This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2017.2665557, IEEE/ACM Transactions on Computational Biology and Bioinformatics

Regularized Non-negative Matrix Factorization for Identifying Differentially Expressed Genes and Clustering Samples: a Survey

Jin-Xing Liu, *Member, IEEE,* Dong Wang, Ying-Lian Gao, Chun-Hou Zheng, *Member, IEEE,* Yong Xu, *Senior Member, IEEE*, and Jiguo Yu, *Member, IEEE*

Abstract—Non-negative Matrix Factorization (NMF), a classical method for dimensionality reduction, has been applied in many fields. It is based on the idea that negative numbers are physically meaningless in various data-processing tasks. Apart from its contribution to conventional data analysis, the recent overwhelming interest in NMF is due to its newly discovered ability to solve challenging data mining and machine learning problems, especially in relation to gene expression data. This survey paper mainly focuses on research examining the application of NMF to identify differentially expressed genes and to cluster samples, and the main NMF models, properties, principles, and algorithms with its various generalizations, extensions, and modifications are summarized. The experimental results demonstrate the performance of the various NMF algorithms in identifying differentially expressed genes and clustering samples.

Index Terms—Constrained optimization, Gene expression, Multivariate statistics, Non-negative matrix factorization, Pattern clustering

1 Introduction

With the rapid development of sequencing technologies, a large amount of biological information has been stored in the gene expression data. It is important to analyze these data to identify the useful information. Under this background and trend, data mining and machine learning have received a large amount of attention. Dimensionality reduction is an effective representation method. Traditional methods in dimensionality reduction, such as Principal Component Analysis (PCA) [1], present some basis vectors that can be used to approximate the original high-dimensional data. Linear Discriminant Analysis (LDA), which is also called Fisher Linear Discriminant (FLD), is a supervised dimensionality reduction method [2]. Locally Linear Embedding (LLE) [3] and Supervised Locally Linear Embedding (SLLE) are non-linear dimensionality reduction methods [4]. Independent Component Analysis (ICA) [5] is a useful extension of PCA, which has been developed in the context of the blind separation of independent sources from their linear mixtures [6]. Partial Least Squares (PLS) regression is a technique that combines features from generalized PCA and multiple linear regression [7]. A highlighted problem with these algorithms is that there are no constraints on the signs of the elements in the factor matrices. Thus, the basis vectors may have both positive and negative components that cannot be reasonably interpreted because the negative components contradict physical realities. In the real world, many data are always non-negative, for example, gene expression data, image pixel values, chemical compound concentration, and signal intensities. Thus, Non-negative Matrix Factorization (NMF) has also been introduced as a dimensionality reduction method. When the observation data are placed in the columns in a matrix with non-negative elements, NMF seeks to identify a lower rank matrix, the elements of which are also non-negative, to approximate the data matrix. The non-negativity allows the intuitive interpretation as the real understanding of the original data.

Because of the outstanding interpretability that is possible with non-negativity, NMF has been extensively studied in bioinformatics. For example, Kim et al. used sparse NMF for microarray data analysis [8]. Brunet et al. applied NMF to cancer microarray data to elucidate tumor subtypes [9]. Devarajan et al. used NMF as an analyt-

- *J.-X. Liu is with the School of Information Science and Engineering, Qufu Normal University, Rizhao, Shandong, 276826 China, and with the Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, Guangdong 518055, China. E-mail: sdcavell@126.com*
- *D. Wang is with School of Information Science and Engineering, Qufu Normal University, Rizhao, Shandong, 276826 China, and with the State Key Laboratory of Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, Hubei, 430079 China. E-mail: dongwshark@126.com*
- *Y.-L. Gao is with the Library of Qufu Normal University, Qufu Normal University, Rizhao, 276826 China. E-mail: yinliangao@126.com*
- *C.-H. Zheng is with the College of Electrical Engineering and Automation, Anhui University, Hefei, Anhui, 230039 China. E-mail: zhengch99@126.com*
- *Y. Xu is with the Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, and with Key Laboratory of Network Oriented Intelligent Computation, Shenzhen, Guangdong, 518055 China. E-mail: yongxu@ymail.com*
- *J. Yu (Corresponding author) is with the School of Information Science and Engineering, Qufu Normal University, Rizhao, Shandong, 276826 China. E-mail: jiguoyu@sina.com*

xxxx-xxxx/0x/$xx.00 © 200x IEEE Published by the IEEE Computer Society

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2017.2665557, IEEE/ACM Transactions on Computational Biology and Bioinformatics

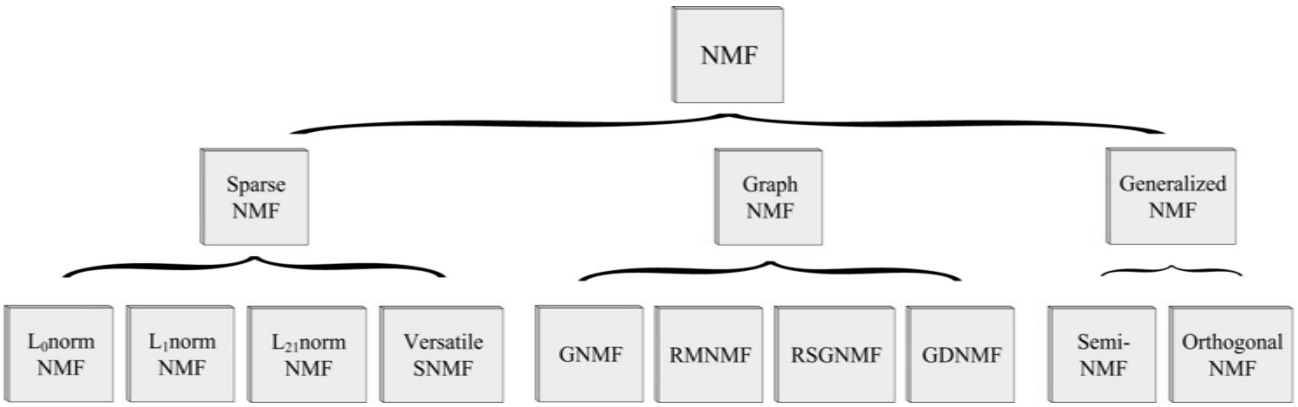


Fig. 1. The hierarchical structure of NMF methods in this paper. The first part is sparse NMF which contains methods with sparse constraint such as L_0 -norm-based NMF, L_1 -norm-based NMF, $L_{2,1}$ -norm-based NMF and Versatile SNMF. The second part is Graph NMF, which contains methods with graph regularized constraints: Graph regularized NMF (GNMF), Robust Manifold NMF (RMNMF), Robust NMF via joint Sparse and Graph regularization model (RSGNMF) and Graph regularized Discriminative NMF (GDNMF). The third part introduces some generalized NMF methods such as Semi-NMF and Orthogonal NMF (Orth-NMF).

ical and interpretive tool in computational biology [10]. Devarajan and Ebrahimi successfully applied NMF as a tool for dimensionality reduction and visualization as well as in kinetic expression profiling to analyze microarray data [11]. Dueck et al. analyzed multi-way clustering of microarray data using sparse methods [12]. Pascual-Montano et al. provided an analytical tool called bio-NMF for simultaneous clustering of genes and samples [13]. Carmona-Saez et al. used NMF to bicluster gene expression data [14]. Kim and Tidor applied NMF as a tool to cluster genes and predict functional cellular relationships in yeast using gene expression data [15], whereas Heger and Holm applied it to the recognition of sequence patterns among related proteins [16]. Frigyesi et al. used NMF to identify clinically relevant tumor subtypes [17]. Numerous contributions have been made to the machine learning and bioinformatics fields by NMF research.

In past years, numerous genomics and proteomics data were presented for biological and biomedical investigation [18]. Concurrently, the identification of differentially expressed genes and clustering samples have become prevalent technologies due to the advances in bioinformatics. As an increasing number of applications are exploited [19-21], gene expression data analysis has become an effective measure for disease diagnosis and treatment, particularly for tumor research. Moreover, to understand the mechanism of cancer cell development, NMF also has been applied to analyze the TCGA (The Cancer Genome Atlas) data [22].

It is well-known that a tumor is a neoplasm or solid lesion that is formed by an abnormal growth of cells. Reliable and precise clustering and identification of the morbigenous genes are essential for effective treatment of tumors. Gene expression data typically contain thousands of genes on each chip (a sample), and the number of samples is much smaller than that of genes, so it is a typical small-sample-size problem [23], i.e., the number of predictor variables m is much greater than that of the available samples n. The particular condition $m >> n$ makes it difficult to most of the standard statistical methods from both analytical and interpretative perspectives. For example, including too many variables may decrease the accuracy for the identification of features and clustering samples, making the cluster rules difficult to establish. The inclusion of irrelevant or noise variables may also degrade the identifying and clustering performance [24]. Thus, NMF algorithms have been used to analyze gene expression data by incorporating the non-negative constraint and thus obtaining the part-based representation as well as correspondingly enhancing the interpretability of the issue. Therefore, a survey is necessary and indispensable. In this paper, we provide an outline of the development of NMF for gene expression analysis and the updated or the comprehensive results [25]. In this paper, we focus on the applications of NMF for identifying differential genes and clustering samples, which differs from other published surveys [26-28]. Furthermore, real data were applied to demonstrate the performance of NMF algorithms. Available NMF methods are divided into three categories, as detailed in Figure 1, which provides a description of the hierarchical structure of NMF methods used in this paper. The first part is sparse NMF, which contains methods with sparse constraints: L_0 -norm-based NMF, L_1 -norm-based NMF, $L_{2,1}$ -norm-based NMF and Versatile SNMF. The second part is Graph NMF, which contains methods with graph regularized constraints: Graph regularized NMF (GNMF), Robust Manifold NMF (RMNMF), Robust NMF via joint Sparse and Graph regularization model (RSGNMF) and Graph regularized Discriminative NMF (GDNMF). The third part introduces some generalized NMF methods such as Semi-NMF and Orthogonal NMF (Orth-NMF). Among them, the codes for NMF L_{21} , RMNMF and RSGNMF were written by our group, and the others are the original existing codes.

The rest of the paper is organized as follows: Section 2 introduces the formulation of basic NMF and efficient algorithms. Section 3 presents the sparse NMF with many different constraints. Models of graph regularized NMF are introduced in Section 4. Section 5 briefly presents other relevant NMF methods. Section 6 describes the results of these algorithms in gene expression data and the TCGA dataset. The conclusions are provided in Section 7.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2017.2665557, IEEE/ACM Transactions on Computational Biology and Bioinformatics

2 BASIC NMF

In this section, we provide a summary of the basic NMF method, with a particular emphasis on the meaning of gene expression analysis and mathematical descriptions.

NMF can be traced back to the 1970s (notes from Golub and Reinsch) [29] and was further studied by Paatero [30]. The work of Lee and Seung [31, 32] has brought much attention to NMF in the data mining and machine learning fields. NMF has been used to tackle many NP-hard problems [33-35]. Algorithmic extensions of NMF have been developed to accommodate a variety of objective functions [36] and a variety of data analysis problems. Many studies have focused on further developing computational methodologies for NMF. The true power of NMF, however, is its ability to solve challenging pattern recognition and data mining problems [27]. For example, Brunet et al. interpreted the gene expression pattern as metagenes that captured gene expression patterns specific to different groups of samples and demonstrated that NMF was more accurate than hierarchical clustering (HC) and more stable than self-organizing maps (SOMs) [9].

NMF is a matrix factorization method that focuses on the analysis of data matrices with factors that are non-negative. Gene expression data are typically presented as a matrix in which the rows correspond to expression levels of genes and the columns correspond to samples, and each entry corresponds to the expression level of a given gene in a given sample.

Given a matrix $\mathbf{X}=(\mathbf{x}_1,\mathbf{x}_2,...,\mathbf{x}_n)\in R^{m\times n}$ with size $m\times n$, its rows contain the expression levels of m genes in the n samples. Without loss of generality, with non-negative elements in $\mathbf{X}$, NMF seeks to decompose $\mathbf{X}$ into a non-negative coefficient matrix $\mathbf{Y}=(\mathbf{y}_1,\mathbf{y}_2,...,\mathbf{y}_n)\in R^{r\times n}$ and non-negative basis matrix $\mathbf{A}=(a_1,a_2,\cdots,a_m)^T\in R^{m\times r}$, such that $\mathbf{X}\approx\mathbf{AY}$, where $\mathbf{A}$ has size $m\times r$, with each of the m rows defining a metagene, and $\mathbf{Y}$ has size $r\times n$, with each of the r rows representing the metapattern of the corresponding sample. This definition states that each column of $\mathbf{X}$ is approximated by a non-negative linear combination of the columns of $\mathbf{A}$, where the coefficients are given by the corresponding column of $\mathbf{Y}$ [37].

