

Research Article

Pattern Expression Nonnegative Matrix Factorization: Algorithm and Applications to Blind Source Separation

Junying Zhang,¹ Le Wei,¹ Xuerong Feng,² Zhen Ma,¹ and Yue Wang³

¹ School of Computer Science and Engineering, Xidian University, Xi'an 710071, China

² Department of Mathematics and Computer Science, Valdosta State University, Valdosta, GA 31698, USA

³ The Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, VA 24061, USA

Correspondence should be addressed to Junying Zhang, jyzhang@mail.xidian.edu.cn

Received 1 November 2007; Accepted 18 April 2008

Recommended by Rafal Zdunek

Independent component analysis (ICA) is a widely applicable and effective approach in blind source separation (BSS), with limitations that sources are statistically independent. However, more common situation is blind source separation for nonnegative linear model (NNLM) where the observations are nonnegative linear combinations of nonnegative sources, and the sources may be statistically dependent. We propose a pattern expression nonnegative matrix factorization (PE-NMF) approach from the view point of using basis vectors most effectively to express patterns. Two regularization or penalty terms are introduced to be added to the original loss function of a standard nonnegative matrix factorization (NMF) for effective expression of patterns with basis vectors in the PE-NMF. Learning algorithm is presented, and the convergence of the algorithm is proved theoretically. Three illustrative examples on blind source separation including heterogeneity correction for gene microarray data indicate that the sources can be successfully recovered with the proposed PE-NMF when the two parameters can be suitably chosen from prior knowledge of the problem.

Copyright © 2008 Junying Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Blind source separation (BSS) is a very active topic recently in signal processing and neural network fields [1, 2]. It is an approach to recover the sources from their combinations (observations) without any understanding of how the sources are mixed. For a linear model, the observations are linear combinations of sources, that is, $X = AS$, where S is an $r \times n$ matrix indicating r source signals each in n -dimensional space, X is an $m \times n$ matrix showing m observations in n -dimensional space, and A is an $m \times r$ mixing matrix. Therefore, BSS problem is a matrix factorization, that is, to factorize observation matrix V into mixing matrix A and source matrix S .

Independent component analysis (ICA) has been found very effective in BSS for the cases where the sources are statistically independent. In fact, it factorizes the observation matrix V into mixing matrix A and source matrix S by searching the most non-gaussianity directions in the scatter plot of observations, and has a very good estimation

performance of the recovered sources when the sources are statistically independent. This is based on the Central Limit Theorem, that is, the distribution of a sum (observations) of independent random variables (sources) tends toward a Gaussian distribution under certain conditions. This induces the two serious constraints of ICA to the application of BSS: (1) the sources should be statistically independent to each other; (2) the sources should not follow Gaussian distribution. The performance of the recovered sources with ICA approach depends on the satisfactory of these two constraints, and decreases very rapidly when either of them is not satisfied. However in real world, there are many applications of blind source separation where the observations are nonnegative linear combinations of nonnegative sources, and the sources are statistically dependent to some extent. This is the model referred to as nonnegative linear model (NNLM), that is, $X = AS$ with elements in both A and S nonnegative, and the rows in S (the sources) may be statistically dependent to some extent. One of the applications of this model is gene expression profiles, where

each of the profiles, which is only in nonnegative values, represents a composite of more than one distinct but partially dependent sources [3], the profiles from normal tissue and from cancer tissue. What needs to be developed is an algorithm to recover dependent sources from the composite observations.

It is easy to recognize that BSS for NNLM is a nonnegative matrix factorization, that is, to factorize X into nonnegative A and nonnegative S , where nonnegative matrix factorization (NMF) technique is applicable. Several approaches have been developed on applying NMF-based technique for BSS of NNLM. For example, we proposed a method for decomposition of molecular signatures based on BSS of nonnegative dependent sources with direct usage of standard NMF [3]; Chichocki and his colleagues proposed a new algorithm for nonnegative matrix factorization in applications to blind source separation [4] by adding two suitable regularizations or penalty terms in the original objective function of the NMF to increase sparseness and/or smoothness of the estimated components. In addition, multilayer NMF was proposed by Cichocki and Zdunek for blind source separation [5], and nonsmooth nonnegative matrix factorization was proposed aiming at finding localized, part-based representations of nonnegative multivariate data items [6]. Some other researches include the work of Zdunek and Cichocki, who proposed to take advantage of the second-order terms of a cost function to overcome the disadvantages of gradient (multiplicative) algorithms for NMF for tackling the slow convergence problem of the standard NMF learning algorithms [7]; the work by Ivica Kopriva and his colleagues, who proposed a single-frame blind image deconvolution approach with nonnegative sparse matrix factorization for blind image deconvolution [8]; and the work by Liu and Zheng who proposed nonnegative matrix factorization-based methods for object recognition [9].

In this paper, we extend NMF to pattern expression NMF (PE-NMF) from the view point that the basis vector is desired to be the one which can express the data most efficiently. Its successful application to blind source separation of extended bar problem, nonnegative signal recovery problem, and heterogeneity correction problem for real gene microarray data indicates that it is of great potential in blind separation of dependent sources for NNLM model. The loss function for the PE-NMF proposed here is a special case of that proposed in [4], and here not only the learning algorithm for the proposed PE-NMF approach is provided, but also the convergence of the learning algorithm is proved by introducing some auxiliary function. For speeding up the learning procedure, a technique based on independent component analysis (ICA) is proposed, and has been verified to be effective for the learning algorithm to converge to desired solutions.

