

Received October 19, 2020, accepted November 2, 2020, date of publication November 17, 2020, date of current version December 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3038383

# Robust Ensemble Manifold Projective Non-Negative Matrix Factorization for Image Representation

PENG LUO<sup>1,3</sup>, XILONG QU<sup>2</sup>, LINA TAN<sup>4</sup>, XIAOLIANG XIE<sup>1</sup>, WEIJIN JIANG<sup>4</sup>, LIRONG HUANG<sup>2</sup>, WAI HUNG IP<sup>5</sup>, (Senior Member, IEEE), AND KAI LEUNG YUNG<sup>5</sup>

<sup>1</sup>School of Mathematics and Statistics, Hunan University of Technology and Business, Changsha 410205, China

<sup>2</sup>School of Information Technology and Management, Hunan University of Finance and Economics, Changsha 410205, China

<sup>3</sup>Key Laboratory of Hunan Province for Statistical Learning and Intelligent Computation, Hunan University of Technology and Business, Changsha 410205, China

<sup>4</sup>School of Computer and Information Engineering, Hunan University of Technology and Business, Changsha 410205, China

<sup>5</sup>Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong

Corresponding author: Xilong Qu (xlq1978@126.com)

This work was supported in part by the Projects of the National Social Science Foundation of China under Grant 19BTJ011, in part by the Outstanding Youth Project of Education Department of Hunan Province, China, under Grant 20B151, in part by the Teaching Reform Key Research Project of Hunan University of Finance and Economics under Grant 2017xjjg012, in part by the Teaching Reform Research Project of Hunan University of Finance and Economics under Grant 2017xjjg015, in part by the 2017 Teaching Reform Research Project of Hunan Province (Research on CDIO Talent Training Model of Computer Science and Technology Major with Local and Industry Characteristics), in part by a Research project of the Education Department of Hunan Province (17K015) and (18A441), in part by the Educational Planning Key Project of Hunan Province under Grant XJK18DJA1, in part by the Research project of Education Department of Hunan Province under Grant 18C0958, and in part by the Department of Industrial and Systems Engineering of the Hong Kong Polytechnic University under Grant H-ZG3K.

**ABSTRACT** Projective non-negative matrix factorization (PNMF) as a variant of NMF has received considerable attention. However, the existing PNMf methods can be further improved from two aspects. On the one hand, the square loss function that is intended to measure the reconstruction error is sensitive to noise. On the other hand, it is non-trivial to estimate the intrinsic manifold of the feature space in a principal manner. So current paper is an attempt that has proposed a new method named as robust ensemble manifold projective non-negative matrix factorization (REPNMF) for image representation. Specifically, REPNMF not only assesses the influence of noise by imposing a spare noise matrix for image reconstruction, but it also assumes that the intrinsic manifold exists in a convex hull of certain pre-given manifold candidates. We aim to remove noise from the data and find the optimized combination of candidate manifolds to approximate the intrinsic manifold simultaneously. We develop iterative multiplicative updating rules for the optimization of REPNMF along with its convergence proof. The experimental results on four image datasets verify that REPNMF is superior as compare to other related state-of-the-art methods.

**INDEX TERMS** Non-negative matrix factorization, projection recovery, image representation, ensemble manifold learning.

## I. INTRODUCTION

Matrix factorization is a highly effective strategy for representation learning. It aims to find two or more low-rank factor matrices for which the product could closely approximate the original high-dimensional data matrix. For one thing, the dimension of decomposed factor matrices is often lower than that of the original data matrix. This provides a compact representation, which is beneficial to many subsequent learning tasks, such as clustering. For another, because the

learned latent representation can be represented by the corresponding basis components, we could purposefully regularize the factor matrices for specific assignments. Non-negative matrix factorization (NMF) [1] is one of the most characteristic techniques among matrix factorization. The non-negative constraints may tend to a part-based representation in which a zero-value represents the absence, and a positive-value represents the presence of a component. Therefore, NMF allows a non-subtractive combination of parts to make a whole, which is highly similar to the human perception mechanism. NMF has demonstrated superior performance in image representation [2]–[7].

The associate editor coordinating the review of this manuscript and approving it for publication was Mu-Yen Chen <sup>5</sup>.

One limitation of NMF is that it fails to consider the out-of-sample problem [3]. That is to say, NMF is not capable to get the coefficient of any unseen instance. To get rid of the limitations of standard NMF while inheriting all of its advantages, Yuan *et al.* [3] developed an improved NMF, named projective non-negative matrix factorization (PNMF), which approximates the data matrix by its non-negative subspace projection. Specially, PNMf learns the non-negative basis matrix of the low-rank subspace and reuses its transpose as the projection matrix. Due to the non-negative constraint of the learned projection matrix, PNMf can obtain non-negative coefficients for any new coming instance because the production of a non-negative matrix and non-negative vector is a non-negative vector. Moreover, PNMf has fewer parameters and generates a much sparser factor matrix, and thus, it is beneficial to subsequent learning tasks, such as data clustering.

Both NMF and PNMf attempt to obtain new basis vectors such that the data can be well represented. Data points are typically sampled from a sub-manifold of the ambient Euclidean space [8]–[10]. Therefore, considering the intrinsic manifold structure will be favourable for learning new data representations. To address this issue, graph regularized non-negative matrix factorization [2] and graph regularized projective non-negative matrix factorization [11] were proposed by adding the manifold regularizer on the new data representations derived from NMF and PNMf, respectively. Lu *et al.* [12] proposed projective robust non-negative matrix factorization (PRNF) to make PNMf more robust. However, in many real-world data analytic problems, data points might be sampled from various distributions; hence, it is crucial to estimate the intrinsic manifold in a principled manner.

To address this problem, we propose a novel method termed Robust Ensemble manifold Projective Non-negative Matrix Factorization (REPNMF) for image representation. On the one hand, a sparse error matrix is introduced into REPNMF to capture the noise such that the factorization can extract more intrinsic information from the scoured data. On the other hand, an assumption is made that intrinsic manifold is embedded in the convex hull of a set of pre-given candidate manifolds [13]. The main purpose is to find the combination of candidate manifolds to approximate the intrinsic manifold. In this way, the local geometrical structure of data can be better preserved because different candidate manifold characters have different structural information of the data. An iterative updating rule is generated that repeatedly updates the projective matrix to optimize the objective function. Furthermore, the convergence proof of updating rule is presented. Experimental results on three face datasets and one science-quality dataset (Chang'e 3 data: Rover Pano-ramic Camera images [14], [15]) have shown the superiority of REPNMF, and it can significantly outperform related algorithms.

The rest of the paper is structured as follows. Section II provides a short review of the related works. In Section III, we introduce the proposed robust ensemble manifold projective non-negative matrix factorization method. In Section IV,

we report the experimental results with analysis. The paper is summed up in section 5 with concluding remarks.

## II. RELATED WORK

The section II is designed to give overview of related works. This section starts with an explanation of commonly mentioned notations in present paper.

### A. COMMON NOTATIONS

In this paper, we use uppercase boldface letters to denote matrices and lowercase boldface letters to denote vectors.  $M_{ij}$  indicates the  $(i, j)$ -th element of matrix  $\mathbf{M}$ . The  $i$ -th element of a vector  $\mathbf{a}$  is denoted by  $a_i$ . In the non-negative data matrix  $\mathbf{X} \in \mathbb{R}_+^{M \times N}$ , each column vector represents the feature vector of the corresponding item. Throughout the entire paper, Frobenius norm of matrix  $\mathbf{M}$  is represented as  $\|\mathbf{M}\|_F$ .

