Correntropy-Based Hypergraph Regularized NMF for Clustering and Feature Selection on Multi-Cancer Integrated Data

Na Yu, Ming-Juan Wu, Jin-Xing Liu, *Member, IEEE*, Chun-Hou Zheng, *Member, IEEE*, and Yong Xu, *Senior Member, IEEE*

***Abstract*—Non-negative matrix factorization (NMF) has become one of the most powerful methods for clustering and feature selection. However, the performance of the traditional NMF method severely degrades when the data contain noises and outliers or the manifold structure of the data is not taken into account. In this article, a novel method called correntropy-based hypergraph regularized NMF (CHNMF) is proposed to solve the above problem. Specifically, we use the correntropy instead of the Euclidean norm in the loss term of CHNMF, which will improve the robustness of the algorithm. And the hypergraph regularization term is also applied to the objective function, which can explore the high-order geometric information in more sample points. Then, the half-quadratic (HQ) optimization technique is adopted to solve the complex optimization problem of CHNMF. Finally, extensive experimental results on multi-cancer integrated data indicate that the proposed CHNMF method is superior to other state-of-the-art methods for clustering and feature selection.**

***Index Terms*—Correntropy, clustering, feature selection, hypergraph regularization, non-negative matrix factorization (NMF).**

I. Introduction

With the advent of the biomolecular era, research on gene expression data has become an unstoppable trend. This is because these data contain important information that regulates gene expression. Research on rich information of genetic activity will lay the foundation for exploring the nature of life and the prevention and treatment of diseases. It is well known that gene expression data also have the characteristics of high-dimensional small samples and sample imbalance. Multiview data formed by integrating data are receiving more attention from scholars [1], [2]. Dimensionality reduction technology is one of the effective tools for mining multiview data. In this article, multiview data consist of different cancers with the same genes. We call them multi-cancer integrated data. These data can effectively solve the problem of sample imbalance when clustering. And multi-cancer integrated data give us an unprecedented opportunity to explore the potential relationships between multiple cancers.

In the field of bioinformatics, important applications for gene expression data include sample clustering and feature selection [3]. Sample clustering is to divide a series of samples according to similarity. In this way, samples with high similarity will be divided into one subset [4]. Sample clustering contributes to the targeted treatment of cancer and the discovery of new cancer subtypes. In general, only a small percentage of genes are involved in the diseases. Therefore, the identification of abnormally expressed genes by feature selection is crucial for the study of cancer. The abnormal genes selected in the multi-cancer integrated data are likely to be associated with a variety of cancers. This provides us with a new direction to study multi-cancer.

As an effective dimension reduction method, non-negative matrix factorization (NMF) [5] has been frequently applied to data analysis tasks. Due to the existence of non-negative constraints, NMF produces a natural part-based representation. To improve the performance of the original NMF, many NMF extensions have been developed from different aspects [6]–[8]. For example, Song *et al.* [9] used the alternating direction method of multiplier (ADMM) to effectively optimize the NMF method. To preserve the intrinsic geometry of the data space, Cai *et al.* [10] proposed the graph regularized NMF (GNMF), which considers the pairwise geometric relationships between samples by constructing a simple graph. Guan *et al.* [11] proposed the manifold regularized discriminative NMF (MD-NMF) by considering the geometry of the data and the discriminative information of different classes. The graph dual regularization NMF (DNMF), which considers both the sample manifold and the feature manifold, was used for co-clustering [12]. After that, Zeng *et al.* [13] proposed the hypergraph regularized NMF (HNMF) that was

Manuscript received August 21, 2019; revised November 30, 2019, March 13, 2020, and May 23, 2020; accepted June 2, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61872220, Grant 61873001, and Grant 61572284. This article was recommended by Associate Editor D. Tao. *(Corresponding author: Jin-Xing Liu.)*

Na Yu is with the School of Information Science and Engineering, Qufu Normal University, Rizhao 276826, China (e-mail: yunacsw@126.com).

Ming-Juan Wu, Jin-Xing Liu, and Chun-Hou Zheng are with the School of Information Science and Engineering, Qufu Normal University, Rizhao 276826, China (e-mail: mingjuansw@163.com; sdcavell@126.com; zhengch99@126.com).

Yong Xu is with the Bio-Computing Research Center, Harbin Institute of Technology, Shenzhen 518055, China (e-mail: yongxu@ymail.com).

This article has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2020.3000799

used to encode high-order relationships between multiple sample points. Because the hyperedge of the hypergraph contains multiple vertices, it can better reflect the complex structure inherent in the data.

As we have introduced above, these methods all use the Euclidean norm to minimize the distance between the original data matrix and the reconstruction matrix. In addition, due to its comprehensibility and mathematical tractability under the Gaussian assumption, the Euclidean norm is also widely used in most NMF-based methods. However, many real-world data contain Gaussian noises, non-Gaussian noises (e.g., in the process of measuring and collecting gene expression data), or outliers. In this case, the performance of the NMF methods based on the Euclidean norm will be drastically reduced. Fortunately, the correntropy has proven to be effective in dealing with noises and outliers [14]. As a measure of nonlinearity and local similarity, the correntropy is related to the similarity probability of two random variables [15]. Unlike the Euclidean norm that can only consider second-order moment, the correntropy can consider the higher order moments [16]. Therefore, the correntropy can effectively improve the robustness of the algorithm, and many variants of the NMF method based on the correntropy have been proposed. For example, maximum correntropy criterion-based NMF (MCC-NMF) was developed for cancer clustering [17]. Wang *et al.* [18] proposed a correntropy GNMF (CGNMF) for image clustering. However, correntropy-based HNMF (CHNMF) method has not yet been studied.

Inspired by the above, we design a novel CHNMF method for clustering and feature selection. Specifically, CHNMF can not only effectively reduce the influence of noises and outliers but also consider the high-order geometric relationships inherent in the data. The main contributions of this article are as follows.

1) A novel method called CHNMF is presented to enhance the performance of traditional NMF. CHNMF uses the correntropy measure instead of the Euclidean norm in the objective function, which facilitates capturing the higher order moments of the data. Furthermore, the robustness of the CHNMF algorithm to noises and outliers is improved. In addition, the CHNMF method is also applied with the hypergraph regularization term to consider high-order geometric information of the data.
2) The half-quadratic (HQ) optimization technique [19] is employed in the proposed CHNMF method. It can transform nonconvex optimization problems into iterative weighted NMF problems. The multiplicative iterative algorithm is then used to solve the CHNMF. In addition, we derive a general algorithm for any data matrix. We also analyze the convergence, computational complexity, relation with the gradient descent method, and robustness of the proposed CHNMF algorithm.
3) We perform comprehensive experiments to validate the CHNMF method on the multi-cancer integrated data. The experimental results show that our method is meaningful and superior to the other four representative methods.

The remainder of this article is organized as follows. NMF, correntropy, and hypergraph regularization are simply reviewed in Section II. In Section III, the CHNMF method is described in detail. The experimental results of clustering and feature selection are demonstrated in Section IV. This article is concluded in Section V.

II. Related Work

A. Non-Negative Matrix Factorization

NMF is an effective data analysis technique that focuses on the fact that data elements are non-negative. In the field of bioinformatics, gene expression datasets can be represented in the form of non-negative matrices. Columns of non-negative matrices are used as samples, and rows are used as expression levels of genes in these samples.

Given a data matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n] \in R^{m\times n}$, where $\mathbf{x}_j$ is a sample vector containing m elements. The goal of NMF is to seek the product of two non-negative matrices $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k] \in R^{m\times k}$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k] \in R^{n\times k}$ to approximate the data matrix $\mathbf{X}$, that is, $\mathbf{X} \approx \mathbf{U}\mathbf{V}^{\mathrm{T}}$ [20]. The most commonly used measure in the NMF objective function is the Euclidean distance, which can be expressed as the following optimization problem:

$$\min_{\mathbf{U},\mathbf{V}} \left\|\mathbf{X} - \mathbf{U}\mathbf{V}^{\mathrm{T}}\right\|_F^2 \quad \text{s.t.} \quad \mathbf{U} \geq 0, \quad \mathbf{V} \geq 0 \tag{1}$$

where $\|\cdot\|_F$ represents the Frobenius norm of the matrix. $[\,\cdot\,]^{\mathrm{T}}$ is the transpose operator. Since $\mathbf{x}_j$ can be approximated a linear combination of the columns of $\mathbf{U}$, where the weight is each row of $\mathbf{V}$ [21]. $\mathbf{U}$ and $\mathbf{V}$ are referred to as the basis matrix and the coefficient matrix, respectively. Then, the multiplicative update rules are shown as follows:

$$\mathbf{U} \leftarrow \mathbf{U}\frac{XV}{\mathbf{U}\mathbf{V}^{\mathrm{T}}\mathbf{V}} \tag{2}$$

$$\mathbf{V} \leftarrow \mathbf{V}\frac{\mathbf{X}^{\mathrm{T}}\mathbf{U}}{\mathbf{V}\mathbf{U}^{\mathrm{T}}\mathbf{U}}. \tag{3}$$