2. Pattern Expression NMF and BSS for NNLM Model

NMF problem is given a nonnegative $n \times m$ matrix V , find nonnegative $n \times r$ and $r \times m$ matrix factors W and H such that

the difference measure between V and WH is the minimum according to some cost function, that is,

$$V \approx WH. \quad (1)$$

NMF is a method to obtain a representation of data using nonnegative constraints. These constraints lead to a part-based representation because they allow only additive, not subtractive, combinations of the original data. For the i th column of (1), that is, $v_i = Wh_i$, where v_i and h_i are the i th column of V and H , the i th datum (observation) is a nonnegative linear combination of the columns of $W = (W_1, W_2, \dots, W_r)$, while the combinatorial coefficients are the elements of h_i . Therefore, the columns of W , that is, $\{W_1, W_2, \dots, W_r\}$, can be viewed as the basis of the data V when V is optimally estimated by its factors.

2.1. Pattern Basis

Let W_1, W_2, \dots, W_r be linearly independent n -dimensional vectors. We refer to the space spanned by arbitrarily non-negatively linear combination of these r vectors the positive subspace spanned by W_1, W_2, \dots, W_r . Then, W_1, W_2, \dots, W_r is the pattern expression of the data in this subspace, and is called the basis of the subspace. Evidently, the basis W_1, W_2, \dots, W_r derived from NMF is the pattern expression of the observation data in columns of V , but this expression may not be unique. Figure 1(a) shows an example of the data V which have two pattern expressions of $\{W_1, W_2\}$ and $\{W'_1, W'_2\}$. Hence, we have the following questions: which basis is more effective in expressing the pattern of the observations in V ? In order for the basis to express the pattern in V effectively, in our opinion, following three requirements should be satisfied:

- (1) the angle between the vectors in the basis should be as large as possible, such that each data in V is a nonnegatively linear combination of the vectors;
- (2) the angles between the vectors in the basis should be as small as possible to make the vectors clamp the data as tightly as possible, such that no space is left for expression of what is not included in V ;
- (3) each vector in the basis should be of the most efficient in expression of the data in V , and the same efficient in this expression compared with any other vector in the basis.

The vectors defined with the above three requirements are what we call the pattern basis of the data, and the number of vectors in the basis, r , is called the pattern dimension of the data. Figures 1(a), 1(b), and 1(c) show, respectively, the too large between-angle, too small between-angles, and too unequally important basis situation with $\{W'_1, W'_2\}$ as basis, where data in Figures 1(a) and 1(b) are assumed to be uniformly distributed in the gray area while those in Figure 1(c) are assumed to be nonuniformly distributed (the data in the dark gray area is denser compared with those in the light gray area). For these three cases, $\{W_1, W_2\}$ is a better basis to express the data.

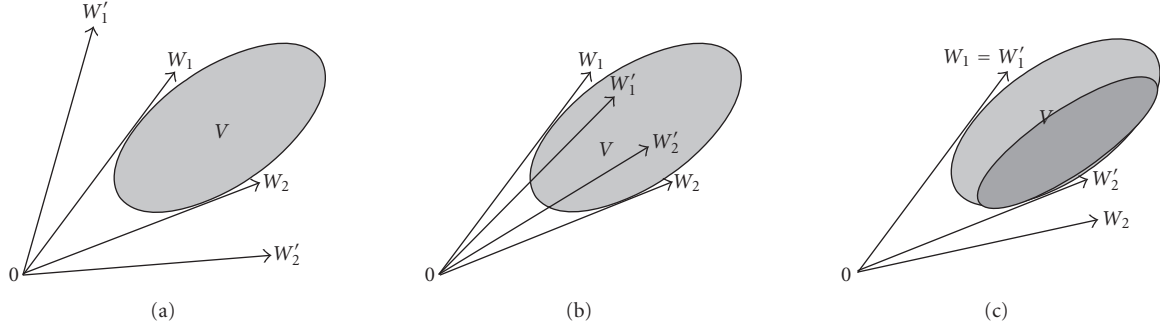


FIGURE 1: The basis $\{W_1, W_2\}/\{W'_1, W'_2\}$ which obeys/violates (a) the first point; (b) the second point; (c) the third point in the definition of the pattern basis.

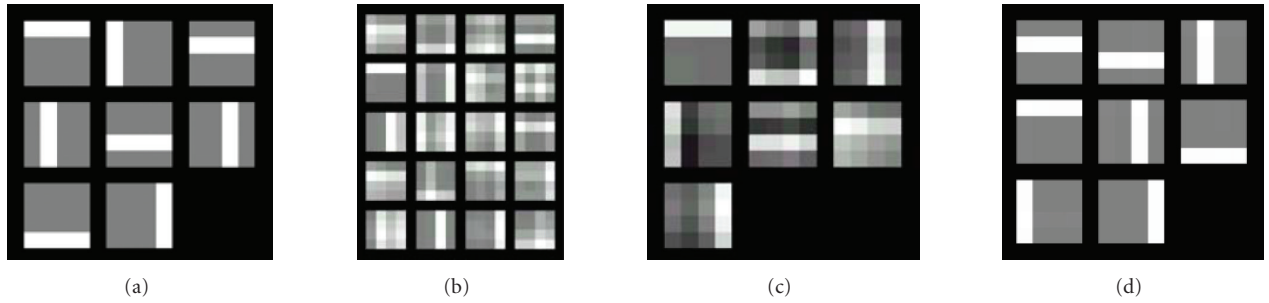


FIGURE 2: Bar problem solution obtained from NMF: (a) source images, (b) mixed images, (c) recovered images from ICA, and (d) recovered images from NMF.

Notice that the second requirement in the definition of the pattern basis readily holds from the constraint of NMF that the elements in H are nonnegative. Then, we can get the three constraints as follows: (1) due to the requirement that the between-angle between each pair of vectors in the basis should be as large as possible, we have $W_i^T W_j \rightarrow \min$, for $i \neq j$, where W_i is the i th column of the matrix W ; (2) due to the requirement that each vector in the basis should be equally efficient in expression of the data in V , while the efficiency of the vector in this expression is measured by the summation of the projection coordinates of all the data in V to this vector, that is, samples v_j , $j = 1, 2, \dots, n$ if expressed in the vector W_i , the efficiency of the vector W_i for expression of v_j , $j = 1, 2, \dots, n$ is $\sum_{j=1}^n h_{ji}$, we have $\sum_{i=1}^r \sum_{j=1}^n h_{ji} \rightarrow \min$. Hence, we formulate PE-NMF problem as minimizing the loss function $E(W, H; \alpha, \beta)$ in the following equation subject to nonnegativity constraints.

