Robust Local Coordinate Non-negative Matrix Factorization via Maximum Correntropy Criteria

Qing Liao†, Xiang Zhang‡, Naiyang Guan‡, Qian Zhang†
† Dept. of Computer Science & Engineering, The Hong Kong University of Science & Technology
‡ College of Computer, National University of Defense Technology
qnature@ust.hk, zhangxiang_43@aliyun.com, ny_guan@nudt.edu.cn, qzhang@ust.hk

***Abstract*—Non-negative matric factorization (NMF) decomposes a given data matrix X into the product of two lower dimensional non-negative matrices U and V. It has been widely applied in pattern recognition and computer vision because of its simplicity and effectiveness. However, existing NMF methods often fail to learn the sparse representation on high-dimensional dataset, especially when some examples are heavily corrupted. In this paper, we propose a robust local coordinate NMF method (RLCNMF) by using the maximum correntropy criteria to overcome such deficiency. Particularly, RLCNMF induces sparse coefficients by imposing the local coordinate constraint over both factors. To solve RLCNMF, we developed a multiplicative update rules and theoretically proved its convergence. Experimental results on popular image datasets verify the effectiveness of RLCNMF comparing with the representative methods.**

***Index Terms*—Local coordinate, Non-negative matrix factorization (NMF), Robust NMF.**

I. Introduction

Many practical tasks suffer from the "curse of dimensions" challenge [1], especially in computer vision [2], pattern recognition [3] [4] and biological tasks [5] [6] [7]. Hence it is demanded to reduce the dimensions of data at the data preprocessing stage. Recently, non-negative matrix factorization (NMF) technique has been proven to be a powerful dimension reduction method, which decomposes a given matrix into to the product of two lower dimensional non-negative matrices.

Since NMF naturally preserves the non-negativity property of real-world dataset, such as pixel values and video frames, in the low-dimensional space, it has been widely used in many data mining [8] [9] and computer vision tasks [2] [10]. For example, Cai *et al.* [11] proposed graph regularized NMF (GNMF) to preserve the geometric structure of the dataset to enhance the image representative capacity of the learned lower dimensional space. Guan *et al.* [12] proposed a manifold regularized discriminative NMF (MD-NMF) to preserves both local geometry and label information of samples simultaneously to boost NMF for classification. Shen and Si [13] proposed a multiple manifolds method (MM-NMF) to model the intrinsic geometrical structure of data on multiple manifolds. Wang *et al.* [14] applied NMF for natural image matting by removing the confused boundaries of images.

Although conventional NMF method has been shown effective in practices, they are sensitive to noisy datasets because the traditional loss function including Frobenius norm and Kullback-Leibler (KL) divergence cannot handle outliers. In order to overcome this deficiency, Zhang *et al.* [15] and Shen *et al.* [16] proposed the sparse robust NMF (SR-NMF) method to decompose the original matrix into a sparse component and a low-rank component, such that the former one can model the outliers and the later one contain the static background. Kong *et al.* [17] [18] proposed the $L_{2,1}$-NMF method which replaces the traditional loss function by $L_{2,1}$-norm to penalize the reconstruction loss for removing outliers from data. Du *et al.* [19] proposed the CIM-NMF method by substituting the squared error on each entry. CIM-NMF is very useful for the non-Gaussian noise and outliers because outliers have light weights in the objective function. Guan *et al.* [20] proposed Manhattan NMF (MahNMF) to filtering out outliers in a non-negative low-rank and sparse matrix decomposition sense. However, both traditional NMF and its variants cannot guarantee the decomposition results of NMF to be sparse in theory.

This point often induces performance degeneration in clustering. To address this problem, Hoyer *et al.* [21] explicitly incorporated sparseness constraints over both factors to present the sparse NMF method (SNMF). Li *et al.* also proposed local NMF (LNMF, [3]) which imposes sparse constraint to guarantee the representation of traditional NMF to be sparse. Besides, Yuan *et al.* [22] developed the projective NMF (PNMF) to induce parts-based representation by implicitly enforcing the orthogonal constraint over the basis. However, since they cannot explicitly guarantee the learned basis, i.e., the clustering center, to be close to the original data, the learned coefficients fail to reveal the clustering memberships of samples. To address this issue, Chen *et al.* [23] proposed the non-negative local coordinate factorization (NLCF) to guarantee its sparseness via the local coordinate constraint. But it often yields the trivial basis. To overcome this deficiency, Liu *et al.* [24] developed the local coordinate concept factorization method (LCF) to learn the effective basis to induce sparse coefficients. Since LCF requires that the learned basis vector to be close to the original data points, each data point can be approximated by a linear combination of as few basis vectors as possible, but the learned basis is often sensitive to outliers.