### B. NON-NEGATIVE MATRIX FACTORIZATION

Assumed a non-negative data matrix  $\mathbf{X} \in \mathbb{R}_+^{M \times N}$  which includes  $N$  data points in  $M$ -dimensional space and an existing positive integer  $K < \min(M, N)$ , NMF aims to obtain two lower-rank non-negative matrices  $\mathbf{U} \in \mathbb{R}_+^{M \times K}$  and  $\mathbf{V} \in \mathbb{R}_+^{K \times N}$  the product of which could closely approximate the original high-dimensional data matrix  $\mathbf{X}$ :

$$\mathcal{J}_{NMF} = \min_{\mathbf{U}, \mathbf{V}} \|\mathbf{X} - \mathbf{UV}\|_F^2 \quad \text{s.t. } U_{ik} \geq 0, V_{kj} \geq 0, \forall i, j, k. \quad (1)$$

Although NMF is jointly non-convex with respect to  $\mathbf{U}$  and  $\mathbf{V}$ , it is convex with respect to  $\mathbf{U}$  and  $\mathbf{V}$ , respectively. Thus, the optimization problem in Eq. (1) could be optimized by efficient multiplicative update rules [16] as follows:

$$U_{ik} = U_{ik} \frac{(\mathbf{XV}^T)_{ik}}{(\mathbf{UVV}^T)_{ik}}, \quad V_{kj} = V_{kj} \frac{(\mathbf{U}^T \mathbf{X})_{kj}}{(\mathbf{U}^T \mathbf{UV})_{kj}} \quad (2)$$

It should be noted that the non-negative coefficient of forecasted data point cannot be figured by NMF, because NMF encounters challenge of out of the sample problem.

### C. PROJECTIVE NON-NEGATIVE MATRIX FACTORIZATION

To eradicate the insufficiency of NMF to deal with out-of-the-sample problem, Yuan *et al.* [3] developed projective non-negative matrix factorization (PNMF). PNMf obtains a non-negative projective matrix and projects the original data matrix into a latent subspace that has lower dimensions. Given the basis matrix  $\mathbf{U}$ , PNMf regards  $\mathbf{U}^T \mathbf{X}$  as the new representation and employs  $\mathbf{UU}^T \mathbf{X}$  for reconstruction data matrix  $\mathbf{X}$ . The objective function of PNMf is formulated as:

$$\mathcal{J}_{PNMF} = \min_{\mathbf{U}} \|\mathbf{X} - \mathbf{UU}^T \mathbf{X}\|_F^2 \quad \text{s.t. } U_{ik} \geq 0, \forall i, k. \quad (3)$$

Since the objective function of  $\mathcal{J}_{PNMF}$  in Eq. (3) is not convex in  $\mathbf{W}$ , achieving the global minimum of PNMf is impossible. The local minima of  $\mathcal{J}_{PNMF}$  is calculated using the following updating rule [3]:

$$U_{ik} = U_{ik} \frac{(\mathbf{VV}^T \mathbf{U})_{ik}}{(\mathbf{UU}^T \mathbf{VV}^T \mathbf{U} + \mathbf{VV}^T \mathbf{UU}^T \mathbf{U})_{ik}} \quad (4)$$

It was revealed by Yuan *et al.* [3] that the local minima can be found by objective function  $\mathcal{J}_{PNMF}$  according to the updating rule Eq. (4).

**D. AUTOMATED GRAPH REGULARIZED PROJECTIVE NON-NEGATIVE MATRIX FACTORIZATION**

NMF and PNMf fit data in Euclidean space, however the intrinsic geometry of data remains unexplored. To address this issue, Pei *et al.* [11] presented automated graph regularized projective non-negative matrix factorization. The purpose behind AGPNMF is to reform the regular PNMf that is done by combining automated graph regularization and the PNMf decomposition. The main feature of this reformed method is that AGPNMF instantaneously calculates the graph weights matrix and operates a dimensionality reduction of raw data. The model of AGPNMF is as follow:

$$\begin{aligned} \mathcal{J}_{AGPNMF} = \min_{\mathbf{H}, \mathbf{S}} & \|\mathbf{X} - \mathbf{X}\mathbf{H}\mathbf{H}^T\|_F^2 + \tau \text{Tr}(\mathbf{H}^T \mathbf{L} \mathbf{H}) \\ & + \|\mathbf{X} - \mathbf{X}\mathbf{S}\|_F^2 + \alpha \text{Tr}(\mathbf{S}^T \mathbf{S}) + \beta \|\mathbf{S}\|_1 \\ \text{s.t. } & H_{ik} \geq 0, S_{ik} \geq 0, \quad \forall i, k. \end{aligned} \quad (5)$$

In above equation  $\tau, \alpha, \beta$  are constants. Additionally, multiplicative update rule was designed by Pei *et al.* [11] proposed to solve AGPNMF:

$$S_{ik} = S_{ik} \frac{(\mathbf{X}^T \mathbf{X} + \frac{1}{2} \tau \mathbf{H}\mathbf{H}^T)_{ik}}{(\mathbf{X}^T \mathbf{X}\mathbf{S})_{ik} + \frac{1}{4} \tau ((\mathbf{H}\mathbf{H}^T))_{ik} + (\mathbf{H}\mathbf{H}^T)_{ii} + \frac{\beta}{2}} \quad (6)$$

$$H_{ik} = H_{ik} \frac{-B_{ik} + \sqrt{B_{ik}^2 - 4A_{ik}C_{ik}}}{2A_{ik}} \quad (7)$$

where

$$\begin{aligned} A_{ik} &= (\mathbf{H}\mathbf{H}^T \mathbf{X}^T \mathbf{X} \mathbf{H} + \mathbf{X}^T \mathbf{X} \mathbf{H}\mathbf{H}^T \mathbf{H} + (\tau \mathbf{D})\mathbf{H})_{ik} \\ B_{ik} &= -2(\mathbf{X}^T \mathbf{X}\mathbf{H})_{ik} H_{ik} \\ C_{ik} &= -\frac{1}{2} \tau (\mathbf{R}\mathbf{H})_{ik} H_{ik}^2 \end{aligned}$$

Pei *et al.* [11] have demonstrated that the objective function  $\mathcal{J}_{AGPNMF}$  will obtain the local minima according to the updating rules Eq. (6),(7).

**E. ROBUST PROJECTIVE NON-NEGATIVE MATRIX FACTORIZATION**

NMF, PNMf and AGPNMF are all susceptible to noise and are unfit for feature extraction when the data are contaminated by noise. To increase the robustness of these methods, projective robust non-negative factorization [12] (PRNF) is developed for robust feature extraction. To capture the geometrical structure of raw data, a graph regularization term was introduced into PRNF. Additionally, PRNF introduced a sparsity-inducing norm as sparsity constraints on the noise matrix. RPNF is formulated as follow:

$$\begin{aligned} \mathcal{J}_{RPNF} &= \min_{\mathbf{U}, \mathbf{E}} \|\mathbf{X} - \mathbf{U}\mathbf{U}^T \mathbf{X} - \mathbf{E}\|_F^2 + \lambda_1 \text{Tr}(\mathbf{U}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{U}) + \lambda_2 \|\mathbf{E}\|_{1/2} \\ \text{s.t. } & U_{ik} \geq 0, E_{ij} \geq 0, \quad \forall i, j, k. \end{aligned} \quad (8)$$

where  $\lambda_1 \geq 0$  and  $\lambda_2 \geq 0$  are regularization parameters,  $\mathbf{L} = \mathbf{D} - \mathbf{S}$  and  $\mathbf{S}$  is the weight matrix of the nearest neighbour graph,  $\mathbf{D}_{ii} = \sum \mathbf{S}_{ij}$ . and  $\|\mathbf{E}\|_{1/2}$  is defined as

$$\|\mathbf{E}\|_{1/2} = \sum_{i=1}^M \sum_{j=1}^N e_{ij}^{1/2} \quad (9)$$

An alternating scheme was designed by Lu *et al.* [12] to optimize the objective function  $\mathcal{J}_{RPNF}$ :

$$U_{ik} = U_{ik} \frac{(2\mathbf{X}\mathbf{X}^T \mathbf{U} + \lambda_1 \mathbf{X}\mathbf{W}\mathbf{X}^T \mathbf{U})_{ik}}{(2\mathbf{E}\mathbf{X}^T \mathbf{U} + 2\mathbf{U}\mathbf{U}^T \mathbf{X}\mathbf{X}^T \mathbf{U} + \lambda_1 \mathbf{X}\mathbf{D}\mathbf{X}^T \mathbf{U})_{ik}}, \quad (10)$$

$$E_{ij} = E_{ij} \frac{(2\mathbf{X})_{ij}}{(2\mathbf{E} + 2\mathbf{U}\mathbf{U}^T \mathbf{X} + \frac{\lambda_2}{2} \mathbf{E} - \frac{1}{2})_{kj}} \quad (11)$$

Lu *et al.* [12] have demonstrated that the objective function  $\mathcal{J}_{RPNF}$  will obtain the local minima according to the updating rules Eq. (10), (11).

