LETTER

A New Approach of Matrix Factorization on Complex Domain for Data Representation

Viet-Hang DUONG[†], Manh-Quan BUI[†], Jian-Jiun DING^{††}, Yuan-Shan LEE[†], Bach-Tung PHAM[†],
Pham The BAO^{†††}, Nonmembers, and Jia-Ching WANG^{†a)}, Member

SUMMARY This work presents a new approach which derives a learned data representation method through matrix factorization on the complex domain. In particular, we introduce an encoding matrix—a new representation of data—that satisfies the simplicial constraint of the projective basis matrix on the field of complex numbers. A complex optimization framework is provided. It employs the gradient descent method and computes the derivative of the cost function based on Wirtinger's calculus.

key words: complex matrix factorization, data representation, image clustering, gradient descent method

1. Introduction

Data representation is to extract the intrinsic structure of data. Matrix factorization (MF), a mathematical tool related to matrix analysis, is a fundamental technique for data representation. Nonnegative matrix factorization (NMF) [1], [2], a well-known MF model, is a part-based representation method resulted from the nonnegative constraint. NMF aims to find a linear approximation to the original matrix using two nonnegative matrices. It has attracted considerable attentions for a wide area of applications, such as image clustering [3]–[5], face and facial expression recognition [6]–[8], etc.

Motivated by the interesting works of Liwicki *et al.* [9] and Duong *et al.* [10], [11], which showed the effectiveness of projecting pixel intensity values to the complex plane, this work develops a new approach for data representation, named simplicial complex matrix factorization (siCMF). Herein, the cosine dissimilarity, used on the reconstruction error function, is exploited judiciously by transferring data from the real-valued space to the complex field. As a result, the real optimization problem with cosine divergence is replaced by optimizing a complex function with the Frobenius norm. In the proposed model, a complex-valued matrix will be decomposed into two matrices of complex bases and real coefficients—the solutions of a constraint complex optimization problem. The optimization of the

real-valued cost function with respect to complex parameters was treated wisely by using the block coordinate descent [12] and the gradient descent method [13]. Without limiting the sign of data, the proposed method, siCMF, can yield extension on real-world applications, particularly the field of complex-valued data processing, such as communication, acoustic, optic, and electromagnetics. In this paper, extensive comparisons over some representative NMF variants and the popular k -means algorithm on image clustering show that the proposed siCMF approach has outstanding performance in image clustering.

2. From NMF to siCMF for Data Representation

2.1 Nonnegative Matrix Factorization

The NMF problem can be formulated as follows: Given a nonnegative data matrix $\mathbf{X} \in \mathbb{R}_+^{N \times M}$, NMF aims to find one lower-rank nonnegative basis matrix $\mathbf{B} \in \mathbb{R}_+^{N \times K}$ and one nonnegative coefficient matrix $\mathbf{E} \in \mathbb{R}_+^{K \times M}$ such that $\mathbf{X} \approx \mathbf{BE}$.

The columns $\{\mathbf{X}_i\}$ ($i = 1, 2, \dots, M$) of $\mathbf{X}$ are expressed as $\mathbf{X}_i = \mathbf{BE}_i$; $\mathbf{X}_i \in \mathbb{R}_+^N$, $\mathbf{E}_i \in \mathbb{R}_+^K$. That is, all the columns of $\mathbf{X}$ can be represented as a non-negative combination of the columns $\{\mathbf{B}_i\}$ ($i = 1, 2, \dots, K$). NMF implicitly maps M -dimensional data to K -dimensional representations. The matrix $\mathbf{E}$ is a new representation of $\mathbf{X}$ in the part-based representation space which is spanned by the base $\mathbf{B}$.

Many modified NMFs have been developed to perform dimension reduction and data representation. One variation of NMF is simplicial nonnegative matrix factorization (siNMF) in which each data instance is a convex combination of latent components. In the siNMF model, each instance is a probabilistic distribution over latent components. Compared to other NMF formulations, it has several advantages such as easy interpretability, low complexity, convexity, sparsity, distributability, and parallelability [14].