PE-NMF problem

Given an n by m nonnegative observation matrix V , find an n by r and an r by m nonnegative matrix factors W and H , such that

$$\begin{aligned} \min_{W, H} E(W, H; \alpha, \beta) &= \frac{1}{2} \|V - WH\|^2 + \alpha \sum_{i, j, j \neq i} W_i^T W_j \\ &+ \beta \sum_{i, j} h_{ij}, \quad \text{s.t. } W \geq 0, H \geq 0, \end{aligned} \quad (2)$$

where $W \geq 0, H \geq 0$ indicates that both W and H are nonnegative matrices, respectively, W_i is i th column of matrix W , and h_{ij} is the element in the i th row and j th column of the matrix H .

This problem is a special case of the constrained optimization problem proposed in [4]:

$$\begin{aligned} \min_{W, H} E(W, H; \alpha, \beta) &= \frac{1}{2} \|V - WH\|^2 + \alpha J_W(W) \\ &+ \beta J_H(H), \quad \text{s.t. } W \geq 0, H \geq 0. \end{aligned} \quad (3)$$

2.2. PE-NMF Algorithm and its Convergence

For the derivation of learning algorithm for W and H , we first present and prove the following lemma.

Lemma 1. For any r by r symmetric nonnegative matrix Q and for any r -dimensional nonnegative row vector \mathbf{w} , the r by r matrix

$$F = \delta_{ab} \frac{(Q\mathbf{w}^T)_a}{w_a} - Q \quad (4)$$

is always semipositive definite, where $\delta_{ab}((Q\mathbf{w}^T)_a/w_a)$ represents a diagonal matrix with diagonal element in the a th row and a th column being $(Q\mathbf{w}^T)_a/w_a$.

Proof. By noticing that w_a and w_b are nonnegative, the definition of the matrix F is the same as that of the matrix S in

which the a th row and b th column element is $S_{ab} = w_a F w_b$. Hence, we consider proving the semipositive definition of the matrix S in the following context.

For any r -dimensional vector \mathbf{V} , we have the following formula:

$$\begin{aligned} V^T S V &= \sum_{ab} V_a S_{ab} V_b = \sum_{ab} V_a w_a F w_b V_b \\ &= \sum_{ab} \left[V_a w_a \delta_{ab} \frac{(Q w^T)_a}{w_a} w_b V_b - V_a w_a Q_{ab} w_b V_b \right] \\ &= \sum_{ab} \left[V_a w_a \delta_{ab} \frac{(Q w^T)_a}{w_a} w_b V_b \right] - \sum_{ab} V_a w_a Q_{ab} w_b V_b \\ &= A - B, \end{aligned} \quad (5)$$

where A denotes the first term and B denotes the second term in the above formula. By noticing that Q is a symmetric matrix, we have $(Q w^T)_a = (w Q)_a$, and hence the first term A becomes

$$\begin{aligned} A &= \sum_a V_a w_a \frac{(w Q)_a}{w_a} w_a V_a = \sum_a (w Q)_a w_a V_a^2 \\ &= \sum_a \left(\sum_b w_b Q_{ba} \right) w_a V_a^2 = \sum_b \left(\sum_a w_a Q_{ab} \right) w_b V_b^2 \quad (6) \\ &= \sum_{ab} w_a Q_{ab} w_b V_a^2, \end{aligned}$$

we substitute the above A into formula (5), and obtain

$$\begin{aligned} V^T S V &= \sum_{ab} (w_a Q_{ab} w_b V_a - w_a Q_{ab} w_b V_a V_b) \\ &= \sum_{ab} w_a Q_{ab} w_b [V_a^2 - V_a V_b] \\ &= \sum_{ab} w_a Q_{ab} w_b \left[\frac{1}{2} V_a^2 + \frac{1}{2} V_b^2 - V_a V_b \right] \quad (7) \\ &= \frac{1}{2} \sum_{ab} w_a Q_{ab} w_b (V_a - V_b)^2. \end{aligned}$$

Due to the fact that w and Q are a row vector and a nonnegative matrix, respectively, hence for any r -dimensional row vector V , we have

$$V^T S V = \frac{1}{2} \sum_{ab} w_a Q_{ab} w_b (V_a - V_b)^2 \geq 0. \quad (8)$$

Hence, the matrix S and therefore $F = \delta_{ab}((Q w^T)_a / w_a) - Q$ is a semipositive definite matrix.

Now in Theorem 1, we derive learning algorithm and prove its convergence for updating each row w in W when H is set to be a fixed nonnegative matrix. The learning algorithm for updating each column h in H when W is set to be a fixed nonnegative matrix is depicted in Theorem 2, and can be proved similarly but skipped due to the limitation of the space. \square

Theorem 1. For the quadratic optimization problem,

$$\min_w E(w; H, \alpha) = \frac{1}{2} \|v - wH\|^2 + \frac{1}{2} \alpha w M w^T, \quad \text{s.t. } w \geq 0, \quad (9)$$

where w is an r -dimensional row vector, v is a given m -dimensional nonnegative row vector, H is an r by m fixed nonnegative matrix, M is an r by r constant matrix with all elements being 1 except diagonal elements being zeros, and α is a fixed nonnegative parameter. The following update algorithm

$$w_a^{t+1} = w_a^t \frac{(v H^T)_a}{(w^t H H^T + \alpha w^t M)_a} \quad (10)$$

converges to its optimal solution from any initialized nonnegative vector w^0 .