**III. ROBUST ENSEMBLE MANIFOLD PROJECTIVE NON-NEGATIVE MATRIX FACTORIZATION**

In this section, we develop a novel robust ensemble manifold projective non-negative factorization (REPNNMF) algorithm, which takes into account the impact of noise and better captures the geometrical structure of the data simultaneously.

**A. THE BASIC OBJECTIVE**

PNMF [3] is an extension of NMF, but the squared loss used in PNMf is sensitive to noise data. Typically, there are two ways to deal with it. One involves using noise-robust loss instead of squared loss as in PNMf. The other is removing noise from the data. In this article, we choose the second method. Inspired by Robust Principal Component Analysis (RPCA) [17], the raw data matrix  $\mathbf{X}$  can be decomposed as a low-rank component and a sparse component. In fact, we can find that the matrix  $\mathbf{U}$  and  $\mathbf{U}^T \mathbf{X}$  are all low-rank matrices. Intuitively, we can design a sparse component matrix  $\mathbf{E}$  to the objective function such that the low-rank component can lead to a better reconstruction of the underlying structure. Thus, the basic objective function can be formulated as:

$$\begin{aligned} \mathcal{J}_{Basic} &= \min_{\mathbf{U}, \mathbf{E}} \|\mathbf{X} - \mathbf{U}\mathbf{U}^T \mathbf{X} - \mathbf{E}\|_F^2 + \beta \|\mathbf{E}\|_1 \\ \text{s.t. } & U_{ik} \geq 0, \quad \forall i, k. \end{aligned} \quad (12)$$

where  $\|\mathbf{E}\|_1 = \sum_{ij} |E_{ij}|$  presents the  $l_1$  norm, which ensures the sparseness of the matrix.

**B. ENSEMBLE MANIFOLD**

Recall that PNMf attempts to obtain a set of projective vectors that could be optimized for the linear approximation of the original data. The  $j$ -th column of matrix  $\mathbf{U}^T \mathbf{X}$ ,  $\mathbf{z}_j = [(\mathbf{U}^T \mathbf{X})_{j1}, \dots, (\mathbf{U}^T \mathbf{X})_{jr}]^T$  can be viewed as the novel representation of each instance according to the new basis. A natural assumption is that if data points  $\mathbf{x}_j, \mathbf{x}_l$  are adjacent in the original feature space, then  $\mathbf{z}_j, \mathbf{z}_l$ , the new representation of these two data points in the new basis should also be close to each other. This assumption is known as local invariance assumption, which plays an essential role in various algorithms, including semi-supervised learning algorithms [18] and dimensionality reduction algorithms [8].

Ensemble manifold theory [13] assumes that the intrinsic manifold is located in a convex hull of the pre-defined manifold candidates, each of which indicates one kind of data distribution. Ensemble manifold essentially learns to combine the diverse manifold candidates whose optimal linear combination can approximate the intrinsic manifold. Let  $\mathbf{L}_{En}$  be the intrinsic manifold,  $\mathbf{L}_i$  be the  $i$ -th candidate manifold, and there are  $q$  manifold candidates corresponding to various data distribution. In our work, we use the  $k$ -nearest neighbour graph to characterize data distribution. Specifically, for data point  $\mathbf{x}_j$ , we seek its  $k$  nearest neighbours and generate an edge between  $\mathbf{x}_j$  and its neighbours. Generally speaking, there are many schemes for defining weight matrix  $\mathbf{S}$ . The three most common methods are as follows:

- 1) **0-1 Weighting.**  $S_{jl} = 1$ , if and only if node  $j$  and node  $l$  are attached by an edge.
- 2) **Hear kernel Weighting.** If node  $j$  and node  $l$  are linked, set

$$S_{jl} = e^{-\frac{\|\mathbf{x}_j - \mathbf{x}_l\|^2}{\sigma}}$$

- 3) **Cosine Similarity Weighting.** If node  $j$  and node  $l$  are linked by an edge, set

$$S_{jl} = \frac{\mathbf{x}_j \mathbf{x}_l}{\|\mathbf{x}_j\| \|\mathbf{x}_l\|}$$

Each candidate graph Laplacian denotes a manifold which can be denoted by  $\mathbf{L}_i = \mathbf{D}_i - \mathbf{S}_i$ , where  $\mathbf{D}_{ii} = \sum_j S_{ij}$ .

Thus, the ensemble learning assumption is equivalent to constraining the search space of candidate graph Laplacians, i.e.,

$$\mathbf{L}_{En} = \sum_{i=1}^q \mu_i \mathbf{L}_i, \text{ s.t. } \sum_{i=1}^m \mu_i = 1, \mu_i \geq 0, \quad i = 1, \dots, q \quad (13)$$

Because  $\mathbf{L}_{En}$  is a linear combination of  $m$  candidate graph Laplacians, it is also a graph Laplacian.

### C. OBJECTIVE FUNCTION

To model the local geometric structure of the data points, we impose the ensemble manifold regularizer  $Tr(\mathbf{W}^T \mathbf{X} \mathbf{L}_{En} \mathbf{X}^T \mathbf{W})$  into a basic objective function. Therefore, the objective function of REPnMF can be written as

$$\begin{aligned} \mathcal{J}_{REPnMF} &= \min_{\mathbf{U}, \mathbf{E}, \mu} \frac{1}{2} \|\mathbf{X} - \mathbf{U}\mathbf{U}^T \mathbf{X} - \mathbf{E}\|_F^2 + \beta \|\mathbf{E}\|_1 \\ &\quad + \lambda Tr(\mathbf{U}^T \mathbf{X} \mathbf{L}_{En} \mathbf{X}^T \mathbf{U}) + \alpha \|\mu\|^2 \\ \text{s.t. } U_{jk} &\geq 0, \mathbf{L}_{En} = \sum_{i=1}^q \mu_i \mathbf{L}_i, \sum_i \mu_i = 1, \mu_i \geq 0 \forall i, j, k. \end{aligned} \quad (14)$$

where the parameter  $\lambda$  controls the contribution of ensemble manifold regularizer,  $\beta$  controls the regularization term  $\|\mathbf{E}\|_1$  and  $\alpha$  controls the regularization term  $\|\mu\|^2$ .

### Algorithm 1 Optimization of REPnMF

---

**Input:** data matrix  $\mathbf{X}$ , parameters  $\beta, \lambda, \mu$   
**Output:** projective matrix  $\mathbf{U}$

```

1 begin
2   Initialize  $\mathbf{U}, \mathbf{E}$ 
3   Initialize  $(\mu_1, \mu_2, \mu_3) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ 
4   repeat
5     Optimize problem (14) with respect to  $\mathbf{U}$  while
      keeping  $\mathbf{E}$  and  $\mu$  fixed.
6     Optimize problem (14) with respect to  $\mathbf{E}$  while
      keeping  $\mathbf{U}$  and  $\mu$  fixed.
7     Optimize problem (14) with respect to  $\mu$  while
      keeping  $\mathbf{U}$  and  $\mathbf{E}$  fixed
8   until convergence or max no. iterations reached
9 end

```

---

### D. OPTIMIZATION

In this section, we investigate how to optimize the objective function of  $\mathcal{J}_{REPnMF}$  in Eq. (14). It is easy to see that  $\mathcal{J}_{REPnMF}$  is non-convex in  $\mathbf{U}, \mathbf{E}$  and  $\mu$  jointly. However it is convex in them, respectively. Therefore, it is impossible to obtain the global minimum because no analytical solution can be found. We design an alternating scheme to optimize the objective function. The procedure is depicted in Algorithm 1. For the sake of convenience, we replace  $\mathbf{L}_{En}$  as  $\mathbf{L}$ . It can be easily observed that  $\mathbf{L} = \mathbf{L}^+ - \mathbf{L}^-$ ,  $L_{ij}^+ = (|L_{ij}| + L_{ij})/2$  and  $L_{ij}^- = (|L_{ij}| - L_{ij})/2$ . We rewrite the objective function as follows:

$$\begin{aligned} \mathcal{O} &= \frac{1}{2} Tr[(\mathbf{X} - \mathbf{E})(\mathbf{X} - \mathbf{E})^T] - Tr(\mathbf{U}\mathbf{U}^T \mathbf{X}(\mathbf{X} - \mathbf{E})^T) \\ &\quad + \frac{1}{2} Tr(\mathbf{U}\mathbf{U}^T \mathbf{X} \mathbf{X}^T \mathbf{U}\mathbf{U}^T) + \lambda Tr(\mathbf{U}^T \mathbf{X} \mathbf{L}^+ \mathbf{X}^T \mathbf{U}) \\ &\quad - Tr(\mathbf{U}^T \mathbf{X} \mathbf{L}^- \mathbf{X}^T \mathbf{U}) + \beta \|\mathbf{E}\|_1 + \alpha \|\mu\|^2 \end{aligned} \quad (15)$$