The error function of the basic NMF is

$$\left\|\mathbf{X}-\mathbf{AY}\right\|_F^2=\sum_{j=1}^{n}\left\|\mathbf{x}_j-\mathbf{A}\mathbf{y}_j\right\|^2,\quad s.t.\ \mathbf{Y}>\mathbf{0},\ \mathbf{A}>\mathbf{0}. \tag{1}$$

There are two commonly used error functions that quantify the quality of the approximation. The first one is the square of the Euclidean distance between two matrices [30]:

$$f_1=\left\|\mathbf{X}-\mathbf{AY}\right\|^2=\sum_{i,j}\left(x_{ij}-\sum_{k=1}^{r}a_{ik}y_{kj}\right)^2. \tag{2}$$

The second one is the Kullback-Leibler divergence between two matrices [32] :

$$f_2=D(\mathbf{X}\,|\,\mathbf{AY})=\sum_{i,j}\left(x_{ij}\log\frac{x_{ij}}{v_{ij}}-x_{ij}+v_{ij}\right), \tag{3}$$

where $\mathbf{V}=[v_{ij}]=\mathbf{AY}$. In this paper, we refer to f_1 and f_2 as F-norm and divergence formulations, respectively. Lee and Seung [32] presented two iterative update algorithms. The algorithm minimizing the objective function f_1 is as follows:

$$a_{ik}\leftarrow a_{ik}\frac{\left(\mathbf{XY}^T\right)_{ik}}{\left(\mathbf{AYY}^T\right)_{ik}},\quad y_{kj}\leftarrow y_{kj}\frac{\left(\mathbf{A}^T\mathbf{X}\right)_{kj}}{\left(\mathbf{YAA}^T\right)_{kj}}. \tag{4}$$

The algorithm minimizing the objective function f_2 is as follows:

$$a_{ik}\leftarrow a_{ik}\frac{\sum_j\left(x_{ij}y_{kj}/\sum_k a_{ik}y_{kj}\right)}{\sum_j y_{kj}},\qquad y_{kj}\leftarrow y_{kj}\frac{\sum_i\left(x_{ij}a_{ik}/\sum_k a_{ik}y_{kj}\right)}{\sum_i a_{ik}}. \tag{5}$$

The above two algorithms can be used to find the local minima of the objective functions f_1 and f_2 [38].

Prior knowledge of two distances with respect to the probability distribution of the noise is different. Euclidean distance minimization can be seen as a maximum likelihood estimator for which the difference is due to additive Gaussian noise, whereas Kullback-Leibler divergence can be considered as likelihood for Poisson processes [39]. The Kullback-Leibler divergence has some deficiencies, especially because the gradients needed for optimization depend heavily on the scales of factorizing matrices, leading to many iterations.

The details of the basic NMF method are summarized as listed in Algorithm 1. The iteration procedure is repeated until the algorithm converges.

Algorithm 1: Basic NMF

Input: $\mathbf{X}\in R^{m\times n}$

Output: $\mathbf{Y}\in R^{r\times n}$, $\mathbf{A}\in\mathbf{R}^{m\times r}$

1: Initialize $\mathbf{A}_0\in\mathbf{R}^{m\times r}$ and $\mathbf{Y}_0\in\mathbf{R}^{r\times n}$ as non-negative matrices.

Set k=0.

2: repeat

Update $\mathbf{A}_{k+1}$ as

$$\mathbf{A}_{k+1}\leftarrow\mathbf{A}_k\frac{\mathbf{X}\mathbf{Y}_k^T}{\mathbf{A}_k\mathbf{Y}_k\mathbf{Y}_k^T}$$

Update $\mathbf{Y}_{k+1}$ as

$$\mathbf{Y}_{k+1}\leftarrow\mathbf{Y}_k\frac{\mathbf{A}_{k+1}^T\mathbf{X}}{\mathbf{Y}_k\mathbf{A}_{k+1}\mathbf{A}_{k+1}^T}$$

r=r+1

Until convergence

Although the basic NMF has been used in the bioinformatics field, it has many drawbacks due to the dense basis and coefficient matrices. Gene expression data have a high-dimensionality and consistently contain some redundant information, i.e., not all features are associated with a special biological process or function. To circumvent this problem, sparsity must be introduced, which means that some elements of the vectors are zero.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2017.2665557, IEEE/ACM Transactions on Computational Biology and Bioinformatics

3 SPARSE NMF

The previous section presented the basic NMF method, but the method has some limitations, for example, the nonsparsity, which can be solved by the sparse methods presented in this section by using sparse constraints.

In genomics, gene expression data typically contain thousands of genes, but the number of samples is much smaller than that of genes. A given disease or biological function is usually associated with a few genes. It is a crucial problem in bioinformatics research to select a few relevant genes among thousands of genes. It is also crucial for disease diagnosis to identify meaningful proteomics features from the large amount of gene expression data [40].

Several variants of NMF have recently been presented to improve the performance of NMF [41-43]. Sparse constraints have been incorporated into NMF to obtain sparse solutions [8, 44-46]. Other authors have also noted that non-negative constraints alone cannot guarantee a sparse representation of the original data, and further, that the degree of sparseness cannot be controlled [24]. Therefore, various investigations have been conducted to incorporate sparse constraints into NMF.

Recently sparse regularization in dimensionality reduction has been widely investigated and applied into feature selection studies [18]. In the following section, we summarize the sparse constraints of NMF algorithms. Given an $m \times n$ matrix $\mathbf{M}$,

$$\left\|\mathbf{M}\right\|_0 = \lim_{p\to 0}\left\|\mathbf{M}\right\|_p^p = \lim_{p\to 0}\sum_{i=1}^{m}\left|m_i\right|^p, \quad (6)$$

when $p \to 0$, the function is denoted as the L_0-norm of matrix $\mathbf{M}$.

$$\left\|\mathbf{M}\right\|_1 = \sum_{i=1}^{m}\sum_{j=1}^{n}\left|m_{j,i}\right| \quad (7)$$

is denoted as the L_1-norm of matrix $\mathbf{M}$. It denotes the sum of the absolute values of elements in the matrix $\mathbf{M}$. The L_1-norm was first introduced as LASSO in [47].

$$\left\|\mathbf{M}\right\|_{2,1} = \sum_{i=1}^{m}\sqrt{\sum_{j=1}^{n}\mathbf{m}_{ij}^2} = \sum_{i=1}^{m}\left\|\mathbf{m}^i\right\|_2 \quad (8)$$

is denoted as the $L_{2,1}$-norm of matrix $\mathbf{M}$. First, we compute the L_2-norm of rows $\mathbf{m}^i$, and then compute the L_1-norm of vector $b(\mathbf{M}) = (\left\|\mathbf{m}^1\right\|_2, \left\|\mathbf{m}^2\right\|_2, ..., \left\|\mathbf{m}^m\right\|_2)$. The $L_{2,1}$-norm of a matrix was first introduced in [48] as rotational invariant L_1-norm and used for multi-task learning [49, 50] and tensor factorization.

The L_0-norm function is a non-convex, non-smooth, discontinuity and global non-differentiable function. The L_1-norm function is a convex, non-smooth and global non-differentiable function. The $L_{2,1}$-norm function is a convex, smooth and global non-differentiable function [51].

3.1 L_0-norm NMF

L_0-norm is the most convenient and intuitive sparseness constraint, which means the number of non-zero elements of the vector or matrix. However, there are two reasons that can justify an L_0-constrained sparse version of NMF. First, the L_0-norm constraint method, although approximate, is useful, because we are able to constrain the basis or coefficient vectors of NMF to acquire the exact desired number of nonzero elements. Second, the joint optimization of $\mathbf{Y}$ and $\mathbf{A}$ is a non-convex problem per se, which means that all NMF methods proposed to date converge to only a local minimum. The sparse representation method with L_0-norm minimization can obtain the fundamental sparse solution of $\mathbf{Y}$ [51, 52]. Peharz et al. presented an NMF method with L_0-constraints on the columns of $\mathbf{A}$ and $\mathbf{Y}$, respectively [53].

The sparse coding problem can be defined as minimization of $\left\|x-\mathbf{A}y\right\|^2$, $s.t.\ L_0(y) \le K$, where $L_0(.)$ denotes the L_0-norm, i.e., the number of nonzero entries and K denotes the maximal allowed number of non-zero entries in y. For all columns of $\mathbf{X}$, the following matrix form can be extended [52]:

$$\min_{\mathbf{A},\mathbf{Y}}\left\|\mathbf{X}-\mathbf{A}\mathbf{Y}\right\|_F^2, \quad s.t.\ L_0(\mathbf{Y}) \le K. \quad (9)$$

It is well-known that the optimal solution for the sparse coding problem is NP-hard [54], where the challenge is to identify the optimal automated data assignment, i.e., the locations of the non-zero entries in $\mathbf{Y}$. Many approximate sparse coding approaches have been proposed (see, e.g., [55-57]), where one of the most well-known comprehensive algorithms is orthogonal matching pursuit (OMP), as described in [58].

3.2 L_1-norm NMF

From the above analysis, the L_0-norm constraint makes this objective function difficult to optimize, so it is approximated using L_1-norm, which has been a popular strategy in prior studies [59].

Since real gene expression data often contain noise, Shen et al. proposed a Robust NMF method [46] that was able to simultaneously learn the basis and coefficient matrices and to estimate the positions and values of noise [60, 61].

A robust method of NMF with the L_1-norm constraint can explicitly model the partial corruption, which can be treated as large additive noises. Let $\mathbf{X} \in R^{m\times n}$ denote the non-negative observation matrix with corrupted data, while each column of $\mathbf{X}$ is a data sample. Let $\hat{\mathbf{X}} \in R^{m\times n}$ denote the clean data without noise. Then, $\mathbf{X} = \hat{\mathbf{X}} + \mathbf{E}$, where $\mathbf{E} \in R^{m\times n}$ is the matrix with large additive noise. Moreover, the partial noise is concerned, where partial indicates that the distribution of noise is sparse. Thus, only a small portion of entries for the noise matrix are nonzero. The clean data $\hat{\mathbf{X}}$ are approximated by $\mathbf{A}\mathbf{Y}(\mathbf{A} \in R^{m\times r}, \mathbf{Y} \in R^{r\times n})$ as in the basic NMF, and thus $\mathbf{X} \approx \mathbf{A}\mathbf{Y} + \mathbf{E}$ [46].

Introducing the L_1-norm into the objective function, it can be written as $O = \left\|\mathbf{X}-\mathbf{A}\mathbf{Y}-\mathbf{E}\right\|_F^2 + \lambda\sum_j\left\|\mathbf{E}_{.j}\right\|_1^2$, where the L_1-norm penalty is used for sparseness, which has been confirmed to be effective and computationally convenient [8, 46, 62, 63].

In addition, Hoyer proposed an efficient algorithm to minimize the objective $\left\|\mathbf{X}-\mathbf{A}\mathbf{Y}\right\|_F^2 + \lambda\sum_{ij}\left|\mathbf{Y}\right|_{ij}$ [58], which

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2017.2665557, IEEE/ACM Transactions on Computational Biology and Bioinformatics

penalized the coefficients matrix $\mathbf{Y}$ by using the L_1-norm constraint. Furthermore, Hoyer [44] defined a sparseness function of a factor vector via the L_1-norm and presented an NMF method, which constrained the columns of $\mathbf{A}$ or $\mathbf{Y}$ to given sparse values. Based on the same objective, Eggert and Koerner [64] proposed an alternative update rule that implicitly normalized the columns of $\mathbf{Y}$ to the unit length [52].

3.3 $L_{2,1}$ -norm NMF

It is well known that gene expression data contain noise and outliers similar to other high-dimensional data. In contrast, in the basic NMF method, the error for each data point enters the objective function as the squared residual error. Thus, a few outliers with large errors easily dominate the objection function. As a result, Kong et al. [65] proposed a robust NMF by using $L_{2,1}$-norm to diminish the impact of the outliers [66, 67]. The optimization problem can be written as follows:

$$\min_{\mathbf{A},\mathbf{Y}} \left\|\mathbf{X}-\mathbf{AY}\right\|_{2,1} \quad s.t.\ \ \mathbf{A}\geq 0, \mathbf{Y}\geq 0. \tag{10}$$

The objective function of robust NMF can be rewritten as follows:

$$\left\|\mathbf{X}-\mathbf{AY}\right\|_{2,1} = \sum_{j=1}^{n}\sqrt{\sum_{i=1}^{m}(\mathbf{X}-\mathbf{AY})_{ij}^2} = \sum_{j=1}^{n}\left\|\mathbf{x}_j-\mathbf{A}\mathbf{y}_j\right\|. \tag{11}$$

In this robust formulation, the error for each data point is $\left\|\mathbf{x}_j-\mathbf{A}\mathbf{y}_j\right\|$, which is not squared. Thus, the errors of the outlies and noise do not dominate the objective function because they are not squared [65]. The impact of the outliers and noise of the data can then be reduced.

