localized, parts-based representation, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001, I–I. https://doi.org/10.1109/CVPR.2001.990477

21. Y. Gao, G. Church, Improving molecular cancer class discovery through sparse nonnegative matrix factorization, *Bioinformatics*, **21** (2005), 3970–3975. https://doi.org/10.1093/bioinformatics/bti653

22. S. Jia, Y. Qian, Constrained nonnegative matrix factorization for hyperspectral unmixing, *IEEE Trans. Geosci. Remote*, **47** (2009), 161–173. https://doi.org/10.1109/TGRS.2008.2002882

23. J. Tang, Z. Wan, Orthogonal dual graph-regularized nonnegative matrix factorization for co-clustering, *J. Sci. Comput.*, **87** (2021), 66. https://doi.org/10.1007/s10915-021-01489-w

24. C. Ding, T. Li, W. Peng, H. Park, Orthogonal nonnegative matrix t-factorizations for clustering, *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2006, 126–135. https://doi.org/10.1145/1150402.1150420

25. J. Yoo, S. Choi, Orthogonal nonnegative matrix tri-factorization for co-clustering: multiplicative updates on stiefel manifolds, *Inform. Process. Manag.*, **46** (2010), 559–570. https://doi.org/10.1016/j.ipm.2009.12.007

26. B. Li, G. Zhou, A. Cichocki, Two efficient algorithms for approximately orthogonal nonnegative matrix factorization, *IEEE Signal Process. Lett.*, **22** (2014), 843–846. https://doi.org/10.1109/LSP.2014.2371895

27. Y. Liu, X. Li, C. Liu, H. Liu, Structure-constrained low-rank and partial sparse representation with sample selection for image classification, *Pattern Recogn.*, **59** (2016), 5–13. https://doi.org/10.1016/j.patcog.2016.01.026

28. F. R. Bach, M. I. Jordan, Spectral clustering for speech separation, In: *Automatic speech and speaker recognition: large margin and kernel methods*, Hoboken: John Wiley & Sons, Ltd., 2009, 221–250. https://doi.org/10.1002/9780470742044.ch13

29. D. Cai, X. He, J. Han, T. Huang, Graph regularized nonnegative matrix factorization for data representation, *IEEE Trans. Pattern Anal.*, **33** (2011), 1548–1560. https://doi.org/10.1109/TPAMI.2010.231

30. P. Li, J. Bu, L. Zhang, C. Chen, Graph-based local concept coordinate factorization, *Knowl. Inf. Syst.*, **43** (2015), 103–126. https://doi.org/10.1007/s10115-013-0715-x

31. W. Hu, K. Choi, P. Wang, Y. Jiang, S. Wang, Convex nonnegative matrix factorization with manifold regularization, *Neural Networks*, **63** (2015), 94–103. https://doi.org/10.1016/j.neunet.2014.11.007

32. Y. Meng, R. Shang, L. Jiao, W. Zhang, Y. Yuan, S. Yang, Feature selection based dual-graph sparse nonnegative matrix factorization for local discriminative clustering, *Neurocomputing*, **290** (2018), 87–99. https://doi.org/10.1016/j.neucom.2018.02.044

33. X. Long, H. Lu, Y. Peng, W. Li, Graph regularized discriminative nonnegative matrix factorization for face recognition, *Multimed. Tools Appl.*, **72** (2014), 2679–2699. https://doi.org/10.1007/s11042-013-1572-z

34. F. Shang, L. Jiao, F. Wang, Graph dual regularization nonnegative matrix factorization for co-clustering, *Pattern Recogn.*, **45** (2012), 2237–2250. https://doi.org/10.1016/j.patcog.2011.12.015

35. C. Liu, S. Wu, R. Li, D. Jiang, H. Wong, Self-supervised graph completion for incomplete multi-view clustering, *IEEE Trans. Knowl. Data Eng.*, **35** (2023), 9394–9406. https://doi.org/10.1109/TKDE.2023.3238416

36. C. Liu, R. Li, H. Che, M. Leung, S. Wu, Z. Yu, et al., Beyond euclidean structures: collaborative topological graph learning for multiview clustering, *IEEE Trans. Neur. Net. Lear.*, in press. https://doi.org/10.1109/TNNLS.2024.3489585

37. X. Yang, H. Che, M. Leung, C. Liu, S. Wen, Auto-weighted multi-view deep nonnegative matrix factorization with multi-kernel learning, *IEEE Trans. Signal Inf. Pr.*, **11** (2024), 23–34. https://doi.org/10.1109/TSIPN.2024.3511262