#### 1) OPTIMIZATION $\mathbf{U}$

Let  $\psi_{jl}$  be the Lagrange multiplier for constraints  $\mathbf{u}_j \geq 0$ , and denote  $\Psi = [\psi_{jl}]$ . Then, we can write the Lagrange

$$\mathcal{L} = \mathcal{O} + Tr(\Psi \mathbf{U}^T) \quad (16)$$

The partial derivatives of  $\mathcal{L}$  with respect to  $\mathbf{U}$  is

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{U}} &= -[(\mathbf{X} - \mathbf{E})\mathbf{X}^T \mathbf{U} + \mathbf{X}(\mathbf{X} - \mathbf{E})^T \mathbf{U}] \\ &\quad + \mathbf{U}\mathbf{U}^T \mathbf{X} \mathbf{X}^T \mathbf{U} + \mathbf{X} \mathbf{X}^T \mathbf{U}\mathbf{U}^T \mathbf{U} + \lambda \mathbf{X} \mathbf{L}^+ \mathbf{X}^T \mathbf{U} + \Psi \end{aligned} \quad (17)$$

According to KKT condition  $\Psi_{ik} \mathbf{U}_{ik} = 0$ , let  $\frac{\partial \mathcal{L}}{\partial \mathbf{U}} = 0$ , we can obtain

$$\mathbf{U}_{ik} = \mathbf{U}_{ik} \frac{(2\mathbf{X} \mathbf{X}^T \mathbf{U} + \lambda \mathbf{X} \mathbf{L}^+ \mathbf{X}^T \mathbf{U})_{ik}}{(\mathbf{D1} + \mathbf{D2} + \mathbf{D3})_{ik}} \quad (18)$$

where  $\mathbf{D1} = \mathbf{E} \mathbf{X}^T \mathbf{U} + \mathbf{X} \mathbf{E}^T \mathbf{U}$ ,  $\mathbf{D2} = \mathbf{U}\mathbf{U}^T \mathbf{X} \mathbf{X}^T \mathbf{U} + \mathbf{X} \mathbf{X}^T \mathbf{U}\mathbf{U}^T \mathbf{U}$  and  $\mathbf{D3} = \lambda \mathbf{X} \mathbf{L}^+ \mathbf{X}^T \mathbf{U}$

## 2) OPTIMIZATION E

It is easy to see that the optimization problem related to  $\mathbf{E}$  is element-wise decoupled. In other words, there is more chance that  $\mathbf{E}$  can be optimized individually. Denote  $\mathbf{R} = \mathbf{X} - \mathbf{U}\mathbf{U}^T\mathbf{X} = [\mathbf{r}_{ij}]$ . Then each sub-problem with respect to  $\mathbf{e}_{ij}$  can be written as:

$$\mathcal{O}_{ij} = (\mathbf{r}_{ij} - \mathbf{e}_{ij})^2 + \beta|\mathbf{e}_{ij}| \quad (19)$$

The unique solution of Eq. (19) can be effectively solved by the soft-thresholding operator [19] as follows:

$$\mathbf{e}_{ij} = \begin{cases} 0, & \text{if } |\mathbf{r}_{ij}| \leq \frac{\beta}{2} \\ \mathbf{r}_{ij} - \frac{\beta}{2} \text{sign}(\mathbf{r}_{ij}), & \text{otherwise} \end{cases} \quad (20)$$

By substituting Eq. (20) into Eq. (19), we may obtain

$$\mathcal{O}_{ij} = \begin{cases} \mathbf{r}_{ij}^2, & \text{if } |\mathbf{r}_{ij}| \leq \frac{\beta}{2} \\ \beta|\mathbf{r}_{ij}| - (\frac{\lambda}{2})^2, & \text{otherwise} \end{cases} \quad (21)$$

It was revealed by the results that REPNMF can adaptively allocate square loss to minor-error entries for accurate reconstruction and  $\ell_1$  loss to major-error entries to reduce the influence of noise.

## 3) OPTIMIZATION $\mu$

After fixing  $\mathbf{U}$  and  $\mathbf{E}$ , the objective function simplified as:

$$\begin{aligned} \min \lambda \text{Tr}[\mathbf{U}^T \mathbf{X} (\sum_{i=1}^q \mu_i \mathbf{L}_i) \mathbf{X}^T \mathbf{U}] + \alpha \|\mu\|^2 \\ \text{s.t. } \sum_{i=1}^q \mu_i = 1, \mu_i \geq 0, \quad i = 1, \dots, q. \end{aligned} \quad (22)$$

This is a classical quadratic programming problem, various convex programming software are very effective designed for solving this problem. Here, this convex optimization problem is solved by CVX<sup>1</sup>, a Matlab-based convex modeling framework for convex optimization.

## E. CONVERGENCE ANALYSIS

In the following, we will analyze the convergence of the updating rules in Eq. (20) and Eq. (18). Apparently  $\mathcal{O}_{ij}$  meets the minimum when  $\mathbf{e}_{ij}$  is updated as Eq. (20). Because  $\mathcal{O}_{ij}$  is non-increasing following Eq. (20), we only require to prove that  $\mathcal{O}_{ij}$  is non-increasing under the updating rule for  $\mathbf{U}$  in Eq. (18). For iterative updating rules (18), we obtain the following theorem.

*Theorem 1:* The objective function (14) does not increase under the update rules in (18).

Firstly, an auxiliary function is introduced to prove Theorem 1, in which the following definitions and lemmas are required:

*Definition 1:* Function  $G(u, u')$  is an auxiliary function for  $F(u)$  if the conditions  $G(u, u') \geq F(u)$ ,  $G(u, u) = F(u)$  are met.

<sup>1</sup><http://cvxr.com/cvx>

*Lemma 1:* If  $G$  is an auxiliary function of  $F$ , then  $F$  is non-increasing under the updating

$$u^{t+1} = \arg \max_u G(u, u') \quad (23)$$

Let  $\mathbf{u}_{ab}$  is any element in  $\mathbf{U}$  and  $F_{\mathbf{u}_{ab}}$  denote the part of (15) relevant to  $\mathbf{u}_{ab}$ .

*Lemma 2:* Function

$$\begin{aligned} G(u, u'_{ab}) \\ = F_{\mathbf{u}_{ab}}(u'_{ab}) + F'_{\mathbf{u}_{ab}}(u'_{ab})(u - u'_{ab}) \\ + \frac{(\mathbf{E}\mathbf{X}^T\mathbf{U} + \mathbf{X}\mathbf{E}^T\mathbf{U} + 6\mathbf{U}\mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U} + \lambda\mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{U})_{ab}}{u'_{ab}}(u - u'_{ab})^2 \end{aligned} \quad (24)$$

is an auxiliary function for  $F_{\mathbf{u}_{ab}}$ , which is the only part of (14) relevant to  $\mathbf{u}_{ab}$ .