The update rules of the robust NMF method with $L_{2,1}$-norm are as follows:

$$a_{ik} \leftarrow a_{ik}\frac{(\mathbf{XDY}^T)_{ik}}{(\mathbf{AYDY}^T)_{ik}}, \quad y_{kj} \leftarrow y_{kj}\frac{(\mathbf{A}^T\mathbf{XD})_{kj}}{(\mathbf{YAA}^T\mathbf{D})_{kj}}, \tag{12}$$

where $\mathbf{D}$ is a diagonal matrix with the elements given by

$$\mathbf{D}_{jj} = 1/\sqrt{\sum_{i=1}^{m}(\mathbf{X}-\mathbf{AY})_{ij}^2} = 1/\left\|\mathbf{x}_j-\mathbf{A}\mathbf{y}_j\right\|. \tag{13}$$

The computational algorithm for robust NMF is unexpectedly simple. It has almost the same computational cost as basic NMF. The convergence proof of the algorithm is provided in [65].

Motivated by the previous studies on the norm-based NMF algorithms, the L_2-norm-based loss function is sensitive to outliers. Thus, Nie et al. [18] proposed an efficient and robust feature selection method to employ a joint $L_{2,1}$-norm minimization for both the loss function and regularization. This method has been adopted in gene expression data to remove outliers, and an $L_{2,1}$-norm regularization is applied to select features across all data points with joint sparsity, i.e., each feature (gene expression in mass spectrometry) either has small scores for all data points or large scores over all data points [18, 68].

3.4 The Versatile Sparse NMF

According to the above analysis of sparse constraints, additional constraints can be added to the basic NMF formulation. By and large, regularized NMF can be written as the following problem [69]:

$$\begin{aligned}\min_{\mathbf{A},\mathbf{Y}} f(\mathbf{A},\mathbf{Y}) = &\frac{1}{2}\left\|\mathbf{X}-\mathbf{AY}\right\|_F^2 \\ &+\sum_{i=1}^{n}\left(\frac{\alpha_2}{2}\left\|a_i\right\|_2^2+\alpha_1\left\|a_i\right\|_1\right) \\ &+\sum_{i=1}^{n}\left(\frac{\lambda_2}{2}\left\|y_i\right\|_2^2+\lambda_1\left\|y_i\right\|_1\right). \\ s.t.\ &\begin{cases}\mathbf{A}\geq 0 & if\ t_1=1\\ \mathbf{Y}\geq 0 & if\ t_2=1\end{cases}\end{aligned} \tag{14}$$

In this optimization problem, the parameters $\alpha_1\geq 0$ and $\alpha_2\geq 0$ control the sparsity and smoothness of the basis vectors, respectively. The parameters $\lambda_1\geq 0$ and $\lambda_2\geq 0$ control the sparsity and smoothness of the coefficient vectors, respectively. The parameters t_1 and t_2 are Boolean variables (0: false, 1: true) that indicate whether non-negativity should or should not be enforced on $\mathbf{A}$ and $\mathbf{Y}$, respectively [70]. This variant of NMF is called Versatile Sparse non-negative Matrix Factorization (VSMF).

From Subsections 3.1 and 3.2, we can see that the L_1-norm cannot make the correlated variables non-zero simultaneously in the induced sparse result, which may explain the ability of L_1-norm to produce a sparse but non-smooth result. While it is known that L_2-norm is able to obtain a smooth but not sparse result, another benefit of L_2-norm is that the scale of each vector can be restricted. This feature can avoid the scale interchange between the basis and coefficient matrices [70]. The advantage of VSMF is that both L_1-norm and L_2-norm can be used on both basis and coefficient matrices. Another advantage of VSMF is that the non-negativity constraint can be switched off/on for either basis or coefficient matrices. In some situations, non-negativity is also needed for the coefficient matrix for better performance and interpretation [70].

The basic NMF, semi-NMF, and sparse-NMF are clearly special cases of VSMF. If α_1=α_2=λ_1=λ_2=0 and $t_1=t_2=1$, VSMF is degraded to the basic NMF, which is proposed in [31]. If α_1=$\lambda_2=0$, α_2=$\lambda_1\neq 0$ and $t_1=t_2=1$, then VSMF is equivalent to the sparse-NMF proposed in [8]. When α_1 is set to zero, VSMF can be kernelized [71]. If α_1=α_2=λ_1=λ_2=0 and $t_1=0$, $t_2=1$, then VSMF becomes semi-NMF as proposed in [72].

Optimization of VSMF is non-convex, as described for most NMF models. The most popular scheme to optimize the model is the block-coordinate descent method [73], which is basically described as follows: in each iteration, $\mathbf{A}$ and $\mathbf{Y}$ are updated iteratively and alternately, that is, $\mathbf{A}$ is updated while keeping $\mathbf{Y}$ fixed, and vice versa. Based on this scheme, the multiplicative update rules for VSMF are as follows.

Multiplicative Update Rules for VSMF

If both $\mathbf{A}$ and $\mathbf{Y}$ are non-negative, we can equivalently rewrite $f(\mathbf{A},\mathbf{Y})$ in Eq. (14) to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2017.2665557, IEEE/ACM Transactions on Computational Biology and Bioinformatics

$$\frac{1}{2}\left\|\mathbf{X}-\mathbf{AY}\right\|_F^2+\frac{\alpha_2}{2}\mathrm{Tr}(\mathbf{A}^T\mathbf{A})+\alpha_1\mathrm{Tr}(\mathbf{E}_1^T\mathbf{A}) +\frac{\lambda_2}{2}\mathrm{Tr}(\mathbf{Y}^T\mathbf{Y})+\lambda_1\mathrm{Tr}(\mathbf{E}_2^T\mathbf{Y}), \tag{15}$$

where $\mathbf{E}_1\in\{1\}^{m\times r}$ and $\mathbf{E}_2\in\{1\}^{r\times n}$. Fixing $\mathbf{A}$ and updating $\mathbf{Y}$, the problem in Eq.(15) can hence be expressed as

$$\min_{\mathbf{Y}} f(\mathbf{Y})=\frac{1}{2}\left\|\mathbf{X}-\mathbf{AY}\right\|_F^2+\frac{\lambda_2}{2}\mathrm{Tr}(\mathbf{Y}^T\mathbf{Y})+\lambda_1\mathrm{Tr}(\mathbf{E}_2^T\mathbf{Y}) \quad s.t.\ \mathbf{Y}\geq 0. \tag{16}$$

Similarly, fixing $\mathbf{Y}$ and updating $\mathbf{A}$, the problem in Eq.(15) can be expressed as

$$\min_{\mathbf{A}} f(\mathbf{A})=\frac{1}{2}\left\|\mathbf{X}-\mathbf{AY}\right\|_F^2+\frac{\alpha_2}{2}\mathrm{Tr}(\mathbf{A}^T\mathbf{A})+\alpha_1\mathrm{Tr}(\mathbf{E}_1^T\mathbf{A}) \quad s.t.\ \mathbf{A}\geq 0. \tag{17}$$

In the case of $t_1=t_2=1$, the multiplicative update rules for the VSMF model are as follows:

$$\mathbf{A}=\mathbf{A}*\frac{\mathbf{XY}^T}{\mathbf{AYY}^T+\alpha_2\mathbf{A}+\alpha_1}, \qquad \mathbf{Y}=\mathbf{Y}*\frac{\mathbf{A}^T\mathbf{X}}{\mathbf{A}^T\mathbf{AY}+\lambda_2\mathbf{Y}+\lambda_1}. \tag{18}$$

The VSMF model is a unified model of many variants of NMF, so we can simultaneously address the robustness, sparsity and non-negativity of gene expression data by using the VSMF.

4 Graph Regularized NMF

NMF methods have shown numerous advantages for the identification of differentially expressed gene and clustering samples with the above analysis of sparsity, which was analyzed in the previous sections, but they fail to discover the intrinsic geometric and discriminating structure of the data space, which is essentially useful for gene expression analysis, especially for gene selection and tumor clustering. Thus, the data are usually sampled from a low dimension manifold embedded in the high dimension [38, 74, 75]. Thus, there is considerable space for improving the performance of NMF, which has received numerous attention due to the geometric perspective [76]. These problems are solved by the methods presented in this section.

Various researchers [3, 77, 78] have considered the case in which data were drawn from sampling a probability distribution with support on or near to a sub-manifold in ambient space [38]. For example, the methods in [62, 79] were proposed to preserve the local structure of the low dimensional manifold. To address the outliers, RSNMF (Robust Sparse Non-negative Matrix Factorization) was proposed [80], which was based on an L_1-norm objective function. An outlier list was maintained in NMF for more robust performance [81]. To detect the underlying manifold structure, many manifold learning methods were proposed to address this case, such as LLE [3], ISOMAP [77], and Laplacian Eigenmap [82].

4 .1 GNMF

Manifold learning is the most popular method to address the embedded structure problem, and the basic idea is that if two data samples are close to each other in the input space, then they are also close to each other in the embedding space. All the manifold algorithms use the so-called locally invariant idea [83, 84], i.e., the nearby points are likely to have similar embeddings. It is known that the gene selection performance can be significantly enhanced if the geometric structure in gene expression data and the local invariance are considered [38]. Motivated by previous progress in matrix factorization and manifold learning [79, 85-87], Cai et al. proposed a method called Graph regularized Non-negative Matrix Factorization (GNMF), which explicitly considered the local invariance. To find a part-based representation space in which two data points are sufficiently close to each other if they are connected in the low dimension graph, a nearest neighbor graph encoding the geometrical information in the original dataset was constructed. A new objective function of matrix factorization was designed by incorporation of the graph structure [38, 88].

The basic assumption here could be that if two original data points x_i and x_j are close in the intrinsic geometry of the data distribution, then s_i and s_j, the representations of the two points with respect to the new basis, are also close to each other [82]. This assumption is usually referred to as the local invariance assumption [86], which plays an important role in the development of various kinds of algorithms [85, 89, 90].

The development of spectral graph theory [91] and manifold learning theory [82] has demonstrated that the local geometric structure embedded in the high dimension can be effectively modeled using a nearest neighbor graph. A graph with n vertices is considered, where each vertex corresponds to a data point [38]. For each data point $\mathbf{x}_i$, its k nearest neighbors can be found, and the edges between $\mathbf{x}_i$ and its neighbors are positioned. The w_{ij} is used to measure the closeness of two points x_i and its neighbor x_j [38]. There are many choices to define the weight matrix $\mathbf{W}$ on the graph.

One method to define the weight matrix can be given as follows:

0-1 weight: $w_{ij}=1$, if and only if node i and j are connected by an edge. This is the simplest weighting method and is very easy to compute.

The $w_{ij}=1$ is only for measuring the closeness of two data points in this method, so we apply the Euclidean distance $O(s_i,s_j)=\left\|s_i-s_j\right\|^2$ to measure the distance between the low dimension of two data points.

The smoothness of the low dimension representation can be measured by

$$\begin{aligned} R&=\frac{1}{2}\sum_{i,j}^{n}\left\|s_i-s_j\right\|^2 w_{ij}\\ &=\sum_{i=1}^{n}s_i^T s_i d_{ii}-\sum_{i,j=1}^{n}s_i^T s_j w_{ij}\\ &=\mathrm{Tr}(\mathbf{YDY}^T)-\mathrm{Tr}(\mathbf{YWY}^T)\\ &=\mathrm{Tr}(\mathbf{YLY}^T), \end{aligned} \tag{19}$$

where $\mathrm{Tr}(\cdot)$ denotes a trace of matrix, $\mathbf{W}$ is a weight matrix of the nearest neighbor graph, and $\mathbf{D}$ is a diagonal matrix with column (or row, since $\mathbf{W}$ is symmetric) sum

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2017.2665557, IEEE/ACM Transactions on Computational Biology and Bioinformatics

entries of $\mathbf{W}$, $d_{ii} = \sum_{j=1}^{n} w_{ij}$. $\mathbf{L} = \mathbf{D} - \mathbf{W}$, which is called the graph Laplacian matrix [38, 92].

GNMF then minimizes the objective function as follows:

$$\min_{\mathbf{A},\mathbf{Y}} \left\|\mathbf{X} - \mathbf{AY}\right\|^2 + \lambda \operatorname{Tr}(\mathbf{YLY}^T), \quad s.t.\ \mathbf{Y} \geq \mathbf{0},\ \mathbf{A} \geq \mathbf{0}, \tag{20}$$

where the regularization parameter $\lambda \geq 0$ controls the smoothness of the new representation [93].

