An Improved Multiplicative Updating Algorithm for Nonnegative Independent Component Analysis

Hui Li, Yue-hong Shen, and Jian-gong Wang

This paper addresses nonnegative independent component analysis (NICA), with the aim to realize the blind separation of nonnegative well-grounded independent source signals, which arises in many practical applications but is hardly ever explored. Recently, Bertrand and Moonen presented a multiplicative NICA (M-NICA) algorithm using multiplicative update and subspace projection. Based on the principle of the mutual correlation minimization, we propose another novel cost function to evaluate the diagonalization level of the correlation matrix, and apply the multiplicative exponentiated gradient (EG) descent update to it to maintain nonnegativity. An efficient approach referred to as the EG-NICA algorithm is derived and its validity is confirmed by numerous simulations conducted on different types of source signals. Results show that the separation performance of the proposed EG-NICA algorithm is superior to that of the previous M-NICA algorithm, with a better unmixing accuracy. In addition, its convergence speed is adjustable by an appropriate user-defined learning rate.

Keywords: Blind source separation (BSS), independent component analysis (ICA), nonnegative constraint, multiplicative update, exponentiated gradient descent.

Manuscript received Apr. 12, 2012; revised Sept. 13, 2012; accepted Sept. 25, 2012.

This work was supported by the National Natural Science Foundation of China under Grant 61172061 and the Natural Science Foundation of JiangSu Province in China under Grant BK2011117.

Hui Li (phone: +86 13815402690, leehoo86@163.com), Yue-hong Shen (chunfeng22259@126.com), and Jian-gong Wang (hw12xian@hotmail.com) are with the Institute of Communication Engineering, PLA University of Science and Technology, Nanjing, China.

http://dx.doi.org/10.4218/etrij.13.0112.0224

I. Introduction

Blind source separation (BSS) is a signal processing problem that arises in many applications. It occurs when one attempts to extract latent sources from their observed mixtures without detailed knowledge of the source signals and the mixing process. Various approaches have been proposed to solve the BSS problem by exploiting available *a priori* information. Often, the mutual statistical independence between the sources is utilized in pioneering literature, which leads to finding a transformation in which the transformed signals are as independent as possible, that is, well-known independent component analysis (ICA) [1], [2].

As a physical condition in the real world, the nonnegativity constraint has attracted growing attention during the last decade, for example, as it applies to natural image [3], [4], spectral data [5]-[7], chemistry [8], music transcription [9], [10], and so on. The case of both nonnegative sources and nonnegative mixing coefficients without the independence assumption has been handled by using nonnegative matrix factorization algorithms [11], [12], wherein a nonnegative matrix is factorized into two smaller nonnegative matrices. However, the nonnegativity constraint alone is insufficient to yield a unique solution [13]. Plumbley and Oja considered the combination of nonnegativity and independence assumptions on the sources (no constraint is imposed on the mixing coefficients concerning the values [positive or negative] of the elements) and introduced nonnegative ICA (NICA) [14]-[17]. Thus, a closely related nonnegative principle component analysis (NPCA) algorithm was proposed to tackle the NICA problem [16]. Essentially, it is a special case of the nonlinear PCA algorithm with a rectification nonlinearity function, searching for an orthogonal rotation in which all of the whitened data fits into the positive region. However, it cannot ensure the negativity during the prewhitening process, resulting in degraded separation performance. Recently, Bertrand and Moonen presented a multiplicative NICA (M-NICA) algorithm [18]. Based on the minimization of mutual correlation, a cost function is constructed and the popular multiplicative update rule [12], [19] is utilized to minimize it under nonnegativity constraints. Though the M-NICA algorithm yields a better unmixing accuracy than the NPCA algorithm, it converges much more slowly since its learning rate cannot be user-defined due to the mechanism of the multiplicative update itself.

In this paper, the exponentiated gradient NICA (EG-NICA) algorithm is proposed. Starting from the same principle of mutual correlation minimization in [18], which means the covariance matrix should be approaching a diagonal matrix as much as possible, we establish another novel cost function to evaluate its diagonalization level. Then, the multiplicative exponentiated gradient descent update [20] is applied to decorrelate the data while at the same time maintain nonnegativity. Consequently, its convergence speed and unmixing accuracy depend on the learning rate, and experiments on different types of signals demonstrate that the unmixing accuracy can be improved significantly by choosing an appropriate parameter. Finally, a correction step based on subspace projection is required to restore the original signal subspace.

The rest of this paper is organized as follows. Section II briefly introduces the data model and the NPCA algorithm as well as the M-NICA algorithm. The proposed EG-NICA algorithm is derived in section III. Some simulation results and performance comparisons are provided in section IV. Finally, a concise summary is presented in section V.