*Proof:* Since  $G(u, u) = F_{\mathbf{u}_{ab}}$ , we only need to prove  $G(u, u'_{ab}) \geq F_{\mathbf{u}_{ab}}(u)$ . We compare  $G(u, u'_{ab})$  with the Talore series expansion of  $F_{\mathbf{u}_{ab}}(u)$

$$\begin{aligned} F_{\mathbf{u}_{ab}}(u) = F_{\mathbf{u}_{ab}}(u'_{ab}) + F'_{\mathbf{u}_{ab}}(u'_{ab})(u - u'_{ab}) \\ + \frac{1}{2}F''_{\mathbf{u}_{ab}}(u'_{ab})(u - u'_{ab})^2, \end{aligned} \quad (25)$$

where  $F''_{\mathbf{u}_{ab}}$  is the second-order derivative with respect to  $\mathbf{U}$ . It is easy to see that

$$\begin{aligned} F'_{\mathbf{u}_{ab}} &= (-4(\mathbf{X} - \mathbf{E})\mathbf{X}^T\mathbf{U} + 4\mathbf{U}\mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U} + 2\lambda\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{U})_{ab}, \\ F''_{\mathbf{u}_{ab}} &= (-4\mathbf{X}(\mathbf{X} - \mathbf{E})^T + 12\mathbf{X}\mathbf{X}^T\mathbf{X}\mathbf{U}\mathbf{U}^T + 2\lambda\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{U})_{ab}. \end{aligned} \quad (26)$$

We find that to prove  $G(u, u'_{ab}) \geq F_{\mathbf{u}_{ab}}(u)$  is equivalent to prove

$$\begin{aligned} \frac{(\mathbf{E}\mathbf{X}^T\mathbf{U} + \mathbf{X}\mathbf{E}^T\mathbf{U} + 6\mathbf{U}\mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U} + \lambda\mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{U})_{ab}}{u'_{ab}} \\ \geq \frac{1}{2}F''_{\mathbf{u}_{ab}}(u'_{ab}). \end{aligned} \quad (27)$$

To prove the above inequality, we have

$$\begin{aligned} (\mathbf{E}\mathbf{X}^T\mathbf{U} + \mathbf{X}\mathbf{E}^T\mathbf{U})_{ab} &\geq \mathbf{u}'_{ab}(\mathbf{X}(\mathbf{E} - \mathbf{X}^T))_{ab} \\ (\mathbf{U}\mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U})_{ab} &\geq \mathbf{u}'_{ab}(\mathbf{X}\mathbf{X}^T\mathbf{U}\mathbf{U}^T)_{ab} \\ (\mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{U})_{ab} &\geq \mathbf{u}'_{ab}(\mathbf{X}\mathbf{L}\mathbf{X}^T)_{ab} \end{aligned} \quad (28)$$

Thus, (27) holds and  $G(u, u'_{ab}) \geq F_{\mathbf{u}_{ab}}(u)$ .

*Proof of Theorem 1:* Replacing  $G(u, u'_{ab})$  in (23) by (24), results in the update rule

$$\begin{aligned} u_{ab}^{(t+1)} \\ = \arg \max_u G(u, u_{ab}^{(t)}) \\ = u_{ab}^{(t)} \frac{(2\mathbf{X}\mathbf{X}^T\mathbf{U} + \lambda\mathbf{X}\mathbf{W}\mathbf{X}^T\mathbf{U})_{ab}}{(\mathbf{E}\mathbf{X}^T\mathbf{U} + \mathbf{X}\mathbf{E}^T\mathbf{U} + 2\mathbf{U}\mathbf{U}^T\mathbf{X}\mathbf{X}^T\mathbf{U} + \lambda\mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{U})_{ab}} \end{aligned} \quad (29)$$

Because (24) is an auxiliary function,  $F_{\mathbf{u}_{ab}}$  is non-increasing under this updating rule.

Give the current solution  $\mathbf{U}'$ , we approximate  $\mathcal{J}_{REP\text{NMF}}$  by its Taylor-series expansion

$$\begin{aligned} \mathcal{J}_{REP\text{NMF}} &\approx \frac{1}{2}Tr(\mathbf{X}\mathbf{X}^T) - Tr(\mathbf{X}\mathbf{E}^T) - Tr(\mathbf{U}^T\mathbf{X}(\mathbf{X} - \mathbf{E})^T\mathbf{U}') \\ &+ \frac{1}{2}Tr(\mathbf{E}\mathbf{E}^T) + \frac{1}{2}Tr(\mathbf{U}^T(\mathbf{X}\mathbf{X}^T\mathbf{U}'\mathbf{U}'^T\mathbf{U}' + \mathbf{U}'\mathbf{U}'^T\mathbf{X}\mathbf{X}^T\mathbf{U}')) \\ &+ \frac{1}{2}\lambda Tr(\mathbf{U}^T\mathbf{X}\mathbf{L}_+\mathbf{X}\mathbf{U}' - \mathbf{U}^T\mathbf{X}\mathbf{L}_-\mathbf{X}\mathbf{U}') \end{aligned} \quad (30)$$

We construct an auxiliary function  $G(\mathbf{U}, \mathbf{U}')$  of  $\mathcal{J}_{REP\text{NMF}}$  as follows:

$$\begin{aligned} G(\mathbf{U}, \mathbf{U}') &= \frac{1}{2}Tr(\mathbf{X}\mathbf{X}^T) - Tr(\mathbf{X}\mathbf{E}^T) + \frac{1}{2}Tr(\mathbf{E}\mathbf{E}^T) \\ &- \frac{1}{2}\sum_{ik}(\mathbf{X}(\mathbf{X} - \mathbf{E})^T\mathbf{U}')_{ik}\mathbf{U}'_{ik}(1 + \log \frac{\mathbf{U}_{ik}}{\mathbf{U}'_{ik}}) \\ &- \frac{1}{2}\sum_{ik}((\mathbf{X} - \mathbf{E})\mathbf{X}^T\mathbf{U}')_{ik}\mathbf{U}'_{ik}(1 + \log \frac{\mathbf{U}_{ik}}{\mathbf{U}'_{ik}}) \\ &+ \frac{1}{2}Tr(\mathbf{U}^T(\mathbf{X}\mathbf{X}^T\mathbf{U}'\mathbf{U}'^T\mathbf{U}' + \mathbf{U}'\mathbf{U}'^T\mathbf{X}\mathbf{X}^T\mathbf{U}')) \\ &+ \frac{1}{2}\lambda Tr(\mathbf{U}^T\mathbf{X}\mathbf{L}_+\mathbf{X}^T\mathbf{U}') \\ &- \frac{1}{2}\sum_{ik}(\mathbf{L}_-\mathbf{X}^T\mathbf{U}')_{ik}(\mathbf{X}\mathbf{U}')_{ik}(1 + \log \frac{(\mathbf{X}^T\mathbf{U})_{ik}}{(\mathbf{X}^T\mathbf{U}')_{ik}}) \end{aligned} \quad (31)$$

It is easy to verify that  $\mathcal{J}_{REP\text{NMF}}(\mathbf{U}') = G(\mathbf{U}', \mathbf{U}')$ .

Next, we will prove that  $\mathcal{J}_{REP\text{NMF}}(\mathbf{U}) \leq G(\mathbf{U}, \mathbf{U}')$  to complete the proof. For any  $z \geq 0$ , we see that  $z \geq 1 + \log z$ . By substituting  $z = \frac{\mathbf{U}_{ik}}{\mathbf{U}'_{ik}}$  into the above inequality, we have

$$\mathbf{U}_{ik} \geq \mathbf{U}'_{ik}(1 + \log \frac{\mathbf{U}_{ik}}{\mathbf{U}'_{ik}}) \quad (32)$$

since  $Tr(\mathbf{U}^T\mathbf{X}(\mathbf{X} - \mathbf{E})^T\mathbf{U}') = \sum_{ik}(\mathbf{X}(\mathbf{X} - \mathbf{E})^T\mathbf{U}')_{ik}\mathbf{U}'_{ik}$ ,  $Tr(\mathbf{U}^T(\mathbf{X} - \mathbf{E})\mathbf{X}^T\mathbf{U}') = \sum_{ik}((\mathbf{X} - \mathbf{E})\mathbf{X}^T\mathbf{U}')_{ik}\mathbf{U}'_{ik}$ ,  $Tr(\mathbf{U}^T\mathbf{X}\mathbf{L}_+\mathbf{X}^T\mathbf{U}') = \sum_{ik}(\mathbf{L}_+\mathbf{X}^T\mathbf{U}')_{ik}(\mathbf{X}\mathbf{U}')_{ik}$ , from (32), we have

$$\begin{aligned} Tr(\mathbf{U}^T\mathbf{X}(\mathbf{X} - \mathbf{E})^T\mathbf{U}') &\geq \sum_{ik}(\mathbf{X}(\mathbf{X} - \mathbf{E})^T\mathbf{U}')_{ik}\mathbf{U}'_{ik}(1 + \log \frac{\mathbf{U}_{ik}}{\mathbf{U}'_{ik}}) \\ Tr(\mathbf{U}^T(\mathbf{X} - \mathbf{E})\mathbf{X}^T\mathbf{U}') &\geq \sum_{ik}((\mathbf{X} - \mathbf{E})\mathbf{X}^T\mathbf{U}')_{ik}\mathbf{U}'_{ik}(1 + \log \frac{\mathbf{U}_{ik}}{\mathbf{U}'_{ik}}) \end{aligned} \quad (33)$$

$$\begin{aligned} Tr(\mathbf{U}^T\mathbf{X}\mathbf{L}_+\mathbf{X}\mathbf{U}') &\geq \sum_{ik}(\mathbf{X}\mathbf{L}_+\mathbf{X}^T\mathbf{U}')_{ik}\mathbf{U}'_{ik}(1 + \log \frac{\mathbf{U}_{ik}}{\mathbf{U}'_{ik}}) \end{aligned} \quad (34)$$

By substituting (34) and (33) into (30), we prove that  $\mathcal{J}_{REP\text{NMF}}(\mathbf{U}) \leq G(\mathbf{U}, \mathbf{U}')$ .