An Efficient Algorithm of GNMF

The error function in Eq.(20) can be rewritten as follows:

$$\begin{aligned} f &= \operatorname{Tr}((\mathbf{X} - \mathbf{AY})(\mathbf{X} - \mathbf{AY})^T) + \lambda \operatorname{Tr}(\mathbf{Y}^T\mathbf{LY}) \\ &= \operatorname{Tr}(\mathbf{XX}^T) - 2\operatorname{Tr}(\mathbf{XY}^T\mathbf{A}^T) + \operatorname{Tr}(\mathbf{AYY}^T\mathbf{A}^T) + \lambda \operatorname{Tr}(\mathbf{YLY}^T). \end{aligned} \tag{21}$$

The optimization problem in Eq.(21) can be solved by using the Lagrangian multipliers ψ_{ik} and ϕ_{kj} under constraints $a_{ik} \geq 0$ and $y_{kj} \geq 0$, respectively [94]. When $\Psi = [\psi_{ik}]$ and $\Phi = [\phi_{kj}]$, the Lagrangian function L is defined as

$$\begin{aligned} L &= \operatorname{Tr}(\mathbf{XX}^T) - 2\operatorname{Tr}(\mathbf{XY}^T\mathbf{A}^T) + \operatorname{Tr}(\mathbf{AYY}^T\mathbf{A}^T) \\ &\quad + \lambda \operatorname{Tr}(\mathbf{YLY}^T) + \operatorname{Tr}(\Psi\mathbf{A}^T) + \operatorname{Tr}(\Phi\mathbf{Y}^T). \end{aligned} \tag{22}$$

Using the KKT conditions [95], we can obtain the update rules which are listed as follows:

$$a_{ik} \leftarrow a_{ik} \frac{(\mathbf{XY}^T)_{ik}}{(\mathbf{AYY}^T)_{ik}}, \tag{23}$$

$$y_{kj} \leftarrow y_{kj} \frac{(\mathbf{A}^T\mathbf{X} + \lambda\mathbf{YW})_{kj}}{(\mathbf{A}^T\mathbf{AY} + \lambda\mathbf{YD})_{kj}}. \tag{24}$$

From the analysis of GNMF, we can see that it has more discriminating power than traditional NMF because it considers the sub-manifold embedded in ambient space, which is essential for tumor clustering and gene selection [88]. The results in Section 6 will demonstrate that the GNMF can identify more differentilly expressed genes than the basic NMF.

4.2 RMNMF

Because the GNMF method cannot address the impact of outliers and noise in the gene expression data, Cai et al. proposed a Robust Manifold Non-negative Matrix Factorization (RMNMF) model in which the mixed-Norm $L_{2,1}$-norm was used to improve the model robustness. This model can be applied to practical data mining applications, such as gene identification and sample clustering. Additionally, a manifold regularization term was incorporated into the geometrical information existing in the high dimension data [74]. The objective function of the RMNMF model is given as follows:

$$\min_{\mathbf{A},\mathbf{Y}} \left\|\mathbf{X} - \mathbf{AY}\right\|_{2,1} + \lambda \operatorname{Tr}(\mathbf{YLY}^T), \quad s.t.\ \mathbf{Y} > \mathbf{0},\ \mathbf{YY}^T = \mathbf{I}. \tag{25}$$

Here the second term is the additional constraint $\mathbf{YY}^T = \mathbf{I}$. The first purpose is to reduce the computation cost for the optimization. Another purpose is to guarantee the uniqueness of the RMNMF solution. Suppose $\mathbf{A}^*$ and $\mathbf{Y}^*$ are the solutions to Eq.(25); then, for any given nonzero constant $c > 1$, $c\mathbf{A}^*$ and $c\mathbf{Y}^*$ would provide an identical value for the first term and a lower value for the second term, whether $\mathbf{A}^*$ and $\mathbf{Y}^*$ are local or global optimum solutions [74]. While inspired by the sparsity, RMNMF has no sparse constraints, making it inefficiency for managing the enormous amount of gene expression data. The next subsection introduces a method that can simultaneously manage the geometric structure, robustness and sparsity.

4.3 RSGNMF

In Subsections 4.1 and 4.2, two methods have been shown to address outliers and embedded structural problems that are essential for the identification of differentially expressed genes. Nonetheless, there is still considerable space to optimize the NMF algorithms with $L_{2,1}$-norm and manifold. Yang et al. proposed a Robust NMF via the joint Sparse and Graph regularization model (RSGNMF), which can simultaneously handle high-dimensional, sparse and noisy data. The $L_{2,1}$-based loss function term can diminish the impact of outliers or noise, and the $L_{2,1}$-norm-based sparse regularization term forces most of the rows in $\mathbf{Y}$ to shrink to zero, which implies that the corresponding features of these zero rows are not important for new representations, and the manifold term preserves the local structures embedded in the original data [96, 97].

Given a data matrix $\mathbf{X}$, we can obtain the final formulation of the RSGNMF model, which simultaneously takes into consideration the robust loss function, sparse regularization and local structures of the data. The formulation is written as follows:

$$\min_{\mathbf{A},\mathbf{Y}} \left\|\mathbf{X} - \mathbf{AY}\right\|_{2,1} + \alpha\left\|\mathbf{Y}\right\|_{2,1} + \beta \operatorname{Tr}(\mathbf{YLY}^T) \quad s.t.\ \mathbf{A} \geq 0,\ \mathbf{Y} \geq 0. \tag{26}$$

We introduce the solution for RSGNMF model via iterative updating algorithm as below,

$$\begin{aligned} \mathbf{Y}_{kj} &\leftarrow \mathbf{Y}_{kj} \frac{(\mathbf{A}^T\mathbf{XD}_1)_{kj}}{(\mathbf{A}^T\mathbf{AYD}_1 + \alpha\mathbf{YD}_2 + 2\beta\mathbf{GL})_{kj}}, \\ \mathbf{A}_{ik} &\leftarrow \mathbf{A}_{ik} \frac{(\mathbf{XD}_1\mathbf{Y}^T)_{ik}}{(\mathbf{AYD}_1\mathbf{Y}^T)_{ik}}, \end{aligned} \tag{27}$$

where $\mathbf{D}_1$ and $\mathbf{D}_2$ are diagonal matrices with diagonal elements given by

$$\begin{aligned} (\mathbf{D}_{jj})_1 &= 1\Big/\sqrt{\textstyle\sum_{j=1}^{n}(\mathbf{X} - \mathbf{AY})_{ij}^2} = 1\Big/\left\|x_j - \mathbf{A}y^j\right\|, \\ (\mathbf{D}_{jj})_2 &= 1\Big/\sqrt{\textstyle\sum_{j=1}^{n}\mathbf{Y}_{ij}^2} = 1\Big/\left\|y^j\right\|. \end{aligned} \tag{28}$$

The additional $L_{2,1}$-norm on both graph and error functions has more advantages than on each other separately. It combines the robust and sparsity characteristics, which can diminish the impact of noise, generate a sparse results and discover the intrinsic geometry for the analysis of gene expression data.

4.4 GDNMF

Another important issue in identifying differentilly expressed genes and clustering samples is the discrimina-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2017.2665557, IEEE/ACM Transactions on Computational Biology and Bioinformatics

tive power of a method. Although the above methods have solved many problems, if the discriminative information can be added to the model, then it will be a more powerful method for managing the gene expression data.

Inspired by the success of NMF based on graph regularization [38] and discriminative dictionary learning [98], Long et al. proposed a method called Graph regularized Discriminative Non-negative Matrix Factorization (GDNMF), which explicitly considered the embedded structure problem and label information. The method encoded the geometric structure of the data in space by constructing a k -nearest-neighbor graph and increased the discriminative power by considering the label information [99, 100]. A matrix decomposition aims to identify a part-based representation discriminative space in which two data points are sufficiently close to each other if they are connected in the original data set graph [101].

First, the class indicator matrix $\mathbf{S} \in \mathbf{R}^{p \times n}$ is defined as follows:

$$\mathbf{S}_{ij} = \begin{cases} 1, & \text{if } w_j = i, \quad j = 1,2,\cdots,n,\ i = 1,2,\cdots,p, \\ 0, & \text{otherwise,} \end{cases} \tag{29}$$

where $w_j \in \{1,2,\cdots,\mathrm{p}\}$ denotes the class label of the j -th sample x_j and p is the total number of classes in matrix $\mathbf{X}$.

The optimization problem of the GDNMF model is

$$\begin{aligned} &\min_{\mathbf{A},\mathbf{Y}} \left\|\mathbf{X} - \mathbf{AY}\right\|_F^2 + \lambda \operatorname{Tr}(\mathbf{YLY}^T) + \beta \left\|\mathbf{S} - \mathbf{CY}\right\|_F^2, \\ &s.t.\ \mathbf{Y} \geq \mathbf{0},\ \mathbf{A} \geq \mathbf{0},\ \mathbf{C} \geq 0, \end{aligned} \tag{30}$$

where the regularization parameter $\lambda \geq 0$ controls the smoothness of the new representation [93], and $\beta \geq 0$ is the regularization parameter. $\mathbf{L}$ is called the graph Laplacian matrix, which is defined in Subsection 4.1 [92]. $\mathbf{C} \in \mathbf{R}^{p \times k}$ is a non-negative matrix and is initialized randomly in this method.

An Efficient Algorithm for GDNMF

The optimization scheme to solve this objective function is based on multiplicative iterative updates of these three factor matrices. The error function in Eq.(30) can be rewritten as follows:

$$\begin{aligned} f &= \operatorname{Tr}((\mathbf{X} - \mathbf{AY})(\mathbf{X} - \mathbf{AY})^T) + \lambda \operatorname{Tr}(\mathbf{Y}^T\mathbf{LY}) \\ &\quad + \beta \operatorname{Tr}((\mathbf{S} - \mathbf{CY})^T(\mathbf{S} - \mathbf{CY})) \\ &= \operatorname{Tr}(\mathbf{XX}^T) - 2\operatorname{Tr}(\mathbf{XY}^T\mathbf{A}^T) + \operatorname{Tr}(\mathbf{AYY}^T\mathbf{A}^T) + \lambda \operatorname{Tr}(\mathbf{YLY}^T) \\ &\quad + \beta \operatorname{Tr}(\mathbf{S}^T\mathbf{S}) - 2\beta \operatorname{Tr}(\mathbf{S}^T\mathbf{CY}) + \beta \operatorname{Tr}(\mathbf{Y}^T\mathbf{C}^T\mathbf{CY}). \end{aligned} \tag{31}$$

To solve the constrained optimization problem in Eq.(31), the Lagrangian multipliers ψ_{ik}, ϕ_{kj}, and Ω_{pk} are introduced under constraints $a_{ik} \geq 0$, $y_{kj} \geq 0$ and $\Omega_{pk} \geq 0$, respectively [94]. If we let $\mathbf{\Psi} = [\psi_{ik}]$, $\mathbf{\Phi} = [\phi_{kj}]$ and $\mathbf{\Omega} = [\Omega_{pk}]$, then the Lagrangian function L is defined as

$$\begin{aligned} L &= \operatorname{Tr}(\mathbf{XX}^T) - 2\operatorname{Tr}(\mathbf{XY}^T\mathbf{A}^T) + \operatorname{Tr}(\mathbf{AYY}^T\mathbf{A}^T) + \lambda \operatorname{Tr}(\mathbf{YLY}^T) \\ &\quad + \beta \operatorname{Tr}(\mathbf{S}^T\mathbf{S}) - 2\beta \operatorname{Tr}(\mathbf{S}^T\mathbf{CY}) + \beta \operatorname{Tr}(\mathbf{Y}^T\mathbf{C}^T\mathbf{CY}) \\ &\quad + \operatorname{Tr}(\mathbf{\Psi A}^T) + \operatorname{Tr}(\mathbf{\Phi Y}^T) + \operatorname{Tr}(\mathbf{\Omega}^T\mathbf{C}). \end{aligned} \tag{32}$$

Using the KKT conditions [95] $\psi_{ik}a_{ik} = 0$, $\phi_{kj}y_{kj} = 0$ and $\Omega_{pk}\mathbf{C}_{pk} = 0$, and multiplying the two sides of the derivatives of L with respect to $\mathbf{A}$, $\mathbf{Y}$ and $\mathbf{C}$ by a_{ik}, y_{kj} and Ω_{pk}, respectively, the update rules are obtained as follows:

$$a_{ik} \leftarrow a_{ik} \frac{(\mathbf{XY}^T)_{ik}}{(\mathbf{AYY}^T)_{ik}}, \tag{33}$$

$$y_{kj} \leftarrow y_{kj} \frac{(\beta\mathbf{C}^T\mathbf{S} + \mathbf{A}^T\mathbf{X} + \lambda\mathbf{YW})_{kj}}{(\beta\mathbf{C}^T\mathbf{CY} + \mathbf{A}^T\mathbf{AY} + \lambda\mathbf{YD})_{kj}}, \tag{34}$$

$$c_{\mathrm{pk}} \leftarrow c_{pk} \frac{(\mathbf{SY}^T)_{pk}}{(\mathbf{CYY}^T)_{pk}}. \tag{35}$$

In conclusion, from Subsections 4.1, 4.2 and 4.3, we can see that combining the $L_{2,1}$ -norm for both the error function and the regularization term can simultaneously solve the outliers and noise and the sparse problems. Manifold regularization can be added to consider the structure embedded in high dimension ambient space and to identify the instinct geometric structure that is essential for the gene expression analysis. Additionally, in Subsection 4.4, incorporating supervised label information into the methods may strengthen the discriminative power of the algorithms.