This paper proposes a based local coordinate NMF method (RLCNMF) to induce sparse representation and robustness in NMF. Particularly, RLCNMF induces sparse coefficients by imposing the local coordinate constraint over both factors. The learned sparse coefficients encourage the energy to be distributed over the whole low-dimensional space. Inspired by

[17] [18] [19], RLCNMF incorporates the correntropy induced metric to measure the reconstruction error rather than the traditional loss function including Frobenius norm or Kullback-Leibler (KL) divergence. To solve RLCNMF, we developed a multiplicative update rule to optimize RLCNMF and theoretically proved its convergence. Experiments of clustering on popular face image datasets suggest the effectiveness of RLCNMF.

The rest of this paper is organized as follows. Section II briefly reviews related works on NMF and its variants. Section III proposed the RLCNMF method and the multiplicative rule (MUR) for optimizing it. Section IV evaluates the effectiveness of RLCNMF by using experiments, and Section V summaries this paper.

II. Related Works

This section briefly reviews conventional non-negative matrix factorization (NMF, [25] [26]) and its variants that are mostly related to this work, including $L_{2,1}$-NMF [17] [18] and CIM-NMF [19].

A. Non-negative Matrix Factorization

Given any non-negative matrix, i.e. $X \in R_{+}^{m\times n}$, NMF [25] approximates X by the product of two lower dimensional non-negative matrices, i.e., $U \in R_{+}^{m\times k}$ and $V \in R_{+}^{n\times k}$, by minimizing the distance between X and UV^T, i.e.,

$$\min_{U\in R_{+}^{m\times k}, V\in R_{+}^{n\times k}} \|X - UV^T\|_F^2, \tag{1}$$

where U and V are two factor matrices, $\|\cdot\|_F$ denotes Frobenius norm, and k denotes the reduced dimensionality which satisfies $k \ll \min\{m, n\}$. The Frobenius norm (1) can be replaced by using Kullback-Leibler (KL) divergence. Since both Frobenius norm and KL-divergence can be dominated by those errors of large magnitudes, traditional NMF methods are non-robust to outliers.

B. NMF Extensions

To enhance the robustness of NMF, Kong *et al.* [17] [18] proposed $L_{2,1}$-NMF whichs is defined as follows

$$\min_{U\in R_{+}^{m\times k}, V\in R_{+}^{n\times k}} \|X - UV^T\|_{2,1}^2, \tag{2}$$

where the Frobenius norm based loss function has been substituted by the $L_{2,1}$-norm, i.e., $\|A\|_{2,1} = \sum_{i=1}^{n}\sqrt{\sum_{j=1}^{m} A_{ji}^2}$. Since the square operator $\|\cdot\|_{2,1}$ in reduces the components occupied by the large magnitude of errors in the loss function, the corrupted samples never dominate the objective function (2). In this sense, $L_{2,1}$-NMF performs more robustly than NMF.

Du *et al.* [19] proposed CIM-NMF based on correntropy induced metric (CIM) because correntropy has been shown robust in information theoretic learning (ITL) to process non-Gaussian and impulsive noise. CIM-NMF introduce CIM to measure the loss of the factorization, i.e.

$$\min_{U\geq 0, V\geq 0} \sum_{i=1}^{N}\sum_{j=1}^{M}(1 - g(X_{ij} - \sum_{k=1}^{K} U_{ik}V_{jk}, \sigma)), \tag{3}$$

where $g(e, \sigma)$ is the Gaussian kernel $\frac{1}{\sqrt{2\pi}\sigma}\exp(\frac{-e^2}{2\sigma^2})$. Since CIM-NMF shrinks the weight of large outliers, it is quite effective for filtering out non-Gaussian noises.

Both $L_{2,1}$-NMF and CIM-NMF are more robust than NMF on corrupted datasets because their loss functions are less sensitive to outliers. However, both NMF and these extensions does not induce sparse representations.

III. Robust Local Coordinate NMF

In this paper, we proposed robust local coordinate NMF method (RLCNMF) by incorporating local coordinate regularization to correntropy induced metric (CIM-NMF) [17]. Since CIM term is mainly determined by small errors, it is more robust for non-Gaussian with large outliers. The local coordinate regularization can guarantee the sparsity of the learned coefficients on noisy datasets especially heavily corrupted ones.