2.2 Fast Robust Correlation and Image Representation in Complex Domain

The use of similarity measure is to quantify the similarity degree between two evaluated objects. Fitch *et al.* [15] introduced the fast robust correlation and expressed a robust matching surface as a series of correlations to improve the performance of matching. Liwicki *et al.* [9] adopted this cosine-based dissimilarity measure as follows:

Manuscript received May 22, 2017.

Manuscript revised September 7, 2017.

Manuscript publicized September 15, 2017.

[†]The authors are with National Central University, Zhongli, Taiwan.

^{††}The author is with National Taiwan University, Taipei, Taiwan.

^{†††}The author is with University of Science, Ho Chi Minh City, Viet Nam.

a) E-mail: jiacwang@gmail.com (Corresponding author)

DOI: 10.1587/transinf.2017EDL8115

$$d_{\cos}(x_k, x_j) = \sum_{c=1}^N [1 - \cos(\alpha\pi(x_k(c) - x_j(c)))] \quad (1)$$

where x_k and x_j denote the N -dimensional vectors obtained by expressing images $\mathbf{X}_k$ and $\mathbf{X}_j$ in the lexicographic order, respectively, $x_k, x_j \in [0, 1]$, and $\alpha \in \mathbb{R}^+$. The robustness of cosine dissimilarity in the real domain has been found. It is able to suppress outliers [9].

An explicit mapping function f from the space pixel intensity ($\mathbb{R}^N$) to the unit sphere ($\mathbb{C}^N$) by Euler's formula [16] is defined as follows:

$$f(\mathbf{x}_k) = \mathbf{z}_k = \frac{1}{\sqrt{2}} e^{i\alpha\pi\mathbf{x}_k} = \frac{1}{\sqrt{2}} \begin{bmatrix} e^{i\alpha\pi\mathbf{x}_k(1)} \\ \vdots \\ e^{i\alpha\pi\mathbf{x}_k(N)} \end{bmatrix} \quad (2)$$

$$\text{where } e^{i\alpha\pi x} = \cos(\alpha\pi x) + i \sin(\alpha\pi x). \quad (3)$$

Observationally, there is an equivalence between the cosine dissimilarity of two real vectors $\mathbf{x}_k, \mathbf{x}_j$ and the squared Frobenius distance of their corresponding images $\mathbf{z}_k, \mathbf{z}_j$ from the complex mapping $f(\cdot)$. In fact:

$$\begin{aligned} \|\mathbf{z}_k - \mathbf{z}_j\|_F^2 &= \|f(\mathbf{x}_k) - f(\mathbf{x}_j)\|_F^2 = \left\| \frac{1}{\sqrt{2}} e^{i\alpha\pi\mathbf{x}_k} - \frac{1}{\sqrt{2}} e^{i\alpha\pi\mathbf{x}_j} \right\|_F^2 \\ &= \frac{1}{2} \|\cos(\alpha\pi\mathbf{x}_k) + i \sin(\alpha\pi\mathbf{x}_k) - \cos(\alpha\pi\mathbf{x}_j) - i \sin(\alpha\pi\mathbf{x}_j)\|_F^2 \\ &= \sum_{c=1}^N [1 - \cos(\alpha\pi\mathbf{x}_k(c) - \alpha\pi\mathbf{x}_j(c))] = d_{\cos}(\mathbf{x}_k, \mathbf{x}_j). \end{aligned}$$

Through this transformation space, the optimization problem using cosine dissimilarity in the real field will be replaced by an equivalent complex optimization problem where the Frobenius norm is employed for reconstruction error measurement [10], [11].