In summary, if we can apply the $L_{2,1}$ -norm of the error function and/or regularization function, then the robustness and sparsity can be obtained. If we can incorporate manifold learning and label information into one method to analyze the gene expression data, the method may simultaneously address the geometric structure, discriminative power, robustness and sparsity. This idea will be addressed in future.

5 GENERALIZED NMF

In this section, in contrast with the above additional constraints as penalty terms, some models of generalized NMF and their iterative update rules are simply introduced.

5.1 Semi-NMF

Conventional NMF restricts every element in the data matrix $\mathbf{X}$ to be non-negative. While many candidate data in practical applications are not always non-negative, the factor features or principal components may also contain some negative elements in the factor matrix reflecting the phase information [25]. Ding et al. suggested an extended version referred to as Semi-NMF [72], which maintained some kernel concepts of NMF, where $\mathbf{A}$ was still restricted to being non-negative while placing no restrictions on the signs of $\mathbf{Y}$ [102, 103].

The formulations are collectively summarized as follows:

$$\begin{aligned} \text{PCA: } &\mathbf{X}_{\pm} \approx \mathbf{A}_{\pm}\mathbf{Y}_{\pm}, \\ \text{NMF: } &\mathbf{X}_{+} \approx \mathbf{A}_{+}\mathbf{Y}_{+}, \\ \text{Semi-NMF: } &\mathbf{X}_{\pm} \approx \mathbf{A}_{+}\mathbf{Y}_{\pm}. \end{aligned} \tag{36}$$

Ding et al. employed an alternating iterative approach to solve the optimization problem, where the positive and negative parts were separated from the mixed-sign matrix.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2017.2665557, IEEE/ACM Transactions on Computational Biology and Bioinformatics

$\mathbf{A}$ is updated using multiplicative rules while holding $\mathbf{Y}$ fixed, and vice versa [72].

5.2 Orthogonal NMF

Orthogonal NMF has an orthogonality constraint on either factor $\mathbf{A}$ and/or $\mathbf{Y}$ [104]. The orthogonality principle was first employed by Li et al. [105] to minimize redundancy between different bases, and then Ding et al. broached the concept of Orthogonal NMF explicitly [106]. In the case of non-negativity, orthogonality will naturally result in sparseness. Thus, it can be viewed as a special case of Sparse NMF. However, there is a notable difference between the optimization models [25, 45, 107].

In this subsection, we emphasize the orthogonality of matrix factors in NMF. The formulation of Orthogonal NMF is given as follows:

$$\min_{\mathbf{A}\geq 0,\mathbf{Y}>0} \left\|\mathbf{X}-\mathbf{AY}\right\|^2, \quad s.t.\ \ \mathbf{Y}^T\mathbf{Y}=\mathbf{I}. \tag{37}$$

The update rules are

$$\mathbf{Y}_{kj} \leftarrow \mathbf{Y}_{kj}\frac{(\mathbf{A}^T\mathbf{X})_{kj}}{(\mathbf{YY}^T\mathbf{A}^T\mathbf{X})_{kj}}, \qquad \mathbf{A}_{ik} \leftarrow \mathbf{A}_{ik}\frac{(\mathbf{XY}^T)_{ik}}{(\mathbf{AYY}^T)_{ik}}. \tag{38}$$

Furthermore, it is natural to consider imposing orthogonality on both $\mathbf{A}$ and $\mathbf{Y}$ in NMF simultaneously, so-called bi-orthogonality,

$$\min_{\mathbf{A}>0,\mathbf{Y}>0} \left\|\mathbf{X}-\mathbf{AY}\right\|^2, \quad s.t.\ \mathbf{Y}^T\mathbf{Y}=\mathbf{I},\ \mathbf{A}^T\mathbf{A}=\mathbf{I}, \tag{39}$$

which nevertheless typically provides poor approximation performance [108-110].

6 Results for the Tumor Dataset

To verify the performance of the algorithms mentioned above, we will perform experiments based on some tumor datasets.

Inspired by a previous study of the NMF algorithms and their various extensions, NMF algorithms have advantages in identifying differentially expressed genes and clustering samples [5, 111]. To evaluate the performance of the above various NMF extended algorithms, in this section, we will analyze gene expression data for identifying differentially expressed genes and clustering samples. Furthermore, we will also briefly apply the methods to the TCGA dataset. Several experiments are performed to demonstrate the performance of these algorithms.

6.1 Gene Identification Results with the Different Methods

In this subsection, the gene identification results of the above algorithms are introduced. The identification of differentially expressed genes based on these NMF methods is described as follows:

1) Gain the data matrix $\mathbf{X}$ according to the gene expression data.

2) Obtain the basis matrix $\mathbf{A}$ by using NMF-based methods.

3) Identify the differentially expressed genes via basis matrix $\mathbf{A}$.

4) Check the identified genes using the Gene Ontology (GO) tool.

The details of how to identify genes from $\mathbf{A}$ are then shown as follows:

The basis matrix $\mathbf{A}$ can be described as follows:

$$\mathbf{A}=\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1r} \\ a_{21} & a_{22} & \cdots & a_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mr} \end{bmatrix}. \tag{40}$$

The non-zero entries in the basis matrix $\mathbf{A}$ may reflect the differential expression of genes. Here, differentially expressed genes are identified by considering the amplitude of the entries in $\mathbf{A}$. Therefore, the absolute values of entries in the basis matrix $\mathbf{A}$ are firstly calculated, and then a vector $\hat{\mathbf{A}}$ can be obtained by summing the basis matrix $\mathbf{A}$ by rows. Mathematically, the above processes can be formulated as follows:

$$\hat{\mathbf{A}}=\left[\sum_{j=1}^{r}\left|a_{1j}\right|,\cdots,\sum_{j=1}^{r}\left|a_{mj}\right|\right]^T. \tag{41}$$

Finally, the evaluating vector $\bar{\mathbf{A}}$ is obtained by sorting $\hat{\mathbf{A}}$ in descending order. Without loss of generality, we suppose that the first k ($k\leq m$) entries in $\hat{\mathbf{A}}$ are non-zero, that is

$$\bar{\mathbf{A}}=\left[a_1,\ \cdots,\ a_k,\ \underbrace{0,\ \cdots,\ 0}_{m-k}\right]^T. \tag{42}$$

Without loss of generality, the larger the element is in $\bar{\mathbf{A}}$, the more differential the gene. Consequently, the genes associated with the first num ($num\leq k$) largest entries in $\bar{\mathbf{A}}$ are selected as differentially expressed genes.

A tumor is a swollen or distended part of the most harmful diseases, and tens of thousands of people die every year because of this condition. Therefore, it is an important challenge for scientists to identify the the related virulence genes among numerous gene expression data. In our experiment, we apply the DLBCL (Diffuse Large B-Cell Lymphoma) dataset, the most common lymphoid malignancy in adults, with a curable rate of less than 50% of patients. Shipp et al. (www.genome.wi.edu/MPR/lymphoma) applied a supervised learning method to an expression profiling dataset of 7139 genes in 58 tumor specimens, and identified 13 genes that were highly predictive of outcomes [112, 113].

For a fair comparison, 100 genes are identified as the differentially expressed genes in the tumor datasets by using these NMF-based methods. The GO enrichment [114] of functional annotation of the identified genes by these methods is detected by ToppFun, which is publicly available at [http://toppgene.cchmc.org/enrichment.jsp].

Table 1 lists the ten closely related terms to the DLBCL, with P-values corresponding to the different methods. The annotation column represents the total number of genes related to the GO terms in the web tool. The input

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2017.2665557, IEEE/ACM Transactions on Computational Biology and Bioinformatics

TABLE 1
THE TEN CLOSELY RELATED TERMS ON DLBCL TUMOR DATASET

		NMF		NMFL21		GNMF		RMNMF		RSGNMF		GDNMF		orth-NMF		
ID	name	p-Value	Input	p-Value	Input	p-Value	Input	p-Value	Input	p-Value	Input	p-Value	Input	p-Value	Input	Annotation
12456497-Table4	Human Leukemia Durig03 88genes	1.37E-10	6	2.53E-10	7	2.56E-102	45	2.33E-102	46	2.79E-101	45	4.12E-107	47	2.82E-74	36	81
M11197	Housekeeping genes identified as expressed across 19 normal tissues.	3.93E-32	30	7.12E-34	29	1.13E-95	57	1.26E-97	57	2.88E-99	59	4.58E-101	60	3.90E-78	51	389
GO:0022626	cytosolic ribosome	1.52E-07	7	1.58E-05	5	5.60E-91	44	5.60E-91	44	5.51E-90	44	8.44E-96	46	4.24E-63	34	96
GO:0006415	translational termination	9.32E-08	8	8.70E-10	11	7.48E-91	44	7.28E-92	46	7.37E-90	44	1.13E-95	46	1.13E-62	34	95
GO:0006414	translational elongation	8.28E-08	8	2.26E-11	11	1.04E-88	46	1.15E-87	48	1.25E-87	46	3.94E-93	48	7.30E-62	36	130
GO:0044391	ribosomal subunit	2.87E-06	7	1.25E-04	5	5.43E-86	46	4.28E-85	46	6.53E-85	46	2.92E-90	48	3.38E-60	36	148
GO:0003735	structural constituent of ribosome	4.13E-06	7	1.84E-04	5	1.51E-81	45	2.63E-80	46	1.63E-80	45	1.06E-85	47	1.47E-56	35	156
GO:0005198	structural molecule activity	1.39E-05	12	8.82E-06	11	1.49E-55	47	1.49E-55	47	1.89E-54	47	2.78E-58	49	4.90E-42	40	641
16872506-SuppTable1	Human Leukemia Yukinawa06 2000genes	2.92E-15	30	1.48E-20	32	1.74E-42	47	1.36E-45	50	1.25E-44	49	1.05E-45	50	2.26E-36	44	1505
GO:0003723	RNA binding	2.24E-04	17	1.16E-05	17	6.92E-42	50	2.63E-40	50	1.16E-40	50	3.08E-42	51	2.77E-29	42	1568

Note: NMF: Basic NMF; NMFL21: $L_{2,1}$ -norm NMF; GNMF: Graph regularized NMF; RMNMF: Robust Manifold NMF; RSGNMF: Robust NMF via joint Sparse and Graph regularized model; GDNMF: Graph regularized Discriminant NMF; orth-NMF: Orthogonal NMF. It lists the ten closely related to the DLBCL terms of P-values corresponding to different methods. The input columns represent the numbers of gene extracted by these methods. The annotation column represents the total number of genes included in the GO terms.

columns represent the numbers of the differentially expressed genes related to the GO terms, which were extracted using these methods.

The results demonstrate that these algorithms performed more efficiently for identifying genes. For example, for the term of "structural constituent of ribosome"(GO: 0003735), the basic NMF can extract 6 genes while GDNMF can extract 47 genes that correspond to the term. The genes selected by the methods were all verified and have relevance to the DLBCL tumor. Accordingly, the more genes identified by the method, the better the method will perform. Furthermore, the gene numbers have a negative correlation with the P-value. Most of the extended NMF algorithms can identify more differential genes (with lower P-values) than the basic NMF algorithm, which demonstrates that the improved NMF algorithms are efficient for identifying genes.

6.2 The Clustering Results of Different Methods

To evaluate the performance of NMF and its extended algorithms in clustering samples, in this subsection, we apply the widely used tumor sample leukemia dataset in the experiments. Acute myelogenous leukemia (AML) and acute lymphoblastic leukemia (ALL) can be easily distinguished. ALL can be further divided into T and B subtypes. The distinction between AML and ALL, as well as the division of ALL into the T and B cell subtypes, is known. The dataset contains 5000 genes in 38 samples, and consists of 19 cases of B cell ALL (ALL_B), 8 cases of T cell ALL (ALL_T), and 11 cases of AML [24], as listed in Table 2. Table 3 lists the clustering results of different NMF methods with the clustering number in $k = 2$ and $k = 3$. The incorrect clustering samples are underlined, in italics and in bold. Table 4 lists the clustering accuracy of these methods.

TABLE 2
THE SAMPLE NUMBERS OF LEUKEMIA DATASET

Types	ALL_B	ALL_T	AML
Number	19	8	11

From Table 3 and Table 4, we can find that NMFL21 has the highest clustering accuracy of 97.38% compared with the other algorithms, and except for the RSGNMF algorithm, other methods have the same clustering accuracy of 92.10% when the clustering number $k = 2$. For the clustering number $k = 3$, the RSGNMF has the same clustering accuracy as orth-NMF, which is lower than the other methods. Since GNMF considers the geometric information, it has the highest clustering accuracy of 94.83% compared with the other methods. Based on the results, we can conclude that NMF methods with manifold learning can achieve a higher clustering accuracy because they consider the geometric information for the original data distribution.