Proof. The convergence proof will be performed by introducing an appropriate auxiliary function $F(w, w^t)$ that satisfies

$$F(w^t, w^t) = E(w^t), \quad (11)$$

$$F(w, w^t) \geq E(w). \quad (12)$$

If such a function can be found, then the update of w by setting

$$w^{t+1} = \arg \min_w F(w, w^t) \quad (13)$$

will make

$$E(w^{t+1}) \leq F(w^{t+1}, w^t) \leq F(w^t, w^t) = E(w^t) \quad (14)$$

which always makes the objective function $E(w)$ to be decreased with respect to iterations in the algorithm, indicating that the algorithm converges with the updating formula (13).

Now we construct the auxiliary function to be

$$\begin{aligned} F(w, w^t) &= E(w^t) + (w - w^t) \nabla E(w^t) \\ &\quad + \frac{1}{2} (w - w^t) J(w^t) (w - w^t)^T, \end{aligned} \quad (15)$$

where $J(w^t)$ is a diagonal matrix

$$J(w^t) = \delta_{ab} \frac{(H H^T w^{tT} + \alpha M w^{tT})_a}{(w^{tT})_a}. \quad (16)$$

Obviously, $F(w^t, w^t) = E(w^t)$, so formula (11) holds.

The Taylor expansion of the loss function $E(w)$, when w approaches w^t , can be written to be

$$\begin{aligned} E(w) &= E(w^t) + (w - w^t) \nabla E(w^t) \\ &\quad + \frac{1}{2} (w - w^t) (H H^T + \alpha M) (w - w^t)^T. \end{aligned} \quad (17)$$

By subtracting $F(w, w^t)$ in (8)–(16) to $E(w)$ in (7)–(17), we have

$$\begin{aligned} F(w, w^t) - E(w) &= \frac{1}{2}(w - w^t) \left[\delta_{ab} \frac{(HH^T w^{tT} + \alpha M w^{tT})_a}{(w^{tT})_a} \right. \\ &\quad \left. - HH^T - \alpha M \right] (w - w^t)^T \\ &= \frac{1}{2}(w - w^t) \left[\delta_{ab} \frac{(Q w^{tT})_a}{(w^{tT})_a} - Q \right] (w - w^t)^T, \end{aligned} \quad (18)$$

where $Q = HH^T + \alpha M$. Due to the fact that Q is a nonnegative symmetric matrix since H is the nonnegative factor of V , and α is always a nonnegative parameter, and the fact that w^t is a nonnegative vector, we have, from Lemma 1, that the matrix $\delta_{ab}((Q w^{tT})_a / (w^{tT})_a) - Q$ is semipositive definite, and therefore we always have $F(w, w^t) - E(w) \geq 0$. Hence, updating w according to $w^{t+1} = \arg \min_w F(w, w^t)$ always leads the iteration process to converge.

We employ the steepest descent search strategy for optimal w . For this purpose, we have w^{t+1} to satisfy $(\partial F(w, w^t)) / \partial w|_{w=w^{t+1}} = 0$, from which we get $\nabla E(w^t) + J(w^t)(w^{t+1} - w^t)^T = 0$, or equally

$$w^{t+1T} = w^{tT} - J^{-1}(w^t) \nabla E(w^t). \quad (19)$$

By the definition of the loss function $E(w)$, we have

$$\begin{aligned} \nabla E(w^t) &= H(H^T w^t - v^T) + \alpha M w^{tT} \\ &= (HH^T + \alpha M) w^{tT} - H v^T. \end{aligned} \quad (20)$$

Since $J(w^t)$ is a diagonal matrix, we only need to compute inversion of each diagonal element in J for J^{-1} . Hence, we have the following updating formula for the a th element of w :

$$\begin{aligned} w_a^{t+1} &= w_a^t - \frac{(w^{tT})_a}{(HH^T w^{tT} + \alpha M w^{tT})_a} \\ &\quad \cdot (HH^T w^{tT} + \alpha M w^{tT} - H v^T)_a \\ &= w_a^t \frac{(H v^T)_a}{(HH^T w^{tT} + \alpha M w^{tT})_a} \\ &= w_a^t \frac{(v H^T)_a}{(w^t H H^T + \alpha w^t M)_a}. \end{aligned} \quad (21)$$

□

Theorem 2. For the quadratic optimization problem.

$$\min_h E(h; W, \beta) = \frac{1}{2} \|v - Wh\|^2 + \beta I^T h, \quad \text{s.t. } h \geq 0, \quad (22)$$

where h is r -dimensional column vector, v is a given n -dimensional nonnegative column vector, W is an n by r fixed nonneg-

Algorithm parameters: α, β ;

Input: an n by m nonnegative observation matrix V ;
Output: an n by r nonnegative matrix W and an r by m nonnegative matrix H .

Step 1: set $t = 0$, and generate nonnegative matrix W^0 and H^0 at random;

Step 2: Update H from H^t to H^{t+1} by

$$H^{t+1} = H^t \otimes \frac{(W^t)^T V}{(W^t)^T (W^t) (H^t) + \beta I},$$

$$W^{t+1} = W^t \otimes \frac{V (H^t)^T}{W^t (H^t) (H^t)^T + \alpha W^t M},$$

where I is an r by m matrix full of elements being 1s, and M is an r by r matrix with all elements being 1s except diagonal elements being zeros.

Step 3: Increment t by $t = t + 1$ and go to step 2 until H^{t+1} and W^{t+1} converge.

ALGORITHM 1: Learning algorithm.

ative matrix, I is an $r \times m$ matrix with all the elements being 1s, and β is a fixed nonnegative parameter. The following rule

$$h_a^{t+1} = h_a^t \frac{(W^T v)_a}{(W^T W h^t + \beta I)_a} \quad (23)$$

converges to its optimal solution from any initialized nonnegative vector h^0 .

This theorem can be proved similarly as the proof of Theorem 1.

By representing (10) and (23) in (elementwise) Hadamard product, one has the following learning algorithm for updating both W and H for the PE-NMF optimization problem in (1).