2.3 Simplicial Complex Matrix Factorization (siCMF)

Considering a dataset of M complex vectors $\mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_M]$, $\mathbf{Z}_i \in \mathbb{C}^N$; $i = 1, 2, \dots, M$. Each of $\mathbf{Z}_i$ represents a data instance. For our proposed siCMF model, we decompose $\mathbf{Z}$ into a product of two matrices $\mathbf{B}$ and $\mathbf{E}$. $\mathbf{B} \in \mathbb{C}^{N \times K}$ and $\mathbf{E} \in \mathbb{R}_+^{K \times M}$ are the solutions of minimizing a cost function $O_{\text{siCMF}}(\mathbf{B}, \mathbf{E})$ such that each instance $\mathbf{Z}_i$ is a convex combination of latent components $\mathbf{B}$. We call $\mathbf{E}$ and $\mathbf{B}$ the encoding matrix and the basis matrix, respectively. The siCMF objective function is defined as follows:

$$\begin{aligned} \min_{\mathbf{B}, \mathbf{E}} O_{\text{siCMF}}(\mathbf{B}, \mathbf{E}) &= \frac{1}{2} \|\mathbf{Z} - \mathbf{B}\mathbf{E}\|_F^2 \\ \text{s.t } \mathbf{B} &\in \mathbb{C}^{N \times K}, \mathbf{E} \in \mathbb{R}_+^{K \times M} \text{ and } \sum_{i=1}^K \mathbf{E}_{ij} = 1 \quad \forall j \end{aligned} \quad (4)$$

$\|\cdot\|_F$ denotes the Frobenius norm and $K \ll \min\{N, M\}$.

3. Gradient Descent Method for siCMF

Equation (4) is a nonconvex minimization problem with respect to both matrices $\mathbf{B}$ and $\mathbf{E}$. Since this is an NP-hard

Algorithm 1: Solving the siCMF problem

Input: A data matrix $\mathbf{Z}$ and the number of latent components K .

Output: The new data representation $\mathbf{E}$ and latent components $\mathbf{B}$.

1. Initialize randomly $\mathbf{B} \in \mathbb{C}^{N \times K}$, $\mathbf{E} \in \mathbb{R}_+^{K \times M}$.

2. Normalize $\mathbf{E}$ such that $\sum_{i=1}^K \mathbf{E}_{ij} = 1 \quad \forall j$

3. **repeat**

4. Fix $\mathbf{B}$ to find $\mathbf{E}$ by minimizing

$$O_1(\mathbf{E}) = \frac{1}{2} \|\mathbf{Z} - \mathbf{B}\mathbf{E}\|_F^2 \text{ s.t } \mathbf{E} \in \mathbb{R}_+^{K \times M} \text{ and } \sum_{i=1}^K \mathbf{E}_{ij} = 1 \quad \forall j \quad (5)$$

5. Fix $\mathbf{E}$ to find $\mathbf{B}$ by minimizing

$$O_2(\mathbf{B}) = \frac{1}{2} \|\mathbf{Z} - \mathbf{B}\mathbf{E}\|_F^2. \quad (6)$$

6. **until** convergence

problem, it is impractical to solve (4) directly. However, it can be tackled by applying the block coordinate descent (BCD) with two matrix blocks [12]. The framework of minimizing over one matrix while keeping the other matrix fixed is described in Algorithm 1.

For minimizing the function of real variables $O_1(\mathbf{E})$, by expressing the constraints $\mathbf{E} \in \mathbb{R}_+^{K \times M}$ and $\sum_{i=1}^K \mathbf{E}_{ij} = 1$ as the column-wise normalized version of $\mathbf{E}$ [17], the constrained problem (5) was simplified as follows:

$$\min_{\bar{\mathbf{E}} \geq 0} O_1(\bar{\mathbf{E}}) = \min_{\bar{\mathbf{E}} \geq 0} \frac{1}{2} \|\mathbf{Z} - \mathbf{B}\bar{\mathbf{E}}\|_F^2 \quad (7)$$

$$\text{where } \bar{\mathbf{E}} = \left[\frac{\mathbf{E}_1}{\|\mathbf{E}_1\|_1}, \frac{\mathbf{E}_2}{\|\mathbf{E}_2\|_1}, \dots, \frac{\mathbf{E}_M}{\|\mathbf{E}_M\|_1} \right]; \bar{\mathbf{E}} \geq 0 \quad (8)$$

$$\text{and } O_1(\bar{\mathbf{E}}) = \frac{1}{2} \text{Trace}(\mathbf{Z}^H \mathbf{Z} - 2\bar{\mathbf{E}}^T \mathbf{B}^H \mathbf{Z} + \bar{\mathbf{E}}^T \mathbf{B}^H \mathbf{B} \bar{\mathbf{E}}). \quad (9)$$