38. H. Che, C. Li, M. Leung, D. Ouyang, X. Dai, S. Wen, Robust hypergraph regularized deep nonnegative matrix factorization for multi-view clustering, *IEEE Transactions on Emerging Topics in Computational Intelligence*, **9** (2025), 1817–1829. https://doi.org/10.1109/TETCI.2024.3451352

39. G. Li, X. Zhang, S. Zheng, D. Li, Semi-supervised convex nonnegative matrix factorizations with graph regularized for image representation, *Neurocomputing*, **237** (2017), 1–11. https://doi.org/10.1016/j.neucom.2016.04.028

40. W. Zhang, M. Tan, Q. Sheng, L. Yao, Q. Shi, Efficient orthogonal nonnegative matrix factorization over stiefel manifold, *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, 2016, 1743–1752. https://doi.org/10.1145/2983323.2983761

41. S. Wang, T. Chang, Y. Cui, J. Pang, Clustering by orthogonal nonnegative matrix factorization: a sequential non-convex penalty approach, *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, 5576–5580. https://doi.org/10.1109/ICASSP.2019.8683466

42. Y. Qin, C. Jia, Y. Li, Community detection using nonnegative matrix factorization with orthogonal constraint, *Proceedings of Eighth International Conference on Advanced Computational Intelligence (ICACI)*, 2016, 49–54. https://doi.org/10.1109/ICACI.2016.7449802

43. P. He, X. Xu, J. Ding, B. Fan, Low-rank nonnegative matrix factorization on stiefel manifold, *Inform. Sciences*, **514** (2020), 131–148. https://doi.org/10.1016/j.ins.2019.12.004

44. H. Abe, H. Yadohisa, Orthogonal nonnegative matrix tri-factorization based on Tweedie distributions, *Adv. Data Anal. Classif.*, **13** (2019), 825–853. https://doi.org/10.1007/s11634-018-0348-8

45. P. Li, J. Bu, C. Chen, C. Wang, D. Cai, Subspace learning via locally constrained A-optimal nonnegative projection, *Neurocomputing*, **115** (2013), 49–62. https://doi.org/10.1016/j.neucom.2012.12.029

46. P. Li, J. Bu, B. Xu, B. Wang, C. Chen, Locally discriminative spectral clustering with composite manifold, *Neurocomputing*, **119** (2013), 243–252. https://doi.org/10.1016/j.neucom.2013.03.034

47. P. Li, J. Bu, Y. Yang, R. Ji, C. Chen, D. Cai, Discriminative orthogonal nonnegative matrix factorization with flexibility for data representation, *Expert Syst. Appl.*, **41** (2014), 1283–1293. https://doi.org/10.1016/j.eswa.2013.08.026

48. Z. Li, X. Wu, H. Peng, Nonnegative matrix factorization on orthogonal subspace, *Pattern Recogn. Lett.*, **31** (2010), 905–911. https://doi.org/10.1016/j.patrec.2009.12.023

49. A. Tosyali, J. Kim, J. Choi, M. Jeong, Regularized asymmetric nonnegative matrix factorization for clustering in directed networks, *Pattern Recogn. Lett.*, **125** (2019), 750–757. https://doi.org/10.1016/j.patrec.2019.07.005

50. D. Lee, H. S. Seung, Algorithms for nonnegative matrix factorization, *Proceedings of the 14th International Conference on Neural Information Processing Systems*, 2000, 535–541.

51. A. Mirzal, A convergent algorithm for orthogonal nonnegative matrix factorization, *J. Comput. Appl. Math.*, **260** (2014), 149–166. https://doi.org/10.1016/j.cam.2013.09.022

52. S. Peng, W. Ser, B. Chen, L. Sun, Z. Lin, Robust nonnegative matrix factorization with local coordinate constraint for image clustering, *Eng. Appl. Artif. Intel.*, **88** (2020), 103354. https://doi.org/10.1016/j.engappai.2019.103354

53. G. Chen, C. Xu, J. Wang, J. W. Feng, J. Q. Feng, Graph regularization weighted nonnegative matrix factorization for link prediction in weighted complex network, *Neurocomputing*, **369** (2019), 50–60. https://doi.org/10.1016/j.neucom.2019.08.068

54. Q. Gu, J. Zhou, Co-clustering on manifolds, *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2009, 359–368. https://doi.org/10.1145/1557019.1557063

55. Q. Wang, M. Chen, F. Nie, X. Li, Detecting coherent groups in crowd scenes by multiview clustering, *IEEE Trans. Pattern Anal.*, **42** (2020), 46–58. https://doi.org/10.1109/TPAMI.2018.2875002

56. J. Guo, Z. Wan, A modified spectral prp conjugate gradient projection method for solving large-scale monotone equations and its application in compressed sensing, *Math. Probl. Eng.*, **2019** (2019), 5261830. https://doi.org/10.1155/2019/5261830