B. Correntropy

In real-world applications, efficient processing of noises or outliers is an intractable problem. Recently, correntropy has been proposed for robustness analysis in information theoretic learning (ITL), and has been widely used in signal processing [14], biological information [17], facial recognition, and other fields [22]. The correntropy is a measure of nonlinearity and local similarity for two random variables x and y. It is defined as follows:

$$C(x, y) = E\big[k(x, y)\big] \tag{4}$$

where $E[\,\cdot\,]$ and $k(\cdot,\cdot)$ are the expectation operator and the kernel function satisfying the Mercer theory, respectively. In this article, the commonly used Gaussian kernel is used as a kernel function of the correntropy, given by

$$k_\sigma(x, y) = g(x - y) = \exp\left(-\frac{(x-y)^2}{2\sigma^2}\right) \tag{5}$$

where σ is the kernel bandwidth parameter and requires $\sigma > 0$. If x and y are vectors, then the Gaussian kernel function is $k_\sigma(x, y) = g(x - y) = \exp(-[\|x - y\|^2/2\sigma^2])$. Since the joint

distribution function of the random variables x and y is usually unknown, the available data samples $\{(\mathbf{x}_n, y_n)\}_{n=1}^N$ are finite. The sample correntropy can be estimated as

$$\hat{C}_\sigma(x, y) = \frac{1}{N}\sum_{n=1}^{N} g(x_n - y_n). \tag{6}$$

The maximization of the correntropy in (6) is called the maximum correntropy criterion (MCC) [17]. Unlike the Euclidean norm that can only consider the second-order moment, the correntropy can consider higher order moments [14], [16]. Therefore, the MCC-based NMF methods have better robustness when processing data containing noises and outliers.

C. Hypergraph Regularization

With the development of graph theory, hypergraph learning has become an important tool for data representation [10], [23]. In reality, there are complex relationships between data samples. However, the graph regularization only considers the pairwise geometric relationship between the two data samples, which is not able to retain the complex structure inherent in the data more effectively. Because hypergraph regularization considers a high-order geometric relationship between multiple samples, it can better capture potential information in data [13].

Typically, triple $G = (V, E, \mathbf{W})$ is used to represent the hypergraph. V is a vertex set and E is a hyperedge set. In addition, $\mathbf{W}$ is a hyperedge weight matrix, which is a diagonal matrix. And $w(e)$ is the corresponding weight of the hyperedge e. The vertex–edge incidence matrix $\mathbf{H} \in R^{|\mathbf{V}|\times|E|}$ can be computed as

$$\mathbf{H}(v, e) = \begin{cases} 1, & \text{if } v \in e \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

where $\mathbf{H}$ is a binary matrix. Then, the weight $\mathbf{W}_i$ of each hyperedge e_i is defined as

$$\mathbf{W}_i = \mathbf{W}(e_i) = \sum_{v_j \in e_i} \exp\left(-\frac{\|v_i - v_j\|_2^2}{\delta}\right) \tag{8}$$

where $\delta = \sum_{v_j \in e_i} \|v_i - v_j\|_2^2 / k$, k is the value of k-nearest neighbors for each vertex. In addition, the degree of a vertex v can be expressed as

$$d(v) = \sum_{e \in E} w(e)\mathbf{H}(v, e). \tag{9}$$

The degree of a hyperedge e can be expressed as

$$f(e) = \sum_{v \in \mathbf{V}} \mathbf{H}(v, e). \tag{10}$$

Finally, hypergraph regularization [13] can be measured by

$$\begin{aligned} &\frac{1}{2}\sum_{e\in E}\sum_{(i,j)\in e}\frac{w(e)}{f(e)}\left\|\mathbf{v}_i - \mathbf{v}_j\right\|^2 \\ &= \mathrm{Tr}\left(\mathbf{V}^{\mathrm{T}}(\mathbf{D}_v - \mathbf{E})\mathbf{V}\right) \\ &= \mathrm{Tr}\left(\mathbf{V}^{\mathrm{T}}\mathbf{L}_{\mathrm{hyper}}\mathbf{V}\right) \end{aligned} \tag{11}$$

where $\mathbf{v}_i$ and $\mathbf{v}_j$ are low-dimensional representations of the original data points $\mathbf{x}_i$ and $\mathbf{x}_j$, respectively. $\mathbf{D}_v$ represents the diagonal matrix composed of $d(v)$. $\mathbf{E} = \mathbf{H}\mathbf{W}(\mathbf{D}_e)^{-1}\mathbf{H}^{\mathrm{T}}$, where $\mathbf{W}$ is a diagonal matrix composed of $w(e)$, and $\mathbf{D}_e$ is a diagonal matrix whose entries are $f(e)$. $\mathbf{L}_{\mathrm{hyper}}$ is called the unnormalized hypergraph Laplacian matrix [23].

III. Correntropy-Based Hypergraph Regularized Non-Negative Matrix Factorization

In this section, we introduce the CHNMF algorithm in detail.

A. Objective Function

The traditional NMF-based algorithms use the Euclidean norm measure, which is easy to solve in mathematics. However, the performance of these algorithms will be severely affected when there are noises and outliers in the original data [24]. In other words, the NMF algorithms based on the Euclidean norm are sensitive to noises and outliers. In addition, it is necessary to consider the low-dimensional manifold structure embedded in high-dimensional space for performance improvement.

To overcome these limitations, an algorithm called CHNMF is proposed. We use the correntropy instead of the Euclidean norm in the loss term of CHNMF, which will improve the robustness of the algorithm. Moreover, to explore high-order geometric information in more sample points, the hypergraph regularization term is also applied to the objective function. Therefore, the objective function of CHNMF can be written as follows:

$$\max_{\mathbf{U},\mathbf{V}} \sum_{i=1}^{M} g\left(\sqrt{\sum_{j=1}^{N}\left(\mathbf{X}_{i,j} - \left(\mathbf{U}\mathbf{V}^{\mathrm{T}}\right)_{i,j}\right)^2}\right) - \lambda\mathrm{Tr}\left(\mathbf{V}^{\mathrm{T}}\mathbf{L}_{\mathrm{hyper}}\mathbf{V}\right) \tag{12}$$

where $\lambda \geq 0$ denotes a regularization parameter for balancing the correntropy of CHNMF in the first term and the hypergraph regularization in the second term. Obviously, the objective function of CHNMF is nonquadratic and nonconvex, which is difficult to directly optimize the solution. Fortunately, the HQ technique [25] based on the convex conjugate function theory can effectively solve the above optimization problem. It converts the correntropy term in the objective function into a quadratic term of the multiplicative form [26]. By using the property of the convex conjugate function, Proposition 1 [22] is defined as follows.

Proposition 1: There exists a convex conjugate function $\phi(\cdot)$ of $g(x)$, such that

$$g(x) = \max_z\left(z\frac{\|x\|^2}{\sigma^2} - \phi(z)\right) \tag{13}$$

and for a fix x, the maximum is reached at $z = -g(x)$.

According to Proposition 1, we substitute (13) into (12). This makes it possible to obtain the augmented objective

function in an enlarged parameter space

$$\max_{\mathbf{U},\mathbf{V},\mathbf{z}} \sum_{i=1}^{M} g\left(\frac{z_i}{\sigma^2}\left(\sqrt{\sum_{j=1}^{N}\left(\mathbf{X}_{i,j}-\left(\mathbf{U}\mathbf{V}^{\mathrm{T}}\right)_{i,j}\right)^2}\right)-\phi(z_i)\right) - \lambda\mathrm{Tr}\left(\mathbf{V}^{\mathrm{T}}\mathbf{L}_{\mathrm{hyper}}\mathbf{V}\right) \tag{14}$$

where $\mathbf{z}=[z_1,\ldots,z_M]^{\mathrm{T}}$ denotes the auxiliary vector in the HQ technique. Then, maximizing the augmented objective function with respect to $\mathbf{z}$ by fixing $\mathbf{U}$ and $\mathbf{V}$, we obtain

$$\begin{aligned} z_i &= -g\left(\sqrt{\sum_{j=1}^{N}\left(\mathbf{X}_{ij}-\left(\mathbf{U}\mathbf{V}^{\mathrm{T}}\right)_{ij}\right)^2}\right) \\ &= -\exp\left(-\frac{\sum_{j=1}^{N}\left(\mathbf{X}_{ij}-\left(\mathbf{U}\mathbf{V}^{\mathrm{T}}\right)_{ij}\right)^2}{2\sigma^2}\right). \end{aligned} \tag{15}$$

The kernel bandwidth parameter σ is usually empirically determined. In this article, we refer to [17] to obtain

$$\sigma=\sqrt{\frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(\mathbf{X}_{i,j}-\left(\mathbf{U}\mathbf{V}^{\mathrm{T}}\right)_{i,j}\right)^2}. \tag{16}$$