II. Data Model and Related Algorithms

Consider the simplest form of the noise-free NICA problem, that is, linear instantaneous mixing. The *M* signals observed by a set of sensors $\mathbf{x}(t)=[x_1(t),\dots,x_M(t)]^T$ at time instant *t* are expressed as the linear mixture of N mutually independent source signals $\mathbf{s}(t)=[s_1(t),\dots,s_N(t)]^T$ without any time delay:

$$\mathbf{x}(t)=\mathbf{A}\cdot\mathbf{s}(t), \tag{1}$$

where the unknown $M\times N$ matrix $\mathbf{A}$ is called the mixing matrix (usually $M\geq N$). We deem the source s_n nonnegative if $\Pr(s_n<0)=0,\ n=1,\dots,N$. As done in [16], [18], here, an additional assumption that the sources are well grounded is made. This means that they have a non-zero probability density function all the way down to zero, that is, $\Pr(s_n<\delta)>0$ for any $\delta>0$. In practice, the sources are often well grounded, for example, when the sources have an on-off behavior or when the sources are sparse [18].

The task of NICA is to figure out an $N\times M$ unmixing matrix $\mathbf{W}$, so that the estimated signals are

$$\mathbf{y}(t)=\mathbf{W}\cdot\mathbf{x}(t)=\mathbf{WA}\cdot\mathbf{s}(t)=\mathbf{G}\cdot\mathbf{s}(t), \tag{2}$$

where $\mathbf{G}=\mathbf{WA}$ is called the global matrix with only one non-zero element in each row and each column, which permutes and scales the sources. Typically, we assume that the sources have unit variance, with any scaling factor being absorbed into the mixing matrix $\mathbf{A}$; then, $\mathbf{y}$ will be a permutation of $\mathbf{s}$ with just a sign ambiguity.

In [15], the following theorem was proven.

Theorem. Suppose that $\mathbf{s}$ is an *N*-dimensional vector of nonnegative and well-grounded mutually independent source signals with unit variance, and let $\mathbf{y}=\mathbf{U}\cdot\mathbf{s}$ be an orthonormal rotation of $\mathbf{s}$, that is, $\mathbf{U}^T\mathbf{U}=\mathbf{U}\mathbf{U}^T=\mathbf{I}_{N\times N}$, where $\mathbf{I}_{N\times N}$ denotes the $N\times N$ identity matrix. Then, $\mathbf{y}$ is a permutation of $\mathbf{s}$ if and only if the signals in $\mathbf{y}$ are nonnegative with probability 1.

Oja and Plumbley used the theorem to derive the aforementioned NPCA algorithm [16] for a simple solution to the NICA problem. The first stage is to whiten the observed data $\mathbf{x}$ by the whitening matrix $\mathbf{V}$ so that $\mathbf{z}=\mathbf{V}\cdot\mathbf{x}$ with $E\{\mathbf{z}\mathbf{z}^T\}=\mathbf{I}$. Therefore, it suffices to find an orthogonal matrix $\mathbf{K}$ for which $\mathbf{y}=\mathbf{K}\cdot\mathbf{z}$ preserves the nonnegativity. The learning rule of the NPCA algorithm becomes

$$\Delta\mathbf{K}=-\eta\left[f(y)y^T-yf(y)^T\right]\mathbf{K}, \tag{3}$$

where the rectification nonlinearity function $f(y_i)=\min(0,y_i)$, with a positive learning rate η.

A corollary from the above theorem was given in [18] as follows.

Corollary. Suppose that $\mathbf{s}$ is an *N*-dimensional vector of nonnegative and well-grounded mutually independent source signals with unit variance. Let $\mathbf{x}=\mathbf{A}\cdot\mathbf{s}$ with a full column rank $M\times N$ mixing matrix $\mathbf{A}$, and let $\mathbf{y}=\mathbf{W}\cdot\mathbf{x}$ with an $N\times M$ unmixing matrix $\mathbf{W}$. Then, $\mathbf{y}$ is a permutation of $\mathbf{s}$ if and only if the signals in $\mathbf{y}$ are mutually uncorrelated and nonnegative with probability 1.

Simplified from the NICA problem, this corollary finds an $N\times M$ unmixing matrix $\mathbf{W}$, which results in N nonnegative uncorrelated signals. Assuming we collect an $M\times L$ data matrix $\mathbf{X}$ that contains L number of $\mathbf{x}[l],\ l=1,\dots,L$ observations in its columns, then the rows of the $N\times L$ matrix $\mathbf{Y}=\mathbf{WX}$ are uncorrelated and only contain nonnegative values. Denote the covariance matrix as $\mathbf{C}_Y=(\mathbf{Y}-\bar{\mathbf{Y}})(\mathbf{Y}-\bar{\mathbf{Y}})^T$, where $\bar{\mathbf{Y}}=\frac{1}{L}\mathbf{Y}\mathbf{1}_{L\times L}$, with $\mathbf{1}_{L\times L}$ being an $L\times L$ matrix in which each element is 1. It is obvious that $\mathbf{Y} \geq 0$ and $\mathbf{C}_Y$ is a diagonal matrix due to the rows of $\mathbf{Y}$ being uncorrelated.