A. The Proposed Model

Given a data matrix $X = [x_1, \ldots, x_n] \in R^{m\times n}$, non-negative local coordinate factorization (NLCF) incorporates the local coordinate into NMF to learn the sparse coefficients:

$$\min_{U,V\geq 0} \|X - UV\|_F^2 + \lambda \sum_{a=1}^{n}\sum_{b=1}^{k} |v_{ba}|\|x_a - u_b\|_2^2, \tag{4}$$

where $\lambda \geq 0$ signifies the regularization parameter, v_{ji} and u_j denote the j-th row and i-th column entry of and the j-th entry of U, respectively. However, the model (4) is vulnerable to the outliers because it assumes that the noise obeys Gaussian distribution. This assumption often violates the practical situation. To address this problem, we propose a robust local coordinate NMF method (RLCNMF) which incorporate the correntropy induced metric to measure the reconstruction loss in (4). Therefore, we obtain

$$\min_{U,V\geq 0} \sum_{ij}(1 - g(X_{ij} - \sum_{k} U_{ik}V_{kj}, \sigma)) + \lambda \sum_{a=1}^{n}\sum_{b=1}^{k} |v_{ba}|\|x_a - u_b\|_2^2, \tag{5}$$

where $g(\cdot)$ signifies the Gaussian kernel function and σ the predefined parameter. Based on correntropy induced metric (CIM) [27], CIM can remove the effect of large outliers in theory, and has been applied in NMF [17], i.e., CIM-NMF. It also shows some useful properties which avoid inducing the trivial solution in (5) while this point often happens in the model (4). This advantage significantly boosts RLCNMF in clustering. Moreover, since RLCNMF learns sparse coefficients and encourages the energy to be distributed over the specific basis, i.e., the clustering center. This property of RLCNMF facilitates finding out the true clustering tasks.

According to Figure 1(a), the learned basis by NMF is so unclear that it fails to represent the true cluster centers of samples. In contrast, the basis of RLCNMF approximates the original data to reveal the cluster memberships of samples, and thus has strong representation capacity. Benefit from this property, RLCNMF induces as more sparse coefficients

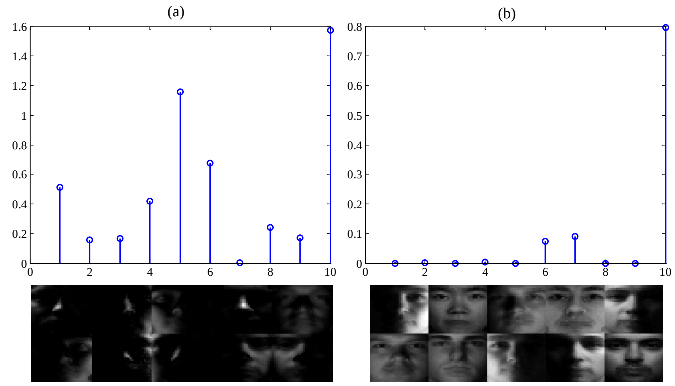

Fig. 1. The coefficient of an image and basis learned by (a) NMF and (b) RLCNMF on YaleB dataset. The first row shows the learned coefficient while the second demonstrates the learned basis. For the illustrated image, the maximum entry of the learned coefficient by RLCNMF corresponds to the true clustering center of this image while the other entries have very small values.

as possible close to their true cluster identity. These merits of RLCNMF have been intuitively illustrated in Figure 1. Interestingly, RLCNMF can be easily extended for multi-modal learning in the similar ways as the works in [28] [29] [30].

B. Optimization Algorithm

The model (5) is not jointly convex over U and V, and thus it is impossible to obtain the global solution. Luckily, it is convex with respect to U with V fixed, and vice versa. Thus, we developed a multiplicative update rule (MUR) to optimize RLCNMF by alternatively updating both factor matrices.

The above optimization problem of RLCNMF is equivalent to minimizing the following augmented objective function in an enlarged parameter space

$$\min_{U,V\geq 0} \sum_{ij} W_{ij}(X_{ij} - \sum_{k} U_{ik}V_{kj})^2 + \phi(W_{ij}) + \lambda \sum_{a=1}^{n}\sum_{b=1}^{k} |v_{ba}| \|x_a - u_b\|_2^2, \tag{6}$$

where $\phi(W_{ij})$ denotes the conjugate function of NMF and W_{ij} indicates the corresponding auxiliary variable.

