Hui LI, Yue-hong SHEN, Jian-gong WANG

Institute of Communications Engineering, PLA University of Science and Technology

Online Blind Separation of Dependent Sources Using Nonnegative Matrix Factorization Based on KL Divergence

***Abstract**. This paper proposes a novel online algorithm for nonnegative matrix factorization (NMF) based on the generalized Kullback-Leibler (KL) divergence criterion, aimed to overcome the high computation problem of large-scale data brought about by conventional batch NMF algorithms. It features stable updating the factors alternately for each new-coming observation, and provides an efficient solution for the blind separation of statistically dependent sources (i.e., the sources are mutually correlated). Our theoretic analysis is validated by simulation examples.*

***Streszczenie.** Przedstawiono nowy algorytm do faktoryzacji nieujemnej macierzy bazujący na kryterium Kullback-Leibler, pozwalający usprawnić problem obliczeń dużej ilości danych. Algorytm sukcesywnie zmienia współczynniki i pozwala na ślepą separację statystycznie zależnych źródeł. (**On-line ślepa separacja zależnych źródeł przy użyciu faktoryzacji nieujemnej matrycy bazująca na kryterium Kullback-Leibler**)*

Keywords: nonnegative matrix factorization (NMF); KL divergence; blind source separation (BSS); dependent sources.
Słowa kluczowe: nieujemna macierz, ślepa separacja, dane zależne.

Introduction

The challenging blind source separation (BSS) consists of recovering unknown source signals only from observed mixtures. This issue has received intensive attention during the past decade with a broad range of applications, such as digital communications [1], biomedical image processing [2], speech enhancement [3], etc. Most of the proposed BSS techniques rely on the assumption that source signals are independent or at least uncorrelated. However, there are many practical applications that require considering sources that usually exhibit slight or strong dependence, such as image processing and spectral data analysis [4-5]. Take a wireless surveillance system for example, images captured by cameras that monitor overlapped areas are mutually correlated. Traditional BSS methods will fail for these cases with statistically dependent sources.

The existing works on blind separation of dependent sources are scare and far from mature, such as the algorithm based on the maximization of a non-Gaussianity measure [6], the method under the assumption that wideband sources are dependent but some of their sub-components are independent [7], and the approach proposed in [8] using properly designed precoders. To settle the dependent source separation, more and more researchers have been paying attention to the newly developed nonnegative matrix factorization (NMF) [9-10], in which a nonnegative matrix is decomposed into the product of two smaller nonnegative matrices. This corresponds to the BSS case where both the mixing matrix and the sources are nonnegative. As NMF does not rely on the statistical features of the sources (such as independence, nonstationarity, etc), it (with some constraints, such as the volume constraint [11]) shows some potential to perform BSS no matter the sources are mutually dependent or independent [12]. Besides, nonnegativity arises in many practical problems, e.g., natural images, the microarray data [13], chemistry [14] and music transcription [15], etc.

It is well known that the computational complexity of traditional batch NMF is proportional to the length of observation data. So the problems of huge storage requirement and high computational complexity become serious especially for large-scale observations. And this paper considers the NMF-based online scheme of dependent source separation. Paper [16] mentioned that an effective online factorization is expected to update its factors without causing much computational effort. Recently, Zhou et al. [17] have derived an INMF-VC (i.e., incremental NMF with volume constraint) algorithm and utilized it for solving adaptive BSS. However, they only gave the derivation for the Euclidean distance cost function. Accordingly to Lee and Seung [10], another mostly used cost function for NMF is based on the generalized Kullback-Leibler (KL) divergence, which intrinsically ensures nonnegativity constraints and naturally affects the reconstruction performance. For example, the divergence is more sensitive to low-energy observations, making it a better approximation of human auditory perception [18]. Here we present an incremental NMF algorithm based on KL divergence (INMF-KL), which features stable updating the factors alternately for each new-coming observation and is suitable for online blind separation of large-scale dependent sources. What is more, it can be deemed as an important complement for the above INMF-VC algorithm.

The rest of this paper is organized as follows. In section II, we briefly review the NMF-based BSS models. The next section presents the INMF-KL algorithm. The simulation experiments are then carried out in section IV. Finally, a concise conclusion is given.