Accordingly, the M-NICA algorithm [18] constructs the cost function as

$$F(\mathbf{Y}) = \sum_{i,j} \frac{[\mathbf{C}_Y]_{ij}^2}{[\mathbf{C}_Y]_{ii}[\mathbf{C}_Y]_{jj}}. \quad (4)$$

Let $\nabla F(\mathbf{Y})$ denote the gradient of the above cost function, and it can be split into a positive part and a negative part, that is,

$$\nabla F(\mathbf{Y}) = \nabla^+ F(\mathbf{Y}) - \nabla^- F(\mathbf{Y}), \quad (5)$$

where $[\nabla^+ F(\mathbf{Y})]_{ij} \geq 0$ and $[\nabla^- F(\mathbf{Y})]_{ij} \geq 0$. Thus, the following multiplicative update rule [19] can be used to maintain nonnegativity for the M-NICA algorithm:

$$[\mathbf{Y}]_{ij} \leftarrow [\mathbf{Y}]_{ij} \frac{[\nabla^- F(\mathbf{Y})]_{ij}}{[\nabla^+ F(\mathbf{Y})]_{ij}}. \quad (6)$$

Note that no user-defined learning rate is required in (6), making it incapable of controlling the slow convergence of the M-NICA algorithm.

III. Proposed EG-NICA Algorithm

Based on the same principle of the mutual correlation minimization, we decide to construct a new cost function to measure the diagonalization level of the covariance matrix $\mathbf{C}_Y$ instead of evaluating the sum of the squared correlation coefficients of the rows of $\mathbf{Y}$ as (4). Hadamard inequation [21] tells us that:

$$\det(\mathbf{H}) \leq \prod_{i=1}^{m} \left(\sum_{j=1}^{m} \left| [\mathbf{H}]_{ij} \right|^2 \right)^{1/2}. \quad (7)$$

Equation (7) holds true for any rectangular $m \times m$ matrix $\mathbf{H}$. If $\mathbf{H}$ happens to be a diagonal matrix, then the equation sign comes into existence. Therefore, we introduce the following cost function:

$$F(\mathbf{Y}) = \ln \left(\prod_{i=1}^{N} \left(\sum_{j=1}^{N} \left| [\mathbf{C}_Y]_{ij} \right|^2 \right)^{1/2} \Big/ \left| \det(\mathbf{C}_Y) \right| \right). \quad (8)$$

Then, NICA is translated into such an optimization problem:

$$\begin{cases} \min\limits_{\mathbf{Y}} F(\mathbf{Y}), \\ \text{s.t.} \quad \mathbf{Y} \geq 0 \text{ and } \mathbf{Y} = \mathbf{W}\mathbf{X}. \end{cases} \quad (9)$$

To derive the learning rule of the proposed EG-NICA algorithm, the multiplicative exponentiated gradient descent update [20] is applied:

$$y_{mn} \leftarrow y_{mn} \exp(-\mu \frac{\partial F(\mathbf{Y})}{\partial y_{mn}} y_{mn}), \quad (10)$$

where the nonnegative learning rates μ can take different forms. If $\mathbf{Y}$ is initialized with nonnegative values, all of its elements y_{mn} will remain nonnegative under the update in (10), so the nonnegativity constraint of (9) is automatically satisfied. The exponentiated gradient descent update can be further improved in terms of convergence, computational efficiency, and numerical stability in several ways.

Here, for expression convenience, let $\mathbf{P} = \mathbf{I}_{L\times L} - \frac{1}{L}\mathbf{1}_{L\times 1}\mathbf{1}_{L\times 1}^T$, then $\mathbf{B} = \mathbf{Y}\mathbf{P} = \mathbf{Y} - \bar{\mathbf{Y}}$ and $\mathbf{C}_Y = \mathbf{Y}\mathbf{P}\mathbf{P}^T\mathbf{Y}^T = \mathbf{Y}\mathbf{P}\mathbf{Y}^T$.

Additionally, let $D(\mathbf{Y}) = \prod_{i=1}^{N} \left(\sum_{j=1}^{N} \left| [\mathbf{C}_Y]_{ij} \right|^2 \right)^{1/2}$,

$$\frac{\partial D(\mathbf{Y})}{\partial \mathbf{Y}} = \left[\begin{array}{l} \left(\left(\dfrac{\mathbf{1}_{N\times N}}{(\mathbf{C}_Y \odot \mathbf{C}_Y)\mathbf{1}_{N\times N}} \right)^T \odot \mathbf{C}_Y \right) \mathbf{B} \\ + \left(\dfrac{\mathbf{1}_{N\times N}}{(\mathbf{C}_Y \odot \mathbf{C}_Y)\mathbf{1}_{N\times N}} \right) \odot (\mathbf{C}_Y \mathbf{B}) \end{array} \right] \times D(\mathbf{Y}), \quad (11)$$