We optimize (6) with respect to one variable with the other fixed as follows:

Computation of W: When U and V are fixed, the optimization problem with respect to W is:

$$W_{ij} = \exp(-\frac{(X_{ij} - \sum_{k=1}^{K} U_{ik}V_{kj})^2}{2\sigma^2}). \tag{7}$$

Computation of U: Given V, we can yield the derivative of (6) with respect to U,

$$\Delta_U = -(W \otimes X)V^T + (W \otimes (UV))V^T + \lambda UH - \lambda XV^T. \tag{8}$$

By setting the derivative of U to zero, we obtain:

$$U = U \otimes \frac{(W \otimes X)V^T + \lambda XV^T}{(W \otimes (UV))V^T + \lambda UH}, \tag{9}$$

Algorithm 1 Optimization algorithm for RLCNMF

Input:Examples $X \in R^{m\times n}$, r number of cluster r, parameter $\lambda > 0$

Output: V and U

1: Initialize U, V with random values.
2: **repeat**
3: Construct W via (7).
4: Update U via (9).
5: Update V via (11).
6: **until** {Convergence.}

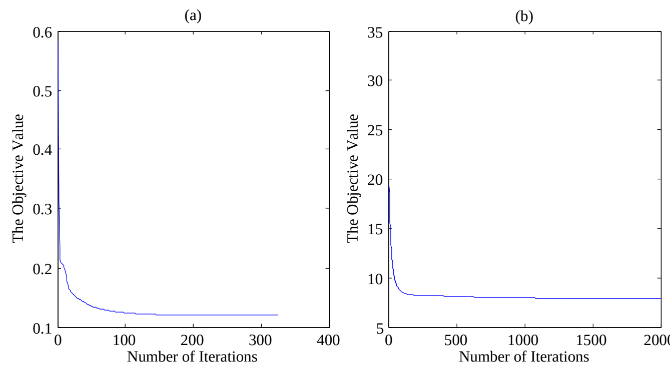

Fig. 2. The convergence curves of RLCNMF on (a) Yale and (b) YaleB datasets, respectively.

where H is diagonal matrix whose entries are row sums of V.

Computation of V: Given U, we can yield the derivative of (6) with respect to V,

$$\Delta_V = -2U^T(W \otimes X) + 2U^T(W \otimes (UV)) + \lambda(C + D) - 2\lambda U^T X. \tag{10}$$

By (8) and setting the derivative of V to zero, we can obtain:

$$V = V \otimes \frac{2U^T(W \otimes X) + 2\lambda U^T X}{2U^T(W \otimes (UV)) + \lambda(C + D)}, \tag{11}$$

where column vectors $c = \mathrm{diag}(X^T X) \in R^n$ and $d = \mathrm{diag}(U^T U) \in R^k$. Let $C = (c, \cdots, c)^T \in R^{k\times n}$, $D = (d, \cdots, d) \in R^{k\times n}$.

For completeness, we summarize the optimization procedure of RLCNMF into **Algorithm 1**. Figure 2 shows the convergence curve of RLCNMF on Yale [27] and YaleB [31] dataset, respectively. We leave the convergent proof of RLCNMF in **Appendix**. The time cost of **Algorithm 1** lies in the line 3, line 4 and line 5, respectively. The line 4 takes $O(mnr)$. The line 5 and 6 take $O(mn + mr + mnr)$ and $O(mn + nr + mnr)$, respectively. Thus, the total time complexity of RLCNMF is $O(mn + mr + nr + mnr)$.

IV. Experiments

We verify the effectiveness by comparing the clustering performance of RLCNMF to the representative methods including NMF [25] [26], $L_{2,1}$-NMF [17] [18], PNMF [22], CIM-NMF [19] and Kmeans on the Yale [31], YaleB [32] and AR [33] datasets. We randomly selected $r = 2, \ldots, 10$ individuals

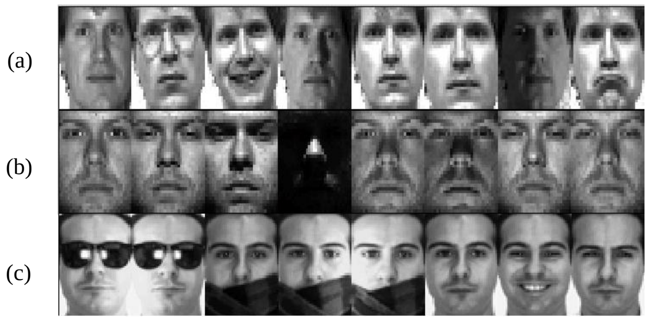

Fig. 3. The image instances of (a) Yale, (b) YaleB and (c) AR datasets, respectively.