Theorem 3. For the optimization problem shown in (1), the above learning algorithm converges to locally optimal solution from any initialized nonnegative vector H^0 and W^0 .

It is evident that the portion relating to the row w in the objective function $E(W, H; \alpha, \beta)$ in (1) is just $E(w; H, \alpha)$ in (9), and the portion relating to the column h in the objective function $E(W, H; \alpha, \beta)$ in (1) is just $E(h; W, \beta)$ in (22). Hence, using formula to update w and h alternatively will make the learning process to converge to the solution of the objective function $E(W, H; \alpha, \beta)$. Hence, the above theorem can be easily proved on the basis of Theorems 1 and 2.

The update of the W and H can also be expressed with MatLab command of $W = W.*(V*H')./(W*H*H' + \text{alfa}*W*M)$ and $H = H.*(W'*V)./(W'*W*H + \text{beta})$.

2.3. Initialization of the Algorithm

To our knowledge, it seems that there are two main reasons for NMF to converge to undesired solutions. One is that

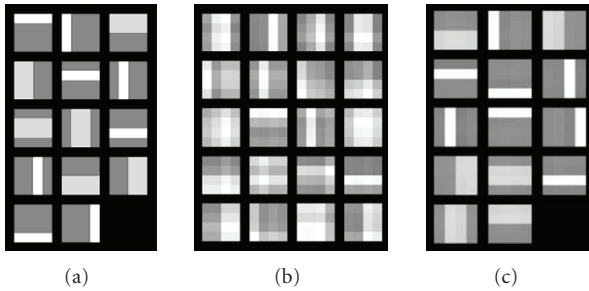


FIGURE 3: Extended bar problem solution obtained from PE-NMF: (a) source images, (b) mixed images, (c) recovered images from PE-NMF.

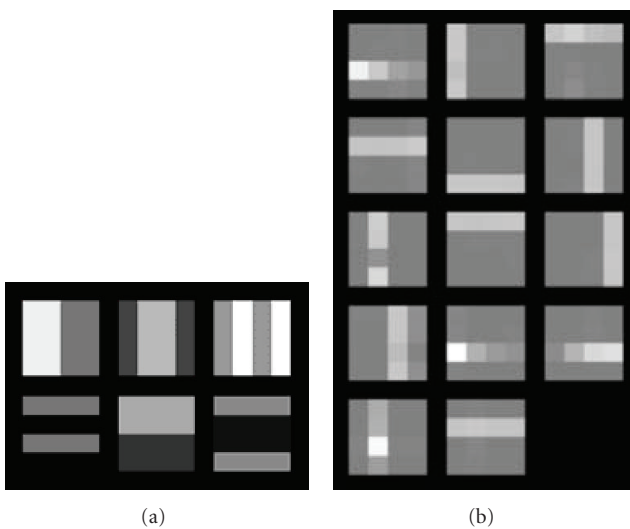


FIGURE 4: Recovered images from (a) ICA, and (b) NMF for the extended bar problem.

the basis of a space may not be unique theoretically, and therefore separate runs of NMF may lead to different results. Another reason may come from the algorithm itself, that the loss function sometimes gets stuck into local minimum during its iteration. By revisiting the loss function of the proposed PE-NMF, it is seen that similar to NMF, the above PE-NMF still sometimes gets stuck into local minimum during its iteration, and/or the number of iterations required for obtaining desired solutions is very large. For the sake of these, an ICA-based technique was proposed for initializing source matrix instead of setting it to be a nonnegative matrix at random: we performed ICA on the observation signals, and set the absolute of the independent components obtained from ICA to be the initialization of the source matrix. In fact, there are reasons that the resultant independent components obtained from ICA are generally not the original sources. One reason is the nonnegativity of the original sources but centering preprocess of the ICA makes each independent component be both positive and negative in its elements: the means of each independent component is zero. Another reason is possibly dependent

or partially independent original sources which does not follow the independence requirement of sources in the ICA study. Hence, the resultant independent components from ICA could not be considered as the recovery of the original sources. Even so, they still provide clues of the original sources: they can be considered as *very rough* estimations of the original sources. From this perspective, and by noticing that the initialization of the source matrix should be nonnegative, we set the absolute of the independent components obtained from ICA as the initialization of the source matrix for the proposed PE-NMF algorithm. Our experiments indicate that such an initialization technique is very effective in speeding up the learning process for getting desired solutions.

3. Experiments and Results

The proposed PE-NMF algorithms have been extensively tested for many difficult benchmarks for signals and images with various statistical distributions. Three examples will be given in the following context for demonstrating the effectiveness of the proposed method compared with standard NMF method and/or ICA method. In ICA approach here, we decentralize the recovered signals/images/microarrays for its nonnegativity property for compensating the centering preprocessing of the ICA approach. The NMF algorithm is simply the one proposed in [10] and the ICA algorithm is simply the FastICA algorithm generally used in many applications in [11]. The examples include blind source separation of extended bar problem, mixed signals, and real microarray gene expression data in which heterogeneity effect occurs.