where the division operation denotes the element-wise division and $\odot$ denotes the element-wise multiplication. The deduction of (11) is presented in the appendix. Note that $\frac{\partial \det(\mathbf{Y}\mathbf{P}\mathbf{Y}^T)}{\partial \mathbf{Y}} = 2\left(\mathbf{Y}\mathbf{P}\mathbf{Y}^T\right)^{-1} \cdot \mathbf{Y}\mathbf{P}$ because $\mathbf{P}$ is a symmetric matrix [21]. Consequently, the derivative of the cost function in (8) with respect to $\mathbf{Y}$ is calculated as

$$\frac{\partial F(\mathbf{Y})}{\partial \mathbf{Y}} = \frac{\partial D(\mathbf{Y})}{\partial \mathbf{Y}} \Big/ D(\mathbf{Y}) - \left(2\mathbf{C}_Y^{-1} \cdot \mathbf{B}\right) \Big/ \left| \det(\mathbf{C}_Y) \right|. \quad (12)$$

By using (11) as a substitute in (12), it can be written as

$$\begin{aligned} \frac{\partial F(\mathbf{Y})}{\partial \mathbf{Y}} = & \left(\left(\frac{\mathbf{1}_{N\times N}}{(\mathbf{C}_Y \odot \mathbf{C}_Y)\mathbf{1}_{N\times N}} \right)^T \odot \mathbf{C}_Y \right) \mathbf{B} \\ & + \left(\frac{\mathbf{1}_{N\times N}}{(\mathbf{C}_Y \odot \mathbf{C}_Y)\mathbf{1}_{N\times N}} \right) \odot (\mathbf{C}_Y \mathbf{B}) \\ & - \left(2\mathbf{C}_Y^{-1} \cdot \mathbf{B}\right) \Big/ \left| \det(\mathbf{C}_Y) \right|. \end{aligned} \quad (13)$$

Though the nonnegativity constraint is satisfied, the second constraint in (9), that is, $\mathbf{Y} = \mathbf{W}\mathbf{X}$, also should be taken into account. The update process results in data that is not in the original signal subspace; hence, after each update in (10), a subspace projection-based correction step is enforced [18]:

$$\mathbf{Y} \leftarrow P_Y\{\mathbf{Y}\}. \quad (14)$$

In addition, the projection can be computed by a heuristic procedure. Represent $\mathbf{X}$ by its best rank N approximation by performing singular value decomposition [21]:

$$\{\mathbf{U}_X, \Sigma, \mathbf{V}_X\} \leftarrow svd(\mathbf{X}), \quad (15)$$

Algorithm 1. Proposed EG-NICA algorithm.

(1) Initialization: Replace $\mathbf{Y}$ with absolute values of the observation $\mathbf{X}$, that is,

$\forall i = 1,\ldots,N,\ j = 1,\ldots,L,\ [\mathbf{Y}]_{ij} \leftarrow |[\mathbf{X}]_{ij}|$

(2) Learning process:

for k=1:iteration times

update $\mathbf{Y}$ by (10) and (13).

project $\mathbf{Y}$ into the original signal subspace by (17).

end;

(3) Estimation: The unmixing matrix $\mathbf{W}$ can be computed as $\mathbf{W} = \mathbf{Y}\bar{\mathbf{V}}_X \bar{\Sigma}^{-1} \bar{\mathbf{U}}_X{}^T$.

$$\mathbf{X} \leftarrow \bar{\mathbf{U}}_X \bar{\Sigma} \bar{\mathbf{V}}_X{}^T, \tag{16}$$

where $\bar{\Sigma}$ is the $N \times N$ diagonal matrix including the N largest singular values of $\mathbf{X}$ and the columns of $\bar{\mathbf{U}}_X$ and $\bar{\mathbf{V}}_X$ are the corresponding left and right singular vectors, respectively. Then, (14) can be calculated by

$$[\mathbf{Y}]_{ij} \leftarrow \max\left([\mathbf{Y}\bar{\mathbf{V}}_X\bar{\mathbf{V}}_X{}^T]_{ij}, 0\right), \tag{17}$$

which can also guarantee that the negative values are rejected.

According to the above description, the main steps of the proposed EG-NICA algorithm are summarized in algorthm 1.

IV. Simulation Results

We provide several sets of simulation results to demonstrate the behavior of the proposed EG-NICA algorithm as well as the NPCA and M-NICA algorithms, and comparisons are also made among them with different types of source signals. In all the experiments of this paper, the three-source three-sensor model is adopted. The sources are nonnegative and scaled to unit variance. The 3×3 mixing matrix $\mathbf{A}$ is randomly generated by the Matlab code "randn," which can generate normally distributed variables, so $\mathbf{A}$ can be either negative or nonnegative. Two different measures, the cross-talk error (CTE) and the signal-to-error ratio (SER), for assessment are utilized:

$$CTE = \frac{1}{N}\sum_{n=1}^{N} E\left\{(s_n - y_n)^2\right\},$$

$$SER=10\log_{10}\left(\frac{1}{N}\sum_{n=1}^{N}\frac{E\left\{s_n{}^2\right\}}{E\left\{(s_n-y_n)^2\right\}}\right)(\text{dB}).$$

To draw conclusions in a general sense, 1,000 Monte-Carlo simulations are performed and the averaged results are presented.