6.3 Analysis of the TCGA Dataset by Different Methods

The TCGA project plans to profile genomic changes in 20 different cancer types and to date, has published results for different cancer types [115, 116]. This project also provides clinical information about the metastatic status of individual patients via clinical stage information. The breadth of the TCGA genomic data sets provides a unique opportunity to consider different categories of genetic aberrations at individual gene resolution that have not been considered in other genomic studies [117-119].

In this subsection, the ColoRectal Cancer (CRC) data in the TCGA dataset are analyzed using NMF and its various extended methods. CRC is one of the leading causes

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2017.2665557, IEEE/ACM Transactions on Computational Biology and Bioinformatics

TABLE 3
THE CLUSTERING RESULTS OF DIFFERENT METHODS ON LEUKEMIA DATASET

Samples	NMF		NMFL21		RMNMF		RSGNMF		GDNMF		GNMF		orth-NMF		K-means	
	k=2	k=3	k=2	k=3	k=2	k=3	k=2	k=3	k=2	k=3	k=2	k=3	k=2	k=3	k=2	k=3
ALL_19769_B-cell	1	1	1	1	1	1	1	1	1	1	1	3	1	***1***	1	3
ALL_23953_B-cell	1	1	1	1	1	1	1	1	1	1	1	3	1	***1***	1	3
ALL_28373_B-cell	1	1	1	1	1	1	1	1	1	1	1	3	1	3	1	3
ALL_9335_B-cell	1	1	1	1	1	1	1	1	1	1	1	3	1	3	1	3
ALL_9692_B-cell	1	1	1	1	1	1	1	1	1	1	1	3	1	***1***	1	3
ALL_14749_B-cell	***2***	***2***	1	1	1	1	1	1	1	1	1	3	1	3	1	3
ALL_17281_B-cell	1	1	1	1	1	1	1	1	1	1	1	3	1	***1***	1	3
ALL_19183_B-cell	1	1	1	1	1	1	1	1	1	1	1	3	1	3	1	3
ALL_20414_B-cell	1	1	1	1	1	1	1	1	1	1	1	3	1	3	1	3
ALL_21302_B-cell	1	***3***	1	***2***	1	***2***	1	***2***	1	***2***	1	1	***2***	3	***2***	3
ALL_549_B-cell	1	1	1	1	1	1	1	1	1	1	1	3	1	3	1	3
ALL_17929_B-cell	1	1	1	1	1	1	1	1	1	1	1	3	1	3	1	3
ALL_20185_B-cell	1	1	1	1	1	1	1	1	1	1	1	3	1	3	1	3
ALL_11103_B-cell	1	1	1	1	1	1	1	1	1	1	1	3	1	***1***	1	3
ALL_18239_B-cell	1	1	1	1	1	1	1	1	1	1	1	3	1	3	1	3
ALL_5982_B-cell	1	1	1	1	1	1	1	1	1	1	1	3	1	3	1	3
ALL_7092_B-cell	***2***	1	1	1	1	1	1	1	1	1	1	3	1	3	1	3
ALL_R11_B-cell	1	1	1	1	1	***2***	1	1	1	1	***2***	3	***2***	3	1	3
ALL_R23_B-cell	1	1	1	1	1	1	1	1	1	1	1	3	1	3	1	3
ALL_16415_T-cell	1	3	1	2	1	2	1	2	1	2	1	1	1	1	2	1
ALL_19881_T-cell	1	3	1	2	1	2	1	2	1	2	1	1	1	1	2	1
ALL_9186_T-cell	1	3	1	2	1	2	1	2	1	2	1	1	1	1	2	1
ALL_9723_T-cell	1	3	1	2	1	2	1	2	1	2	1	1	1	1	2	1
ALL_17269_T-cell	1	3	1	2	1	2	1	2	1	2	1	1	1	1	2	1
ALL_14402_T-cell	1	3	1	2	***2***	2	1	2	1	2	1	1	1	1	2	1
ALL_17638_T-cell	1	3	***2***	2	1	***3***	1	2	1	2	***2***	***2***	1	1	2	1
ALL_22474_T-cell	1	3	1	2	1	2	1	2	1	2	1	1	1	1	2	1
AML_12	2	2	1	***2***	***1***	3	2	3	***1***	***2***	2	2	2	2	2	***3***
AML_13	***1***	***3***	1	3	***1***	3	***1***	3	***1***	3	***1***	***3***	2	2	***1***	***3***
AML_14	2	2	2	3	2	***2***	2	***1***	2	3	2	2	2	2	2	2
AML_16	2	2	2	3	2	3	2	***2***	2	3	2	2	2	2	2	2
AML_20	2	2	2	3	2	3	***1***	***2***	2	3	2	2	2	2	2	***3***
AML_1	2	2	2	***2***	2	3	***1***	3	2	3	2	2	2	2	2	***3***
AML_2	2	2	2	***2***	2	3	2	3	2	3	2	2	2	2	2	***3***
AML_3	2	2	2	3	2	3	2	3	2	3	2	2	2	2	2	2
AML_5	2	***3***	2	3	2	3	***1***	***2***	***1***	***2***	2	2	***1***	***1***	2	***3***
AML_6	2	***1***	2	3	2	3	***1***	***2***	2	3	2	2	2	2	2	***3***
AML_7	2	2	2	3	2	3	2	3	2	3	2	2	2	2	2	2

of cancer deaths worldwide, with mortality primarily resulting from metastatic disease [120]. Identifying the genetic and genomic basis of CRC has significant clinical implications. TCGA is the most comprehensive CRC genomic survey conducted to date [121], and its project relies on a combination of next generation sequencing and microarray genomic platforms to characterize different CRC genetic aberration features and individual affected genes. Genomic data were obtained from the Broad Firehose (http://gdac.broadinstitute.org) which is one of the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2017.2665557, IEEE/ACM Transactions on Computational Biology and Bioinformatics

TABLE 4
THE ACCURACY OF CLUSTERING ON LEUKEMIA DATASET

Number of types	samples	NMF	NMFL21	RMNMF	RSGNMF	GDNMF	GNMF	orth-NMF	K-means
K=2	38	92.10%	**97.38%**	92.10%	86.84%	92.10%	92.10%	92.10%	94.70%
K=3	38	86.84%	89.47%	92.10%	84.21%	92.10%	**94.83%**	84.21%	81.50%

Genome Data Analyses (GDACs) for TCGA project. In this subsection, we conduct an analysis of 197 samples with 5188 genomic features from 1325 genes which integrated exome sequences, DNA copy numbers, methylation and mRNA expression data [122].

NMF-series methods are used to identify differentially expressed genes in CRC. We select 200 genes for each method to analyze overlap among the sets of differential genes identified by using different methods.

Figure 2 shows the overlap among the sets of differentially expressed genes. From this figure, we note that only 2 genes identified by RSGNMF are not shared with other methods. In contrast, GDNMF and RMNMF identify a fair amount of differentially expressed genes that are not shared with other methods (GDNMF and RMNMF have 100 and 99 genes, respectively). Five genes are shared by all the five methods. Eventually, the five genes shared by the five methods are input into the GO tool and demonstrated to play an essential role in CRC development. The details are listed in Table 5. The first gene overlapped with these methods is FABP1 which is commonly found in liver patients. While many studies have demonstrated that approximately 50% of patients with CRC have liver metastases (CLM), patients with respected CRC and CLM can experience a 5-year survival of up to 50–60%. In conclusion, identification of the characteristic genes corresponding to the disease plays an important role in therapy and remains the best prognostic indicator.

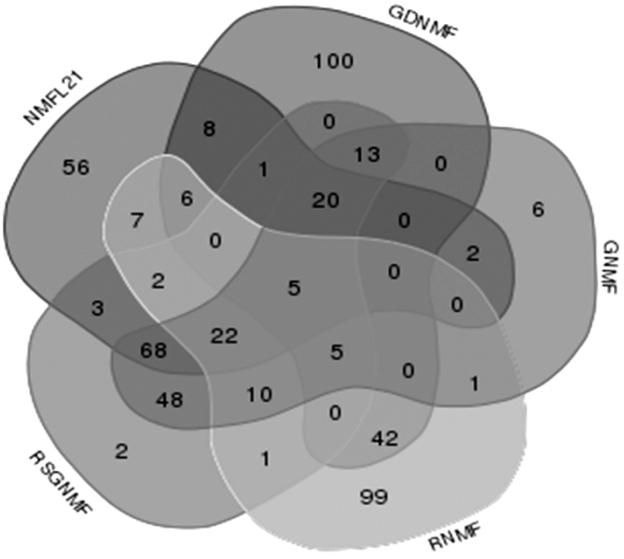


Fig. 2. Overlap among the set of genes identified by the different NMF methods. Only 2 genes identified by RSGNMF are not shared with other methods. In contrast, GDNMF and RMNMF identify a fair amount of differential genes that are not share with other methods (GDNMF and RMNMF have 100 and 99 genes, respectively). There are 5 genes shared by all the five methods.

7 CONCLUSION AND FUTURE WORKS

The basic NMF method has been widely used in data mining and machine learning fields, with its characteristic in data representation. In gene expression data in particular, we can interpret non-zero entries in the basis matrix as differentially expressed genes that capture gene expression patterns specific to different groups of samples. While the characteristic genes are identified, the problems of noisy and redundant information must be addressed. In sample clustering, the label information and geometric structure in the original data must also be considered. Thus, many researchers have enforced desirable properties on factor matrices, such as sparsity and smoothness. The sparse constraint results facilitate the interpretation to an even greater extent. The graph NMF methods consider the geometric structure and the label information in the original dataset. Thus, this paper provides a comprehensive review of NMF methods for the identification of differentially expressed genes and clustering samples. The real gene expression data and TCGA data are also examined to verify the performance of NMF algorithms.

Although many modified NMF methods have been proposed, there is still considerable space for improving the performance of NMF. For example, the following unsolved issues are very important: (1) most existing methods only provide local optimal solutions, so it is essential to introduce some global optimization techniques; (2) the scalability of NMF algorithms for large scale datasets should be improved; (3) a deep understanding of the clustering capability of NMF should be provided together with the theoretical results; (4) the applicability of NMF is not just limited to biological problems but encompasses diverse areas.

Furthermore, with the development of biological data, the TCGA Data Portal provides a platform for researchers to search, download, and analyze data sets generated by the TCGA project. The data contain clinical information, genomic characterization data, and high level sequence analysis of tumor genomes. Consequently, NMF algorithms can be applied for a further analysis of tumors.

TABLE 5
THE FIVE GENES SHARED BY DIFFERENT METHODS

Official gene symbol	Gene Name	Gene function
FABP1	fatty acid binding protein 1, liver	This gene encodes the fatty acid binding protein found in liver. Fatty acid binding proteins are a family of small, highly conserved, cytoplasmic proteins that bind long-chain fatty acids and other hydrophobic ligands. [123]
IL22RA1	interleukin 22 receptor, alpha 1	The protein encoded by this gene belongs to the class II cytokine receptor family, and has been shown to be a receptor for interleukin 22 (IL22). [124]
MALL	mal, T-cell differentiation protein-like	This gene encodes an element of the machinery for raft-mediated trafficking in endothelial cells. The encoded protein, a member of the MAL proteolipid family, predominantly localizes in glycolipid- and cholesterol-enriched membrane (GEM) rafts. [125]
RUNX3	runt-related transcription factor 3	This gene encodes a member of the runt domain-containing family of transcription factors. It functions as a tumor suppressor, and the gene is frequently deleted or transcriptionally silenced in cancer. Multiple transcript variants encoding different isoforms have been found for this gene. [126]
XDH	xanthine dehydrogenase	Xanthine dehydrogenase belongs to the group of molybdenum-containing hydroxylases involved in the oxidative metabolism of purines. [127]

ACKNOWLEDGMENT

This work was supported in part by the NSFC under grant Nos. 61572284, 61502272, 61672321, 61373027 and 61272339.

REFERENCES

[1] W. K. Härdle, and Z. Hlávka, "Principal Component Analysis," *Multivariate Statistics*, pp. 183-203: Springer, 2015.

[2] B. Scholkopft, and K.-R. Mullert, "Fisher discriminant analysis with kernels," *Neural networks for signal processing IX,* vol. 1, pp. 1, 1999.

[3] S. T. Roweis, and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science,* vol. 290, no. 5500, pp. 2323-2326, 2000.

[4] D. de Ridder, O. Kouropteva, O. Okun, M. Pietikäinen, and R. P. Duin, "Supervised locally linear embedding," *Artificial Neural Networks and Neural Information Processing—ICANN/ICONIP 2003*, pp. 333-341: Springer, 2003.

[5] D.-S. Huang, and C.-H. Zheng, "Independent component analysis-based penalized discriminant method for tumor classification using gene expression data," *Bioinformatics,* vol. 22, no. 15, pp. 1855-1862, 2006.