NMF-based BSS Model

Considering the linear instantaneous mixng BSS model without noise, the nonnegative observations (or mixtures) can be represented by an $m \times l$ matrix $\mathbf{X} = [x_{ij}] \in_{+}^{m \times l}$, with all of its entries being nonnegative. The goal of NMF is to find the nonnegative mixing matrix $\mathbf{W} = [w_{ij}] \in_{+}^{m \times n}$ and the nonnegative latent sources $\mathbf{H} = [h_{ij}] \in_{+}^{n \times l}$, such that

(1) $$\mathbf{X} \approx \mathbf{WH}$$

For simplicity, we will mainly discuss the determined case of $m = n > 2$. It is also assumed that $\mathbf{W}$ is of full column rank, i.e., $\mathrm{rank}(\mathbf{W}) = n$. Besides, to remove the effect of scaling indeterminacy, the columns of $\mathbf{W}$ are assumed to have unit length with respect to the L1-norm, i.e., $\sum_{i=1}^{n} w_{ij} = 1$ [17]. Thus,

(2) $$\sum_{j=1}^{n} h_{jt} = \sum_{j=1}^{n} \sum_{i=1}^{n} w_{ij} h_{jt} = \sum_{i=1}^{n} x_{it} \quad (\forall t \in \{1,2,\ldots,l\})$$

Let $\tilde{x}_{it} = x_{it} \big/ \sum_{i=1}^{n} x_{it}$, $\tilde{h}_{jt} = h_{jt} \big/ \sum_{j=1}^{n} x_{jt}$, then Eq.(1) can be rewritten as

(3) $$\tilde{\mathbf{X}} \approx \mathbf{W}\tilde{\mathbf{H}}$$

The cost function of standard NMF using the KL divergence measure [10] is

(4) $$D_{KL}(\tilde{\mathbf{X}} \| \mathbf{W}\tilde{\mathbf{H}}) = \sum_{ij} \left(\tilde{x}_{ij} \ln \left(\tilde{x}_{ij} \big/ [\mathbf{W}\tilde{\mathbf{H}}]_{ij} \right) - \tilde{x}_{ij} + [\mathbf{W}\tilde{\mathbf{H}}]_{ij} \right)$$

To enforce uniqueness of the factorization, the objective function of modified NMF with volume constraint is as follows

(5) $$\begin{cases} \text{Min}: J = D_{KL}(\tilde{\mathbf{X}} \| \mathbf{W}\tilde{\mathbf{H}}) + \mu \ln|\det \mathbf{W}| \\ \text{s.t.} \quad w_{ij} \geq 0, h_{jt} \geq 0, \sum_{i=1}^{n} w_{ij} = 1 \end{cases}$$

where $\mu > 0$ is the balanced parameter.

INMF-KL Algorithm

The computational complexity of NMF is $O(mnl)$ per iteration, which implies that computational load linearly increases as each sample arrives. Obviously, it is impractical to execute the whole batch NMF process repeatedly. Considering the effect of each coming sample, a new column should be added to both $\mathbf{X}$ and $\mathbf{H}$, and the mixing matrix $\mathbf{W}$ needs to be updated [16].

The cost function of INMF-KL

Let $\mathbf{W}_k$ and $\tilde{\mathbf{H}}_k$ denote the optimized factor matrices for the initial k observation samples, which can be obtained by normal batch NMF algorithm. And we can get the corresponding cost function J_k as

(6) $$J_k = \sum_{i=1}^{n}\sum_{j=1}^{k}\left(\tilde{x}_{ij}\ln(\tilde{x}_{ij}/[\mathbf{W}_k\tilde{\mathbf{H}}_k]_{ij}) - \tilde{x}_{ij} + [\mathbf{W}_k\tilde{\mathbf{H}}_k]_{ij}\right) + \mu\ln|\det \mathbf{W}_k|$$

As the sample number increases, the effect of a new sample on $\mathbf{W}$ decreases, because the new sample would be unable to significantly influence the optimality of $\mathbf{W}$ for previous samples, i.e., $\mathbf{W}_{k+1} \approx \mathbf{W}_k$ if k is large enough [17]. Hence,

(7) $$J_k \approx \sum_{i=1}^{n}\sum_{j=1}^{k}\left(\tilde{x}_{ij}\ln(\tilde{x}_{ij}/[\mathbf{W}_{k+1}\tilde{\mathbf{H}}_k]_{ij}) - \tilde{x}_{ij} + [\mathbf{W}_{k+1}\tilde{\mathbf{H}}_k]_{ij}\right) + \mu\ln|\det \mathbf{W}_k|$$