Assuming that  $\mathbf{U}''$  is the minimum of  $G(\mathbf{U}, \mathbf{U}')$ , we get the following inequalities:

$$\mathcal{J}_{REP\text{NMF}}(\mathbf{U}'') \leq G(\mathbf{U}'', \mathbf{U}') \leq G(\mathbf{U}', \mathbf{U}') = \mathcal{J}_{REP\text{NMF}}(\mathbf{U}'). \quad (35)$$

TABLE 1. Statistics of the datasets.

Dataset	# Example	# Feature	# Classes
ORL	400	1024	40
YALE	165	1024	15
UMIST	575	644	20
Chang'e 3	200	1024	10

The remaining items involve calculating  $\mathbf{U}''$  and verifying its nonnegativity constraint. To this end, we set the gradient of  $G(\mathbf{U}, \mathbf{U}')$  to zero, that is to say i.e.,

$$\begin{aligned} \frac{\partial G(\mathbf{U}, \mathbf{U}')}{\partial \mathbf{U}_{ik}} &= -(\mathbf{X}(\mathbf{X} - \mathbf{E})^T\mathbf{U}')_{ik}\frac{\mathbf{U}'_{ik}}{\mathbf{U}_{ik}} - ((\mathbf{X} - \mathbf{E})\mathbf{X}^T\mathbf{U}')_{ik}\frac{\mathbf{U}'_{ik}}{\mathbf{U}_{ik}} \\ &+ ((\mathbf{X}\mathbf{X}^T\mathbf{U}'\mathbf{U}'^T\mathbf{U}')_{ik} + \mathbf{U}'\mathbf{U}'^T\mathbf{X}\mathbf{X}^T\mathbf{U}')_{ik} \\ &+ \lambda(\mathbf{U}^T\mathbf{X}\mathbf{L}_+\mathbf{X}\mathbf{U}')_{ik} - \lambda(\mathbf{U}^T\mathbf{X}\mathbf{L}_-\mathbf{X}\mathbf{U}')_{ik}\frac{\mathbf{U}'_{ik}}{\mathbf{U}_{ik}} = 0. \end{aligned} \quad (36)$$

Eq. (36) gives

$$\mathbf{U}'' = \frac{\mathbf{A}_{ik}}{\mathbf{B1}_{ik} + \mathbf{B2}_{ik}} \quad (37)$$

where  $\mathbf{A} = 2\mathbf{X}\mathbf{X}^T\mathbf{U}' + \lambda\mathbf{X}\mathbf{L}_-\mathbf{X}^T\mathbf{U}'$ ,  $\mathbf{B1} = \mathbf{E}\mathbf{X}^T\mathbf{U}' + \mathbf{X}\mathbf{E}^T\mathbf{U}' + \mathbf{U}'\mathbf{U}'^T\mathbf{X}\mathbf{X}^T\mathbf{U}'$  and  $\mathbf{B2} = \mathbf{X}\mathbf{X}^T\mathbf{U}'\mathbf{U}'^T\mathbf{U}' + \lambda\mathbf{X}\mathbf{L}_+\mathbf{X}^T\mathbf{U}'$

Because (37) contains multiplications and divisions of non-negative entries,  $\mathbf{U}''$  is a non-negative matrix.

It is clear that (37) is equivalent to (18), and thus (35) implies that (18) decreases the objective function of REP\text{NMF}. It completes the proof.

#### IV. EXPERIMENT

We use extensive experiments on four image datasets: ORL, YALE, UMIST and Chang'e 3 image database to validate the efficacy of REP\text{NMF}.

##### 1) DATASETS AND METRICS

Four datasets are chosen in the experiment. Table 1 summarizes the statistical information of these datasets.

###### a: ORL

This dataset comprises 400 images of 40 different people. 10 different images were taken from each person at different times, in varying lighting and with changed facial expressions. These images are cropped to  $32 \times 32$  pixel grey-level images and then reconstructed into a 1024-dimensional vector.

###### b: YALE

Yale is the second face database which is made of total 165 images, belonging to 15 different people. Each person has 11 images with diverse facial expressions or configurations. All images are reshaped to  $32 \times 32$  pixel grey-level images and then adjusted into a 1024-dimensional vector.

###### c: UMIST

There are 575 images of 20 individuals in UMIST face databases. Like other dataset images are cut to  $23 \times 28$  pixel

grey-scale images, and each image is reshaped into a 644-dimensional vector.

*d: Chang'e 3*

Chang'e 3 is a Chinese robotic lunar exploration mission, started by China National Space Administration (CNSA). This mission is the first Chinese mission that has robotic lander and lunar rover. From past decades China has been successful in most of the space and lunar missions. Prof. KL Yung [20]–[22] has led many space mission successfully and his team was conducted at the Hong Kong Polytechnic University in which they designed the camera system to install it the lunar surface lander for getting images. For current experience 200 images of 10 classes from Change 3 images are randomly selected. All images are cropped to pixel grey-scale images and then reconstructed into a 1024-dimensional vector.

To evaluate the clustering performance, we use two commonly used metrics: Clustering accuracy (ACC) and normalized mutual information (NMI). They are defined as follows:

$$\text{Accuracy} = \frac{\sum_{i=1}^n \delta(s_i, \text{map}(r_i))}{n}, \quad (38)$$

$$\text{NMI}(C, C^\dagger) = \frac{MI(C, C^\dagger)}{\max(H(C), H(C^\dagger))}, \quad (39)$$

where  $s_i$  serves as the cluster label of item  $i$  and  $r_i$  is the outcome of clustering of item  $i$ . If  $x$  is equal to  $y$ , then  $\delta(x, y)$  is equal to 1, otherwise  $\delta(x, y) = 0$ .  $\text{map}(r_i)$  is the permutation mapping function that maps  $r_i$  to the equivalent cluster label in ground truth.  $H(C)$  denotes the entropy of cluster set  $C$ .  $MI(C, C^\dagger)$  is the mutual information between  $C$  and  $C^\dagger$ :

$$MI(C, C^\dagger) = \sum_{c_i \in C, c_j^\dagger \in C^\dagger} p(c_i, c_j^\dagger) \log_2 \frac{p(c_i, c_j^\dagger)}{p(c_i)p(c_j^\dagger)}. \quad (40)$$

$p(c_i)$  shows the likelihood of the item that is nominated randomly fits in cluster  $c_i$  among testing items and  $p(c_i, c_j^\dagger)$  is the joint probability of the randomly selected item in both  $c_i$  and  $c_j^\dagger$ . If the two cluster sets are identical,  $\text{NMI}(C, C^\dagger) = 1$ . If the two cluster sets are fully independent,  $\text{NMI}(C, C^\dagger) = 0$ .

2) COMPARED ALGORITHMS

To highlight the effects of the new method over other baseline algorithms and its significance, REPnMF is compared with the following methods:

- Traditional kmeans clustering method (**Kmeans** in short).
- NMF [1]-based clustering (**NMF** in short).
- PNMf [3]-based clustering (**PNMF** in short).
- Automated graph regularized projective non-negative matrix factorization [11]-based clustering (**AGPNMF** in short).
- Projective robust non-negative factorization [12]-based clustering (**PRNF** in short).

TABLE 2. Clustering accuracy on the 4 datasets.