Given fixed $\mathbf{B}$, the updating rule of (7) is [13]:

$$\bar{\mathbf{E}}^{(t+1)} = P[\bar{\mathbf{E}}^{(t)} - \beta^{(t)} \nabla_{\mathbf{E}} O_1(\bar{\mathbf{E}})] \quad (10)$$

where the projective rule $P[\cdot] = \max(0, \cdot)$ ensures that the elements of $\bar{\mathbf{E}}^{(t+1)}$ are nonnegative and the step size $\beta^{(t)}$ was estimated in the same way as in [13].

The computation of the gradient of $O_1(\bar{\mathbf{E}})$ with respect to $\mathbf{E}$ is more complicated. Assume that $\bar{O}_1(\mathbf{E}) = O_1(\bar{\mathbf{E}})$, then

$$\nabla_{\mathbf{E}} \bar{O}_1(\mathbf{E}) = \frac{1}{\|\mathbf{E}\|_1} (\mathbf{I} - \mathbf{F}\bar{\mathbf{E}}^T) \nabla_{\mathbf{E}} O_1(\bar{\mathbf{E}}) \quad (11)$$

where $\mathbf{F} = \partial \|\mathbf{S}\|_1 / \partial \mathbf{S}|_{\mathbf{S}=\mathbf{E}}$ (Its proof can be seen from [17]). Since $\partial \|\mathbf{S}\|_1 / \partial \mathbf{S} = 1$, therefore,

$$\nabla_{\mathbf{E}} \bar{O}_1(\mathbf{E}) = \frac{1}{\|\mathbf{E}\|_1} (\mathbf{I} - \bar{\mathbf{E}}^T) \nabla_{\mathbf{E}} O_1(\bar{\mathbf{E}}). \quad (12)$$

Here,

$$\nabla_{\mathbf{E}} O_1(\bar{\mathbf{E}}) = \mathbf{B}^H \mathbf{B} \bar{\mathbf{E}} - \mathbf{B}^H \mathbf{Z}. \quad (13)$$

About the real-valued function of complex variables $O_2(\mathbf{B})$, the Wirtinger calculus [18], which uses a conjugate coordinate system, can simplify the gradient evaluation of this nonholomorphic activation function. It means that $O_2(\mathbf{B})$ is treated as $O_2(\mathbf{B}, \mathbf{B}^*)$ and the complex-gradient of a real-valued function is evaluated as:

$$\begin{aligned} \nabla_{\mathbf{B}^*} O_2(\mathbf{B}, \mathbf{B}^*) &= 2 \frac{\partial O_2(\mathbf{B}, \mathbf{B}^*)}{\partial \mathbf{B}^*} \\ &= \frac{\partial O_2(\mathbf{B}, \mathbf{B}^*)}{\partial \text{Re} \mathbf{B}} + i \frac{\partial O_2(\mathbf{B}, \mathbf{B}^*)}{\partial \text{Im} \mathbf{B}}. \end{aligned} \quad (14)$$

The optimization of $O_2(.,.)$ can be performed by the following update rule [18], [19]:

$$\Delta \mathbf{B} = \mathbf{B}^{(t+1)} - \mathbf{B}^{(t)} = -\alpha^{(t)} \nabla_{\mathbf{B}^*} O_2(\mathbf{B}^{(t)}, \mathbf{B}^{*(t)}). \quad (15)$$

Then, there is a nonpositive increment of the function for $O_2(.,.)$ which is given by:

$$\Delta O_2(\mathbf{B}^{(t)}, \mathbf{B}^{*(t)}) = -\alpha^{(t)} \|\nabla_{\mathbf{B}^*} O_2(\mathbf{B}^{(t)}, \mathbf{B}^{*(t)})\|^2. \quad (16)$$