Then, for a fixed σ, the augmented objective function in (14) can be rewritten as

$$\max_{\mathbf{U},\mathbf{V}} \sum_{i=1}^{M}\left(\frac{z_i}{\sigma^2}\left(\sum_{j=1}^{N}\left(\mathbf{X}_{i,j}-\left(\mathbf{U}\mathbf{V}^{\mathrm{T}}\right)_{i,j}\right)^2\right)\right) - \lambda\mathrm{Tr}\left(\mathbf{V}^{\mathrm{T}}\mathbf{L}_{\mathrm{hyper}}\mathbf{V}\right) \tag{17}$$

which can also be equivalent to the following optimization problem:

$$\min_{\mathbf{U},\mathbf{V}} \sum_{i=1}^{M}\left(-\frac{z_i}{\sigma^2}\left(\sum_{j=1}^{N}\left(\mathbf{X}_{i,j}-\left(\mathbf{U}\mathbf{V}^{\mathrm{T}}\right)_{i,j}\right)^2\right)\right) + \lambda\mathrm{Tr}\left(\mathbf{V}^{\mathrm{T}}\mathbf{L}_{\mathrm{hyper}}\mathbf{V}\right). \tag{18}$$

B. Optimization of CHNMF

Since we use the HQ technique to optimize the CHNMF algorithm, the optimization problem in (18) becomes a weighted NMF problem [27]. But it is nonconvex with respect to $\mathbf{U}$ and $\mathbf{V}$ together. The alternate iterative strategy can be utilized to derive a local optimal solution. Therefore, the optimization problem (18) can be further rewritten as follows:

$$\begin{aligned} f &= \mathrm{Tr}\left(\left(\mathbf{X}-\mathbf{U}\mathbf{V}^{\mathrm{T}}\right)^{\mathrm{T}}\mathbf{D}(\mathbf{X}-\mathbf{U}\mathbf{V})\right)+\lambda\mathrm{Tr}\left(\mathbf{V}^{\mathrm{T}}\mathbf{L}_{\mathrm{hyper}}\mathbf{V}\right) \\ &= \mathrm{Tr}\left(\mathbf{X}^{\mathrm{T}}\mathbf{D}\mathbf{X}\right)-2\mathrm{Tr}\left(\mathbf{V}\mathbf{U}^{\mathrm{T}}\mathbf{D}\mathbf{X}\right)+\mathrm{Tr}\left(\mathbf{V}\mathbf{U}^{\mathrm{T}}\mathbf{D}\mathbf{U}\mathbf{V}^{\mathrm{T}}\right) \\ &\quad+\lambda\mathrm{Tr}\left(\mathbf{V}^{\mathrm{T}}\mathbf{L}_{\mathrm{hyper}}\mathbf{V}\right) \end{aligned} \tag{19}$$

where $\mathbf{D}$ represents a diagonal matrix whose entries are

$$\mathbf{D}_{ii}=-\frac{z_i}{\sigma^2}=\sigma^{-2}\exp\left(-\frac{1}{2\sigma^2}\sum_{j=1}^{N}\left(\mathbf{X}_{i,j}-\left(\mathbf{U}\mathbf{V}^{\mathrm{T}}\right)_{i,j}\right)^2\right). \tag{20}$$

The multiplicative iteration method is applied to solve (19). Suppose $\boldsymbol{\psi}=[\psi_{ik}]$ and $\boldsymbol{\varphi}=[\varphi_{jk}]$ are Lagrange multipliers for constraining $\mathbf{U}\geq 0$ and $\mathbf{V}\geq 0$, respectively. Then, the Lagrange function $\mathbf{L}$ is defined as

$$\begin{aligned} \mathbf{L} &= \mathrm{Tr}\left(\mathbf{X}^{\mathrm{T}}\mathbf{D}\mathbf{X}\right)-2\mathrm{Tr}\left(\mathbf{V}\mathbf{U}^{\mathrm{T}}\mathbf{D}\mathbf{X}\right)+\mathrm{Tr}\left(\mathbf{V}\mathbf{U}^{\mathrm{T}}\mathbf{D}\mathbf{U}\mathbf{V}^{\mathrm{T}}\right) \\ &\quad+\lambda\mathrm{Tr}\left(\mathbf{V}^{\mathrm{T}}\mathbf{L}_{\mathrm{hyper}}\mathbf{V}\right)+\mathrm{Tr}\left(\boldsymbol{\psi}\mathbf{U}^{\mathrm{T}}\right)+\mathrm{Tr}\left(\boldsymbol{\varphi}\mathbf{V}^{\mathrm{T}}\right). \end{aligned} \tag{21}$$

Then, the partial derivative of $\mathbf{L}$ with respect to $\mathbf{U}$ and $\mathbf{V}$ is

$$\frac{\partial\mathbf{L}}{\partial\mathbf{U}}=-2\mathbf{D}\mathbf{X}\mathbf{V}+2\mathbf{D}\mathbf{U}\mathbf{V}^{\mathrm{T}}\mathbf{V}+\boldsymbol{\psi} \tag{22}$$

$$\frac{\partial\mathbf{L}}{\partial\mathbf{V}}=-2\mathbf{X}^{\mathrm{T}}\mathbf{D}\mathbf{U}+2\mathbf{V}\mathbf{U}^{\mathrm{T}}\mathbf{D}\mathbf{U}+2\lambda\mathbf{L}_{\mathrm{hyper}}\mathbf{V}+\boldsymbol{\varphi}. \tag{23}$$

By using the KKT conditions [28] $\boldsymbol{\psi}\mathbf{U}=0$ and $\boldsymbol{\varphi}\mathbf{V}=0$, we have

$$-\left(\mathbf{D}\mathbf{X}\mathbf{V}\right)_{ik}u_{ik}+\left(\mathbf{D}\mathbf{U}\mathbf{V}^{\mathrm{T}}\mathbf{V}\right)_{ik}u_{ik}=0 \tag{24}$$

$$-\left(\mathbf{X}^{\mathrm{T}}\mathbf{D}\mathbf{U}\right)_{jk}v_{jk}+\left(\mathbf{V}\mathbf{U}^{\mathrm{T}}\mathbf{D}\mathbf{U}\right)_{jk}v_{jk}+\lambda\left(\mathbf{L}_{\mathrm{hyper}}\mathbf{V}\right)_{jk}v_{jk}=0. \tag{25}$$

According to the above equations, the update rules are as follows:

$$u_{ik}\leftarrow u_{ik}\frac{\left(\mathbf{D}\mathbf{X}\mathbf{V}\right)_{ik}}{\left(\mathbf{D}\mathbf{U}\mathbf{V}^{\mathrm{T}}\mathbf{V}\right)_{ik}} \tag{26}$$

$$v_{jk}\leftarrow v_{jk}\frac{\left(\mathbf{X}^{\mathrm{T}}\mathbf{D}\mathbf{U}\right)_{jk}}{\left(\mathbf{V}\mathbf{U}^{\mathrm{T}}\mathbf{D}\mathbf{U}+\lambda\mathbf{L}_{\mathrm{hyper}}\mathbf{V}\right)_{jk}}. \tag{27}$$

Finally, the detailed steps of the proposed CHNMF method are summarized in Algorithm 1.

Algorithm 1 CHNMF

Data Input: $\mathbf{X}\in R^{m\times n}$
Parameters: λ
Output: $\mathbf{U}\in R^{m\times k}$, $\mathbf{V}\in R^{n\times k}$, $\mathbf{L}_{\mathrm{hyper}}\in R^{n\times n}$
Initialization: $\mathbf{U}\geq 0$, $\mathbf{V}\geq 0$
Set $r=1$.
Repeat
  Update σ by (16);
  Update $\mathbf{D}$ by (20);
  Update $\mathbf{U}$ by (26);
  Update $\mathbf{V}$ by (27);
  $r=r+1$;
Until convergence

C. Complexity Analysis

In this section, we discuss the computational complexity of the CHNMF algorithm. O notation is used to describe the computational cost. Based on update rules (26) and (27), the computational complexity of the proposed CHNMF algorithm is $O(mnk)$. Moreover, both the kernel bandwidth parameter σ and the diagonal matrix $\mathbf{D}$ need to be updated in (16) and (20), and their computational cost is also $O(mnk)$. Hypergraph requires $O(n^2m)$ to build. Assuming t is the number of iterations of the algorithm, then the total computational complexity of CHNMF is $O(tmnk+n^2m)$. The computational cost required for the graph is also $O(n^2m)$. The computational complexity of the five methods used in this article is listed in Table I. From Table I, we can observe that the computational costs of the five methods are comparable. In other words, the

TABLE I
COMPARISON OF COMPUTATIONAL COMPLEXITY OF FIVE METHODS

Methods	Computational Complexity
NMF	$O(tmnk)$
GNMF	$O(tmnk+n^2m)$
HNMF	$O(tmnk+n^2m)$
CGNMF	$O(tmnk+n^2m)$
CHNMF	$O(tmnk+n^2m)$

proposed CHNMF method does not increase the computational complexity compared to other baseline methods.

D. Connection With the Gradient Descent Method

In this section, we use the gradient descent method [29] to solve the optimization problem of CHNMF. So the update rules of the objective function (18) are as follows:

$$u_{ik} \leftarrow u_{ik} + \eta_{ik}\frac{\partial f}{\partial u_{ik}}, \quad v_{jk} \leftarrow v_{jk} + \xi_{jk}\frac{\partial f}{\partial v_{jk}} \tag{28}$$

where η_{ik} and ξ_{jk} denote the step size parameters.