57. T. Li, Z. Wan, New adaptive Barzilai-Borwein step size and its application in solving large-scale optimization problems, *ANZIAM J.*, **61** (2019), 76–98. https://doi.org/10.1017/S1446181118000263

58. J. Lv, S. Deng, Z. Wan, An efficient single-parameter scaling memoryless broyden-fletcher-goldfarb-shanno algorithm for solving large scale unconstrained optimization problems, *IEEE Access*, **8** (2020), 85664–85674. https://doi.org/10.1109/ACCESS.2020.2992340

59. A. Martinez, R. Benavente, The AR face database: Cvc technical report, 24, Commissioned report, 1998.

60. F. Li, R. Fergus, P. Perona, Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories, *Proceedings of Conference on Computer Vision and Pattern Recognition Workshop*, 2004, 178–178. https://doi.org/10.1109/CVPR.2004.383

61. J. D. Wang, J. Wang, Q. Ke, G. Zeng, S. Li, Fast approximate k-means via cluster closures, In: *Multimedia data mining and analytics: disruptive innovation*, Cham: Springer, 2015, 373–395. https://doi.org/10.1007/978-3-319-14998-1_17

62. P. O. Hoyer, Non-negative sparse coding, *Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing*, 2002, 557–565. https://doi.org/10.1109/NNSP.2002.1030067

63. J. Huang, F. Nie, H. Huang, C. Ding, Robust manifold nonnegative matrix factorization, *ACM Trans. Knowl. Discov. D.*, **8** (2014), 11. https://doi.org/10.1145/2601434

64. D. Tolić, N. Antulov-Fantulin, I. Kopriva, A nonlinear orthogonal nonnegative matrix factorization approach to subspace clustering, *Pattern Recogn.*, **82** (2018), 40–55. https://doi.org/10.1016/j.patcog.2018.04.029

65. V. Leplat, A. Ang, N. Gillis, Minimum-volume rank-deficient nonnegative matrix factorizations, *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, 3402–3406. https://doi.org/10.1109/ICASSP.2019.8682280

66. C. Peng, Z. Kang, Y. Hu, J. Cheng, Q. Cheng, Nonnegative matrix factorization with integrated graph and feature learning, *ACM Trans. Intel. Syst. Tec.*, **8** (2017), 42. https://doi.org/10.1145/2987378

67. S. Boyd, L. Vandenberghe, *Convex optimization*, Cambridge: Cambridge University Press, 2004. https://doi.org/10.1017/CBO9780511804441

Appendix

Convergence analysis of Algorithm 1

In this section, we will investigate the convergence of Algorithm 1. Since the objective function of the GNMFOSV model (3.3) has a lower bound, we only need to prove (3.3) is nonincreasing with respect to the update rules (3.14)–(3.16). Before that, we first introduce a definition given by Lee and Seung [50], as well as a lemma with a concise proof.

Definition 1. *[50] Let $F : R_+^{\upsilon\times\omega} \to R$ be continuously differentiable with $\upsilon, \omega \in \mathbb{N}$, and $G : R_+^{\upsilon\times\omega} \times R_+^{\upsilon\times\omega} \to R$ also be continuously differentiable. Then G is called an auxiliary function of F if it satisfies the following conditions:*

$$G(U, U) = F(U),\ G(U, U^t) \geq F(U),$$

for any $U, U^t \in R_+^{\upsilon\times\omega}$.

Lemma 1. *Let G be an auxiliary function of F. Then F is nonincreasing under the sequence of matrices $\{U^t\}_{t=0}^{+\infty}$ generated by the following update rule:*

$$U^{t+1} = \arg\min_{U \in R^{v \times \omega}} G(U, U^t). \tag{5.1}$$

Proof. By Definition 1 and (5.1), the following formula

$$G(U^{t+1}, U^{t+1}) = F(U^{t+1}) \leq G(U^{t+1}, U^t) \leq G(U^t, U^t) = F(U^t) \tag{5.2}$$

holds naturally. □

Next, we will provide the convergence proofs for the updated rules (3.14)–(3.16), which is also an important innovation of this paper. To this end, we first present the objective function F, and it is introduced to replace the symbols in the objective function $J(W, H)$ in (3.4), i.e., $F = J(W, V, H)$. Moreover, let $F_{W_{ik}}$, $F_{V_{kj}}$, and $F_{H_{kj}}$ represent the parts of the objective function (3.4) associated with the elements W_{ik}, V_{kj}, and H_{kj}, respectively. In addition, due to the complexity of representation of $F(H)$ based only on H, the function $F(H)$ will be rewritten as $F(H) = F_1(H) + F_2(H)$ with

$$F_1(H) = \frac{1}{2}\|X - WH\|_F^2 + \frac{\lambda}{2}Tr(HLH^T),\ F_2(H) = \frac{\alpha_1}{2}\|I - HV^T\|_F^2 + \frac{\alpha_2}{2}\|V - H\|_F^2. \tag{5.3}$$