In this paper, the Armijo rule is used to estimate the step size α for the t^{th} iteration [13]. Specifically, $\alpha^{(t)} = \mu^{s_t}$, $0 < \mu < 1$, and s_t is the first non-negative integer such that:

$$\begin{aligned} O_2(\mathbf{B}^{(t+1)}, \mathbf{B}^{*(t+1)}) - O_2(\mathbf{B}^{(t)}, \mathbf{B}^{*(t)}) \\ \leq 2\sigma \text{Re} \left\{ \left\langle \nabla_{\mathbf{B}^*} O_2(\mathbf{B}^{(t)}, \mathbf{B}^{*(t)}), \mathbf{B}^{(t+1)} - \mathbf{B}^{(t)} \right\rangle \right\} \end{aligned} \quad (17)$$

$$\text{where } \nabla_{\mathbf{B}^*} O_2(\mathbf{B}, \mathbf{B}^*) = \mathbf{B} \mathbf{E} \mathbf{E}^H - \mathbf{Z} \mathbf{E}^H. \quad (18)$$

4. SiCMF for Image Clustering

SiCMF is utilized to represent data in a lower dimensional subspace. The proposed algorithm was applied for image clustering in an unsupervised setting. The overall procedure is described in Algorithm 2 where a new encoding matrix is estimated by siCMF and clustering is employed on the obtained subspace. Images are categorized based on the k -means clustering method.

Algorithm 2: Algorithm for image clustering

Input: Data matrix $\mathbf{X}$.

Output: Clustering results.

1. Normalize and transform $\mathbf{X}$ into the complex field using (2) to obtain the complex data matrix $\mathbf{Z}$,

$$\mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_M], \mathbf{Z}_i \in \mathbb{C}^N; i = 1, 2, \dots, M.$$

2. Dimensionality reduction: Apply siCMF on the complex data matrix $\mathbf{Z}$ to get the encoding matrix $\mathbf{E}$.

3. Clustering in the low dimension space: Apply k -means in the siCMF space to cluster $\mathbf{E}$ into categories.

4. Evaluate the accuracy (AC) and the normalized mutual information (NMI) of the clustering results.

It is noticed that (2) involves a parameter α in which a smaller α results in a dissimilarity function similar to the L_2 -norm. Theoretically, α represents the frequency of the cosine function and is optimized to suppress the values caused by outliers [9]. However, this parameter is not critical as we can see from CMF [9], EE-CMF [10], and siCMF, which simply set α to 1 and still had competitive performance. To fasten computation, we set $\alpha = 1$.

5. Experiments

In this section, extensive experiments are carried out to examine the effectiveness of the proposed siCMF model on clustering. Traditional methods such as canonical k -means clustering on the original data space (KM) [19], graph regularized NMF (GNMF) [3], a weighted NMF with different error weights for different entries (weNMF) [20], nonnegative graph embedding (NGE) [6], complex matrix factorization (CMF) [9], and exemplar-embed complex matrix factorization (EE-CMF) [10] are compared.

5.1 Datasets and Experimental Settings

We conducted clustering experiments on two popular datasets including the object dataset COIL-100 [22] and the face dataset ORL [23]. The COIL-100 dataset comes from the Columbia object image library for validating object clustering algorithms. The ORL dataset consists of 400 face images from 40 different persons. For each person, there are 10 sample images captured at different conditions.

All images are resized to 64×64 pixels with 256 possible gray level values.

As most prospective NMF studies for clustering, in our experiments, the number of bases K is set to be the same as the number of clusters.

The evaluations are conducted with different numbers of clusters K varying in the range of {4; 6; 8; 10; 12; 14}. For each fixed cluster number K , we randomly selected K categories from COIL-100 and ORL datasets as the data matrix $\mathbf{X}$. At least ten subsets $\mathbf{X}$ were randomly created and we repeated the clustering experiments ten times for each of the data matrices.