When the $(k+1)$-th sample $\tilde{\boldsymbol{x}}^{(k+1)}$ arrives, the first k columns of $\tilde{\mathbf{H}}_{k+1}$ are approximately equal to $\tilde{\mathbf{H}}_k$, and the cost function J_{k+1} based on Eq.(6) is updated as

(8) $$\begin{aligned} J_{k+1} &= \sum_{i=1}^{n}\sum_{j=1}^{k+1}\left(\tilde{x}_{ij}\ln(\tilde{x}_{ij}/[\mathbf{W}_{k+1}\tilde{\mathbf{H}}_{k+1}]_{ij}) - \tilde{x}_{ij} + [\mathbf{W}_{k+1}\tilde{\mathbf{H}}_{k+1}]_{ij}\right) + \mu\ln|\det \mathbf{W}_{k+1}| \\ &\approx \sum_{i=1}^{n}\sum_{j=1}^{k}\left(\tilde{x}_{ij}\ln(\tilde{x}_{ij}/[\mathbf{W}_{k+1}\tilde{\mathbf{H}}_k]_{ij}) - \tilde{x}_{ij} + [\mathbf{W}_{k+1}\tilde{\mathbf{H}}_k]_{ij}\right) + \mu\ln|\det \mathbf{W}_{k+1}| \\ &\quad + \sum_{i=1}^{n}\left(\tilde{x}_{i,k+1}\ln(\tilde{x}_{i,k+1}/[\mathbf{W}_{k+1}\tilde{\boldsymbol{h}}^{(k+1)}]_i) - \tilde{x}_{i,k+1} + [\mathbf{W}_{k+1}\tilde{\boldsymbol{h}}^{(k+1)}]_i\right) \\ &\approx J_k + \sum_{i=1}^{n}\left(\tilde{x}_{i,k+1}\ln(\tilde{x}_{i,k+1}/[\mathbf{W}_{k+1}\tilde{\boldsymbol{h}}^{(k+1)}]_i) - \tilde{x}_{i,k+1} + [\mathbf{W}_{k+1}\tilde{\boldsymbol{h}}^{(k+1)}]_i\right) \\ &\quad + \mu\ln|\det \mathbf{W}_{k+1}| - \mu\ln|\det \mathbf{W}_k| \end{aligned}$$

Let

(9) $$\begin{aligned} \Delta J_{k+1} &= \sum_{i=1}^{n}\left(\tilde{x}_{i,k+1}\ln(\tilde{x}_{i,k+1}/[\mathbf{W}_{k+1}\tilde{\boldsymbol{h}}^{(k+1)}]_i) - \tilde{x}_{i,k+1} + [\mathbf{W}_{k+1}\tilde{\boldsymbol{h}}^{(k+1)}]_i\right) \\ &\quad + \mu\ln|\det \mathbf{W}_{k+1}| - \mu\ln|\det \mathbf{W}_k| \end{aligned}$$

Then

(10) $$J_{k+1} = J_k + \Delta J_{k+1}$$

Adopting the same amnestic average method in [17] to realize satistical efficiency, Eq.(10) can be modified into

(11) $$J_{k+1} = \alpha J_k + \beta \Delta J_{k+1}$$

where α, β denote the smoothing parameters, typically, $\alpha = 1 - L/k, \beta = L/k$ ($L \in \{1,2,3,4\}$, and $L = 1$ in this paper).

Multiplicative updating rules of INMF-KL

After constructing the cost function given by Eq.(11), gradient descent optimization will be performed to update the mixing matrix $\mathbf{W}$ and the corresponding new column of $\tilde{\mathbf{H}}$ (i.e., $\tilde{\boldsymbol{h}}$) whenever a new sample is acquired.