The first and second-order derivatives of F with respect to W_{ik} are listed below:

$$F'_{W_{ik}} = \left(\frac{\partial J(W, V, H)}{\partial W}\right)_{ik} = (WHH^T - XH^T)_{ik},\ F''_{W_{ik}} = (HH^T)_{kk}.$$

Meanwhile, the first and second-order derivatives of $F_1(H)$ with respect to H_{kj} will be shown below:

$$F'_{1_{H_{kj}}} = \left(\frac{\partial J(W, V, H)}{\partial H}\right)_{kj} = (W^T WH - W^T X + \lambda HL)_{kj},\ F''_{1_{H_{kj}}} = (W^T W)_{kk} + (\lambda L)_{jj}.$$

Note that we have omitted the partial derivatives of F_2 and $F_{V_{kj}}$ because they are not required when constructing the auxiliary functions.

We can now define auxiliary functions. Since the update rules (3.14)–(3.16) are all updated in the form of each element of the matrix, according to Lemma 1, it is sufficient to prove that $F_{W_{ik}}$, $F_{V_{kj}}$ and $F_{H_{kj}}$ do not increase under the sequences $\{W_{ik}^t\}_{t=0}^{+\infty}$, $\{V_{kj}^t\}_{t=0}^{+\infty}$, and $\{H_{kj}^t\}_{t=0}^{+\infty}$ generated by (3.14)–(3.16), respectively. Then, the specific expressions of auxiliary functions are

$$G_W(W_{ik}, W_{ik}^t) = F_{W_{ik}}(W_{ik}^t) + F'_{W_{ik}}(W_{ik}^t)(W_{ik} - W_{ik}^t) + \frac{1}{2}\frac{(W^t HH^T)_{ik}}{W_{ik}^t}(W_{ik} - W_{ik}^t)^2, \tag{5.4}$$

$$G_V(V_{kj}, V_{kj}^t) = \frac{\alpha_1}{2}\sum_{a=1}^{k}\sum_{b=1}^{k}\frac{1}{(HV^{tT})_{ab}}\sum_{r=1}^{n}H_{ar}V^{tT}_{rb}\left[\delta_{ab} - \frac{V_{br}}{V_{br}^t}(HV^{tT})_{ab}\right]^2 + \frac{\alpha_2}{2}\|V - H\|_F^2, \tag{5.5}$$

$$G_{H_1}(H_{kj}, H_{kj}^t) = F_{1_{H_{kj}}}(H_{kj}^t) + F'_{1_{H_{kj}}}(H_{kj}^t)(H_{kj} - H_{kj}^t) + \frac{1}{2}\frac{(W^T WH^t + \lambda H^t D)_{kj}}{H_{kj}^t}(H_{kj} - H_{kj}^t)^2, \tag{5.6}$$

$$G_{H_2}(H_{kj}, H_{kj}^t) = \frac{\alpha_1}{2}\sum_{a=1}^{k}\sum_{b=1}^{k}\frac{1}{(H^tV^T)_{ab}}\sum_{r=1}^{n}H_{ar}^tV_{rb}^T\left[\delta_{ab} - \frac{H_{ar}}{H_{ar}^t}(H^tV^T)_{ab}\right]^2 + \frac{\alpha_2}{2}\|V - H\|_F^2. \tag{5.7}$$

We now have the following auxiliary function for H via (5.6) and (5.7):

$$G_H(H_{kj}, H_{kj}^t) = G_{H_1}(H_{kj}, H_{kj}^t) + G_{H_2}(H_{kj}, H_{kj}^t). \tag{5.8}$$

Next, we will provide one by one in the form of lemmas that (5.4), (5.5), and (5.8) are indeed auxiliary functions. The first is to show that $G_W(W_{ik}, W_{ik}^t)$ is an auxiliary function of $F_{W_{ik}}$.