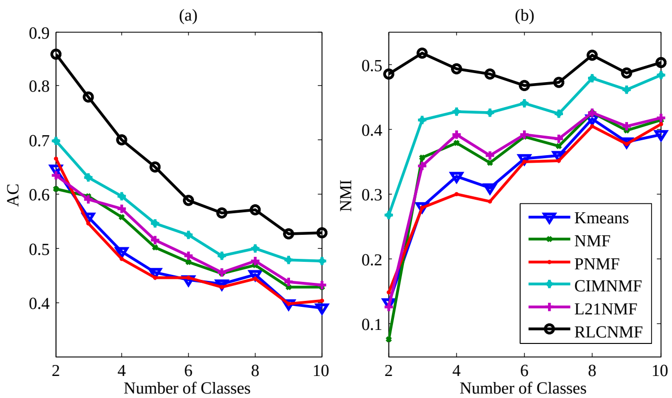

Fig. 4. Average accuracy (AC) and normalized mutual information (NMI) versus different numbers of classes on Yale dataset.

from both datasets and the selected images are clustered by RLCNMF and compared methods. RLCNMF learns the sparse coefficients and thus selects the index of the maximum coefficient values of each sample as the clustering identity. Then, we conducted the experiments 10 trails and evaluate the compared methods based on both average accuracy (AC) and the normalized mutual information (NMI). The details of AC and NMI can be found in [32]. Their image instances are shown in Figure 3.

A. Yale Dataset

The Yale face dataset [31] contains 165 images of 15 individuals. Each subject has 11 different images under various facial expressions and lighting conditions. All images are cropped to 32x32-pixel grayscale images and reshaped into a 1024-dimensional vector. We set the parameter $\lambda = 0.001$ for RLCNMF on this dataset. The compared methods involve no parameters.

Figure 4 shows that RLCNMF consistently outperforms the compared methods in terms of clustering accuracy and normalized mutual information (NMI) under different number of classes [34].

B. YaleB Dataset

The YaleB face dataset [32] is an extension of the Yale face database. By contrast with Yale dataset, it is more tough to perform clustering tasks on YaleB dataset. There are totally 2424 face images of 38 individuals. Each individual has 59 images at least and 64 images at most under different illumination conditions. All images are still cropped to 32x32-pixel grayscale images and reshaped into a 1024-dimensional vector. We set the parameter $\lambda = 0.005$ for RLCNMF on this dataset. The compared methods involve none of parameters. Figure 5 reports the clustering accuracy and normalized mutual information (NMI) of the compared methods on YaleB dataset. This also implies that RLCNMF is superior to the representative methods in quantities.

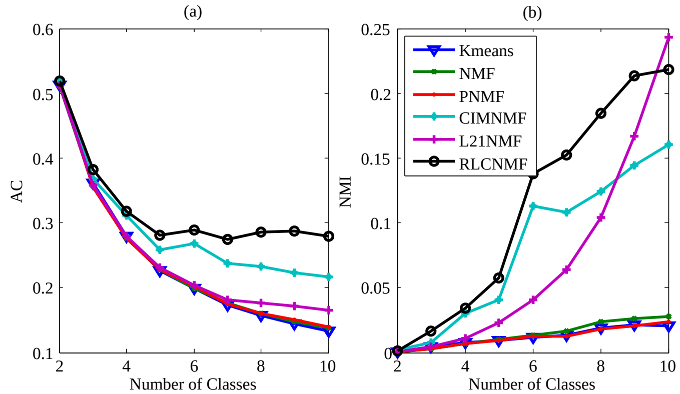

Fig. 5. Average accuracy (AC) and normalized mutual information (NMI) versus different numbers of classes on YaleB dataset.

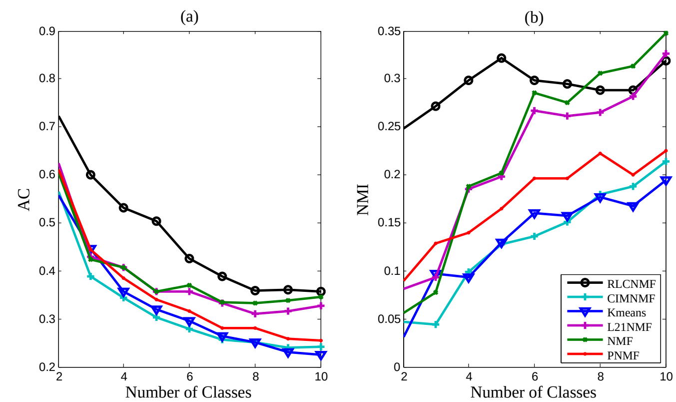

Fig. 6. Average accuracy (AC) and normalized mutual information (NMI) versus different numbers of classes on AR dataset.