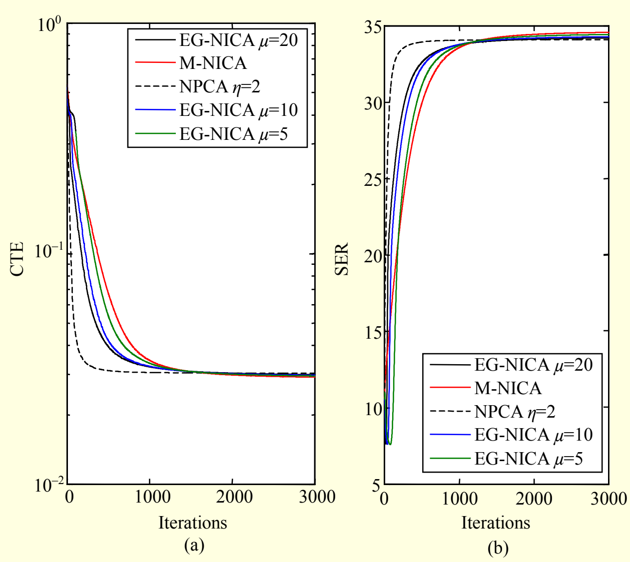

Fig. 1. (a) CTE and (b) SER performance for uniformly distributed random signals on unit interval, averaged over 1,000 independent runs.

1. Experiment 1: Uniformly Distributed Random Signals on the Unit Interval

In this experiment, we use a uniformly distributed random process on the unit interval (that is, [0, 1]) to generate L=1,000 samples of the N=3 nonnegative source signals. Figures 1(a) and 1(b) respectively show the CTE and the SER performance for the three algorithms versus the iteration number. Both the NPCA and the EG-NICA algorithm depend on a user-defined learning rate, that is, η and μ, respectively. Here, the learning rate for NPCA is set to be $\eta = 2$, which is observed to provide the best results (in terms of convergence speed and unmixing accuracy). Otherwise, NPCA will either have extremely slow convergence if the chosen η is too small or undesired oscillation of the separation performance and might not converge at all if η is too large.

From Fig. 1, we can see that the differences between the unmixing accuracy of the three algorithms are almost indistinguishable. However, NPCA converges much faster than M-NICA and EG-NICA, with the former algorithm based on rectified nonlinear PCA and the latter two algorithms based on mutual correlation minimization, which may explain the inherent reason for this phenomenon. On the other hand, EG-NICA under different learning rates outperforms M-NICA, considering the convergence speed, especially when $\mu = 20$. Additionally, as we might expect, it becomes slower as μ decreases.

What should be noticed is that the random signals disaccord with the well-grounded assumption in section II. In Experiment

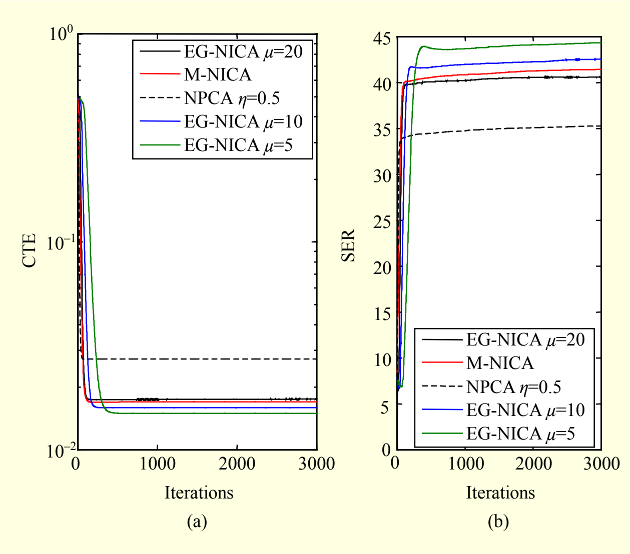

Fig. 2. (a) CTE and (b) SER performance for sparse signals on unit interval, averaged over 1,000 independent runs.

Fig. 3. Image separation for proposed EG-NICA algorithm, showing (a) three source images, (b) three mixed images, and (c) three recovered images.

2, sparse signals that can be deemed as well-grounded signals are used.

2. Experiment 2: Sparse Signals on the Unit Interval

To satisfy the well-grounded constraint, that is, $\forall\delta > 0$, $\Pr(0 \le s_n < \delta) > 0$, sparse signals on the unit interval containing clusters of zero-valued samples can be generated by modeling the on-off behavior of the sources. For the details of this process, refer to [18].