Set $\eta_{ik} = -(u_{ik}/2(\mathbf{DUV}^{\mathrm{T}}\mathbf{V})_{ik})$, we obtain

$$\begin{aligned} u_{ik} + \eta_{ik}\frac{\partial f}{\partial u_{ik}} &= u_{ik} - \frac{u_{ik}}{2\left(\mathbf{DUV}^{\mathrm{T}}\mathbf{V}\right)_{ik}}\frac{\partial f}{\partial u_{ik}} \\ &= u_{ik} - \frac{u_{ik}}{2\left(\mathbf{DUV}^{T}\mathbf{V}\right)_{ik}}\left(-2\mathbf{DXV} + 2\mathbf{DUV}^{T}\mathbf{V}\right)_{ik} \\ &= u_{ik}\frac{(\mathbf{DXV})_{ik}}{\left(\mathbf{DUV}^{T}\mathbf{V}\right)_{ik}}. \end{aligned} \tag{29}$$

Similarly, set $\xi_{jk} = -(v_{jk}/2(\mathbf{VU}^{\mathrm{T}}\mathbf{DU} + \lambda\mathbf{L}_{\mathrm{hyper}}\mathbf{V})_{jk})$, then we obtain

$$\begin{aligned} v_{jk} + \xi_{jk}\frac{\partial f}{\partial v_{jk}} &= v_{jk} - \frac{v_{jk}}{2\left(\mathbf{VU}^{\mathrm{T}}\mathbf{DU} + \lambda\mathbf{L}_{\mathrm{hyper}}\mathbf{V}\right)_{jk}}\frac{\partial f}{\partial v_{jk}} \\ &= v_{jk} - \frac{v_{jk}}{2\left(\mathbf{VU}^{T}\mathbf{DU} + \lambda\mathbf{L}_{\mathrm{hyper}}\mathbf{V}\right)_{jk}} \\ &\quad \times \left(-2\mathbf{X}^{T}\mathbf{DU} + 2\mathbf{VU}^{T}\mathbf{DU} + 2\lambda\mathbf{L}_{\mathrm{hyper}}\mathbf{V}\right)_{jk} \\ &= v_{jk}\frac{\left(\mathbf{X}^{T}\mathbf{DU}\right)_{jk}}{\left(\mathbf{VU}^{T}\mathbf{DU} + \lambda\mathbf{L}_{\mathrm{hyper}}\mathbf{V}\right)_{jk}}. \end{aligned} \tag{30}$$

Clearly, using some tricks to choose the step size parameters in the gradient descent method, the multiplicative updating rules (26) and (27) can be regarded as the special cases of the gradient descent method. However, the gradient descent method cannot guarantee the non-negativity of the decomposed matrices $\mathbf{U}$ and $\mathbf{V}$. Fortunately, the multiplicative iteration method overcomes this shortcoming.

E. Robustness Analysis

In this section, we analyze the robustness of the proposed CHNMF method. Because we use the HQ technology to solve CHNMF, this transforms the nonconvex optimization problem into an iteratively weighted NMF problem. According to Proposition 1, when matrices $\mathbf{U}$ and $\mathbf{V}$ are fixed, each entry of each sample in the data matrix $\mathbf{X}$ is assigned a weight. The weight can be calculated as

$$\begin{aligned} z_i &= -g\left(\sqrt{\sum_{j=1}^{N}\left(\mathbf{X}_{ij} - \left(\mathbf{UV}^{\mathrm{T}}\right)_{ij}\right)^2}\right) \\ &= -\exp\left(-\frac{\sum_{j=1}^{N}\left(\mathbf{X}_{ij} - \left(\mathbf{UV}^{\mathrm{T}}\right)_{ij}\right)^2}{2\sigma^2}\right). \end{aligned} \tag{31}$$

Let $\mathbf{R} = \mathbf{X} - \mathbf{UV}^{\mathrm{T}}$ be the error matrix. When entries are severely damaged, this can produce large reconstruction errors. According to (31), they will be assigned a smaller weight [27]. Further, they make a smaller contribution to the objective function. Therefore, the CHNMF method has better robustness. As we all know, the NMF algorithms based on the Euclidean norm seriously suffer from noisy data. Fortunately, the correntropy is insensitive to noises and outliers, which helps improve the performance of the algorithm.

Next, we design an experiment to further illustrate the robustness of CHNMF. NMF and HNMF are used as comparison methods. They are applied together with CHNMF to a synthetic dataset consisting of 200 2-D data points. All data points are distributed in a 1-D subspace. Fig. 1(a) reveals the ability of these three methods to learn subspace on a clean dataset. In Fig. 1(b)–(d), 50, 100, and 150 data points are randomly selected as contaminated data points, respectively. It can be seen that the ability of the NMF and HNMF to recover the subspace is severely limited as the number of contaminated points increases. The ability of NMF and GNMF to recover subspace is similar. In Fig. 1(b)–(d), CHNMF can still successfully explore the subspace structure, even in the case of extreme noise points. Therefore, our proposed CHNMF method has better robustness. Comparing CHNMF with NMF and HNMF, we can get the correntropy is a robust metric. The NMF method based on correntropy has better quality.

IV. RESULTS ON MULTI-CANCER INTEGRATED DATA

In this section, clustering samples and selection of abnormally expressed genes are performed to analyze the performance of the CHNMF method. In addition, NMF [5], GNMF [10], HNMF [13], and CGNMF [18] are used as comparison methods to verify the effectiveness of the proposed method.

A. Datasets

The datasets used in these experiments are described as follows. As the largest database of cancer multiomics information, The Cancer Genome Atlas (TCGA) aims to apply high-throughput genomic analysis techniques to help people better understand cancer. At the same time, it also contains a lot of valuable information, so the in-depth study of TCGA is very necessary. In this article, we use integrated gene expression data from a variety of cancers to analyze the performance of the CHNMF method. Each cancer data can be downloaded from the TCGA (https://tcgadata.nci.nih.gov/tcga/). These data include pancreatic adenocarcinoma (PAAD_GE),

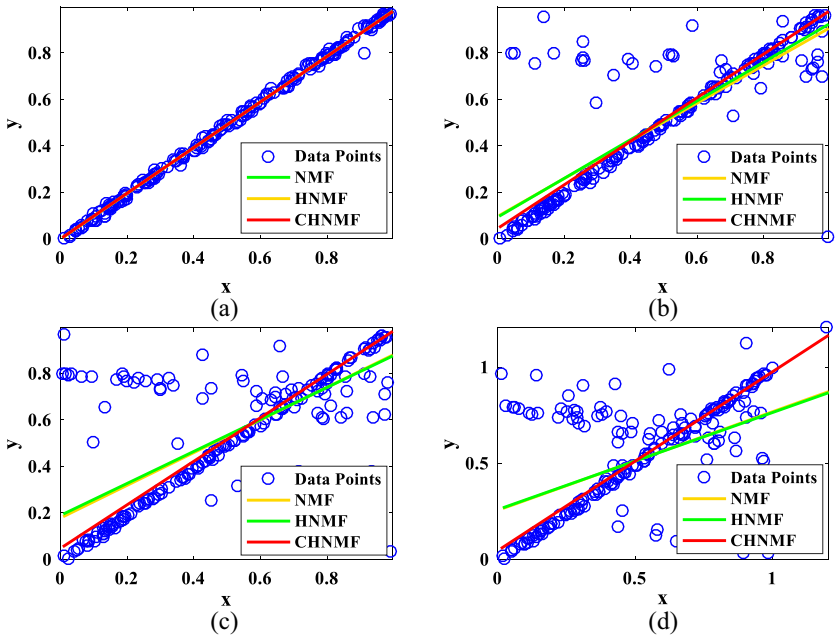


Fig. 1. Robust comparison of NMF, HNMF, and CHNMF on the synthetic dataset. (a) Clean dataset. (b) 50 contaminated data points. (c) 100 contaminated data points. (d) 150 contaminated data points.

TABLE II
SUMMARY OF FOUR MULTI-CANCER INTEGRATED DATA

Datasets	Number of		
	Classes	Samples	Genes
PAAD_HNSC_ESCA_GE	3	757	20502
PAAD_HNSC_ESCA_COAD_GE	4	1019	20502
PAAD_HNSC_ESCA_COAD_CHOL_GE	5	1055	20502
PAAD_HNSC_ESCA_COAD_CHOL_BRCA_GE	6	2157	20502

head and neck squamous cell carcinoma (HNSC_GE), esophagus carcinoma (ESCA_GE), colon adenocarcinoma (COAD), cholangiocarcinoma (CHOL_GE), and breast invasive carcinoma (BRCA_GE). The six cancers after removing the normal sample above are integrated into four datasets. Then we use principal component analysis (PCA) to reduce the dimensions of the datasets to 2000. This still retains the main information in the data. Specific information on multi-cancer integrated data is listed in Table II.