C. AR Dataset

The AR face dataset [33] contains of 2600 frontal images of 100 individuals. Each individual has 26 images with different facial expressions, illumination conditions and facial disguises, such as sun glasses or scarf. All images are still cropped to 55x40-pixel grayscale images and then reshaped into a 2,200-dimensional vector. We set the parameter $\lambda = 10$ for RLCNMF on this dataset. The compared methods involve none of parameters. Figure 6 reports the clustering accuracy and normalized mutual information (NMI) of the compared methods on AR dataset. It also shows that RLCNMF is superior to the representative methods in quantities.

V. Conclusion

This paper proposes a correntropy induced metric local coordinate NMF (RLCNMF) to induce the sparse coefficients under the real noise datasets. RLCNMF incorporates the local coordinate constraint over both the basis and coefficients to learn sparse representation and further enhances the representation ability of NMF. Moreover, RLCNMF utilizes the correntropy induced metric as the loss function to be robust to the outliers. Experimental results of image clustering on three popular face image datasets verify the effectiveness of RLCNMF in quantities.

Appendix

In this subsection, we use the auxiliary function technique to prove the convergence of RLCNMF. Let objective function (5) the be $F(U,V)$.

Lemma 1. If there exists an function G for $F(x)$ which satisfies $G(x,x') \geq F(x)$ and $G(x,x) = F(x)$, then we call it an auxiliary function, and F is non-increasing under the update rule

$$x^{t+1} = \arg\min_{x} G(x,x'). \tag{12}$$

Let $J(U) = F_u(U,V)$ denote the function with over U with V fixed, and $J(V) = F_v(U,V)$ denote the function with respect to V with U fixed.

Lemma 2. The following function is the auxiliary function of $J_{u_{ab}}$:

$$\begin{aligned} G(u,u_{ab}^t) &= J_{u_{ab}}(u_{ab}^t) + J'_{u_{ab}}(u_{ab}^t)(u - u_{ab}^t) \\ &+ \frac{(W \otimes (UV)V^T)_{ab} + \lambda \sum_{r=1}^{n}(UH)_{ab}}{u_{ab}^t}. \end{aligned} \tag{13}$$

where $J'_{u_{ab}}$ is the first order derivative with respect to U.

Proof. Obviously, $G(u,u) = J_{u_{ab}}(u)$. The Taylor series expansion of $J_{u_{ab}}$ is

$$J_{u_{ab}}(u) = J_{u_{ab}}(u_{ab}^t) + J'_{u_{ab}}(u - u_{ab}^t) + \frac{1}{2}J''_{u_{ab}}(u - u_{ab}^t)^2, \tag{14}$$

and we have

$$J''_{u_{ab}} = 2(W(V \otimes V)^T)_{ab} + 2\lambda \sum_{r=1}^{n}(H_r)_{bb} \tag{15}$$

with (13) to find that $G(u,u_{ab}^t) \geq F_{ab}(u)$ is equivalent to

$$\begin{aligned} &(W \otimes (UV)V^T)_{ab} + \lambda \sum_{r=1}^{n}(UH_r)_{ab} \\ &\geq ((W(V \otimes V)^T)_{ab} + \lambda \sum_{r=1}^{n}(H_r)_{bb})u_{ab}^t. \end{aligned} \tag{16}$$

We have

$$\begin{aligned} (W \otimes (UV)V^T)_{ab} &= \sum_k w_{ak}(UV)_{ak}v_{bk} \\ \geq \sum_k w_{ak}v_{bk}v_{bk}u_{ab} &\geq ((W(V \otimes V)^T)_{ab}u_{ab}^t \end{aligned} \tag{17}$$

and

$$\lambda \sum_{r=1}^{n}(UH_r)_{ab} = \lambda \sum_{r=1}^{n}(H_r)_{bb}u_{ab}^t \tag{18}$$

So we have the $G(u,u_{ab}^t) \geq J_{u_{ab}}(u)$.