The next step is to map the original space $\mathbf{X}$ to a low-dimensional (K -dimensional) space. There are two cases on this learning subspace phase. The first case is to apply directly NMF methods (GNMF, weNMF and NGE) on the real set $\mathbf{X}$. Transforming the pixel intensive matrix $\mathbf{X}$ to the complex data matrix $\mathbf{Z}$ by (2) and then employing complex matrix factorization techniques (CMF, EE-CMF and proposed siCMF) is the second case of learning new data representation $\mathbf{E}$. The dimensionality of the learn space is herein set to be the same as the number of clusters.

A clustering tool, k -means, was applied thirty times to the new data representation $\mathbf{E}$ for images clustering with different initializations. The best results were reported in terms of two popular evaluation metrics for clustering, including accuracy (AC) and normalized mutual information (NMI).

Table 1 Clustering performance (mean±std) on the COIL100 dataset.

K	Accuracy (%)							
	KM	NMF	GNMF	weNMF	NGE	CMF	EE-CMF	siCMF
4	82.87±15.05	76.16±11.74	81.60±15.59	79.80±13.38	83.07±14.37	70.43±13.02	72.30±10.60	83.39±13.83
6	81.12±10.04	75.56±07.97	81.94±10.42	75.57±10.08	81.73±09.44	67.79±10.38	64.14±06.40	83.38±08.68
8	71.95±13.05	69.51±12.51	75.23±11.34	70.56±11.85	74.10±11.61	60.26±09.14	65.56±10.55	75.89±11.42
10	71.26±11.02	69.99±10.86	73.40±13.97	70.45±11.20	74.47±12.51	64.57±08.54	68.75±09.58	77.26±14.80
12	71.48±10.54	70.51±10.98	73.58±12.20	70.15±10.79	73.20±11.19	64.58±08.57	69.39±09.53	74.18±13.84
14	63.05±05.16	62.01±04.59	69.67±05.87	62.50±04.95	69.27±08.09	62.96±06.90	68.07±07.60	63.09±06.57
K	Normalized mutual information (%)							
	KM	NMF	GNMF	weNMF	NGE	CMF	EE-CMF	siCMF
4	78.58±18.14	70.71±15.97	80.23±10.49	73.08±17.13	78.09±17.74	56.74±17.39	72.35±06.88	80.55±14.67
6	81.23±9.15	75.66±08.07	81.75±07.16	75.57±09.40	81.88±9.28	62.51±10.37	75.54±07.79	83.19±08.43
8	75.40±12.09	71.13±11.53	80.02±09.99	72.11±11.20	77.21±10.61	60.43±09.29	71.28±09.57	78.16±11.53
10	76.99±10.30	74.11±09.88	80.99±10.51	74.14±10.14	78.91±10.63	66.68±09.10	75.00±08.67	81.62±10.68
12	78.22±9.86	75.95±09.86	80.65±06.14	75.81±09.51	79.13±8.98	69.00±07.75	75.08±08.25	79.47±11.37
14	73.21±4.75	70.79±04.22	78.61±06.20	71.14±04.54	78.60±6.46	70.04±06.04	75.19±07.68	74.95±07.53

Table 2 Clustering performance (mean±std) on the ORL dataset.