B. Parameter Setting

In the proposed CHNMF method, hypergraph regularization parameter λ needs to be selected. Its value will directly affect the experimental results. The value of λ means the degree to which high-order geometric relationships of data are explored. In order to reasonably select hypergraph regularization parameters, the five-fold cross-validation method is used. In the experiment, regularization parameter λ is automatically adjusted in $\{10^r : r \in \{-5, -4, -3, \ldots, 3, 4, 5\}\}$. Fig. 2 depicts the effect of parameter changes on CHNMF clustering performance.

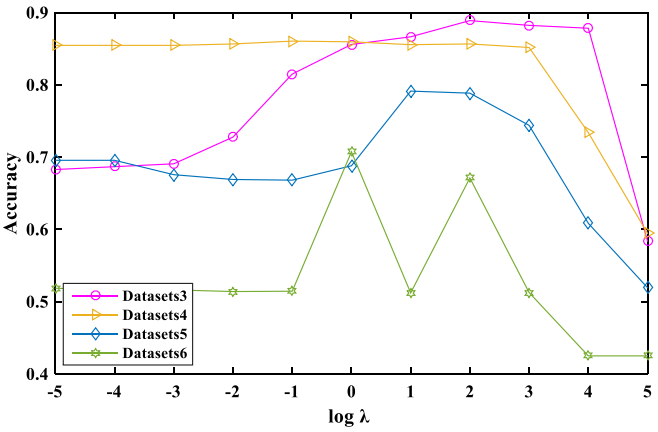


Fig. 2. Performance of the CHNMF set with different values of λ.

We can see from Fig. 2 that CHNMF is sensitive to regularization parameters. Fig. 2 suggests that the hypergraph regularization parameters λ are 10^2, 10^{-1}, 10^1, and 10^0 on PAAD_HNSC_ESCA_GE, PAAD_HNSC_ESCA_COAD_GE, PAAD_HNSC_ESCA_COAD_CHOL_GE, and PAAD_HNSC_ESCA_COAD_CHOL_BRCA_GE, respectively. For convenience, in Fig. 2, PAAD_HNSC_ESCA_GE, PAAD_HNSC_ESCA_COAD_GE, PAAD_HNSC_ESCA_COAD_CHOL_GE, and PAAD_HNSC_ESCA_COAD_CHOL_BRCA_GE are replaced with Datasets3, Datasets4, Datasets5, and Datasets6, respectively (in Fig. 3, we also use this shorthand).

C. Convergence Analysis

In this section, we design an experiment to demonstrate the convergence of the proposed CHNMF algorithm. The error value of the y-axis is the loss function value, and the iteration

TABLE III
CLUSTERING PERFORMANCE ON MULTI-CANCER INTEGRATED DATA

Datasets	Metrics	NMF	GNMF	HNMF	CGNMF	CHNMF
PAAD_HNSC_ESCA_GE	AC(%)	73.80±0.92	74.91±0.70	77.84±0.68	85.47±0.10	**88.85±0.02**
	NMI(%)	47.66±0.75	51.96±0.67	50.04±0.86	59.42±0.12	**62.89±0.05**
	SC(%)	59.92±0.22	60.35±0.17	60.78±0.24	61.38±0.01	**62.85±0.01**
PAAD_HNSC_ESCA_COAD_GE	AC(%)	75.00±0.93	72.51±0.46	76.39±0.64	81.88±0.34	**83.72±0.28**
	NMI(%)	54.18±0.42	55.79±0.36	56.06±0.32	61.18±0.09	**61.81±0.02**
	SC(%)	55.75±0.24	56.16±0.25	56.24±0.21	57.45±0.05	**58.25±0.17**
PAAD_HNSC_ESCA_COAD_CHOL_GE	AC(%)	68.23±0.15	69.09±0.48	69.98±0.55	73.95±0.17	**75.19±0.12**
	NMI(%)	52.10±0.22	52.91±0.37	55.69±0.11	57.19±0.01	**58.82±0.05**
	SC(%)	48.88±0.15	50.37±0.34	51.93±0.16	54.08±0.01	**55.46±0.04**
PAAD_HNSC_ESCA_COAD_CHOL_BRCA_GE	AC(%)	55.81±1.41	58.28±1.72	62.45±1.38	64.55±1.16	**66.03±0.11**
	NMI(%)	45.62±0.07	48.97±0.01	50.62±0.06	51.43±0.16	**53.47±0.07**
	SC(%)	66.71±0.76	69.80±1.29	70.04±1.33	71.96±1.54	**73.23±1.06**

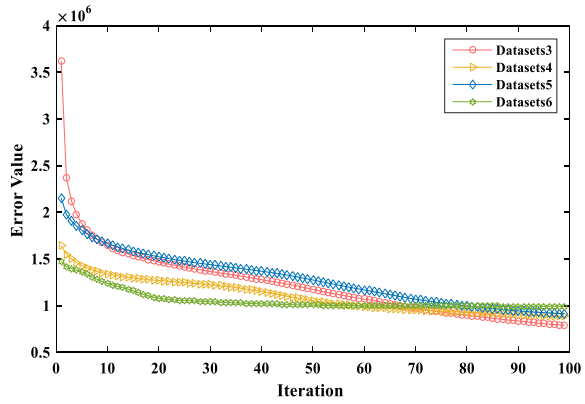


Fig. 3. Convergence curves of CHNMF on multi-cancer integrated data.

number of x-axis is set to 100. From Fig. 3, we can see that the error value gradually decreases as the number of iterations increases. This proves that CHNMF is convergent on all datasets. The CHNMF algorithm converges quickly within 100 iterations.

D. Clustering Results

In this section, to illustrate the superiority of our proposed method, clustering experiments are performed on the multi-cancer integrated data. We use the K-means algorithm to cluster on the decomposed coefficient matrix.

1) Evaluation Metrics: In clustering experiments, three commonly used metrics are adopted to strictly analyze clustering performance, namely, accuracy (AC), normalized mutual information (NMI), and silhouette coefficient (SC) [30], [31]. AC represents the percentage of samples that are correctly clustered. The AC can be calculated by

$$\mathrm{AC} = \frac{\sum_{i=1}^{n} \delta(s_i, \mathrm{map}(r_i))}{n} \times 100\% \tag{32}$$

where n represents the total number of samples, s_i is the ground-truth label provided by the dataset, and r_i is the clustering label obtained by our algorithm. map$(\cdot)$ denotes the optimal permuting function that maps the clustering label to the ground-truth label using the Kuhn–Munkres algorithm [32]. Then, $\delta(x, y)$ is a delta function. $\delta(x, y)$ is equal to 1 if $x = y$, otherwise, $\delta(x, y)$ is 0.

NMI represents the similarity of two cluster sets. Given two cluster sets C and C', C is the ground-truth cluster and C' is the cluster obtained by our algorithm. Then their mutual information (MI) is expressed as

$$\mathrm{MI}(C, C') = \sum_{c_i \in C, c'_j \in C'} p\left(c_i, c'_j\right) \log \frac{p\left(c_i, c'_j\right)}{p(c_i) p\left(c'_j\right)} \tag{33}$$

where $p(c_i)$ and $p(c'_j)$ represent the probability that any sample points belong to C and C', respectively. $p(c_i, c'_j)$ is the probability that the sample points belong to both cluster sets C and C'. Then, NMI is formulated as

$$\mathrm{NMI}(C, C') = \frac{MI(C, C')}{\max(H(C), H(C'))} \tag{34}$$

where $H(C)$ and $H(C')$ are the entropies of C and C', respectively.

SC combines the tightness and separation of clusters to evaluate clustering quality. Let i denote any sample. Assuming it belongs to cluster A, then the SC $S(i)$ of i is defined as follows:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \tag{35}$$

where $a(i)$ represents the average distance from sample i to the other samples in cluster A. $b(i)$ denotes the minimum of the average distance of sample i to all samples in other clusters C. Then, SC can be computed by

$$\mathrm{SC} = \frac{1}{n} \sum_{i=1}^{n} S(i) \tag{36}$$

where n is the total number of samples. As we all know, the values of AC and NMI range from 0 to 1. The value of SC is -1 to 1. And the larger their values, the better the clustering performance of the method.

2) Comparison of Clustering Performance: To reduce the impact of random initialization on the experimental results, we run each method 50 times. The mean and variance of AC, NMI, and SC are then calculated as the final experimental results. The clustering performance of all methods on the multi-cancer integrated data is recorded in Table III.

Based on Table III, we can draw the following conclusions.