Figures 2(a) and 2(b) respectively show the CTE and the SER performance for the three algorithms versus the iteration number. The learning rate for NPCA is set to be $\eta = 0.5$, which is observed to provide the best results in this sparse signal case. It is easy to observe that compared to the previous experiment, all three algorithms achieve a faster convergence speed, and the improvement is especially obvious for M-NICA and EG-NICA. The signal sparsity contributes to this phenomenon. Furthermore, though NPCA still converges the fastest, the unmixing accuracy of M-NICA and EG-NICA is significantly better than that of NPCA. For example, the SER of M-NICA is nearly 6 dB higher than that of NPCA, as shown in Fig. 2(b).

As for EG-NICA, its unmixing accuracy enhances as the learning rate μ decreases. The SER of EG-NICA with $\mu = 5$ can achieve about 5 dB higher than M-NICA, which is 11 dB higher than NPCA. The advantage appears to be of vital usefulness when the separation accuracy is emphasized in practical applications. Otherwise, if the convergence speed is considered, EG-NICA with $\mu = 20$ behaves almost the same as M-NICA, that is, the differences in convergence speed and unmixing accuracy are negligible. In a word, the proposed EG-NICA is able to act superiorly to M-NICA with a properly chosen user-defined learning rate.

3. Experiment 3: Image Signals

The blind separation of image signals is suitable for the NICA problem since the pixel values of images are nonnegative integers, which are located between 0 and 255. The three source images used in this experiment are shown in Fig. 3(a). They are natural images of size 128×128 (downsampled by a factor of four from the 512×512 original images [22]). Each source is the sequence of pixel values of length L=128^2, which is obtained as we scan across each image from top left to bottom right. After scaling each source into unit variance, the source covariance matrix is

$$cov = \begin{bmatrix} 1.0000 & 0.0552 & -0.1060 \\ 0.0552 & 1.0000 & -0.0449 \\ -0.1060 & -0.0449 & 1.0000 \end{bmatrix}, \qquad (18)$$

which indicates that the source image signals are slightly correlated, and the small correlation is validated to be acceptable for the NICA problem.

The mixed images and the recovered images by the proposed EG-NICA algorithm in a typical run are shown in Figs. 3(b) and 3(c), respectively. After learning over 3×10^3 steps, the global matrix $\mathbf{G}$ becomes

$$\mathbf{G}=\begin{bmatrix} 0.0968 & \mathbf{0.9594} & -0.1974 \\ \mathbf{0.9911} & -0.0991 & 0.2704 \\ -0.1505 & 0.2730 & \mathbf{0.9488} \end{bmatrix} \tag{19}$$

with a dominant element in each row and each column, which indicates that the group of recovered signals is a permutation of the group of corresponding source signals. Therefore, we can say that the proposed EG-NICA algorithm is able to realize the successful blind separation of images.

V. Conclusion

In this paper, we considered the nonnegative ICA problem and proposed a novel EG-NICA algorithm for well-grounded source signals. The EG-NICA algorithm is derived by applying the multiplicative exponentiated gradient descent update rule to the constructed cost function, proposed based on the mutual correlation minimization principle. Simulations on different types of signals were carried out to illustrate its separation performance. Compared with the NPCA algorithm, it is gradient-based and depends on the learning rate. Though EG-NICA converges more slowly than NPCA, its SER to evaluate the unmixing accuracy can achieve nearly 11 dB higher than that of NPCA for sparse signals, which is rather considerable. Compared to another related algorithm, referred to as M-NICA, EG-NICA has a flexible learning rate and has superior behavior, both in terms of convergence speed and unmixing accuracy if an appropriate learning rate is selected. Therefore, the proposed EG-NICA is best suited for those practical applications calling for accurate separation. In future research, it would be interesting and valuable to explore NICA algorithms for signals without the requirement that the signals be well grounded.

Appendix. The Deduction of (11)

Let us review the related mathematical expressions, $\mathbf{B}=\mathbf{YP}=\mathbf{Y}-\bar{\mathbf{Y}}$, $\mathbf{C}_Y=\mathbf{YPP}^T\mathbf{Y}^T=\mathbf{YPY}^T$, and $D(\mathbf{Y})=\prod_{i=1}^{N}\left(\sum_{j=1}^{N}\left|[\mathbf{C}_Y]_{ij}\right|^2\right)^{1/2}$. Then,

$$\frac{\partial D(\mathbf{Y})}{\partial y_{mn}}=\sum_{i=1}^{N}\left[\frac{\partial\left(\sum_{j=1}^{N}\left|[\mathbf{C}_Y]_{ij}\right|^2\right)}{\partial y_{mn}}\cdot\frac{1}{2}\left(\sum_{j=1}^{N}\left|[\mathbf{C}_Y]_{ij}\right|^2\right)^{-\frac{1}{2}}\prod_{\substack{k=1\\k\neq i}}^{N}\left(\sum_{j=1}^{N}\left|[\mathbf{C}_Y]_{kj}\right|^2\right)^{1/2}\right]$$

$$=\frac{D(\mathbf{Y})}{2}\times\sum_{i=1}^{N}\left[\left(\sum_{j=1}^{N}\left|[\mathbf{C}_Y]_{ij}\right|^2\right)^{-1}\frac{\partial\left(\sum_{j=1}^{N}\left|[\mathbf{C}_Y]_{ij}\right|^2\right)}{\partial y_{mn}}\right]. \tag{A1}$$