Data Sets	Kmeans	NMF	PNMF	PRNF	AGPNMF	REPnMF
ORL	0.5063	0.5093	0.5190	0.5650	0.5635	<b>0.5853</b>
YALE	0.3879	0.3888	0.4073	0.4324	0.4318	<b>0.4485</b>
UMIST	0.3976	0.4044	0.4048	0.5043	0.5037	<b>0.5461</b>
Chang'e 3	0.4138	0.4117	0.4213	0.4268	0.4352	<b>0.4575</b>

TABLE 3. Normalized mutual information on the 4 datasets.

Data Sets	Kmeans	NMF	PNMF	PRNF	AGPNMF	REPnMF
ORL	0.7020	0.7111	0.7152	0.7549	0.7523	<b>0.7698</b>
YALE	0.4413	0.4431	0.4538	0.4781	0.4758	<b>0.4854</b>
UMIST	0.5845	0.5947	0.5953	0.6692	0.6637	<b>0.6955</b>
Chang'e 3	0.4502	0.4511	0.4587	0.4731	0.4862	<b>0.4953</b>

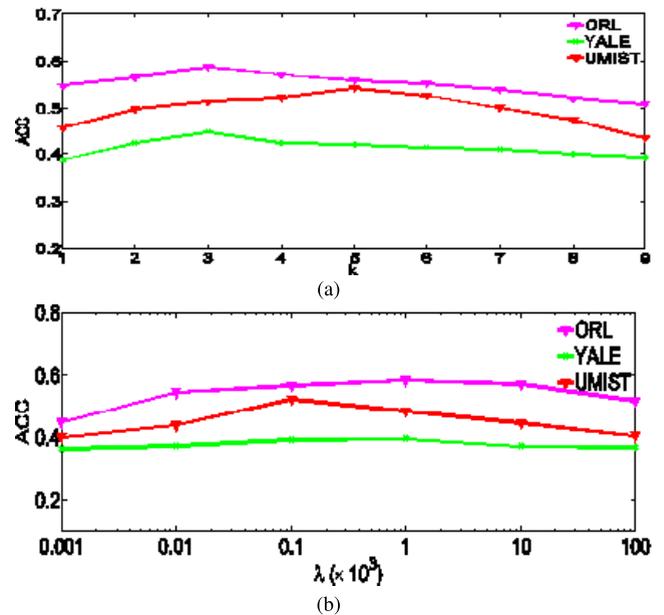


FIGURE 1. Parameters Sensitivity Analysis: (a) nearest neighbour  $k$  (b) ensemble manifold regularization  $\lambda$ .

3) PERFORMANCE COMPARISON

Table 2 is the illustration of clustering accuracy that confirms the accuracy of all the procedures operated on four data sets, whereas Table 3 shows the normalized mutual information. The observations are analyzed as follows. Firstly, methods that made use of parted-based representation outperformed Kmeans. These results are consistent to preceding work on NMF. Secondly, PNMf-based methods outperformed NMF methods, which indicated that learning spatially localized, sparse and part-based subspace representation can improve discriminative structures of latent space. Thirdly, AGPNMF and PRNF performed better than PNMf. The reason behind this is aforesaid in Section II, which is that AGPNMF and PRNF consider native manifold structure of data that ultimately enrich the performance. Finally, REPnMF left behind all the baseline methods in almost all cases. This is due to the fact that REPnMF integrates the ensemble manifold into PNMf, which can be better to capture the intrinsic information of the dataset. Furthermore, REPnMF impose  $\ell_1$  norm

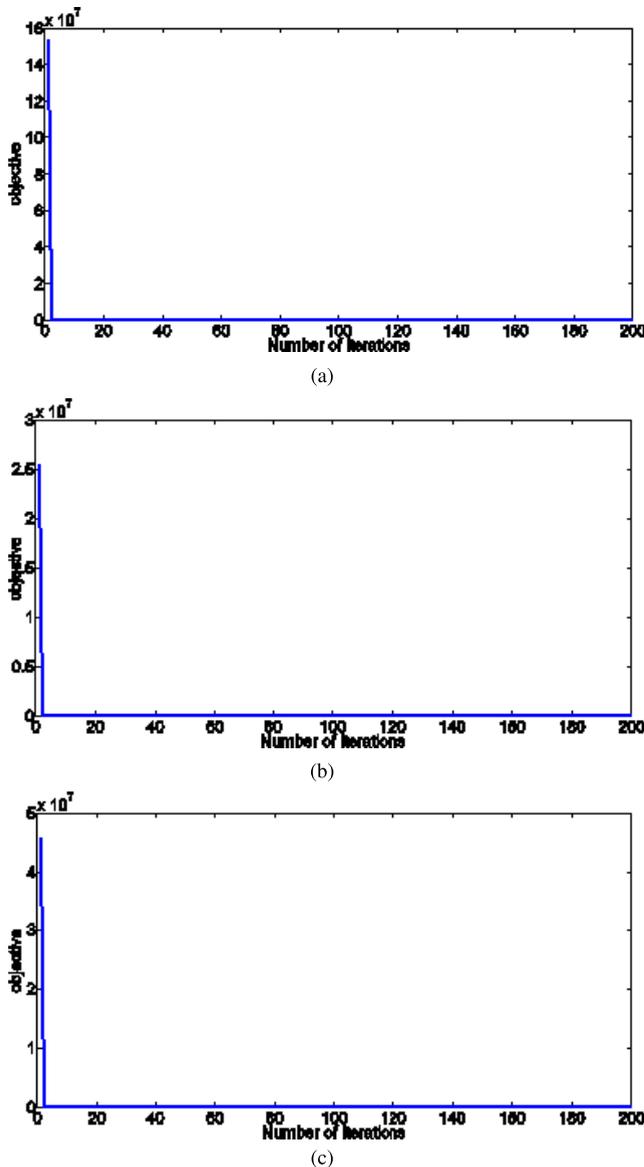


FIGURE 2. Convergence Analysis: (a) ORL, (b) YALE, (c) UMIST.

sparsity constraint on noise matrix which can effectively weaken the noise.

#### 4) PARAMETER STUDY

There are four parameters in REPNMF which are as follow:  $\lambda$ ,  $\beta$ ,  $\alpha$  and the number of nearest neighbours  $k$ .  $\lambda$  measures the importance of the ensemble manifold regularized part of REPNMF,  $\beta$  administers the amount of sparsity of the noise matrices, and  $\alpha$  is used to control the smoothness of weight vector  $\mu$ . In this paper the impact of these parameters on REPNMF is investigated by changing one parameter while others parameters remain unchanged. REPNMF was executed 10 times for each set individually along with record of average performance.

The outcomes of REPNMF in relation to  $k$  are outlined in Fig 1 (a).  $k$  tracks the number of nearest neighbors.  $k$  must be not too small, otherwise the native structure will not be entirely manipulated. If  $k$  is too large, the nearest

neighbor graph may attach two samples with dissimilar labels. REPNMF has a tendency to perform better when  $k \in [2, 6]$ .

The performance of REPNMF with respect to  $\lambda$  is shown in Fig 1 (b).  $\lambda$  measures the importance of the ensemble manifold regularization. A small  $\lambda$  cannot affect the objective function because the geometrical structure of the data cannot be fully exploited. On the contrary, a large  $\lambda$  might dominate the objective function and lead to a trivial solution. REPNMF shows superior performance when  $\lambda = 100$  for YALE and  $\lambda = 1000$  for ORL and UMIST.

It is interesting to note that the performance of REPNMF is stable when  $\beta \in [0.01, 0.1]$  and  $\alpha \in [1000, 10000]$ .

The optimization method for REPNMF unties the subproblems for  $\mathbf{U}$  and  $\mathbf{E}$  iteratively to obtain the local minimum of (14). We analyze empirical convergence properties on three datasets. Fig 2(a), Fig 2(b) and Fig 2(c) presents the objective function values against the number of iterations computed for ORL, YALE and UMIST, respectively. Results indicated that at the beginning, the objective function value decreased sharply and the performance increased dramatically. The optimization procedure is quickly convergent around dozens of iterations for 3 datasets.