3.1. Extended Bar Problem

The linear bar problem [12] is a blind separation of bars from their combinations. 8 nonnegative feature images (sources) sized 4×4 including 4 vertical and 4 horizontal thin bar images, shown in Figure 2(a), are randomly mixed to form 1000 observation images, the first 20 shown in Figure 2(b). The solution obtained from ICA and NMF with $r = 8$ are shown in Figures 2(c) and 2(d), respectively, indicating that NMF can fulfill the task very well compared with ICA. However, when we extended this bar problem into the one which is composed of two types of bars, thin one and thick one, NMF failed to estimate the original sources. For example, fourteen source images sized 4×4 with four thin vertical bars, four thin horizontal bars, three wide vertical bars, and three wide horizontal bars, shown in Figure 3(a), are nonnegative and evidently statistically dependent. These source images were randomly mixed with mixing matrix of elements arbitrarily chosen in $[0, 1]$ to form 1000 mixed images, the first 20 shown in Figure 2(b). The PE-NMF with parameter $\alpha = 4$ and $\beta = 1$ was performed on these mixed images for $r = 14$. The resultant images, which are shown in Figure 2(c), indicate that the sources were recovered successfully with the proposed PE-NMF. For comparison, many times we tried using ICA and NMF on this problem for avoiding obtaining local minimum solutions, but always

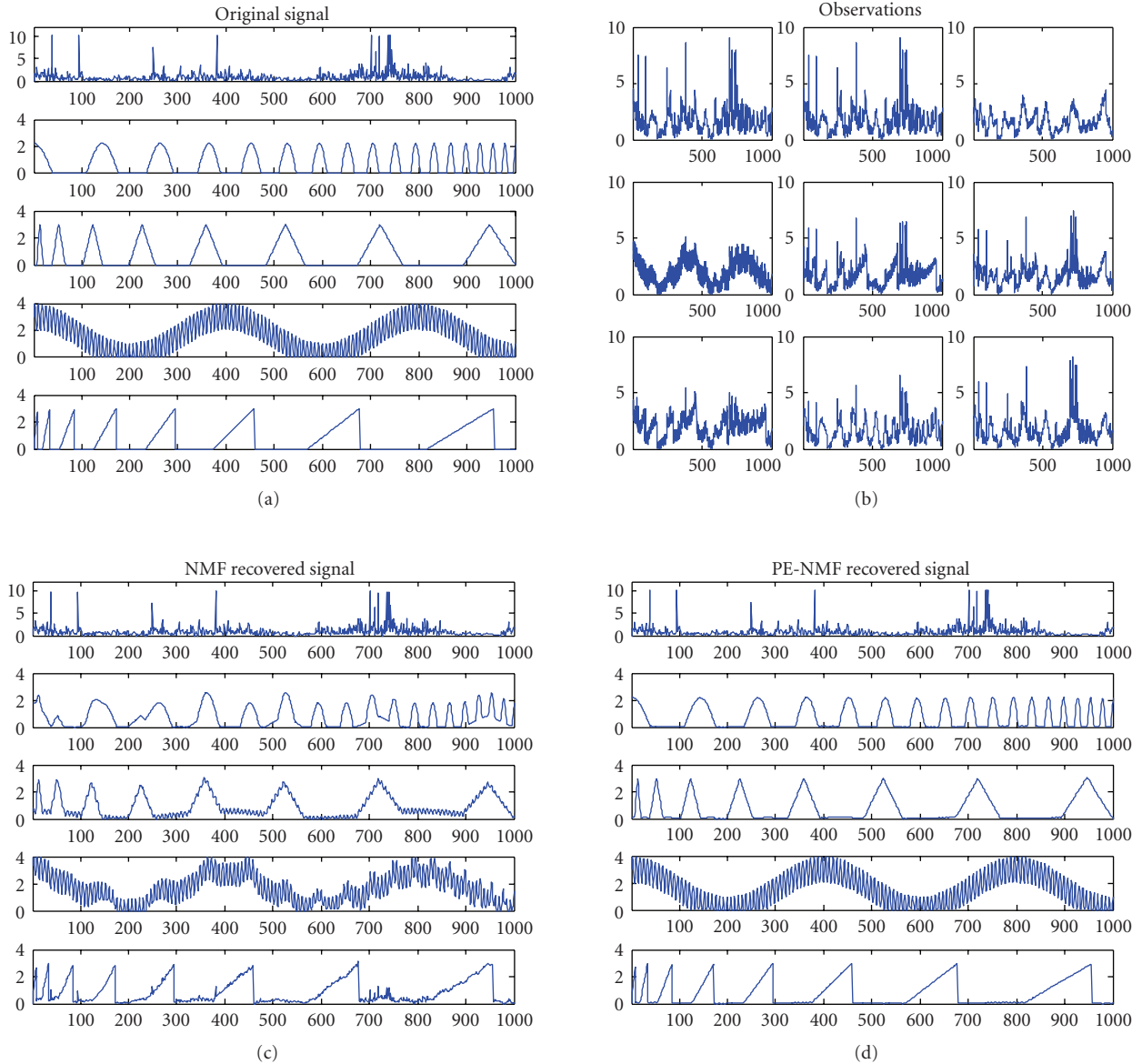


FIGURE 5: Blind signal separation example: (a) 5 original signals, (b) 9 observations, (c) recovered signals from NMF, and (d) recovered signals from PE-NMF.

failed to recover the original sources. Shown in Figures 4(a) and 4(b) are the examples of the recovered images with these two approaches. Notice that both the ones recovered from ICA and NMF are very far from the original sources, and even the number of sources estimated from the ICA is only 6, rather than 14. It is noticeable that the recovered images from the PE-NMF with some other parameter such as $\alpha = 4.2$ and $\beta = 0.1$ are comparable to the ones shown in Figure 3(c), indicating that the proposed method is not very sensitive to the parameter selection for this example.

3.2. Recovery of Mixed Signals

We performed experiments on recovering 5 nonnegative signals from 9 mixtures of 5 nonnegative dependent source sig-

nals, which is the one in [4]. The 9 mixture observation signals come from arbitrarily nonnegative linear combinations of the 5 nonnegative source signals shown in Figure 5(a). The difficulty to recover the sources is a very small number of observations compared with the number of sources. Both NMF and our proposed PE-NMF (where α and β are taken to be 0.001 and 17.6, resp.) were employed for recovery of the sources. By comparison of the resultant signals obtained by NMF shown in Figure 5(c) and these obtained by PE-NMF shown in Figure 5(d), it is evident that the PE-NMF can recover the sources with a higher recovery performance. In fact, the signal-to-interference ratios (SIRs) for the recovered sources from NMF is only 22.17, 11.13, 10.98, 14.91, and 14.15 while that from PE-NMF increases to 47.10, 28.89, 26.67, 83.44, and 28.75 for the 5 source signals.