Note that

$$\mathbf{B}_{ij}=[\mathbf{YP}]_{ij}=\sum_{k=1}^{N}y_{ik}p_{kj}, \tag{A2}$$

$$[\mathbf{C}_Y]_{ij}=[\mathbf{YPY}^T]_{ij}=\sum_{k=1}^{N}\sum_{q=1}^{N}y_{ik}p_{kq}y_{jq}. \tag{A3}$$

Hence, $\forall i\neq m$,

$$\frac{\partial\sum_{j=1}^{N}\left(\left|[\mathbf{C}_Y]_{ij}\right|^2\right)}{\partial y_{mn}}=2[\mathbf{C}_Y]_{im}\frac{\partial\left([\mathbf{C}_Y]_{im}\right)}{\partial y_{mn}}=2[\mathbf{C}_Y]_{im}\left(\sum_{k=1}^{N}y_{ik}p_{kn}\right)=2[\mathbf{C}_Y]_{im}\mathbf{B}_{in}. \tag{A4}$$

Particularly, $\forall i=m$,

$$\begin{aligned}\frac{\partial\sum_{j=1}^{N}\left(\left|[\mathbf{C}_Y]_{mj}\right|^2\right)}{\partial y_{mn}}&=\sum_{\substack{j=1\\j\neq m}}^{N}\left(2[\mathbf{C}_Y]_{mj}\frac{\partial\left([\mathbf{C}_Y]_{mj}\right)}{\partial y_{mj}}\right)+2[\mathbf{C}_Y]_{mm}\frac{\partial\left([\mathbf{C}_Y]_{mm}\right)}{\partial y_{mn}}\\&=2\sum_{\substack{j=1\\j\neq m}}^{N}\left[[\mathbf{C}_Y]_{mj}\left(\sum_{q=1}^{N}p_{nq}y_{jq}\right)\right]\\&\quad+2[\mathbf{C}_Y]_{mm}\left[\left(\sum_{k=1}^{N}y_{mk}p_{kn}\right)+\left(\sum_{q=1}^{N}p_{nq}y_{mq}\right)\right]\\&\xrightarrow{(\because\mathbf{P}=\mathbf{P}^T)}=2\sum_{j=1}^{N}\left[[\mathbf{C}_Y]_{mj}\left(\sum_{q=1}^{N}y_{jq}p_{qn}\right)\right]\\&\quad+2[\mathbf{C}_Y]_{mm}\left(\sum_{k=1}^{N}y_{mk}p_{kn}\right)\\&=2\sum_{j=1}^{N}\left([\mathbf{C}_Y]_{mj}\mathbf{B}_{jn}\right)+2[\mathbf{C}_Y]_{mm}\mathbf{B}_{mn}.\end{aligned} \tag{A5}$$

Substituting (A4) and (A5) into (A1),

$$\begin{aligned}\frac{\partial D(\mathbf{Y})}{\partial y_{mn}}&=\frac{D(\mathbf{Y})}{2}\times\sum_{i=1}^{N}\left[\left(\sum_{j=1}^{N}\left|[\mathbf{C}_Y]_{ij}\right|^2\right)^{-1}\frac{\partial\left(\sum_{j=1}^{N}\left|[\mathbf{C}_Y]_{ij}\right|^2\right)}{\partial y_{mn}}\right]\\&=D(\mathbf{Y})\times\left(\begin{aligned}&\sum_{i=1}^{N}\left[\left(\sum_{j=1}^{N}\left|[\mathbf{C}_Y]_{ij}\right|^2\right)^{-1}[\mathbf{C}_Y]_{im}\mathbf{B}_{in}\right]\\&+\left(\sum_{j=1}^{N}\left|[\mathbf{C}_Y]_{mj}\right|^2\right)^{-1}\sum_{j=1}^{N}\left([\mathbf{C}_Y]_{mj}\mathbf{B}_{jn}\right)\end{aligned}\right).\end{aligned} \tag{A6}$$

Thus, it is obvious that (A6) can be extended as follows:

$$\frac{\partial D(\mathbf{Y})}{\partial\mathbf{Y}}=\left[\begin{aligned}&\left(\left(\frac{\mathbf{1}_{N\times N}}{(\mathbf{C}_Y\odot\mathbf{C}_Y)\mathbf{1}_{N\times N}}\right)^T\odot\mathbf{C}_Y\right)\mathbf{B}\\&+\left(\frac{\mathbf{1}_{N\times N}}{(\mathbf{C}_Y\odot\mathbf{C}_Y)\mathbf{1}_{N\times N}}\right)\odot(\mathbf{C}_Y\mathbf{B})\end{aligned}\right]\times D(\mathbf{Y}), \tag{A7}$$

where the division operation denotes the element-wise division and $\odot$ denotes the element-wise multiplication.

The deduction is completed.