The update rule of $\tilde{\boldsymbol{h}}^{(k+1)}$ can be formulated as

(12) $$\tilde{\boldsymbol{h}}^{(k+1)} \leftarrow \tilde{\boldsymbol{h}}^{(k+1)} - \eta_{\tilde{h}}\frac{\partial J_{k+1}}{\partial \tilde{\boldsymbol{h}}^{(k+1)}}$$

Due to the fact that $\frac{\partial J_k}{\partial \tilde{\boldsymbol{h}}^{(k+1)}} = 0$, and the partial derivation in Eq.(12) can be calculated as

(13) $$\frac{\partial J_{k+1}}{\partial \tilde{\boldsymbol{h}}^{(k+1)}} = \beta\frac{\partial \Delta J_{k+1}}{\partial \tilde{\boldsymbol{h}}^{(k+1)}} = \beta\left(-\mathbf{W}_{k+1}{}^T\frac{\tilde{\boldsymbol{x}}^{(k+1)}}{\mathbf{W}_{k+1}\tilde{\boldsymbol{h}}^{(k+1)}} + \mathbf{W}_{k+1}{}^T\mathbf{1}_{n\times 1}\right)$$

where $\mathbf{1}_{n\times 1}$ represents $n\times 1$ vector with all the entries equal to unity, and the division operation denotes the component-wise division without special explanation. Select the step size as

(14) $$\eta_{\tilde{h}} = \frac{\tilde{\boldsymbol{h}}^{(k+1)}}{\mathbf{W}_{k+1}{}^T\mathbf{1}_{n\times 1}} = \tilde{\boldsymbol{h}}^{(k+1)} \text{ (notice that } \textstyle\sum_{i=1}^{n} w_{ij} = 1\text{)}$$

Then

(15) $$\tilde{\boldsymbol{h}}^{(k+1)} \leftarrow \tilde{\boldsymbol{h}}^{(k+1)} \otimes \beta\left(-\mathbf{W}_{k+1}{}^T\frac{\tilde{\boldsymbol{x}}^{(k+1)}}{\mathbf{W}_{k+1}\tilde{\boldsymbol{h}}^{(k+1)}}\right)$$

where $\otimes$ denotes the component-wise multiplication.

On the other hand, we can get the update rule of $\mathbf{W}_{k+1}$ as

(16) $$\mathbf{W}_{k+1} \leftarrow \mathbf{W}_{k+1} - \eta_{\mathrm{W}}\frac{\partial J_{k+1}}{\partial \mathbf{W}_{k+1}}$$

Considering that $\frac{\partial \ln|\det \mathbf{W}|}{\partial \mathbf{W}} = (\mathbf{W}^{-1})^T$, Eq.(16) can be modified by natural gradient to avoid negative values resulted by the inversion operation as

(17) $$\mathbf{W}_{k+1} \leftarrow \mathbf{W}_{k+1} - \eta_{\mathrm{W}}\frac{\partial J_{k+1}}{\partial \mathbf{W}_{k+1}}\mathbf{W}_{k+1}{}^T\mathbf{W}_{k+1}$$

and the partial derivation in Eq.(17) can be calculated as

(18) $$\begin{aligned} \frac{\partial J_{k+1}}{\partial \mathbf{W}_{k+1}} &= \alpha\frac{\partial J_k}{\partial \mathbf{W}_{k+1}} + \beta\frac{\partial \Delta J_{k+1}}{\partial \mathbf{W}_{k+1}} \\ &= \alpha\left(-\frac{\tilde{\mathbf{X}}_k}{\mathbf{W}_{k+1}\tilde{\mathbf{H}}_k}\tilde{\mathbf{H}}_k{}^T + \mathbf{1}_{n\times k}\tilde{\mathbf{H}}_k{}^T\right) \\ &\quad + \beta\left(-\frac{\tilde{\boldsymbol{x}}^{(k+1)}}{\mathbf{W}_{k+1}\tilde{\boldsymbol{h}}^{(k+1)}}\left(\tilde{\boldsymbol{h}}^{(k+1)}\right)^T + \mathbf{1}_{n\times 1}\left(\tilde{\boldsymbol{h}}^{(k+1)}\right)^T + \mu(\mathbf{W}_{k+1}{}^{-1})^T\right) \end{aligned}$$

Select the step size as

(19) $$\eta_{\mathrm{W}} = \frac{\mathbf{W}_{k+1}}{\left(\alpha\mathbf{1}_{n\times k}\tilde{\mathbf{H}}_k{}^T + \beta\mathbf{1}_{n\times 1}\left(\tilde{\boldsymbol{h}}^{(k+1)}\right)^T\right)\mathbf{W}_{k+1}{}^T\mathbf{W}_{k+1} + \beta\mu\mathbf{W}_{k+1}}$$