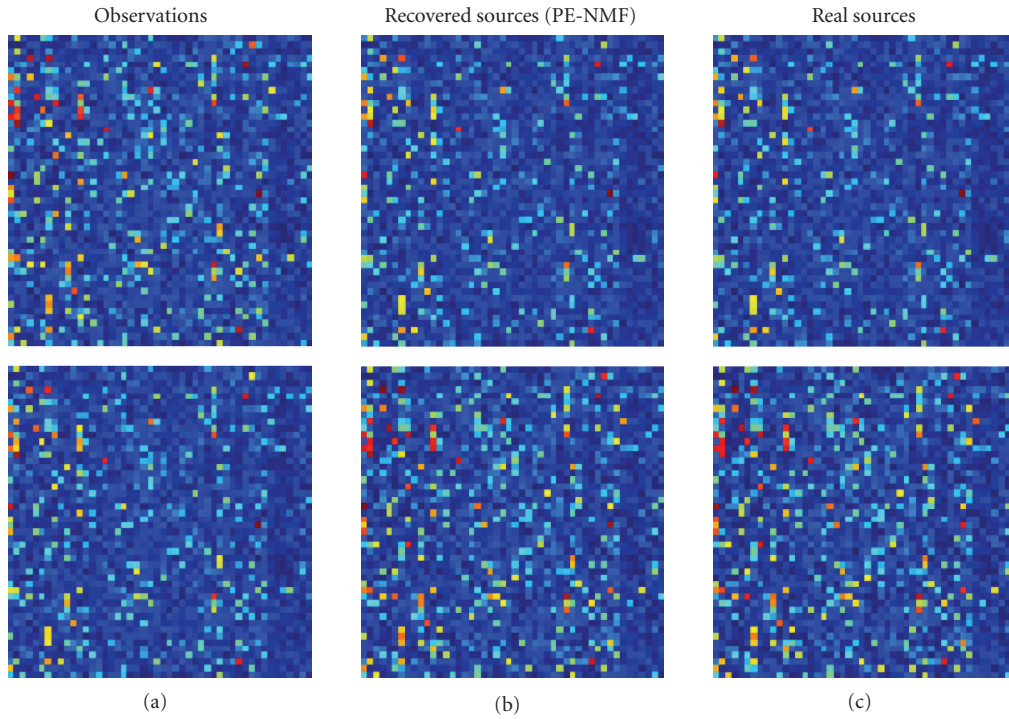


FIGURE 6: Heterogeneity correction result: (a) observations, (b) recovered sources from PE-NMF, and (c) real sources.

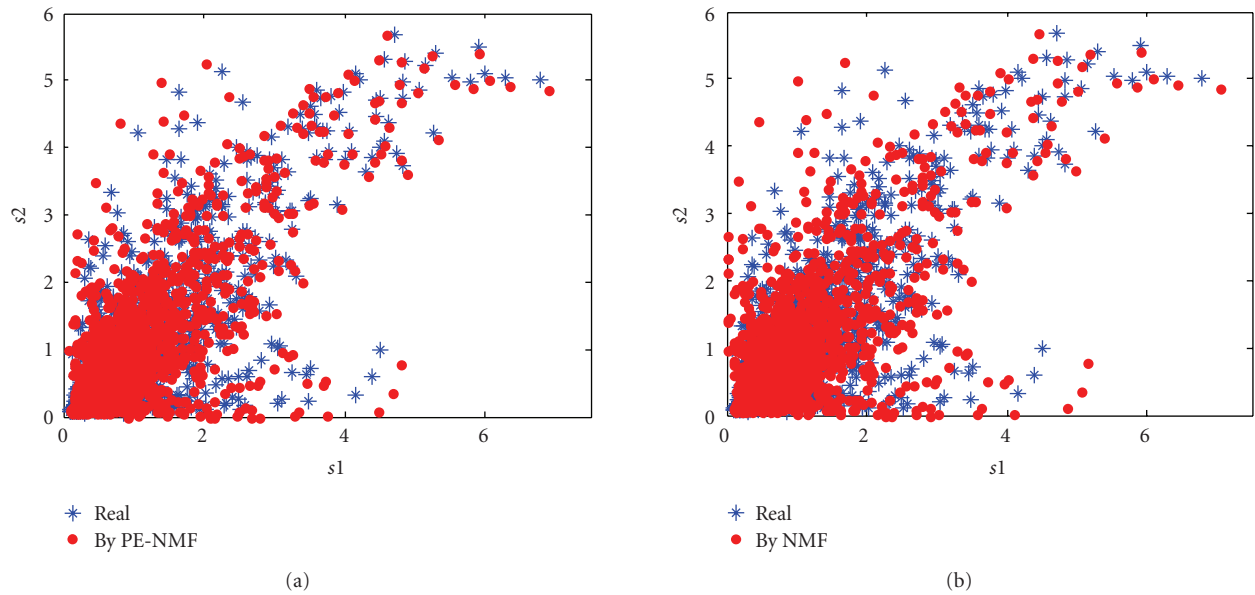


FIGURE 7: The scatter plots of the real sources (blue stars) and the recovered sources (red dots) from (a) PE-NMF, and (b) NMF.