References

[1] P. Comon, "Independent Component Analysis, A New Concept?" *Signal Process.*, vol. 36, no. 3, Apr. 1994, pp. 287-314.

[2] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, Oxford: Academic Press, 2010.

[3] J. Nascimento and J. Dias, "Does Independent Component Analysis Play a Role in Unmixing Hyperspectral Data?" *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 1, Jan. 2005, pp. 175-187.

[4] C.-I. Chang, *Hyperspectral Data Exploitation: Theory and Applications*, Hoboken, NJ: John Wiley & Sons, Inc., 2007.

[5] V.P. Pauce et al., "Unmixing Spectral Data for Space Objects Using Independent Component Analysis and Nonnegative Matrix Factorization," *Proc. ATC*, 2004.

[6] N. Dobigeon et al., "Bayesian Separation of Spectral Sources under Non-negativity and Full Additivity Constraints," *Signal Process.*, vol. 89, no. 12, Dec. 2009, pp. 2657-2669.

[7] Y. Sun et al., "Postprocessing and Sparse Blind Source Separation of Positive and Partially Overlapped Data," *Signal Process.*, vol. 91, no. 8, Aug. 2011, pp. 1838-1851.

[8] R.-C. Henry, "Multivariate Receptor Models – Current Practice and Future Trends," *Chemom. Intell. Lab. Syst.*, vol. 60, no. 1-2, Jan. 2002, pp. 43-48.

[9] M.-D. Plumbley et al., "Automatic Music Transcription and Audio Source Separation," *Cybern. Syst.*, vol. 33, no. 6, 2002, pp. 603-627.

[10] S. Abdallah and M. Plumbley, "Polyphonic Transcription by Non-negative Sparse Coding of Power Spectra," *Proc. ISMIR*, Barcelona, Spain, 2004, pp. 318-325.

[11] D.-D. Lee and H.-S. Seung, "Learning the Parts of Objects by Non-negative Matrix Factorization," *Nature*, vol. 401, Oct. 1999, pp. 788-791.

[12] D.-D. Lee and H.-S. Seung, "Algorithms for Non-negative Matrix Factorization," *Proc. NIPS*, Cambridge, MA: MIT Press, 2001, pp. 556-562.

[13] G.-F. Harpur, *Low Entropy Coding with Unsupervised Neural Networks*, doctoral dissertation, University of Cambridge, 1997.

[14] M.-D. Plumbley, "Algorithms for Nonnegative Independent Component Analysis," *IEEE Trans. Neural Netw.*, vol. 4, no. 3, May 2003, pp. 534-543.

[15] M.-D. Plumbley, "Conditions for Nonnegative Independent Component Analysis," *IEEE Signal Process. Lett.*, vol. 9, no. 6, June 2002, pp. 177-180.

[16] E. Oja and M.-D. Plumbley, "Blind Separation of Positive Sources Using Non-negative PCA," *Proc. ICABSS*, Nara, Japan, 2003, pp. 11-16.

[17] M.-D. Plumbley and E. Oja, "A 'Nonnegative PCA' Algorithm for Independent Component Analysis," *IEEE Trans. Neural Netw.*, vol. 15, no. 1, Jan. 2004, pp. 66-76.

[18] A. Bertrand and M. Moonen, "Blind Separation of Non-negative Source Signals Using Multiplicative Updates and Subspace Projection," *Signal Process.*, vol. 90, no. 10, Oct. 2010, pp. 2877-2890.

[19] Z. Yang and J. Laaksonen, "Multiplicative Updates for Non-negative Projections," *Neurocomputing*, vol. 71, no. 1-3, Dec. 2007, pp. 363-373.

[20] A. Cichocki and S. Amari, "Extended SMART Algorithms for Non-negative Matrix Factorization," *Proc. AISC*, Zakopane, Poland, 2006, pp. 548-562.

[21] H. Lutkepohl, *Handbook of Matrices*, Chichester: John Wiley & Sons Ltd., 1996.

[22] A. Cichocki et al., ICALAB Toolboxes. http://www.bsp.brain.riken.jp/ICALAB

Hui Li received her BS in communication engineering from PLA University of Science and Technology in 2007, Nanjing, China. She is currently pursuing her PhD in the same university. Her main research interests are blind signal processing, high-efficiency spectrum utilization, and wireless resource management.

Yue-hong Shen received his PhD in communication engineering from Nanjing University of Science and Technology (NUST) in 1999. He is a professor and doctoral advisor in the Wireless Communication Department at the Institute of Communication Engineering, PLA University of Science and Technology, Nanjing, China. His current research interests include mobile communication systems, high-efficiency spectrum utilization, multiple antenna techniques, and multicarrier modulation techniques. Dr. Shen has authored many journal papers, as well as conference papers.

Jian-gong Wang received his MS in communication engineering from PLA University of Science and Technology in Nanjing, China, in 2008. He is now working toward his PhD at the same university. His research interests include anti-jamming techniques, cooperative communication, and multiuser information theory.