K	Accuracy (%)							
	KM	NMF	GNMF	weNMF	NGE	CMF	EE-CMF	siCMF
4	90.70±10.23	87.25±10.27	90.06±06.95	82.33±08.19	90.06±06.54	66.82±08.79	38.18±04.41	90.63±07.98
6	92.63±08.51	79.31±06.72	86.53±08.86	78.98±05.94	90.23±07.30	65.24±07.40	41.19±20.77	90.80±09.15
8	82.10±20.58	76.28±05.49	82.57±10.36	76.20±06.40	86.69±07.58	64.29±10.54	62.23±15.38	87.17±08.25
10	81.52±07.89	73.38±05.82	82.38±08.77	74.36±08.14	84.23±07.67	59.43±07.36	61.50±11.52	84.48±08.46
12	78.61±06.00	72.58±05.23	75.46±08.07	71.81±05.32	81.70±05.91	61.86±05.31	57.95±15.75	81.83±08.55
14	71.91±04.76	66.75±15.36	78.24±06.70	70.16±05.15	80.83±06.17	65.46±06.12	59.59±05.94	81.20±06.86
K	Normalized mutual information (%)							
	KM	NMF	GNMF	weNMF	NGE	CMF	EE-CMF	siCMF
4	90.07±8.79	70.71±15.97	86.40±10.12	76.77±09.07	87.74±07.33	51.44±08.79	49.87±04.17	88.40±06.56
6	93.57±6.85	75.66±08.07	86.41±08.13	78.36±06.04	90.54±05.69	58.72±07.53	46.70±25.14	90.67±08.24
8	90.33±5.21	71.13±11.53	86.66±07.63	79.10±05.19	78.31±10.96	65.19±08.32	65.18±19.09	88.65±05.63
10	87.58±5.56	74.11±09.88	86.50±06.39	79.14±06.60	88.17±05.38	63.68±06.94	67.19±14.07	88.99±05.19
12	86.73±4.18	75.95±09.86	82.90±06.42	78.64±04.15	86.84±05.44	68.58±03.72	64.95±17.45	88.38±04.49
14	83.97±3.47	70.79±04.22	84.80±05.38	78.94±03.61	88.40±04.27	71.48±05.35	69.05±04.24	88.20±03.82

tion (NMI) [24], [25]. The average performance as well as the standard deviation were recorded as the final result.

5.2 Performance Evaluation and Comparisons

Tables 1 and 2 show the clustering results on the COIL-100 and ORL datasets, respectively. The best results are highlighted in the bold font. Because AC and NMI are both positive measurements, the larger value means the better clustering performance. From these results, some interesting points can be observed as follows.

The AC metric is more stable than the NMI metric and a smaller number of clusters always leads to a higher accuracy rate. By contrast, the changes of NMI is fluctuation.

On the COIL-100 dataset, the proposed siCMF method is superior to the baseline methods in most cases.

On the ORL dataset, the proposed siCMF algorithm outperforms the complex models CMF and EE-CMF as well as the compared real NMFs in both the AC and NMI . SiCMF is even better than GNMF and NGE, which show

that the underlying manifold structure of the data is useful in data clustering. Here, the k -means method with raw high-dimensional features also have comparable good performance in terms of NMI on the case of small cluster numbers. For the AC metric, the k -means method outperforms CMF, EE-CMF, and real NMFs. However, its performance is still not as good as that of the proposed siCMF method.

The promising results demonstrate that siCMF can learn a compact low dimensional representation by leveraging the power of convex combination on the complex domain.

6. Conclusion

This work explored a new approach of matrix factorization on the complex domain for learning representation. It aims to learn a lower-dimensional data representation that preserves excellent properties of NMF and satisfies simplicial constraints of the projective basis matrix on the complex domain. Experimental results on two real-world datasets

reveal that the proposed method, siCMF, yields promising results by extending the core concept of siNMF from the real number field to the complex field. Utilizing this model for other applications (e.g., face recognition, facial expression recognition, and sound processing) and how to further improve the effectiveness and robustness by adding more constraints are challenges and future research directions.