Thus

(20) $$\mathbf{W}_{k+1} \leftarrow \mathbf{W}_{k+1} \otimes \frac{\alpha\frac{\tilde{\mathbf{X}}_k}{\mathbf{W}_{k+1}\tilde{\mathbf{H}}_k}\tilde{\mathbf{H}}_k{}^T\mathbf{W}_{k+1}{}^T\mathbf{W}_{k+1} + \beta\frac{\tilde{\boldsymbol{x}}^{(k+1)}}{\mathbf{W}_{k+1}\tilde{\boldsymbol{h}}^{(k+1)}}\left(\tilde{\boldsymbol{h}}^{(k+1)}\right)^T\mathbf{W}_{k+1}{}^T\mathbf{W}_{k+1}}{\left(\alpha\mathbf{1}_{n\times k}\tilde{\mathbf{H}}_k{}^T + \beta\mathbf{1}_{n\times 1}\left(\tilde{\boldsymbol{h}}^{(k+1)}\right)^T\right)\mathbf{W}_{k+1}{}^T\mathbf{W}_{k+1} + \beta\mu\mathbf{W}_{k+1}}$$

where $\mu = 70\exp(-0.006(k+1))$ is choosed [17].

Notes: Since both $\tilde{\mathbf{X}}_k$ and $\tilde{\mathbf{H}}_k$ do not change throughout the process, instead of storing $\tilde{\mathbf{X}}_k$ and $\tilde{\mathbf{H}}_k$, the multiplication $\tilde{\mathbf{X}}_k\tilde{\mathbf{H}}_k{}^T$ can be stored. Though their dimensions increase as a new sample arrive, only the last columns are added, and instead of implementing the multiplication repeatedly, we can make use of the following calculation version given by

(21) $$\tilde{\mathbf{X}}_{k+1}\tilde{\mathbf{H}}_{k+1}{}^T = \tilde{\mathbf{X}}_k\tilde{\mathbf{H}}_k{}^T + \tilde{\boldsymbol{x}}^{(k+1)}\left(\tilde{\boldsymbol{h}}^{(k+1)}\right)^T$$

Summary

Table 1 provides a summary of our proposed INMF-KL algorithm for online blind separation of dependent sources.

Table 1. The proposed INMF-KL algorithm

1. Achieve $\mathbf{W}_k$ and $\tilde{\mathbf{H}}_k$

Step1: collect k obervation samples $\mathbf{X}_k$ and do rescaling, i.e., $\tilde{x}_{it} = x_{it} \big/ \sum_{i=1}^{n} x_{it}$;

Step2: perform normal batch NMF algorithm to obtain the corresponding $\mathbf{W}_k$ and $\tilde{\mathbf{H}}_k$.

2. Update $\mathbf{W}$ and $\tilde{\boldsymbol{h}}$

for $t = k+1 : l$;

initialize $\mathbf{W}_t = \mathbf{W}_{t-1}$, $\tilde{\boldsymbol{h}}^{(t)} = \mathbf{W}_t^{-1} \tilde{\boldsymbol{x}}^{(t)}$;

project any negative values in $\tilde{\boldsymbol{h}}^{(t)}$ to be zero;

alternately update $\tilde{\boldsymbol{h}}^{(t)}$ by Eq.(15) and $\mathbf{W}_t$ by Eq.(20);

end;

Simulation Results

In this section, the proposed INMF-KL algorithm have been extensively tested for many different benchmarks for dependent signals and images with various statistical distributions. Comparisions are made between it and the EASI algorithm [19], a famous online BSS algorithm based on the dependence of sources.

The following performance index (PI) [20] and the signal-to-interference ratio (SIR) [21] are utilized to evalutate respectively the eatimation of the mixing matrix and the recovery of the latent sources.

(22) $$PI(dB) = -10\log_{10}\left\{\frac{1}{n}\left[\sum_{i=1}^{n}\left(\sum_{j=1}^{n}\frac{g_{ij}^{2}}{\max_i g_{ij}^{2}}-1\right)+\sum_{j=1}^{n}\left(\sum_{i=1}^{n}\frac{g_{ij}^{2}}{\max_j g_{ij}^{2}}-1\right)\right]\right\}$$

where $g_{ij} = [\hat{\mathbf{W}}^{-1}\mathbf{W}]_{ij}$, $\hat{\mathbf{W}}$ denotes the estimated mixing matrix.

(23) $$\mathrm{SIR}_j = 10\log_{10}\frac{E[\boldsymbol{h}_j]^2}{E[\boldsymbol{h}_j - \boldsymbol{h}_j']^2}$$

where $\boldsymbol{h}_j$ denotes the *j*-th source and $\boldsymbol{h}_j'$ denotes its estimation.