Lemma 2. *Let $G_W(W_{ik}, W_{ik}^t)$ be defined by the formula (5.4). Then, $G_W(W_{ik}, W_{ik}^t)$ is an auxiliary function of $F_{W_{ik}}$.*

Proof. By expanding $F_{W_{ik}}(W_{ik})$ based on the Taylor series, we have the following expansion:

$$F_{W_{ik}}(W_{ik}) = F_{W_{ik}}(W_{ik}^t) + F_{W_{ik}}^{'}(W_{ik}^t)(W_{ik} - W_{ik}^t) + \frac{1}{2}(HH^T)_{kk}(W_{ik} - W_{ik}^t)^2.$$

To prove that $G_W(W_{ik}, W_{ik}^t) \geq F_{W_{ik}}(W_{ik})$, it is sufficient to prove the following inequality:

$$\frac{(W^tHH^T)_{ik}}{W_{ik}^t} \geq (HH^T)_{kk}. \tag{5.9}$$

In fact, (5.9) is true via

$$(W^tHH^T)_{ik} = \sum_{r=1}^{K}W_{ir}^t(HH^T)_{rk} \geq W_{ik}^t(HH^T)_{kk}. \tag{5.10}$$

In addition, (5.10) can also lead to

$$\frac{1}{2}\frac{(W^tHH^T)_{ik}}{W_{ik}^t}(W_{ik} - W_{ik}^t)^2 \geq \frac{1}{2}\frac{W_{ik}^t(HH^T)_{kk}}{W_{ik}^t}(W_{ik} - W_{ik}^t)^2 = \frac{1}{2}(HH^T)_{kk}(W_{ik} - W_{ik}^t)^2.$$

Therefore, $G_W(W_{ik}, W_{ik}^t) \geq F_{W_{ik}}(W_{ik})$ follows. Meanwhile, $G_W(W_{ik}, W_{ik}) = F_{W_{ik}}(W_{ik})$ holds automatically. In summary, one has that $G_W(W_{ik}, W_{ik}^t)$ is an auxiliary function of $F_{W_{ik}}$. □

The following lemmas are needed to prove $G_V(V_{kj}, V_{kj}^t)$ is an auxiliary function of $F_{V_{kj}}$.

Lemma 3. *[67] Let $\Omega \subseteq R^N$ be a convex set, $F : \Omega \to R$ a convex function, and $\lambda_i \in [0, 1]$ with $i \in \{1, \cdots, k\}$ and $\sum_{i=1}^k \lambda_i = 1$. Then, for all $x_i \in \Omega$, $F\left(\sum_{i=1}^k \lambda_i x_i\right) \leq \sum_{i=1}^k \lambda_i F(x_i)$.*

Lemma 4. *Suppose $F(X) = \frac{1}{2}\|Y - KX\|_F^2$ with $X \in R_+^{p\times m}$. Let $A \in R_+^{p\times m}$ and $K \in R_+^{n\times p}$), then,*

$$F(X) \leq \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{1}{(KA)_{ij}}\sum_{k=1}^{p}K_{ik}A_{kj}\left(Y_{ij} - \frac{X_{kj}}{A_{kj}}(KA)_{ij}\right)^2.$$

Proof. First, let's expand the Frobenius norm of $F(X)$ in the following way:

$$F(X) = \frac{1}{2}\|Y - KX\|_F^2 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}(Y_{ij} - (KX)_{ij})^2.$$

Second, by Lemma 3, letting $\lambda_k = \frac{K_{ik}A_{kj}}{(KA)_{ij}}$ and using the auxiliary variable A, we have that

$$[Y_{ij} - (KX)_{ij}]^2 = [Y_{ij} - \sum_{k=1}^{p} \frac{K_{ik}A_{kj}}{(KA)_{ij}} \frac{X_{kj}(KA)_{ij}}{A_{kj}}]^2 \leq \sum_{k=1}^{p} \frac{K_{ik}A_{kj}}{(KA)_{ij}} \left(Y_{ij} - \frac{X_{kj}}{A_{kj}}(KA)_{ij}\right)^2.$$

Therefore, it leads to

$$F(X) = \frac{1}{2}\|Y - KX\|_F^2 \leq \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{1}{(KA)_{ij}} \sum_{k=1}^{p} K_{ik}A_{kj} \left(Y_{ij} - \frac{X_{kj}}{A_{kj}}(KA)_{ij}\right)^2,$$

and the result holds. □

So, we have the following conclusion that $G_V(V_{kj}, V_{kj}^t)$ is an auxiliary function of $F_{V_{kj}}$.