[6] P. Comon, "Independent component analysis, a new concept?," *Signal processing,* vol. 36, no. 3, pp. 287-314, 1994.

[7] D. V. Nguyen, and D. M. Rocke, "Tumor classification by partial least squares using microarray gene expression data," *Bioinformatics,* vol. 18, no. 1, pp. 39-50, 2002.

[8] H. Kim, and H. Park, "Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis," *Bioinformatics,* vol. 23, no. 12, pp. 1495-1502, 2007.

[9] J.-P. Brunet, P. Tamayo, T. R. Golub, and J. P. Mesirov, "Metagenes and molecular pattern discovery using matrix factorization," *Proceedings of the national academy of sciences,* vol. 101, no. 12, pp. 4164-4169, 2004.

[10] K. Devarajan, "Nonnegative matrix factorization: an analytical and interpretive tool in computational biology," *PLoS Comput Biol,* vol. 4, no. 7, pp. e1000029, 2008.

[11] K. Devarajan, and N. Ebrahimi, "Class discovery via nonnegative matrix factorization," *American Journal of Mathematical and Management Sciences,* vol. 28, no. 3-4, pp. 457-467, 2008.

[12] D. Dueck, Q. D. Morris, and B. J. Frey, "Multi-way clustering of microarray data using probabilistic sparse matrix factorization," *Bioinformatics,* vol. 21, no. suppl 1, pp. i144-i151, 2005.

[13] A. Pascual-Montano, P. Carmona-Saez, M. Chagoyen, F. Tirado, J. M. Carazo, and R. D. Pascual-Marqui, "bioNMF: a versatile tool for non-negative matrix factorization in biology," *BMC bioinformatics,* vol. 7, no. 1, pp. 366, 2006.

[14] P. Carmona-Saez, R. D. Pascual-Marqui, F. Tirado, J. M. Carazo, and A. Pascual-Montano, "Biclustering of gene expression data by non-smooth non-negative matrix factorization," *BMC bioinformatics,* vol. 7, no. 1, pp. 78, 2006.

[15] P. M. Kim, and B. Tidor, "Subsystem identification through dimensionality reduction of large-scale gene expression data," *Genome research,* vol. 13, no. 7, pp. 1706-1718, 2003.

[16] A. Heger, and L. Holm, "Sensitive pattern discovery with 'fuzzy'alignments of distantly related proteins," *Bioinformatics,* vol. 19, no. suppl 1, pp. i130-i137, 2003.

[17] A. Frigyesi, and M. Höglund, "Non-negative matrix factorization for the analysis of complex gene expression data: identification of clinically relevant

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2017.2665557, IEEE/ACM Transactions on Computational Biology and Bioinformatics

tumor subtypes," *Cancer informatics,* vol. 6, pp. 275, 2008.

[18] F. Nie, H. Huang, X. Cai, and C. H. Ding, "Efficient and robust feature selection via joint ℓ2, 1-norms minimization." pp. 1813-1821, 2010.

[19] N. Gillis, and S. Vavasis, "Fast and robust recursive algorithmsfor separable nonnegative matrix factorization," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 36, no. 4, pp. 698-714, 2014.

[20] D.-C. Lee, H. Jans, W. Martin, M. Wieler, S. McEwan, and T. Riauka, "Application of the non-negative matrix factorization technique to [[C-11]-DTBZ dynamic PET data for the early detection of Parkinson's disease." pp. 7-7, 2014.

[21] Z. Ji, G. Meng, D. Huang, X. Yue, and B. Wang, "NMFBFS: a NMF-based feature selection method in identifying pivotal clinical symptoms of Hepatocellular carcinoma," *Computational and mathematical methods in medicine,* vol. 2015, 2015.

[22] Y. Yuan, E. M. Van Allen, L. Omberg, N. Wagle, A. Amin-Mansour, A. Sokolov, L. A. Byers, Y. Xu, K. R. Hess, and L. Diao, "Assessing the clinical utility of cancer genomic and proteomic data across tumor types," *Nature biotechnology,* vol. 32, no. 7, pp. 644-652, 2014.

[23] J. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith, and M. West, "Bayesian factor regression models in the "large p, small n" paradigm," *Bayesian statistics,* vol. 7, pp. 733-742, 2003.

[24] C.-H. Zheng, D.-S. Huang, L. Zhang, and K. Xiang-Zhen, "Tumor clustering using nonnegative matrix factorization with gene selection," *IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society,* vol. 13, no. 4, 2009.

[25] Y.-X. Wang, and Y.-J. Zhang, "Nonnegative matrix factorization: A comprehensive review," *Knowledge and Data Engineering, IEEE Transactions on,* vol. 25, no. 6, pp. 1336-1353, 2013.

[26] J. A. Tropp, "Literature survey: Nonnegative matrix factorization," *University of Texas at Asutin*, 2003.

[27] T. Li, and C. H. Ding, "Nonnegative Matrix Factorizations for Clustering: A Survey," 2013.

[28] H. B. Saber, and M. ELLOUMI, "DNA Microarray Data Analysis: A New Survey on Biclustering," *International Journal for Computational Biology (IJCB),* vol. 4, no. 1, pp. 21-37, 2015.

[29] G. H. Golub, and C. Reinsch, "Singular value decomposition and least squares solutions," *Numerische mathematik,* vol. 14, no. 5, pp. 403-420, 1970.

[30] P. Paatero, and U. Tapper, " Positive matrix factorization: A non - negative factor model with optimal utilization of error estimates of data values," *Environmetrics,* vol. 5, no. 2, pp. 111-126, 1994.

[31] 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, 1999.

[32] D. D. Lee, and H. S. Seung, "Algorithms for non-negative matrix factorization." pp. 556-562, 2001.

[33] M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, "Algorithms and applications for approximate nonnegative matrix factorization," *Computational statistics & data analysis,* vol. 52, no. 1, pp. 155-173, 2007.

[34] C. Ding, T. Li, and M. Jordan, "Nonnegative matrix factorization for combinatorial optimization: Spectral clustering, graph matching, and clique finding." pp. 183-192, 2006.

[35] N. Gillis, " Nonnegative matrix factorization: Complexity, algorithms and applications," UCL, 2011.

[36] S. Sra, and I. S. Dhillon, "Generalized nonnegative matrix approximations with Bregman divergences." pp. 283-290, 2005.

[37] R. Gaujoux, and C. Seoighe, "A flexible R package for nonnegative matrix factorization," *BMC bioinformatics,* vol. 11, no. 1, pp. 367, 2010.

[38] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 33, no. 8, pp. 1548-1560, 2011.

[39] A. Cichocki, R. Zdunek, A. H. Phan, and S.-i. Amari, *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*: John Wiley & Sons, 2009.

[40] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *bioinformatics,* vol. 23, no. 19, pp. 2507-2517, 2007.

[41] Y. Gao, and H. M. S. B. M. A. U. S. A. g. m. c. r. m. h. e. Church George Department of Genetics, "Improving molecular cancer class discovery through sparse non-negative matrix factorization," *Bioinformatics,* vol. 21, no. 21, 2005.

[42] M. Ye, Y. Qian, and J. Zhou, "Multitask Sparse Nonnegative Matrix Factorization for Joint Spectral–Spatial Hyperspectral Imagery Denoising," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 53, no. 5, pp. 2621-2639, 2015.

[43] A. Sotiras, S. M. Resnick, and C. Davatzikos, "Finding imaging patterns of structural covariance via Non-Negative Matrix Factorization," *NeuroImage,* vol. 108, pp. 1-16, 2015.

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

[45] J. J.-Y. Wang, J. Z. Huang, Y. Sun, and X. Gao, "Feature selection and multi-kernel learning for adaptive graph regularized nonnegative matrix factorization," *Expert Systems with Applications,* vol. 42, no. 3, pp. 1278-1286, 2015.

[46] B. Shen, L. Si, R. Ji, and B. Liu, "Robust Nonnegative Matrix Factorization via L_1 Norm Regularization," *arXiv preprint arXiv:1204.2311*, 2012.

[47] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological),* vol. 58, no. 1, pp. 267-288, 1996.

[48] C. Ding, D. Zhou, X. He, and H. Zha, "R 1-PCA:

rotational invariant L 1-norm principal component analysis for robust subspace factorization." pp. 281-288, 2006.

[49] A. Evgeniou, and M. Pontil, "Multi-task feature learning," *Advances in neural information processing systems,* vol. 19, pp. 41, 2007.

[50] G. Obozinski, B. Taskar, and M. Jordan, "Multi-task feature selection," *Statistics Department, UC Berkeley, Tech. Rep*, 2006.

[51] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A survey of sparse representation: algorithms and applications," *IEEE Acess*, 2015.

[52] R. Peharz, and F. Pernkopf, "Sparse nonnegative matrix factorization with ℓ0-constraints," *Neurocomputing,* vol. 80, 2012.

[53] T. Gao, Y. Guo, C. Deng, S. Wang, and Q. Yu, "Hyperspectral Unmixing based on Constrained Nonnegative Matrix Factorization via Approximate L0," 2015.

[54] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations," *Constructive approximation,* vol. 13, no. 1, pp. 57-98, 1997.

[55] B. D. Rao, and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *Signal Processing, IEEE Transactions on,* vol. 47, no. 1, pp. 187-200, 1999.

[56] S. G. Mallat, and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *Signal Processing, IEEE Transactions on,* vol. 41, no. 12, pp. 3397-3415, 1993.

[57] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM journal on scientific computing,* vol. 20, no. 1, pp. 33-61, 1998.

[58] P. O. Hoyer, "Non-negative sparse coding." pp. 557-565, 2002.

[59] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal " *Communications on pure and applied mathematics,* vol. 59, no. 6, pp. 797-829, 2006.

[60] B. Shen, B.-D. Liu, Q. Wang, and R. Ji, "Robust nonnegative matrix factorization via L 1 norm regularization by multiplicative updating rules." pp. 5282-5286, 2014.

[61] M. Mouri, I. Takumi, H. Yasukawa, and A. Cichocki, "Revising algorithm for nonnegative matrix factorization based on minimizing quasi-L1 norm." pp. 767-770, 2014.

[62] B. Shen, and L. Si, "Non-Negative Matrix Factorization Clustering on Multiple Manifolds," 2010.

[63] H. Kim, and H. Park, "Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method," *SIAM Journal on Matrix Analysis and Applications,* vol. 30, no. 2, pp. 713-730, 2008.

[64] J. Eggert, and E. Körner, "Sparse coding and NMF." pp. 2529-2533, 2004.

[65] D. Kong, C. Ding, and H. Huang, "Robust nonnegative matrix factorization using l21-norm." pp. 673-682, 2011.

[66] D. Kong, R. Fujimaki, J. Liu, F. Nie, and C. Ding, "Exclusive Feature Learning on Arbitrary Structures via L1, 2 norm." pp. 1655-1663, 2014.

[67] Z. Lu, Y. Sun, Y. Hu, and B. Yin, "Robust Face Recognition Based l21-Norm Sparse Representation." pp. 25-29, 2014.

[68] D. Wang, J.-X. Liu, Y.-L. Gao, J. Yu, C.-H. Zheng, and Y. Xu, "An NMF-L 2, 1-Norm Constraint Method for Characteristic Gene Selection," *PloS one,* vol. 11, no. 7, pp. e0158494, 2016.

[69] H. Zou, and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* vol. 67, no. 2, pp. 301-320, 2005.

[70] Y. Li, and A. Ngom, "Versatile sparse matrix factorization and its applications in high-dimensional biological data analysis," *Pattern Recognition in Bioinformatics,* pp. 91-101: Springer, 2013.

[71] Y. Li, and A. Ngom, "Sparse representation approaches for the classification of high-dimensional biological data," *BMC systems biology,* vol. 7, no. Suppl 4, pp. S6, 2013.

[72] C. Ding, T. Li, and M. Jordan, "Convex and semi-nonnegative matrix factorizations," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 32, no. 1, pp. 45-55, 2010.

[73] D. P. Bertsekas, *Nonlinear programming*: Athena scientific Belmont, 1999.

[74] J. Huang, F. Nie, H. Huang, and C. Ding, "Robust manifold nonnegative matrix factorization," *ACM Transactions on Knowledge Discovery from Data (TKDD),* vol. 8, no. 3, pp. 11, 2014.

[75] W. Hu, K.-S. Choi, P. Wang, Y. Jiang, and S. Wang, "Convex nonnegative matrix factorization with manifold regularization," *Neural Networks,* vol. 63, pp. 94-103, 2015.

[76] H.-J. Yu, and D.-S. Huang, "Graphical representation for DNA sequences via joint diagonalization of matrix pencil," *Biomedical and Health Informatics, IEEE Journal of,* vol. 17, no. 3, pp. 503-511, 2013.

[77] 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, 2000.

[78] T. Roughgarden, and F. Schoppmann, "Local smoothness and the price of anarchy in splittable congestion games," *Journal of Economic Theory,* vol. 156, pp. 317-342, 2015.