According to **Lemma 1**, $F_u(\arg\min_u G(U,U'),V) \leq F_U(U,V)$. Let $\dfrac{\partial G(U,U')}{\partial u_{ab}} = 0$, we have the following update rule

$$U = U \otimes \frac{(W \otimes X)V^T + \lambda XV^T}{(W \otimes (UV))V^T + \lambda UH}. \tag{19}$$

Lemma 3. Let the auxiliary function for $J_{v_{ab}}$ be the following form,

$$\begin{aligned} G(v,v_{ab}^t) &= J_{v_{ab}}(v_{ab}^t) + J'_{v_{ab}}(v_{ab}^t)(v - v_{ab}^t) \\ &+ \frac{(U^T(W \otimes UV))_{ab} + \lambda(C+D)_{ab}}{v_{ab}^t}(v - v_{ab}^t)^2, \end{aligned} \tag{20}$$

where $J'_{v_{ab}}$ is the first order derivative with respect to V.

Proof. Obviously, $G(v,v) = J_{v_{ab}}$. Taylor series expansion of $J_{v_{ab}}(v)$, we obtain

$$J_{v_{ij}}(v) = J_{v_{ij}}(v_{ij}^t) + J'_{v_{ij}}(v - v_{ij}^t) + \frac{1}{2}J''_{v_{ij}}(v - v_{ij}^t)^2, \tag{21}$$

and we have

$$J''_{v_{ab}} = ((U \otimes U)^T W)_{ab}, \tag{22}$$

with (20) to find that $G(v,v_{ab}^t) \geq F_{ab}(v)$ is equivalent to

$$\begin{aligned} &(U^T(W \otimes UV))_{ab} + \lambda(C+D)_{ab} \\ &\geq ((U \otimes U)^T W)_{ab}v_{ab}^t, \end{aligned} \tag{23}$$

We have

$$\begin{aligned} (U^T(W \otimes UV))_{ab} &= \sum_k u_{ka}w_{kb}\sum_r u_{kr}v_{rb} \\ &\geq \sum_k u_{ka}u_{ka}w_{kb}v_{ab}^t \geq ((U \otimes U)^T W)_{ab}v_{ab}^t \end{aligned} \tag{24}$$

So we have $G(v,v_{ij}^t) \geq J_{v_{ab}}(v)$.

Like **Lemma 1**, $F_v(U,\arg\min_v G(V,V')) \leq F_v(U,V)$. Let $\dfrac{\partial G(V,V')}{\partial v_{ab}} = 0$, we have the following update rule

$$V = V \otimes \frac{2U^T(W \otimes X) + 2\lambda U^T X}{2U^T(W \otimes (UV)) + \lambda(C+D)}, \tag{25}$$

In summary, we know that $F(U,V)$ is non-increasing under the multiplicative update rules (19) and (25).

References

[1] Richard Ernest and Bellman, *Adaptive control processes: a guided tour*, vol. 4, Princeton university press Princeton, 1961.

[2] Haifeng Liu, Zhaohui Wu, Deng Cai, and Thomas S. Huang, "Constrained nonnegative matrix factorization for image representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1299–1311, 2012.

[3] Stan Z. Li, XinWen Hou, HongJiang Zhang, and QianSheng Cheng, "Learning spatially localized, parts-based representation," in *Proceedings of the 2001 IEEE Computer Vision and Pattern Recognition*, 2001, vol. 1, pp. 207–212.

[4] Alberto D. Pascual-Montano, José María Carazo, Kieko Kochi, Dietrich Lehmann, and Roberto D. Pascual-Marqui, "Nonsmooth nonnegative matrix factorization (nsnmf)," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 403–415, 2006.

[5] Stephen E Palmer, "Hierarchical structure in perceptual representation," *Cognitive psychology*, vol. 9, no. 4, pp. 441–474, 1977.

[6] Nikos K Logothetis and David L Sheinberg, "Visual object recognition," *Annual review of neuroscience*, vol. 19, no. 1, pp. 577–621, 1996.

[7] E Wachsmuth, MW Oram, and DI Perrett, "Recognition of objects and their component parts: responses of single units in the temporal cortex of the macaque," *Cerebral Cortex*, vol. 4, no. 5, pp. 509–522, 1994.

[8] Chao Liu, Hung-chih Yang, Jinliang Fan, Li-Wei He, and Yi-Min Wang, "Distributed nonnegative matrix factorization for web-scale dyadic data analysis on mapreduce," in *Proceedings of the 19th International Conference on World Wide Web*, 2010, pp. 681–690.

[9] Naiyang Guan, Dacheng Tao, Zhigang Luo, and Bo Yuan, "Nenmf: an optimal gradient method for nonnegative matrix factorization," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2882–2898, 2012.