1) On the four multi-cancer integrated datasets, compared with the GNMF method, the HNMF method can achieve 3%, 1%, and 1% improvement on the metrics AC, NMI, and SC, respectively, in average. The CHNMF method is also superior to the CGNMF method, by about 2%, 2%, and 1% on the metrics AC, NMI, and SC, respectively, on average. That is to say, the methods based on hypergraph regularization are better than that based on the graph regularization. The reason is that the hypergraph regularization preserves high-order geometric relationships whereas the graph regularization considers pairs of geometric relationships. There are complex relationships between samples in multi-cancer integration data. So considering the manifold structure between multiple samples is necessary in clustering applications.
2) With the four multi-cancer integrated datasets as a whole, the CGNMF method performs better than the GNMF method by about 8%, 5%, and 2% in terms of the mean of AC, NMI, and SC. The CHNMF method outperforms the HNMF method approximately 7%, 6%, and 3% on the metrics AC, NMI, and SC, respectively, on average. It can be seen from Table III that the methods of using correntropy measure to the objective function exceed that of using the Euclidean norm. This is because the correntropy is a local measure that takes into account the higher order moments of the data, which in turn makes the methods based on correntropy insensitive to noises and outliers.
3) The clustering result of the NMF method is not always the worst, for example on the PAAD_HNSC_ESCA_COAD_GE dataset. This implies that the improvement of the traditional NMF method may result in the loss of useful information in the data, which in turn affects the clustering results.
4) From Table III, we can see that the proposed CHNMF method has the best performance compared to other methods by more than 2%, 2%, and 1%, with respect to the average values of the metrics AC, NMI, and SC. The reason is that the correntropy measure and hypergraph regularization are used in the objective function. Therefore, the CHNMF method not only has better robustness but also encodes high-order geometric information present in the data.

E. Feature Selection

1) Abnormally Expressed Genes Selection Results: Cancer production is a consequence of genomic changes and genetic mutations, and cancer research is imminent [33], [34]. To save space, the PAAD_HNSC_ESCA_GE dataset is used to select abnormally expressed genes to test the validity of the proposed method. The selected abnormal genes may be expressed in PAAD, HNSC, and ESCA. Therefore, these genes can directly reflect the relationship between various cancers, and then more valuable biological information can be mined.

TABLE IV
ABNORMAL GENES SELECTION RESULTS ON FIVE METHODS

Methods	NMF	GNMF	HNMF	CGNMF	CHNMF
Num	113	124	128	130	**133**
IA(%)	22.6	24.8	25.6	26	**26.6**

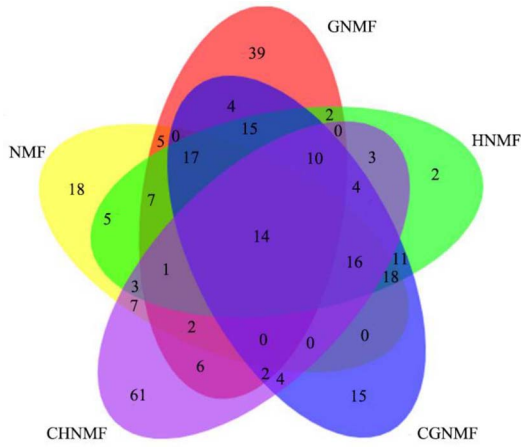


Fig. 4. Overlap among the abnormal genes identified by GeneCards.

In the experiment, we score all the genes and then arranged them in descending order. A gene with a high score is considered to be an abnormal gene. For a fair comparison, the top 500 genes for each method are selected. Then these genes are placed on the GeneGards (http://www.genecards.org/) for analysis. GeneCards is a comprehensive, searchable comprehensive database for predicting and annotating human genes. The results of the selection of abnormal genes are listed in Table IV.

In Table IV, Num is the number of abnormal genes obtained by matching the common virulence gene pools of PAAD, HNSC, and ESCA. The identification accuracy (IA) represents the proportion of abnormal genes selected by our method in the GeneCards of pathogenic genes. It can be seen from Table IV that the number of abnormal genes obtained by CHNMF is 5 more than that of HNMF. The IA of CHNMF is 1% higher than HNMF. This is because the correntropy as a robust measure is not susceptible to noises and outliers. Furthermore, the robustness of the enhancement algorithm is important for improving the performance of the feature selection experiment. From Table IV, we can observe that CHNMF is 3 and 0.6% more than CGNMF on Num and IA, respectively. Hypergraph regularization can capture more complex relationships among samples than graph regularization. It will make the potential geometric information considered more comprehensive. This in turn contributes to the improvement of the experimental results. In addition, the feature selection performance of CHNMF is better than NMF and GNMF. It has the highest Num and IA. This shows that CHNMF is reasonable to

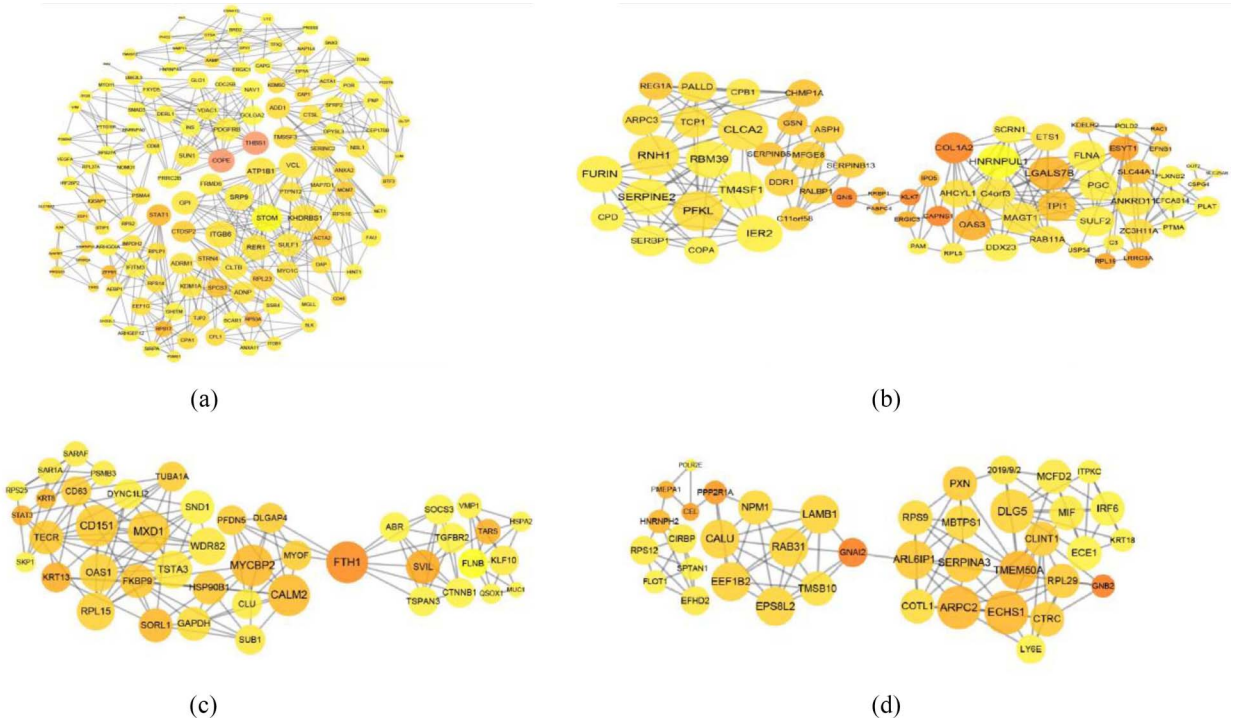


Fig. 5. Gene co-expression network of four modules. The number of nodes in the four modules (a)–(d) decreases in turn.

enhance the robustness of the algorithm and to consider the high-order manifold structure in the feature selection.

Here, we use the Venn diagram to overlap the abnormal genes identified in the GeneCards. We can see from Fig. 4, the number of abnormal genes that can be selected by CHNMF without being selected by other methods is 61. In other words, the CHNMF method has the largest number of unique genes. At the same time, these genes are highly correlated with PAAD, HNSC, and ESCA according to GeneCards. As shown in Table IV, CHNMF has the highest IA. Therefore, the proposed CHNMF is effective in selecting abnormal genes. In addition, the number of abnormal genes selected by the five methods is 14. In the next section, we conduct an in-depth study of the selected abnormal genes.

2) Discussion of Abnormal Genes: Some genes selected by CHNMF that have been ignored by the other methods makes a meaningful contribution to the study of cancer. Table V lists the details of unique abnormal genes with the relevance scores greater than 11. Relevance scores indicate the correlation between genes and diseases. When the score of a gene is higher, it means that the gene is likely to be a pathogenic gene. In Table V, the relevance scores for STAT3 are 46.11, 100.46, and 33.6 in PAAD, HNSC, and ESCA, respectively. The protein encoded by this gene is a member of the STAT protein family. STAT3 plays a key role in many cellular processes, such as cell growth and apoptosis. Published articles confirmed the relationship between STAT3 and the emergence of PAAD, HNSC, and ESCA [35]–[37]. VEGFA induces proliferation and migration of vascular endothelial cells, and is essential for both physiological and pathological angiogenesis. And its expression is correlated with tumor stage and progression. VEGFA has been shown to be associated with PAAD, HNSC, and ESCA [38]–[40]. H19 is an RNA Gene, and is affiliated with the noncoding RNA class. H19 promotes the metastasis of PAAD by antagonizing let-7 [41]. And overexpression of H19 causes HNSC and ESCA production [42], [43]. Through the above analysis, we can get the mutation of one gene that may cause many kinds of cancers. So the analysis of the unique abnormal genes selected for CHNMF helps to explore the link between PAAD, HNSC, and ESCA.