[79] D. Cai, X. He, X. Wang, H. Bao, and J. Han, "Locality Preserving Nonnegative Matrix Factorization." pp. 1010-1015, 2009.

[80] A. B. Hamza, and D. J. Brady, "Reconstruction of reflectance spectra using robust nonnegative matrix factorization," *Signal Processing, IEEE Transactions on,* vol. 54, no. 9, pp. 3637-3642, 2006.

[81] P. Fogel, S. S. Young, D. M. Hawkins, and N. Ledirac, "Inferential, robust non-negative matrix factorization analysis of microarray data," *Bioinformatics,* vol. 23, no. 1, pp. 44-49, 2007.

[82] M. Belkin, and P. Niyogi, "Laplacian eigenmaps and

spectral techniques for embedding and clustering." pp. 585-591, 2001.

[83] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping." pp. 1735-1742, 2006.

[84] Z.-H. You, Y.-K. Lei, J. Gui, D.-S. Huang, and X. Zhou, "Using manifold embedding for assessing and predicting protein interactions from high-throughput experimental data," *Bioinformatics,* vol. 26, no. 21, pp. 2744-2751, 2010.

[85] M. Belkin, P. Niyogi, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from labeled and unlabeled examples," *The Journal of Machine Learning Research,* vol. 7, pp. 2399-2434, 2006.

[86] D. Cai, X. Wang, and X. He, "Probabilistic dyadic data analysis with local and global consistency." pp. 105-112, 2009.

[87] D. Cai, X. He, X. Wu, and J. Han, "Non-negative matrix factorization on manifold." pp. 63-72, 2008.

[88] J. J.-Y. Wang, and X. Gao, "Learning manifold to regularize nonnegative matrix factorization," *arXiv preprint arXiv:1410.2191*, 2014.

[89] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," *Advances in neural information processing systems,* vol. 16, no. 16, pp. 321-328, 2004.

[90] X. Zhu, and J. Lafferty, "Harmonic mixtures: combining mixture models and graph-based methods for inductive and scalable semi-supervised learning." pp. 1052-1059, 2005.

[91] F. R. Chung, *Spectral graph theory*: American Mathematical Soc., 1997.

[92] X. Liu, D. Zhai, D. Zhao, G. Zhai, and W. Gao, "Progressive image denoising through hybrid graph laplacian regularization: a unified framework," *IEEE transactions on image processing: a publication of the IEEE Signal Processing Society,* vol. 23, no. 4, pp. 1491-1503, 2014.

[93] D. Cai, X. He, J. Han, and S. H. Thomas, "Graph Regularized Nonnegative Matrix Factorization for Data Representation," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. W CNKI, 2011.

[94] G.-C. Wu, and D. Baleanu, "Variational iteration method for the Burgers' flow with fractional derivatives—new Lagrange multipliers," *Applied Mathematical Modelling,* vol. 37, no. 9, pp. 6183-6190, 2013.

[95] F. Facchinei, C. Kanzow, and S. Sagratella, "Solving quasi-variational inequalities via their KKT conditions," *Mathematical Programming,* vol. 144, no. 1-2, pp. 369-412, 2014.

[96] S. Yang, C. Hou, C. Zhang, Y. Wu, and S. Weng, "Robust non-negative matrix factorization via joint sparse and graph regularization." pp. 1-5, 2013.

[97] D. Wang, J.-X. Liu, Y.-L. Gao, C.-H. Zheng, and Y. Xu, "Characteristic Gene Selection based on Robust Graph Regularized Non-negative Matrix Factorization," *IEEE Transactions on Computational Biology and Bioinformatics DOI:10.1109/TCBB.2015.2505294*, 2015.

[98] Q. Zhang, and B. Li, "Discriminative K-SVD for dictionary learning in face recognition." pp. 2691-2698, 2010.

[99] J. Dou, J. Li, Q. Qin, and Z. Tu, "Robust visual tracking based on incremental discriminative projective non-negative matrix factorization," *Neurocomputing*, pp. 210–228, 2015.

[100] Y. Ghanbari, A. R. Smith, R. T. Schultz, and R. Verma, "Identifying group discriminative and age regressive sub-networks from DTI-based connectivity via a unified framework of non-negative matrix factorization and graph embedding," *Medical image analysis,* vol. 18, no. 8, pp. 1337-1348, 2014.

[101] X. Long, H. Lu, Y. Peng, and W. Li, "Graph regularized discriminative non-negative matrix factorization for face recognition," *Multimedia Tools and Applications,* vol. 72, no. 3, pp. 2679-2699, 2014.

[102] Y. Yi, Y. Shi, H. Zhang, J. Wang, and J. Kong, "Label propagation based semi-supervised non-negative matrix factorization for feature extraction," *Neurocomputing,* vol. 149, pp. 1021-1037, 2015.

[103] D. Wang, X. Gao, and X. Wang, "Semi-Supervised Nonnegative Matrix Factorization via Constraint Propagation," 2015.

[104] B. Li, G. Zhou, and A. Cichocki, "Two Efficient Algorithms for Approximately Orthogonal Nonnegative Matrix Factorization," *Signal Processing Letters, IEEE,* vol. 22, no. 7, pp. 843-846, 2015.

[105] S. Z. Li, X. W. Hou, H. Zhang, and Q. Cheng, "Learning spatially localized, parts-based representation." pp. I-207-I-212 vol. 1, 2001.

[106] C. Ding, T. Li, W. Peng, and H. Park, "Orthogonal nonnegative matrix t-factorizations for clustering." pp. 126-135, 2006.

[107] K. Kimura, Y. Tanaka, and M. Kudo, "A Fast Hierarchical Alternating Least Squares Algorithm for Orthogonal Nonnegative Matrix Factorization." pp. 129-141, 2014.

[108] P. Li, J. Bu, Y. Yang, R. Ji, C. Chen, and D. Cai, "Discriminative orthogonal nonnegative matrix factorization with flexibility for data representation," *Expert Systems With Applications,* vol. 41, no. 4, pp. 1283-1293, 2014.

[109] A. Mirzal, "A convergent algorithm for orthogonal nonnegative matrix factorization," *Journal of Computational and Applied Mathematics,* vol. 260, pp. 149-166, 2014.

[110] Z. XiuJun, and L. Chang, "Generalized Discriminant Orthogonal Nonnegative Tensor Factorization for Facial Expression Recognition," *The Scientific World Journal,* vol. 2014, 2014.

[111] C.-H. Zheng, D.-S. Huang, L. Zhang, and X.-Z. Kong, "Tumor clustering using nonnegative matrix factorization with gene selection," *Information Technology in Biomedicine, IEEE Transactions on,* vol. 13, no. 4, pp. 599-607, 2009.

[112] L. Veltri, B. N. Savani, M. A. Kharfan - Dabaja, M.

Hamadani, and A. S. Kanate, "Diffuse large B - cell lymphoma, " *Clinical Guide to Transplantation in Lymphoma*, pp. 145-159, 2015.

[113] R. Tembhare Prashant, P. G. Subramanian, K. Sehgal, B. Yajamanam, and A. Kumar, "Hypergranular precursor B-cell acute lymphoblastic leukemia in a 16-year-old boy," *Indian Journal of Pathology & Microbiology,* vol. 52, no. 3, 2009.

[114] G. O. Consortium, "Gene ontology consortium: going forward," *Nucleic acids research,* vol. 43, no. D1, pp. D1049-D1056, 2015.

[115] R. G. Verhaak, K. A. Hoadley, E. Purdom, V. Wang, Y. Qi, M. D. Wilkerson, C. R. Miller, L. Ding, T. Golub, and J. P. Mesirov, "Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1," *Cancer cell,* vol. 17, no. 1, pp. 98-110, 2010.

[116] C. G. A. R. Network, "Integrated genomic analyses of ovarian carcinoma," *Nature,* vol. 474, no. 7353, pp. 609-615, 2011.

[117] R. Shen, A. B. Olshen, and M. Ladanyi, "Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis," *Bioinformatics,* vol. 25, no. 22, pp. 2906-2912, 2009.

[118] E. A. Vucic, K. L. Thu, K. Robison, L. A. Rybaczyk, R. Chari, C. E. Alvarez, and W. L. Lam, "Translating cancer 'omics' to improved outcomes," *Genome research,* vol. 22, no. 2, pp. 188-195, 2012.

[119] L. Chin, W. C. Hahn, G. Getz, and M. Meyerson, "Making sense of cancer genomic data," *Genes & development,* vol. 25, no. 6, pp. 534-555, 2011.

[120] R. Siegel, C. DeSantis, and A. Jemal, "Colorectal cancer statistics, 2014," *CA: a cancer journal for clinicians,* vol. 64, no. 2, pp. 104-117, 2014.

[121] C. G. A. Network, "Comprehensive molecular characterization of human colon and rectal cancer," *Nature,* vol. 487, no. 7407, pp. 330-337, 2012.

[122] H. Lee, P. Flaherty, and H. P. Ji, "Systematic genomic identification of colorectal cancer genes delineating advanced from early clinical stage and metastasis," *BMC medical genomics,* vol. 6, no. 1, pp. 54, 2013.

[123] J. Storch, and A. E. Thumser, "Tissue-specific functions in the fatty acid-binding protein family," *Journal of Biological Chemistry,* vol. 285, no. 43, pp. 32679-32683, 2010.

[124] R. P. Donnelly, F. Sheikh, S. V. Kotenko, and H. Dickensheets, "The expanded family of class II cytokines that share the IL-10 receptor-2 (IL-10R2) chain," *Journal of leukocyte biology,* vol. 76, no. 2, pp. 314-321, 2004.

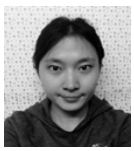
[125] A. Mungall, S. Palmer, S. Sims, C. Edwards, J. Ashurst, L. Wilming, M. Jones, R. Horton, S. Hunt, and C. Scott, "The DNA sequence and analysis of human chromosome 6," *Nature,* vol. 425, no. 6960, pp. 805-811, 2003.

[126] J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, and R. A. Holt, "The sequence of the human genome," *science,* vol. 291, no. 5507, pp. 1304-1351, 2001.

[127] T. Nishino, K. Okamoto, B. T. Eger, E. F. Pai, and T. Nishino, "Mammalian xanthine oxidoreductase–mechanism of transition from xanthine dehydrogenase to xanthine oxidase," *FEBS journal,* vol. 275, no. 13, pp. 3278-3289, 2008.

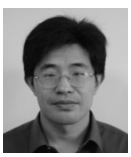
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCBB.2017.2665557, IEEE/ACM Transactions on Computational Biology and Bioinformatics

Jin-Xing Liu received the B.S. degree in electronic information and electrical engineering from Shandong University, China, in 1997; the M.S. degree in control theory and control engineering from QuFu Normal University, China, in 2003; and the Ph.D. degree in computer simulation and control from the South China University of Technology, China, in 2008. He is an associate professor at School of Information Science and Engineering, Qufu Normal University. His research interests include pattern recognition, machine learning, and bioinformatics.



Dong Wang received the B.S. degree in Electronic information science and technology from HeZe University, China, in 2013; the M.S degree in communication and information system from QuFu Normal University, China, in 2016. the Ph.D. degree candidate in the State Key Laboratory of Engineering in Surveying, Mapping and Remote Sensing from Wuhan University, China. Her research interests include pattern recognition, bioinformatics, data mining and information safety.

Ying-Lian Gao received her B.S. and M.S. degrees at QuFu Normal University, China, in 1997 and 2000, respectively.
Now she is currently with the Library of Qufu Normal University. Her current interests include data mining and pattern recognition.



Chun-Hou Zheng received the B.S. degree in physics education and the M.S. degree in control theory and control engineering from QuFu Normal University, China, in 1995 and 2001, respectively, and the Ph.D. degree in pattern recognition and intelligent system from the University of Science and Technology of China in 2006. He is currently with the College of Electrical Engineering and Automation, Anhui University, Hefei, Anhui, China. His research interests include pattern recognition and bioinformatics.

Yong Xu received his B.S. and M.S. degrees at Air Force Institute of Meteorology (China) in 1994 and 1997, respectively. He then received his Ph.D. degree in pattern recognition and intelligence system at the Nanjing University of Science and Technology in 2005. From May 2005 to April 2007, he worked at Shenzhen graduate school, Harbin Institute of Technology as a postdoctoral research fellow. Now he is a professor at Shenzhen graduate school, HIT. His current interests include pattern recognition, machine learning, and bioinformatics.

Jiguo Yu received the PhD degree in School of mathematics from Shandong University in 2004. From 2007, he has been a professor in the School of Computer Science, Qufu Normal University, Shandong, China. He is currently a professor in the School of Information Science and Engineering, Qufu Normal University. His main research interests include wireless networks, distributed algorithms, peer-to-peer computing and graph theory. In particular, he is interested in designing and analyzing algorithms for many computationally hard problems in networks. He is a senior member of the CCF (China Computer Federation).