Simulation 1: dependent synthetic data

Four mutually statistically dependent nonnegative source signals shown in Fig.1(a) are selected from NMFLAB [22] ($l = 1000$), where its correlation coefficient matrix is

(24) $$Corr1 = \begin{bmatrix} 1.0000 & 0.1753 & 0.3475 & 0.4878 \\ 0.1753 & 1.0000 & 0.2098 & 0.2162 \\ 0.3475 & 0.2098 & 1.0000 & 0.5328 \\ 0.4878 & 0.2162 & 0.5328 & 1.0000 \end{bmatrix}$$

For INMF-KL, 30% of the observation samples are collected for the achievement of $\mathbf{W}_k$ and $\tilde{\mathbf{H}}_k$, i.e., $k = 0.3l$. And for EASI, the adaptation step $\lambda = 7\times10^{-3}$.

The mixtures and the corresponding recovered sources of a typical run are shown in Fig.1(b)-(d), where the mixing matrix is randomly generated as

(25) $$\mathbf{W} = \begin{bmatrix} 0.0786 & 0.2138 & 0.2801 & 0.1784 \\ 0.0959 & 0.5579 & 0.1845 & 0.2497 \\ 0.4773 & 0.1664 & 0.3060 & 0.2691 \\ 0.3483 & 0.0620 & 0.2294 & 0.3028 \end{bmatrix}$$

Simulation 2: high correlated face images

Four highly correlated face images of size 128×164 shown in Fig.2(a) are tested [23] ($l = 20992$), where its correlation coefficient matrix is

(26) $$Corr2 = \begin{bmatrix} 1.0000 & 0.9437 & 0.9120 & 0.9026 \\ 0.9437 & 1.0000 & 0.8908 & 0.9072 \\ 0.9120 & 0.8908 & 1.0000 & 0.9161 \\ 0.9026 & 0.9072 & 0.9161 & 1.0000 \end{bmatrix}$$

For INMF-KL, 30% of the observation samples are collected for the achievement of $\mathbf{W}_k$ and $\tilde{\mathbf{H}}_k$, i.e., $k = 0.3l$. And for EASI, the adaptation step $\lambda = 7\times10^{-9}$.

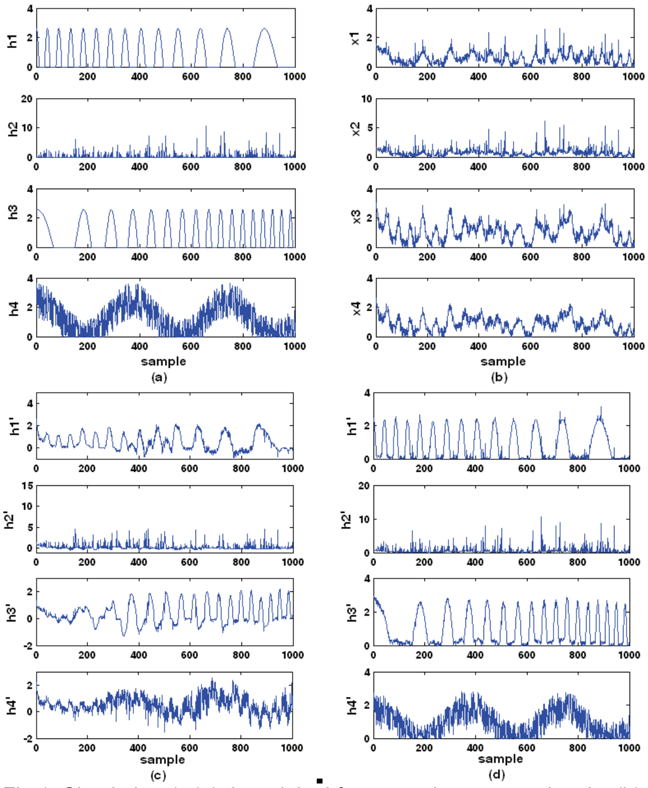

Fig.1. Simulation 1: (a) the original four negative source signals; (b) the four negative mixtures; (c) the four recovered sources using EASI; (d) the four recovered sources using INMF-KL.