Lemma 5. *Let $G_V(V_{kj}, V_{kj}^t)$ be defined by* (5.5)*, then it serves as an auxiliary function for $F_{V_{kj}}$.*

Proof. According to Lemmas 3 and 4, we can estimate $G_V(V_{kj}, V_{kj}^t)$ as follows:

$$\begin{aligned}
G_V(V_{kj}, V_{kj}^t) &= \frac{\alpha_1}{2} \sum_{a=1}^{k} \sum_{b=1}^{k} \frac{1}{(HV^{t^T})_{ab}} \sum_{r=1}^{n} H_{ar} V^{t^T}_{rb} \left[\delta_{ab} - \frac{V_{br}}{V_{br}^t}(HV^{t^T})_{ab}\right]^2 + \frac{\alpha_2}{2}\|V - H\|_F^2 \\
&\geq \frac{\alpha_1}{2} \sum_{a=1}^{k} \sum_{b=1}^{k} \left[\delta_{ab} - \sum_{r=1}^{n} \frac{H_{ar}V_{rb}^{t^T}}{(HV^{t^T})_{ab}} \frac{V_{br}}{V_{br}^t}(HV^{t^T})_{ab}\right]^2 + \frac{\alpha_2}{2}\|V - H\|_F^2 \\
&= \frac{\alpha_1}{2} \sum_{a=1}^{k} \sum_{b=1}^{k} \left[\delta_{ab} - (HV^T)_{ab}\right]^2 + \frac{\alpha_2}{2}\|V - H\|_F^2 \\
&= F(V).
\end{aligned}$$

In addition, it is easy to see that $G_V(V_{kj}, V_{kj}) = F_{V_{kj}}(V_{kj})$ and the result holds. □

The third is to prove that $G_H(H_{kj}, H_{kj}^t)$ can serve as an auxiliary function of $F_{H_{kj}}$.

Lemma 6. *Let $G_H(H_{kj}, H_{kj}^t)$ be defined as in* (5.8)*, then it can serve as an auxiliary function for $F_{H_{kj}}$.*

Proof. According to the definitions of $F_1(H)$ and $F_2(H)$ in (5.3), to prove

$$G_H(H_{kj}, H_{kj}^t) = G_{H_1}(H_{kj}, H_{kj}^t) + G_{H_2}(H_{kj}, H_{kj}^t) \geq F_1(H_{kj}) + F_2(H_{kj}) = F(H_{kj}),$$

it is sufficient to show that

$$G_{H_1}(H_{kj}, H_{kj}^t) \geq F_1(H_{kj}),\ G_{H_2}(H_{kj}, H_{kj}^t) \geq F_2(H_{kj}). \tag{5.11}$$

The proof of the first part in (5.11) is similar to Lemma 2. To eliminate unnecessary misunderstandings, we will provide it in detail. First, by a Taylor series expansion of $F_{1_{H_{kj}}}(H_{kj})$, we have

$$F_1(H_{kj}) = F_1(H_{kj}^t) + F_1^{'}(H_{kj}^t)(H_{kj} - H_{kj}^t) + \frac{1}{2}[(W^TW)_{kk} + (\lambda L)_{jj}](H_{kj} - H_{kj}^t)^2.$$

Note that $(W^TWH^t)_{kj} = \sum_{r=1}^{K}(W^TW)_{kr}(H^t)_{rj} \geq (W^TW)_{kk}H^t_{kj}$, $(\lambda H^tD)_{kj} = \lambda\sum_{r=1}^{K}H^t_{kr}D_{rj} \geq \lambda H^t_{kj}D_{jj} \geq \lambda H^t_{kj}(D-S)_{jj} = \lambda H^t_{kj}L_{jj}$, then, $\frac{(W^TWH^t+\lambda H^tD)_{kj}}{H^t_{kj}} \geq (W^TW)_{kk} + (\lambda L)_{jj}$. Therefore,

$$\begin{aligned}\frac{1}{2}\frac{(W^TWH^t+\lambda H^tD)_{kj}}{H^t_{kj}}(H_{kj}-H^t_{kj})^2 &\geq \frac{1}{2}\frac{(W^TW)_{kk}(H^t)_{kj}+\lambda(H^t)_{kj}L_{jj}}{H^t_{kj}}(H_{kj}-H^t_{kj})^2\\ &= \frac{1}{2}[(W^TW)_{kk}+(\lambda L)_{jj}](H_{kj}-H^t_{kj})^2\end{aligned}$$

holds and $G_{H_1}(H_{kj}, H^t_{kj}) \geq F_1(H_{kj})$ follows.