3.3. Heterogeneity Correction of Gene Microarrays

Gene expression microarrays promise powerful new tools for the large-scale analysis of gene expression. Using this technology, the relative mRNA expression levels derived from tissue samples can be assayed for thousands of genes simultaneously. Such global views are likely to reveal previously

unrecognized patterns of gene regulation and generate new hypotheses warranting further study (e.g., new diagnostic or therapeutic biomarkers). However, as a common feature in microarray profiling, gene expression profiles represent a composite of more than one distinct but partially dependent sources (i.e., the observed signal intensity will consist of the weighted sum of activities of the various sources). More specifically, in the case of solid tumors, the related issue is

called partial volume effect (PVE), that is, the heterogeneity within the tumor samples caused by stromal contamination. Blind application of microarray profiling could result in extracting signatures reflecting the proportion of stromal contamination in the sample, rather than underlying tumor biology. Such “artifacts” would be real, reproducible, and potentially misleading, but would not be of biological or clinical interest, while can severely decrease the sensitivity and specificity for the measurement of molecular signatures associated with different disease processes. Despite their critical importance to almost all the followup analysis steps, this issue, called partial volume correction (PVC), is often less emphasized or at least has not been rigorously addressed as compared to the overwhelming interest and effort in pheno/gene-clustering and class prediction.

The effectiveness of the proposed PE-NMF method was tested with real-world data set, microarray gene expression data set, for PVC. The data set consists of 2308 effective gene expressions from two samples of neuroblastoma and non-Hodgkin lymphoma cell tumors [13]. Two observation microarrays, recovered microarrays from PE-NMF, and two pure source microarrays are shown in Figures 6(a), 6(b), and 6(c), respectively. Notice that the true sources are determined, in our present case, by separately profiling the pure cell lines that provide the ground truth of the gene expression profiles from each cell populations. In our clinical case, we use laser-capture microdissection (LCM) technique to separate cell populations from real biopsy samples. By comparison of Figures 6(b) and 6(c), the blind source separation by PE-NMF method recovered the pure microarray successfully. Figures 7(a) and 7(b) show the scatter plots of the recovered microarrays from PE-NMF and from NMF compared with these of the pure microarrays. These scatter plots and the SIRs of being 56.79 and 31.73 for the PE-NMF approach and of being only 21.20 and 32.81 for the NMF approach also indicate that the proposed PE-NMF is effective in recovering the sources successfully. Many other independent trials using other gene sets reached a similar result.

4. Conclusions

This paper proposes a pattern expression nonnegative matrix factorization (PE-NMF) approach for efficient pattern expression and applies it to blind source separation for nonnegative linear model (NNLM). Its successful application to blind source separation of extended bar problem, nonnegative signal recovery problem, and heterogeneity correction problem for real microarray gene data indicates that it is of great potential in blind source separation problem for NNLM model. The loss function for the PE-NMF proposed here is in fact an extension of the multiplicative update algorithm proposed in [10], with the two terms introduced with parameters α and β , respectively, in which β is for update rule for the matrix H , which is similar to some sparse NMF algorithms [14], and α as the regularization term added to HH^T in the update rule for matrix W . The loss function for the PE-NMF is a special case of that proposed in [4]. However, in this approach, not only the learning algorithm is

motivated by expressing patterns more effectively and more efficiently, and experimented successfully in a wide range of applications, but also the convergence of the learning algorithm is proved by introducing some auxiliary function. In addition, a technique based on independent component analysis (ICA) is proposed for speeding up the learning procedure, and has been verified to be effective for the learning algorithm to converge to desired solutions.

Same as what has been mentioned in [4], the optimal choice of PE-NMF parameters depends on the distribution of data and a priori knowledge about the hidden (latent) components. However, our experimental results on extended bar problem indicate that the parameter choice is not so sensitive to some problems.

Acknowledgment

This work was supported by the National Science Fund of China under Grant nos. 60574039, 60371044, and Sino-Italian joint cooperation fund.

References

- [1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, New York, NY, USA, 2001.
- [2] P. O. Hoyer and A. Hyvärinen, “Independent component analysis applied to feature extraction from colour and stereo images,” *Network: Computation in Neural Systems*, vol. 11, no. 3, pp. 191–210, 2000.
- [3] J. Zhang, L. Wei, and Y. Wang, “Computational decomposition of molecular signatures based on blind source separation of non-negative dependent sources with NMF,” in *Proceedings of the 13th IEEE Workshop on Neural Networks for Signal Processing (NNSP ’03)*, pp. 409–418, Toulouse, France, September 2003.
- [4] A. Cichocki, R. Zdunek, and S. Amari, “New algorithms for non-negative matrix factorization in applications to blind source separation,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP ’06)*, vol. 5, pp. 621–624, Toulouse, France, May 2006.
- [5] A. Cichocki and R. Zdunek, “Multilayer nonnegative matrix factorization,” *Electronics Letters*, vol. 42, no. 6, pp. 947–948, 2006.
- [6] A. Pascual-Montano, J. M. Carazo, K. Kochi, D. Lehmann, and R. D. Pascual-Marqui, “Nonsmooth nonnegative matrix factorization (nsNMF),” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 403–415, 2006.
- [7] R. Zdunek and A. Cichocki, “Nonnegative matrix factorization with constrained second-order optimization,” *Signal Processing*, vol. 87, no. 8, pp. 1904–1916, 2007.
- [8] I. Kopriva, D. J. Garrood, and V. Borjanović, “Single frame blind image deconvolution by non-negative sparse matrix factorization,” *Optics Communications*, vol. 266, no. 2, pp. 456–464, 2006.
- [9] W. Liu and N. Zheng, “Non-negative matrix factorization based methods for object recognition,” *Pattern Recognition Letters*, vol. 25, no. 8, pp. 893–897, 2004.
- [10] 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.
- [11] <http://www.cis.hut.fi/projects/ica/fastical/>.

- [12] P. Földiák, “Forming sparse representations by local anti-Hebbian learning,” *Biological Cybernetics*, vol. 64, no. 2, pp. 165–170, 1990.
- [13] J. Khan, J. S. Wei, M. Ringnér, et al., “Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks,” *Nature Medicine*, vol. 7, no. 6, pp. 673–679, 2001.
- [14] M. Mørup, L. K. Hansen, and S. M. Arnfred, “Algorithms for sparse non-negative TUCKER (also named HONMF),” Tech. Rep., Technical University of Denmark, Lyngby, Denmark, 2007, http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/4658/pdf/imm4658.pdf.