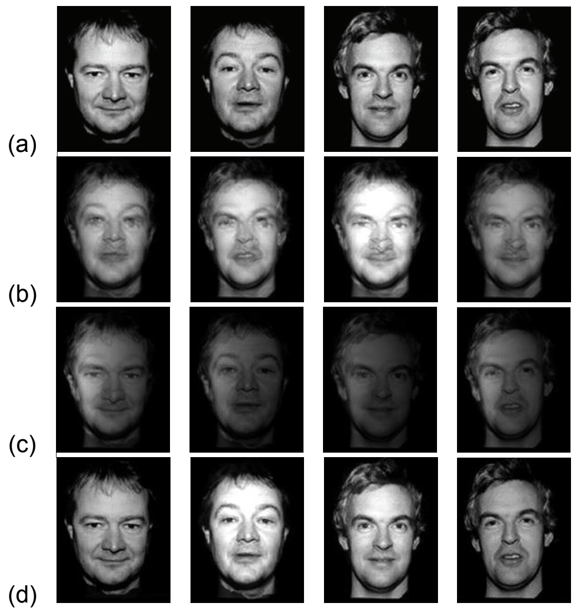

Fig.2. Simulation 2: (a) the original four face images; (b) the four mixtures; (c) the four estimations using EASI; (d) the four estimations using INMF-KL.

The mixtures and the corresponding recovered sources of a typical run are shown in Fig.2(b)-(d), where the mixing matrix is randomly generated as

$$\mathbf{W} = \begin{bmatrix} 0.1455 & 0.4214 & 0.2010 & 0.1412 \\ 0.2018 & 0.1853 & 0.2187 & 0.4351 \\ 0.4804 & 0.2830 & 0.2850 & 0.2641 \\ 0.1723 & 0.1103 & 0.2953 & 0.1597 \end{bmatrix} \quad (27)$$

The estimation performance including PI(dB) and SIR(dB) compared between EASI and the proposed INMF-KL algorithm for both Simulation 1 and Simulation 2 are presented in Table 2, where M-SIR denotes the mean of SIR_1, SIR_2, SIR_3, and SIR_4.

From the simulation results in Fig.1, Fig.2 and Table 2, it is easy to know that our INMF-KL algorithm have much better performance and the reconstructed signals are more similar to the original signals than the EASI algorithm. Therefore the INMF-KL has great potential to separate correlated sources. Furthermore, as the incremental update process yields an online mode, INMF-KL is suitable for online BSS. It is not surprsing that EASI behaves poorly to solve the dependent source separation (the estimated sources even have negative values) since it relies heavily on the indepence of sources.

Table 2. The estimation performance comparison between EASI and INMF-KL

	Simulation 1		Simulation 2	
Algorithms	EASI	INMF-KL	EASI	INMF-KL
PI (dB)	7.9256	15.0470	7.8978	15.7160
SIR_1(dB)	6.9858	18.1726	6.2509	15.9877
SIR_2(dB)	6.4901	11.5244	4.2538	10.2092
SIR_3(dB)	5.3494	17.2258	3.1421	20.6632
SIR_4(dB)	4.0345	11.7905	5.3834	12.8811
M-SIR(dB)	5.7150	14.6783	4.7576	14.9353

Conclusion

In this paper, an incremental NMF algorithm based on KL divergence is derived to trackle the online blind separation of statistically dependent sources, especially for data of large scale. Instead of performing the batch NMF repeatedly whenever a new sample arrives, the incremental scheme would ensure a steady updating process and reduce the computational cost as well as the storage requirement. The proposed INMF-KL algorithm can be regarded as an important complement to the INMF-VC algorithm [17] which is based on Euclidean distance. Our simulation experiments on synthetic signals and real-wold images support the outstanding effiency of INMF-KL. Future research includes the initialization parameter setup and the covergence analysis.

Acknowledgment

The author would like to thank Doctor Guoxu Zhou and Doctor Zuyuan Yang for their valuable advice and suggestion during the preparation of the manuscript.
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'.

REFERENCES

[1] T. Routtenberg and J. Tabrikian, Blind MIMO-AR system identification and source separation with finite-alphabet, IEEE Trans. Signal Process., vol.58, no.3, pp.990-1000, 2010.

[2] V. V. Nikulin, G. Nolte and G. Curio, A novel method for reliable and fast extraction of neuronal EGG/MEG oscillations on the basis of spatio-spectral decomposition, NeuroImage, vol. 55, pp.1528-1535, 2011.