In recent years, gene co-expression networks have become an important tool for disease research. It can reveal the relationship between genes and diseases from a macro perspective. And the pathogenic genes and pathogenic modules in the network contain a wealth of information. Here, we use 500 abnormal genes selected by CHNMF to construct a gene co-expression network. Cytoscape is then used to visualize the network. Fig. 5 shows only modules with at least 40 nodes.

In Fig. 5, the larger the node is, the higher the degree is. The degree of a node represents the number of edges in the network that are directly connected to this node. A node with a large degree is called a hub node. So we use the degree measure for node mining. To save space, the first five nodes with the largest degree are then analyzed in GeneCards. Table VI summarizes the details of these five genes.

In Table VI, ATP1B1 is a protein coding gene. It can encode a beta 1 subunit. And it is related to the emergence of PAAD, ESCA, and breast cancer [44]–[46]. KHDRBS1 may be involved in a variety of cellular processes, including tumorigenesis, and regulation of human immunodeficiency virus gene expression. Therefore, its mutation can cause cancer. Gene ontology annotations related to ITGB6 include virus receptor activity. And it has lots to do with the transfer of ESCA [47]. RER1 is involved in the retention of endoplasmic

TABLE V
SUMMARY OF THE ABNORMAL GENES SELECTED ONLY BY THE CHNMF METHOD

Gene ID	Gene ED	Related Go Annotations	Related Diseases	Relevance scores
6774	STAT3	DNA binding transcription factor activity and sequence-specific DNA binding	Autoimmune Disease, Multisystem, Infantile-Onset, 1 and Hyper-Ige Recurrent Infection Syndrome, Autosomal Dominant	46.11, 100.46, 33.6
7422	VEGFA	rotein homodimerization activity and protein heterodimerization activity	Microvascular Complications Of Diabetes 1 and Poems Syndrome	40.28, 81.7, 45.44
283120	H19	None	Wilms Tumor 2 and Beckwith-Wiedemann Syndrome	38.33, 82.01, 45.22
3480	IGF1R	identical protein binding and protein kinase activity	Insulin-Like Growth Factor I and Ring Chromosome 15	25.2, 46.41, 27.09
4088	SMAD3	DNA binding transcription factor activity and sequence-specific DNA binding	Loeys-Dietz Syndrome 3 and Cutaneous Polyarteritis Nodosa	21.93, 24.93, 13.77
5268	SERPINB5	serine-type endopeptidase inhibitor activity	Syringoma and Bone Squamous Cell Carcinoma	20.57, 26.11, 11.37
4318	MMP9	identical protein binding and metalloendopeptidase activity	Metaphyseal Anadysplasia 2 and Metaphyseal Anadysplasia	19.32, 99.31, 35.54
4363	ABCC1	transporter activity and ATPase activity, coupled to transmembrane movement of substances	Dubin-Johnson Syndrome and Pseudoxanthoma Elasticum	13.34, 29.96, 11.15

Note: 'None' indicates that related GO annotation is not given in GeneCards. The unique abnormal genes with the relevance scores greater than 11 are listed in the Table V.

TABLE VI
TOP 5 GENES WITH HIGHER DEGREE

Gene	Summary
ATP1B1	The protein encoded by this gene belongs to the family of Na+/K+ and H+/K+ ATPases beta chain proteins, and to the subfamily of Na+/K+ -ATPases.
KHDRBS1	This gene encodes a member of the K homology domain-containing, RNA-binding, signal transduction-associated protein family.
ITGB6	This gene encodes a protein that is a member of the integrin superfamily. Members of this family are adhesion receptors that function in signaling from the extracellular matrix to the cell.
STOM	This gene encodes a member of a highly conserved family of integral membrane proteins. The encoded protein localizes to the cell membrane of red blood cells and other cell types.
RER1	The protein encoded by this gene is a multi-pass membrane protein that is localized to the golgi apparatus.

reticulum (ER) membrane proteins. At the same time, RER1 is confirmed to be associated with breast cancer [48]. The full name of STOM is stomatin. Currently, there are few studies on STOM. But the importance of this gene cannot be ignored. It is called a new oncogene. This suggests that biologists need to further study the relationship between STOM and cancer.

V. CONCLUSION

In this article, we propose a robust NMF method called CHNMF. The proposed CHNMF method uses correntropy measure to reduce the negative impact of noises and outliers on experimental results. In addition, the CHNMF method also uses with the hypergraph regularization term to consider high-order geometric information of the data. The experimental results on multi-cancer integrated data demonstrate the effectiveness of the CHNMF method superiority to other four methods. The proposed CHNMF may provide new insights for the improvement of the next NMF method. At the same time, it is a general method that can be used in various applications. The clustering results on the two image datasets are simply shown in the supplementary material.

Because the CHNMF method combines the correntropy and hypergraph regularization constraints, it will outperform other representative methods when the data contain noises and outliers or manifold structures. Which part contributes more to the improvement of CHNMF depends on the characteristics of the dataset itself. However, the proposed CHNMF method also has limitations. For example, the proposed CHNMF method needs to update the kernel bandwidth parameter σ and the diagonal matrix $\mathbf{D}$ during the solution process, its computation time is longer than the NMF method based on the Euclidean norm. The value of σ will affect $\mathbf{U}$ and $\mathbf{V}$. We perform the clustering experiment on the coefficient matrix $\mathbf{V}$ and perform feature selection experiment on the basis matrix $\mathbf{U}$. This will affect the experimental results of clustering and feature selection. That is to say, the update of σ will affect the stability of the algorithm. The value of σ will also affect other applications of the CHNMF method in the real world. Presently, the kernel bandwidth parameter σ is generally empirically determined. We will seek a more computationally efficient algorithm for CHNMF.

REFERENCES

[1] Y. Li, F.-X. Wu, and A. Ngom, "A review on machine learning principles for multi-view biological data integration," *Briefings Bioinformat.*, vol. 19, no. 2, pp. 325–340, 2016.

[2] S.-G. Ge, J. Xia, W. Sha, and C.-H. Zheng, "Cancer subtype discovery based on integrative model of multigenomic data," *IEEE/ACM Trans. Comput. Biol. Bioinformat.*, vol. 14, no. 5, pp. 1115–1121, Sep./Oct. 2017.

[3] J. X. Liu, D. Wang, Y. L. Gao, C. H. Zheng, Y. Xu, and J. Yu, "Regularized non-negative matrix factorization for identifying differential genes and clustering samples: A Survey," *IEEE/ACM Trans. Comput. Biol. Bioinformat.*, vol. 15, no. 3, pp. 974–987, May/Jun. 2018.

[4] A. Thalamuthu, I. Mukhopadhyay, X. Zheng, and G. C. Tseng, "Evaluation and comparison of gene clustering methods in microarray analysis," *Bioinformatics*, vol. 22, no. 19, pp. 2405–2412, 2006.

[5] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2001, pp. 556–562.

[6] J. Wang, F. Tian, H. Yu, C. H. Liu, K. Zhan, and X. Wang, "Diverse non-negative matrix factorization for multiview data representation," *IEEE Trans. Cybern.*, vol. 48, no. 9, pp. 2620–2632, Sep. 2018.

[7] S. Nikitidis, A. Tefas, and I. Pitas, "Projected gradients for subclass discriminant nonnegative subspace learning," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2806–2819, Dec. 2014.

[8] X. Li, G. Cui, and Y. Dong, "Refined-graph regularization-based non-negative matrix factorization," *ACM Trans. Intell. Syst. Technol.*, vol. 9, no. 1, pp. 1–21, 2017.

[9] D. Song, D. A. Meyer, and M. R. Min, "Fast nonnegative matrix factorization with rank-one ADMM," in *Proc. NIPS Workshop Optim. Mach. Learn. (OPT)*, 2014, pp. 1–6.

[10] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1548–1560, Aug. 2011.

[11] N. Guan, D. Tao, Z. Luo, and B. Yuan, "Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 2030–2048, Jul. 2011.

[12] F. Shang, L. C. Jiao, and W. Fei, "Graph dual regularization non-negative matrix factorization for co-clustering," *Pattern Recognit.*, vol. 45, no. 6, pp. 2237–2250, 2012.

[13] K. Zeng, Y. U. Jun, L. I. Cuihua, J. You, and T. Jin, "Image clustering by hyper-graph regularized non-negative matrix factorization," *Neurocomputing*, vol. 138, pp. 209–217, Aug. 2014.

[14] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: Properties and applications in non-Gaussian signal processing," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5286–5298, Nov. 2007.

[15] S. Peng, W. Ser, Z. Lin, and B. Chen, "Robust sparse nonnegative matrix factorization based on maximum correntropy criterion," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, 2018, pp. 1–5.

[16] S. Peng, W. Ser, B. Chen, L. Sun, and Z. Lin, "Correntropy based graph regularized concept factorization for clustering," *Neurocomputing*, vol. 316, pp. 34–48, Nov. 2018.

[17] J. J.-Y. Wang, X. Wang, and X. Gao, "Non-negative matrix factorization by maximizing correntropy for cancer;clustering," *BMC Bioinformat.*, vol. 14, no. 1, p. 107, 2013.

[18] Y. Wang, S. Wu, B. Mao, Z. Xiang, and Z. Luo, "Correntropy induced metric based graph regularized non-negative matrix factorization," *Neurocomputing*, vol. 204, pp. 172–182, Sep. 2016.