## V. CONCLUSIONS

In this work, we proposed robust ensemble manifold projective non-negative matrix factorization (REPNMF), a novel non-negative latent representation learning algorithm for image representation. REPNMF attempts to learn a projective subspace of items by exploiting both cleaned data and ensemble manifold. On the one hand, the  $\ell_1$  norm is added to REPNMF as the constraint of sparsity for noise, because it can efficiently weaken the influence of noise. On the other hand, ensemble manifold is incorporated into REPNMF, which could be able to better model the locally geometrical structure of the data. To optimize the objective function, we provided an alternative method to learn the variable and presented the convergence proof of the optimization scheme. Finally, the experimental results on three face datasets (ORL, YALE, UMIST) and one science-quality dataset (Chang'e 3 data: Rover Panoramic Camera images [14]) have demonstrated that the novel method obtains more competitive performance compared to some alternatives.

## 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. 788–791, Oct. 1999.
- [2] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.
- [3] Z. Yuan and E. Oja, "Projective nonnegative matrix factorization for image compression and feature extraction," in *Proc. Scand. Conf. Image Anal.*, 2005, pp. 333–342.
- [4] P. Luo, J. Peng, Z. Guan, and J. Fan, "Dual regularized multi-view non-negative matrix factorization for clustering," *Neurocomputing*, vol. 294, pp. 1–11, Jun. 2018.
- [5] Y.-L. Gao, M.-X. Hou, J.-X. Liu, and X.-Z. Kong, "An integrated graph regularized non-negative matrix factorization model for gene co-expression network analysis," *IEEE Access*, vol. 7, pp. 126594–126602, 2019.

[6] X. He, Q. Wang, and X. Li, "Robust adaptive graph regularized non-negative matrix factorization," *IEEE Access*, vol. 7, pp. 83101–83110, 2019.

[7] Z. Shu, X.-J. Wu, P. Huang, H. Fan, Z. Liu, and F. Ye, "Multiple graph regularized concept factorization with adaptive weights," *IEEE Access*, vol. 6, pp. 64938–64945, 2018.

[8] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Proc. Int. Conf. Neural Inf. Process. Syst., Natural Synth.*, 2001, pp. 585–591.

[9] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, p. 2323, 2000.

[10] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2010.

[11] X. Pei, T. Wu, and C. Chen, "Automated graph regularized projective nonnegative matrix factorization for document clustering," *IEEE Trans. Cybern.*, vol. 44, no. 10, pp. 1821–1831, Oct. 2014.

[12] Y. Lu, Z. Lai, Y. Xu, J. You, X. Li, and C. Yuan, "Projective robust nonnegative factorization," *Inf. Sci.*, vols. 364–365, pp. 16–32, Oct. 2016.

[13] B. Geng, D. Tao, C. Xu, L. Yang, and X.-S. Hua, "Ensemble manifold regularization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 6, pp. 1227–1233, Jun. 2012.

[14] China National Space Administration, Chinese Academy of Sciences, The Science and Application Center for Moon and Deepspace Exploration. *Chang'e 3 Data: Rover Panoramic Camera*. Accessed: Jan. 13, 2014. [Online]. Available: <http://planetary.s3.amazonaws.com/data/change3/pcam.html>

[15] Y. Zhang, S. Y. Ma, X. Zhang, L. Li, W. H. Ip, and K. L. Yung, "EDGAN: Motion deblurring algorithm based on enhanced generative adversarial networks," *J. Supercomput.*, vol. 76, pp. 1–16, Feb. 2020.

[16] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. NIPS*, 2000, pp. 556–562.

[17] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis," *J. ACM*, vol. 58, no. 3, p. 11, 2011.

[18] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from examples," *J. Mach. Learn. Res.*, vol. 7, no. 1, pp. 2399–2434, 2004.

[19] E. T. Hale, W. Yin, and Y. Zhang, "Fixed-point continuation for  $\ell_1$ -minimization: Methodology and convergence," *SIAM J. Optim.*, vol. 19, no. 3, pp. 1107–1130, 2008.

[20] Polyu. (2015). *Life on Mars? Award-Wining Polyu Device Could Dig Up the Answer*. [Online]. Available: <https://www.polyu.edu.hk/openingminds/en/story.php?sid=3>

[21] Polyu. (2017). *Polyu 70th Anniversary*. [Online]. Available: [https://www.polyu.edu.hk/cpa/70thanniversary/memories\\_achievement10.htm](https://www.polyu.edu.hk/cpa/70thanniversary/memories_achievement10.htm)

[22] *The Hong Kong Polytechnic University-the Space Exploration Journey*. Accessed: Mar. 2019. [Online]. Available: [https://www.polyu.edu.hk/web/filemanager/en/content\\_155/1960/Appendix\\_Milestones\\_SpaceProjects\\_.pdf](https://www.polyu.edu.hk/web/filemanager/en/content_155/1960/Appendix_Milestones_SpaceProjects_.pdf)



**PENG LUO** received the B.S. degree from the National University of Defence Technology, Hunan, China, in 2003, the M.S. degree from Xiangtan University, Hunan, China, in 2007, and the Ph.D. degree from North-West University of China, in 2017. He is currently an Associate Professor with the School of Mathematics and Statistics, Hunan Technology and Business University. His research interests include machine learning and data mining. He serves as a Journal Reviewer

for the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS and *Knowledge-Based Systems*.



**XILONG QU** received the Ph.D. degree from Southwest Jiaotong University. He finished his postdoctoral study with the Postdoctoral Research Center of Computer Science and Technology, South China University of Technology. He is currently the Dean of the School of Information Technology and Management, Hunan University of Finance and Economics. He is also the master's degree student advisor for both Xiangtan University and the Hunan Institute of Engineering. He is the young Core Teacher of Hunan and a member of the "121" talents program of Hunan. He joined four projects from the National High Technology Research and Development Program of China (863 Program) and the National Natural Science Foundation of China. He also led the National Natural Science Foundation of Hunan, the Research Foundation of Education Bureau of Hunan Province, China, and the National Science Foundation for Postdoctoral Scientists of China on 3, 1, and 1 occasion, respectively. He has published four books and 35 articles (five articles are indexed by SCI, and the rest are indexed by EI). He has also been invited as a chair and keynote speaker at numerous conferences and forums in China.



**LINA TAN** received the B.S. degree in computer science and the M.S. degree in computer software and theory from Xiangtan University, Hunan, China, in 2003 and 2006, respectively, and the Ph.D. degree in computer application technology from Hunan University, Hunan, China, in 2012. Her research interests include information security, digital watermarking, image processing, and pattern recognition.



**XIAOLIANG XIE** received the M.Sc. degree in mathematics from Hunan Normal University, Changsha, China, in 1998, and the Ph.D. degree in mathematics from Central South University, Changsha, in 2010. He is currently a Professor with the School of Mathematics and Statistics, Hunan Technology and Business University. His research interests include optimization theory, rough set theory, and information systems.



**WEIJIN JIANG** received the M.S. degree in computer science from the National University of Defence Technology, Hunan, China, in 2004, and the Ph.D. degree in computer science from the Wuhan University of Technology, Hubei, China, in 2013. He is currently a Professor with the College of Computer and Information Engineering, Hunan University of Commerce, Hunan. His research interests include information security, distributed computing, mobile Internet computing,

and trust computing. He is a Senior Member of the CCF.

**LIRONG HUANG** received the master's degree from Hunan Normal University. He is currently an Associate Professor with the School of Information Technology and Management, Hunan University of Finance and Economics. He joined four Research projects at the Education Department of Hunan Province. He has published five articles indexed by SCI or EI.



**WAI HUNG IP** (Senior Member, IEEE) received the LL.B. degree (Hons.) from the University of Wolverhampton, the M.Sc. degree in industrial engineering from Cranfield University, the M.B.A. degree from Brunel University London, and the Ph.D. degree from Loughborough University, U.K.

He is currently a Principal Research Fellow with the Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University; the Professor Emeritus and an Adjunct Professor of Mechanical Engineering with the University of Saskatchewan, Canada; and an Industrial Fellow of the Warwick Manufacturing Group, University of Warwick. He is a member of the Hong Kong Institution of Engineering.



**KAI LEUNG YUNG** received the B.Sc. degree in electronic engineering from Brighton University, in 1975, the M.Sc. and DIC degrees in automatic control systems from the Imperial College of Science, Technology and Medicine, University of London, in 1976, and the Ph.D. degree in micro-processor applications in process control from Plymouth University, U.K., in 1985. He became a Chartered Engineer (C.Eng., MIEE), in 1982.

He is currently an Associate Head and the Chair Professor of the Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University. His research interests include space systems, mechatronics, robotics and automaton, AI, and big data.

...