[10] Naiyang Guan, Dacheng Tao, Zhigang Luo, and Bo Yuan, "Non-negative patch alignment framework," *IEEE Transactions on Neural Networks*, vol. 22, no. 8, pp. 1218–1230, 2011.

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

[12] Naiyang Guan, Dacheng Tao, Zhigang Luo, and Bo Yuan, "Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent," *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 2030–2048, 2011.

[13] Bin Shen and Luo Si, "Non-negative matrix factorization clustering on multiple manifolds," in *Proceedings of the 24th AAAI Conference on Artificial Intelligence*, 2010.

[14] Wang Kun, Zhang Nanning, and Liu Weixiang, "Natural image matting with non-negative matrix factorization," in *IEEE International Conference on Image Processing*. IEEE, 2005, vol. 2, pp. II–1186.

[15] Lijun Zhang, Zhengguang Chen, Miao Zheng, and Xiaofei He, "Robust non-negative matrix factorization," *Frontiers of Electrical and Electronic Engineering in China*, vol. 6, no. 2, pp. 192–200, 2011.

[16] Bin Shen, Luo Si, Rongrong Ji, and Baodi Liu, "Robust nonnegative matrix factorization via l_1 norm regularization," *arXiv preprint arXiv:1204.2311*, 2012.

[17] Deguang Kong, Chris Ding, and Heng Huang, "Robust nonnegative matrix factorization using l21-norm," in *Proceedings of the 20th ACM international conference on Information and knowledge management*. ACM, 2011, pp. 673–682.

[18] Chris Ding, Ding Zhou, Xiaofeng He, and Hongyuan Zha, "R1-pca: rotational invariant l 1-norm principal component analysis for robust subspace factorization," in *Proceedings of the 23rd international conference on Machine learning*. ACM, 2006, pp. 281–288.

[19] Liang Du, Xuan Li, and Yi-Dong Shen, "Robust nonnegative matrix factorization via half-quadratic minimization," 2012, pp. 201–210.

[20] Naiyang Guan, Dacheng Tao, Zhigang Luo, and John Shawe-Taylor, "Mahnmf: Manhattan non-negative matrix factorization," *arXiv preprint arXiv:1207.3438*, 2012.

[21] Patrik O Hoyer, "Non-negative matrix factorization with sparseness constraints," *The Journal of Machine Learning Research*, vol. 5, pp. 1457–1469, 2004.

[22] Zhijian Yuan and Erkki Oja, "Projective nonnegative matrix factorization for image compression and feature extraction," in *Proceedings of 14th Scandinavian Conference on Image Analysis (SCIA),*, pp. 333–342. Springer, 2005.

[23] Yan Chen, Jiemi Zhang, Deng Cai, Wei Liu, and Xiaofei He, "Non-negative local coordinate factorization for image representation," *IEEE Transactions on Image Processing*, vol. 22, no. 3, pp. 969–979, 2013.

[24] Haifeng Liu, Zheng Yang, and Zhaohui Wu, "Locality-constrained concept factorization," in *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*. Citeseer, 2011, vol. 22, p. 1378.

[25] Daniel D. Lee and H. Sebastian Seung, "Algorithms for non-negative matrix factorization," in *Neural Information Processing Systems (NIPS)*, 2000, pp. 556–562.

[26] Daniel D Lee and H Sebastian Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[27] Weifeng Liu, Puskal P Pokharel, and José C Príncipe, "Correntropy: properties and applications in non-gaussian signal processing," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5286–5298, 2007.

[28] Jun Yu, Yong Rui, and Dacheng Tao, "Click prediction for web image reranking using multimodal sparse coding," *IEEE Transactions on Image Processing*, vol. 23, no. 5, pp. 2019–2032, 2014.

[29] Jun Yu, Yong Rui, Yuan Yan Tang, and Dacheng Tao, "High-order distance-based multiview stochastic learning in image classification," *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2431–2442, 2014.

[30] Xiang Zhang, Naiyang Guan, Dacheng Tao, Xiaogang Qiu, and Zhigang Luo, "Online multi-modal robust non-negative dictionary learning for visual tracking," *PLoS ONE*, vol. 10, no. 5, 2015.

[31] Peter N. Belhumeur, João P Hespanha, and David Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[32] Athinodoros S. Georghiades, Peter N. Belhumeur, and David Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[33] Aleix M Martinez, "The ar face database," *CVC Technical Report*, vol. 24, 1998.

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