[19] N. Zhou, Y. Xu, H. Cheng, Z. Yuan, and B. Chen, "Maximum correntropy criterion based sparse subspace learning for unsupervised feature selection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 2, pp. 404–417, Feb. 2019.

[20] 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.

[21] X. Li, G. Cui, and Y. Dong, "Graph regularized non-negative low-rank matrix factorization for image clustering," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3840–3853, Nov. 2017.

[22] R. He, W. S. Zheng, and B. G. Hu, "Maximum correntropy criterion for robust face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1561–1576, Aug. 2011.

[23] D. Zhou, J. Huang, and B. Schölkopf, "Learning with hypergraphs: Clustering, classification, and embedding," in *Proc. Adv. Neural Inf. Process. Syst.*, 2007, pp. 1601–1608.

[24] N. Guan, T. Liu, Y. Zhang, D. Tao, and L. S. Davis, "Truncated cauchy non-negative matrix factorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 1, pp. 246–259, Jan. 2019.

[25] N. Mila, and R. H. Chan, "The equivalence of half-quadratic minimization and the gradient linearization iteration," *IEEE Trans. Image Process.*, vol. 16, no. 6, pp. 1623–1627, Jun. 2007.

[26] R. He, W. S. Zheng, T. Tan, and Z. Sun, "Half-quadratic-based iterative minimization for robust sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 261–275, Feb. 2014.

[27] Y. Wang, C. Pan, S. Xiang, and F. Zhu, "Robust hyperspectral unmixing with correntropy-based metric," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4027–4040, Nov. 2015.

[28] A. Dreves, F. Facchinei, C. Kanzow, and S. Sagratella, "On the solution of the KKT conditions of generalized Nash equilibrium problems," *SIAM J. Optim.*, vol. 21, no. 3, pp. 1082–1108, 2011.

[29] J. Kivinen and M. K. Warmuth, "Additive versus exponentiated gradient updates for linear prediction," in *Proc. 27th Annu. ACM Symp. Theory Comput.*, 1995, pp. 209–218.

[30] G. Cui, X. Li, and Y. Dong, "Subspace clustering guided convex non-negative matrix factorization," *Neurocomputing*, vol. 292, pp. 38–48, May 2018.

[31] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, Nov. 1987.

[32] L. Lovász and M. D. Plummer, *Matching Theory*. Providence, RI, USA: Amer. Math. Soc., 2009.

[33] B. A. J. Ponder, "Cancer genetics," *Nature*, vol. 411, no. 6835, pp. 336–341, 2001.

[34] J. X. Liu, Y. L. Gao, C. H. Zheng, Y. Xu, and J. Yu, "Block-constraint robust principal component analysis and its application to integrated analysis of TCGA data," *IEEE Trans Nanobiosci.*, vol. 15, no. 6, pp. 510–516, Sep. 2016.

[35] D. Wei *et al.*, "STAT3 activation regulates the expression of vascular endothelial growth factor and human pancreatic cancer angiogenesis and metastasis," *Oncogene*, vol. 22, no. 3, pp. 319–329, 2003.

[36] C. H. Squarize, R. M. Castilho, V. Sriuranpong, D. S. Pinto, Jr, and J. S. Gutkind, "Molecular cross-talk between the NFkappaB and STAT3 signaling pathways in head and neck squamous cell carcinoma," *Neoplasia*, vol. 8, no. 9, pp. 733–746, 2006.

[37] C. M. Leu, F. H. Wong, C. Chang, S. F. Huang, and C. Hu, "Interleukin-6 acts as an antiapoptotic factor in human esophageal carcinoma cells through the activation of both STAT3 and mitogen-activated protein kinase pathways," *Oncogene*, vol. 22, no. 49, pp. 7809–7818, 2003.

[38] Y. Doi, M. Yashiro, R. Amano, S. Noda, and K. Hirakawa, "VEGF-A/VEGFR-2 signaling plays an important role for the motility of pancreas cancer cells," *Ann. Surg. Oncol.*, vol. 19, no. 8, pp. 2733–2743, 2012.

[39] P. A. Kyzas, D. Stefanou, A. Batistatou, and N. J. Agnantis, "Prognostic significance of VEGF immunohistochemical expression and tumor angiogenesis in head and neck squamous cell carcinoma," *J. Cancer Res. Clin. Oncol.*, vol. 131, no. 9, pp. 624–630, 2005.

[40] W. Wei, Y. Wang, X. Yu, L. Ye, Y. Jiang, and Y. Cheng, "Expression of TP53, BCL-2, and VEGFA genes in esophagus carcinoma and its biological significance," *Med. Sci. Monitor Int. Med. J. Exp. Clin. Res.*, vol. 21, pp. 3016–3022, Oct. 2015.

[41] C. Ma *et al.*, "H19 promotes pancreatic cancer metastasis by derepressing let-7's suppression on its target HMGA2-mediated EMT," *Tumour Biol.*, vol. 35, no. 9, pp. 9163–9169, 2014.

[42] G. F. Guan *et al.*, "Overexpression of lncRNA H19/miR-675 promotes tumorigenesis in head and neck squamous cell carcinoma," *Int. J. Med. Sci.*, vol. 13, no. 12, pp. 914–922, 2016.

[43] K. Hibi *et al.*, "Loss of H19 imprinting in esophageal cancer," *Cancer Res.*, vol. 56, no. 3, pp. 480–482, 1996.

[44] S. Ponniah *et al.*, "Epigenetic silencing of Na,K-ATPase β 1 subunit gene ATP1B1 by methylation in clear cell renal cell carcinoma," *Epigenetics*, vol. 9, no. 4, pp. 579–586, 2014.

[45] J. Hesong, Z. Guang, W. Jun-Hua, and J. Chun-Ping, "Diverse roles of miR-29 in cancer (review)," *Oncol. Rep.*, vol. 31, no. 4, pp. 1509–1516, 2014.

[46] D. R. Cochrane, B. M. Jacobsen, K. D. Connaghan, E. N. Howe, D. L. Bain, and J. K. Richer, "Progestin regulated miRNAs that mediate progesterone receptor action in breast cancer," *Mol. Cellular Endocrinol.*, vol. 355, no. 1, pp. 15–24, 2012.

[47] C. Jing *et al.*, "MicroRNA-17/20a impedes migration and invasion via TGF-β/ITGB6 pathway in esophageal squamous cell carcinoma," *Amer. J. Cancer Res.*, vol. 6, no. 7, pp. 1549–1562, 2016.

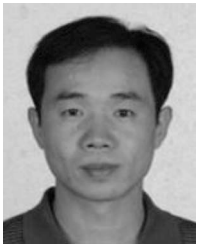
[48] J. Füllekrug, J. Boehm, S. Röttger, T. Nilsson, and H. D. Schmitt, "Human Rer1 is localized to the Golgi apparatus and complements the deletion of the homologous Rer1 protein of Saccharomyces cerevisiae," *Eur. J. Cell Biol.*, vol. 74, no. 1, pp. 31–40, 1997.

Na Yu is currently pursuing the master's degree with the School of Information Science and Engineering, Qufu Normal University, Rizhao, China.

Her research interests include feature selection, pattern recognition, and bioinformatics.

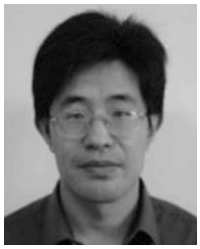
Ming-Juan Wu is currently pursuing the master's degree with the School of Information Science and Engineering, Qufu Normal University, Rizhao, China.

Her research interests include feature selection, pattern recognition, and bioinformatics.



Jin-Xing Liu (Member, IEEE) received the B.S. degree in electronic information and electrical engineering from Shandong University, Jinan, China, in 1997, the M.S. degree in control theory and control engineering from Qufu Normal University, Rizhao, China, in 2003, and the Ph.D. degree in computer simulation and control from the South China University of Technology, Guangzhou, China, in 2008.

He is a Professor with the School of Information Science and Engineering, Qufu Normal University. His research interests include pattern recognition, machine learning, and bioinformatics.



Chun-Hou Zheng (Member, IEEE) received the B.S. degree in physics education and the M.S. degree in control theory and control engineering from Qufu Normal University, Rizhao, China, in 1995 and 2001, respectively, and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2006.

He is currently with the College of Computer Science and Engineering, Anhui University, Hefei, and the School of Information Science and Engineering, Qufu Normal University. His research interests include pattern recognition and bioinformatics.

Yong Xu (Senior Member, IEEE) received the B.S. and M.S. degrees from the Air Force Institute of Meteorology, Nanjing, China, in 1994 and 1997, respectively, and the Ph.D. degree in pattern recognition and intelligence systems from the Nanjing University of Science and Technology, Nanjing, in 2005.

From 2005 to 2007, he worked as a Postdoctoral Research Fellow with the Shenzhen Graduate School, Harbin Institute of Technology, Harbin, China. He is currently a Professor with the Harbin Institute of Technology, Shenzhen, China. His current interests include pattern recognition, machine learning, and bioinformatics.