The second part of (5.11) is also similar to Lemma 5. By Lemmas 3 and 4,

$$\begin{aligned}G_{H_2}(H_{kj}, H^t_{kj}) &= \frac{\alpha_1}{2}\sum_{a=1}^{k}\sum_{b=1}^{k}\frac{1}{(H^tV^T)_{ab}}\sum_{r=1}^{n}H^t_{ar}V^T_{rb}\left[\delta_{ab}-\frac{H_{ar}}{H^t_{ar}}(H^tV^T)_{ab}\right]^2+\frac{\alpha_2}{2}\|V-H\|_F^2\\ &\geq \frac{\alpha_1}{2}\sum_{a=1}^{k}\sum_{b=1}^{k}\left[\delta_{ab}-\sum_{r=1}^{n}\frac{H^t_{ar}V^T_{rb}}{(H^tV^T)_{ab}}\frac{H_{ar}}{H^t_{ar}}(H^tV^T)_{ab}\right]^2+\frac{\alpha_2}{2}\|V-H\|_F^2\\ &= \frac{\alpha_1}{2}\sum_{a=1}^{k}\sum_{b=1}^{k}\left[\delta_{ab}-(HV^T)_{ab}\right]^2+\frac{\alpha_2}{2}\|V-H\|_F^2 = F_2(H_{kj}).\end{aligned}$$

follows and $G_{H_2}(H_{kj}, H^t_{kj}) \geq F_2(H_{kj})$ holds. Then the result follows. □

Based on the above preparations, we now provide the main conclusion of this paper.

Theorem 1. *Let $X \in R_+^{m\times n}$, $W^0 \in R_+^{m\times k}$, $V^0 \in R_+^{k\times n}$, and $H^0 \in R_+^{k\times n}$. Then, the objective function in* (3.4) *is nonincreasing under sequences $\{W^t, V^t, H^t\}_{t=0}^{+\infty}$ generated by* (3.14)–(3.16).

Proof. First, since $\frac{\partial G_W(W_{ik}, W^t_{ik})}{\partial W_{ik}} = F'_{W_{ik}}(W^t_{ik}) + \frac{(W^tHH^T)_{ik}}{W^t_{ik}}(W_{ik}-W^t_{ik})$ and letting $\frac{\partial G_W(W_{ik}, W^t_{ik})}{\partial W_{ik}} = 0$, it then goes to $F'_{W_{ik}}(W^t_{ik})W^t_{ik} + (W^tHH^T)_{ik}W_{ik} - (W^tHH^T)_{ik}W^t_{ik} = 0$. According to Lemma 2,

$$W_{ik} = \frac{W^t_{ik}(W^tHH^T)_{ik} - W^t_{ik}F'_{W_{ik}}(W^t_{ik})}{(W^tHH^T)_{ik}} = \frac{W^t_{ik}(W^tHH^T)_{ik} - W^t_{ik}(W^tHH^T - XH^T)_{ik}}{(W^tHH^T)_{ik}} = W^t_{ik}\frac{(XH^T)_{ik}}{(W^tHH^T)_{ik}}.$$

By Lemma 1, if $W^{t+1}_{ik} = \arg\min_{W_{ik}\in R} G_W(W_{ik}, W^t_{ik}) = W^t_{ik}\frac{XH^T}{(W^tHH^T)_{ik}}$, then, (3.14) is nonincreasing. This with (3.4) has a nonnegative lower bound, then, (3.14) is a convergent iterative formula.

Second, due to the fact that

$$\begin{aligned}\frac{\partial G_V(V_{kj}, V^t_{kj})}{\partial V_{kj}} =& \frac{\alpha_1}{2}\sum_{a=1}^{k}\frac{H_{aj}V^{t^T}_{jk}}{(HV^{t^T})_{ak}}2\left[1-\frac{V_{kj}}{V^t_{kj}}(HV^{t^T})_{ak}\right]\left[-\frac{(HV^{t^T})_{ak}}{V^{t^T}_{kj}}\right]+\frac{\alpha_2}{2}2(V_{kj}-H_{kj})\\ =&\alpha_1\sum_{a=1}^{k}\left[-H_{aj}+\frac{H_{aj}}{V^t_{kj}}(HV^{t^T})_{ak}V_{kj}\right]+\alpha_2(V_{kj}-H_{kj})\\ =&\alpha_1\sum_{a=1}^{k}H_{aj}\left[\frac{V_{kj}}{V^t_{kj}}(HV^{t^T})_{ak}-\delta_{ak}\right]+\alpha_2(V_{kj}-H_{kj})\\ =&\alpha_1\left[\frac{V_{kj}}{V^t_{kj}}(V^tH^TH)_{kj}-H_{kj}\right]+\alpha_2(V_{kj}-H_{kj}).\end{aligned}$$