References

- [1] D.D. Lee and H.S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol.401, no.6755, pp.755–791, 1999.
- [2] D.D. Lee and H.S. Seung, "Algorithms for nonnegative matrix factorization," *Proc. NIPS*, pp.556–562, 2000.
- [3] J. Kim and H. Park, "Sparse nonnegative matrix factorization for clustering," *CSE Technical Reports*, Georgia Institute of Technology, 2008.
- [4] D. Cai, X.F. He, J.W. Han, and T.S. Huang, "Graph regularized non-negative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.33, no.8, pp.1548–1560, 2011.
- [5] H. Zhang, Z.-J. Zha, Y. Yang, S. Yan, and T.-S. Chua, "Robust (semi) nonnegative graph embedding," *IEEE Trans. Image Process.*, vol.23, no.7, pp.2996–3012, 2014.
- [6] P. Li, J. Bu, Y. Yang, R. Ji, C. Chen, and D. Cai, "Discriminative orthogonal nonnegative matrix factorization with flexibility for data representation," *Expert Syst. Appl.*, vol.41, no.4, pp.1283–1293, 2014.
- [7] J. Yang, S. Yang, Y. Fu, X. Li, and T. Huang, "Non-negative graph embedding," *Proc. IEEE CVPR*, pp.1–8, 2008.
- [8] R. Zhi, M. Flierl, Q. Ruan, and W.B. Kleijn, "Graph-preserving sparse nonnegative matrix factorization with application to facial expression recognition," *IEEE Trans. Syst., Man, Cybern.*, vol.41, no.1, pp.38–52, 2011.
- [9] S. Liwicki, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "Euler principal component analysis," *Int. J. Comput. Vis.*, vol.1, pp.498–518, 2013.
- [10] V.H. Duong, Y.S. Lee, B.T. Pham, S. Mathulapragans, P.T. Bao, and J.C. Wang, "Complex matrix factorization for face recognition," available in <https://arxiv.org/ftp/arxiv/papers/1612/1612.02513.pdf>
- [11] V.H. Duong, Y.S. Lee, J.J. Ding, B.T. Pham, P.T. Bao, and J.C. Wang, "Exemplar-embed complex matrix factorization for facial expression recognition," *IEEE ICASSP*, 2017.
- [12] J. Kim, Y. He, and H. Park, "Algorithms for nonnegative matrix and tensor factorizations: a unified view based on block coordinate descent framework," *Global Optimization*, vol.58, pp.285–319, 2013.
- [13] C.-J. Lin, "Projected gradient methods for nonnegative matrix factorization," *Neural Comput.*, vol.19, no.10, pp.2756–2779, 2007.
- [14] D.K. Nguyen, K. Than, and T.B. Ho, "Simplicial nonnegative matrix factorization," *Proc. Int. Conf. on Research, Innovation and Vision for Future (RIVF)*, pp.47–52, 2013.
- [15] A.J. Fitch, A. Kadyrov, W.J. Christmas, and J. Kittler, "Fast robust correlation," *IEEE Trans. Image Process.*, vol.14, no.8, pp.1063–1073, 2005.
- [16] M.A. Moskowitz, *A Course in Complex Analysis in One Variable*, World Scientific Publishing, p.7, 2002.
- [17] J. Le Roux, J.R. Hershey, and F. Weninger, "Sparse NMF half-baked or well done?" *Mitsubishi Electric Research Labs (MERL), Tech. Rep.*, no. TR2015-023, Cambridge, MA, USA, March 2015.
- [18] M.F. Amin, M.I. Amin, A. Al-Nuaimi, and K. Murase, "Wirtinger calculus based gradient descent and Levenberg Marquardt learning algorithms in complex-valued neural networks," *Proc. Int. Conf. Neural Information Processing (ICONIP)*, pp.550–559, 2011.
- [19] H. Li and T. Adali, "Complex-valued adaptive signal processing using nonlinear functions," *EURASIP on Adv. in Sig. Proc.*, pp.1–9, 2008.
- [20] M. Hua, M.K. Lau, J. Pei, and K. Wu, "Continuous K-means monitoring with low reporting cost in sensor networks," *IEEE Trans. Knowl. Data Eng.*, vol.21, no.12, pp.1679–1691, 2009.
- [21] N.D. Ho, P.V. Dooren, and V. Blondel, "Weighted nonnegative matrix factorization and face feature extraction," 2007, available in <http://www.academia.edu/download/30778865/476447.pdf>.
- [22] S.A. Nene, S.K. Nayar, and H. Murase, "Columbia object image library (COIL-100)," *Technical Report CUCS-006-96*, Feb. 1996.
- [23] The ORL dataset of Face, available in <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>.
- [24] W. Xu and Y. Gong, "Document clustering by concept factorization," *ACM Conf. Research and Development in Information Retrieval*, pp.202–209, 2004.
- [25] D. Cai, X.F. He, and J.W. Han, "Locally consistent concept factorization for document clustering," *IEEE Trans. Knowl. Data Eng.*, vol.23, no.6, pp.902–913, 2011.