[3] F. Nesta, T. S. Wada and B.-H. Juang, Batch-online semi-blind source separation applied to multi-channel acoustic echo cancellation, IEEE Trans. Audio, Speech and Lang. Process., vol. 19, no.3, pp. 583-599, 2011.

[4] A. Plaza, et al, Recent advances in techniques for hyperspectral image processing, Remote Sens. Envion., vol.113, pp.S110-S122, 2009.

[5] J. Wang and C.-I. Chang, Applications of independent component analysis in endmember extraction and abundance quantification for hyperspectral imagery, IEEE Trans. Geosci. Remote Sens., vol.44, no.9, pp.2601-2616, 2006.

[6] C. F. Caiafa and A. N. Proto, Separation of statistically dependent sources using an L^2-distance non-Gaussianity measure, Signal Process., vol.86, pp.3404-3420, 2006.

[7] I. Kopriva and D. Sersic, Wavelet packets approach to blind separation of statistically dependent sources, Neurocomputing, vol.71, pp.1642-1655, 2008.

[8] Y. Xiang, S. K. Ng and V. K. Nguyen, Blind separation of mutually correlated sources using precoders, IEEE Trans. Neural Networks, vol.21, no.1, pp.82-90, 2010.

[9] D. D. Lee, H. S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature, vol.401, pp.788-791, 1999.

[10] D. D. Lee, H. S. Seung, Algorithms for non-negative matrix factorization, in: Advances in Neural Information Processing Systems, vol.13, pp.556-562, 2001.

[11] F. Y. Wang, C. Y. Chi, T. H. Chan and Y. Wang, Nonnegative least-correlated component analysis for separation of dependent sources by volume maximization, IEEE Trans. Pattern Anal. Mach. Intell., vol.32, pp.875-888, 2010.

[12] A. Cichocki, R. Zdunek and S. Amari, New algorithm for non-negative matrix factorization in applications to blind source separation, in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process., vol.5, pp.621-624,2006.

[13] H. Kim and H. P. Denning, Sparse nonnegative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis, Bioinformatics, vol.23, no.12, pp.1495-1502, 2007.

[14] R. C. Henry, Multivariate receptor models-current practice and future trends, Chemometrics and Intelligent Laboratory Systems, 60(1-2), pp.43-48, 2002.

[15] S. Abdallah, M. Plumbley, Polyphonic transcription by non-negative sparse coding of power spectra, in: Proc. of Int. Conf. Music Information Retrieval, pp.318-325, 2004.

[16] S. S. Bucak and B. Gunsel, Incremental subspace learning via non-negative matrix factorization, Pattern Reconit., vol.42, no.5, pp.788-797,2009.

[17] G.-X. Zhou, Z.-Y. Yang, S. Xie and J.-M. Yang, Online blind source separation using incremental nonnegative matrix factorization with volume constraint, IEEE Trans. Neural Netw., vol.22, no.4, pp.550-560, 2011.

[18] T. Virtanen, Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria, IEEE Trans. Audio, Speech and Lang. Process., vol.15, no.3, pp.1066-1074, 2007.

[19] J.-F. Cardoso and B.-H. Laheld, Equivariant adaptive source separation, IEEE Trans. Signal Process., vol.44, no.12, pp.3017-3030, 1996.

[20] F. Yin, T.Mei and J. Wang, Blind-source separation based on decorrelation and nonstationarity, IEEE Trans. Circuits Syst. I, vol.54, no.5, pp.1150-1158, 2007.

[21] A. Cichocki and S. Amari, Adaptive blind signal and image processing: learning algorithms and applications. New York: Wiley, 2003.

[22] A. Cichocki and R. Zdunek, The NMFLAB package: for signal processing, version 1.1, RIKEN Brain Science Institute, Wako shi, Saitnama, Japan, 2006.

[23] A. Cichocki, S. Amari, K. Siwek, T. Tanaka, and A.H. Phan, The ICALAB package: for image processing, version 1.2, RIKEN Brain Science Institute, Wako shi, Saitama, Japan, 2006.

***Authors**: Dr. Hui Li, Wireless Communication Faculty, Institute of Communications Engineering, E-mail: leehoo86@163.com; Prof. Yue-hong Shen, Wireless Communication Faculty, Institute of Communications Engineering, E-mail: chunfeng22259@126.com; Dr. Jian-gong Wang, Wireless Communication Faculty, Institute of Communications Engineering, E-mail: hw12xian@hotmail.com.*