Let $\frac{\partial G_V(V_{kj},V_{kj}^t)}{\partial V_{kj}} = 0$, then, $\alpha_1\frac{V_{kj}}{V_{kj}^t}(V^tH^TH)_{kj} + \alpha_2 V_{kj} = (\alpha_1+\alpha_2)H_{kj}, \alpha_1 V_{kj}(V^tH^TH)_{kj} + \alpha_2 V_{kj}V_{kj}^t = (\alpha_1+\alpha_2)H_{kj}V_{kj}^t$. According to Lemma 5, we can get that $V_{kj} = V_{kj}^t\frac{[(\alpha_1+\alpha_2)H]_{kj}}{[\alpha_1 V^tH^TH+\alpha_2 V^t]_{kj}}$. So, by Lemma 1, if $V_{kj}^{t+1} = \arg\min_{V_{kj}\in R} G_V(V_{kj},V_{kj}^t) = V_{kj}^t\frac{[(\alpha_1+\alpha_2)H]_{kj}}{[\alpha_1 V^tH^TH+\alpha_2 V^t]_{kj}}$, then (3.15) is nonincreasing. Besides, the objective function in (3.4) has a nonnegative lower bound, and it can be concluded that the formula in (3.15) is a convergent iterative formula.

Third, it is true that $\frac{\partial G_{H_1}(H_{kj},H_{kj}^t)}{\partial H_{kj}} = F'_{1_{H_{kj}}}(H_{kj}^t) + \frac{(W^TWH^t+\lambda H^tD)_{kj}}{H_{kj}^t}(H_{kj}-H_{kj}^t)$ and

$$\begin{aligned}\frac{\partial G_{H_2}(H_{kj},H_{kj}^t)}{\partial H_{kj}} =& \frac{\alpha_1}{2}\sum_{b=1}^{k}\frac{H_{kj}^tV_{jb}^T}{(H^tV^T)_{kb}}2\left[1-\frac{H_{kj}}{H_{kj}^t}(H^tV^T)_{kb}\right]\left[-\frac{(H^tV^T)_{kb}}{H_{kj}^t}\right]+\frac{\alpha_2}{2}2(H_{kj}-V_{kj})\\ =&\alpha_1\sum_{b=1}^{k}\left[-V_{jb}^T+\frac{V_{jb}^T}{H_{kj}^t}(H^tV^T)_{kb}H_{kj}\right]+\alpha_2(H_{kj}-V_{kj})\\ =&\alpha_1\sum_{b=1}^{k}V_{jb}^T\left[\frac{H_{kj}}{H_{kj}^t}(H^tV^T)_{kb}-\delta_{kb}\right]+\alpha_2(H_{kj}-V_{kj})\\ =&\alpha_1\frac{H_{kj}}{H_{kj}^t}(H^tV^TV)_{kj}+\alpha_2H_{kj}-(\alpha_1+\alpha_2)V_{kj}.\end{aligned}$$

Let $\frac{\partial G_H(H_{kj},H_{kj}^t)}{\partial H_{kj}} = \frac{\partial G_{H_1}(H_{kj},H_{kj}^t)}{\partial H_{kj}} + \frac{\partial G_{H_2}(H_{kj},H_{kj}^t)}{\partial H_{kj}} = 0$, then,

$$\begin{aligned}&F'_{1_{H_{kj}}}(H_{kj}^t)+\frac{(W^TWH^t+\lambda H^tD)_{kj}}{H_{kj}^t}(H_{kj}-H_{kj}^t)+\alpha_1\frac{H_{kj}}{H_{kj}^t}(H^tV^TV)_{kj}+\alpha_2H_{kj}-(\alpha_1+\alpha_2)V_{kj}=0\\ &F'_{1_{H_{kj}}}(H_{kj}^t)H_{kj}^t+(W^TWH^t+\lambda H^tD)_{kj}H_{kj}-(W^TWH^t+\lambda H^tD)_{kj}H_{kj}^t+\alpha_1H_{kj}(H^tV^TV)_{kj}\\ &+\alpha_2H_{kj}H_{kj}^t-(\alpha_1+\alpha_2)V_{kj}H_{kj}^t=0.\end{aligned}$$

According to Lemma 6, it leads to $H_{kj} = H_{kj}^t\frac{[W^TX+\lambda H^tS+(\alpha_1+\alpha_2)V]_{kj}}{[W^TWH^t+\lambda H^tD+\alpha_1H^tV^TV+\alpha_2H^t]_{kj}}$. By Lemma 1, if

$$H_{kj}^{t+1} = \arg\min_{H_{kj}\in R} G_H(H_{kj},H_{kj}^t) = H_{kj}^t\frac{\left[W^TX+\lambda H^tS+(\alpha_1+\alpha_2)V\right]_{kj}}{\left[W^TWH^t+\lambda H^tD+\alpha_1H^tV^TV+\alpha_2H^t\right]_{kj}}$$

holds, then (3.16) is nonincreasing. Once again, since the objective function in (3.4) has a nonnegative lower bound, it can be concluded that the formula (3.16) is a convergent iterative formula. □