ON CORRELATION MATCHING APPROACH TO BLIND SEPARATION OF NONSTATIONARY SOURCES

Seungjin CHOI, Heonseok HONG, Junwhan OH

Department of Electrical Engineering, Chungbuk National University, KOREA
{schoi,hshong,jh}@lisa.chungbuk.ac.kr

ABSTRACT

This paper addresses a new method of blind source separation (BSS) when sources are nonstationary signals. Our method requires only multiple correlation matrices of the observed data at several time-windowed data frames to estimate the mixing matrix. In contrast to most existing BSS methods where higher-order statistics is necessary, our method is based on only second-order statistics. In the framework of correlation matching, we develop a new BSS algorithm. The useful behavior of the proposed method is verified by numerical experiments.

1. INTRODUCTION

Blind source separation (BSS) is a fundamental problem that is encountered in many practical applications such as telecommunications, image processing, feature extraction, pattern recognition, and biomedical signal analysis, etc. The task of BSS is to recover source signals from their linear instantaneous mixtures without resorting to any prior knowledge except for statistical independence of sources. One popular approach to BSS might be independent component analysis (ICA) which aims at decomposing the multivariate observations into a linear sum of statistically independent components. A variety of methods have been developed for BSS via ICA. Recent review articles can be found in [3, 7].

In most existing methods stationary sources are considered, hence, higher-order statistics is necessary either explicitly or implicitly. In typical ICA algorithms, the selection of nonlinear activation functions plays an important role since the optimal nonlinear functions depend on the probability distributions of sources which are unknown *a priori* [12, 10].

Many natural signals are nonstationary (in the sense that their variances are slowly time varying). A typical example may be speech signal. In contrast to the case of stationary sources, it is possible to perform BSS using only second-order statistics. Nonstationary sources in the task of BSS, were first considered by Matsuoka *et al.* [14]. It was further elaborated by Choi [8]. The key idea for BSS when sources are nonstationary lies in the fact that the correlation matrices are time-varying, so they carry sufficient statistics for the identification of the mixing matrix. In this paper we formulate the BSS task problem as a correlation matching problem and develop efficient iterative algorithms.

2. BLIND SOURCE SEPARATION

Here we describe the data model and explain what the BSS is. The conventional BSS or ICA algorithms were briefly reviewed in the framework of laten variable model with probability density matching method.

2.1. Data Model

Let us assume that the n dimensional vector of sensor signals, $\boldsymbol{x}(t) = [x_1(t), \ldots, x_n(t)]^T$ is generated by an unknown linear generative model,

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

where $\boldsymbol{s}(t) = [s_1(t), \ldots, s_n(t)]^T$ is the n dimensional vector whose elements are called sources. The matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is called a mixing matrix that is assumed to be of full rank. It is assumed that source signals $\{s_i(t)\}$ are statistically independent.

The task of BSS is to recover source vector $\boldsymbol{s}(t)$ from the observation vector $\boldsymbol{x}(t)$ without the knowledge of $\boldsymbol{A}$ nor $\boldsymbol{s}(t)$. In other words, BSS aims at finding a linear mapping (recognition model) which transforms sensor signals $\{x_i(t)\}$ to the output signals $\{y_i(t)\}$ such that the signals $\{y_i(t)\}$ are possibly scaled estimates of sources $\{s_i(t)\}$.

2.2. Probability Density Matching

Most methods of BSS have focused on statistically independent stationary non-Gaussian sources, so higher-order statistics was necessary. In such a case, BSS can be formulated as an ICA problem [13]. BSS or ICA can be illustrated as a probability density matching problem [4] which is identical to nonlinear infomax [5], maximum likelihood estimation [6], mutual information minimization [3].

Let us denote the observed density and model density by $p^o(\boldsymbol{x})$ and $p(\boldsymbol{x})$, respectively. The model density $p(\boldsymbol{x})$ satisfies the following relation.

$$\log p(\boldsymbol{x}) = -\log|\det \boldsymbol{A}| + \sum_{i=1}^{n} \log p_i(s_i). \tag{2}$$

As an optimization function to find $\boldsymbol{A}$ which best match $p^o(\boldsymbol{x})$ and $p(\boldsymbol{x})$, the Kullback-Leibler divergence is considered [4]. This gives the risk R that has the form

$$\begin{aligned} R &= KL\left[p^o(\boldsymbol{x})||p(\boldsymbol{x})\right] \\ &= \int p^o(\boldsymbol{x}) \log \frac{p^o(\boldsymbol{x})}{p(\boldsymbol{x})} d\boldsymbol{x}. \end{aligned} \tag{3}$$

Invoking the relation in (2), the loss function L is

$$L = \log|\det \boldsymbol{A}| - \sum_{i=1}^{n} \log p_i(s_i), \tag{4}$$

where $\log p^o(\boldsymbol{x})$ was neglected since it does not depend on $\boldsymbol{A}$. The loss function (4) can also be obtained from nonlinear infomax, maximum likelihood estimation, and mutual information minimization. Popular ICA algorithms were derived from the minimization of the loss function (4) using the natural gradient [1]. The adaptation algorithm for the mixing matrix $\boldsymbol{A}$ (see [1, 11] for more details) has the form

$$\Delta \boldsymbol{A} = -\eta \boldsymbol{A} \left\{ \boldsymbol{I} - \varphi(\hat{\boldsymbol{s}})\hat{\boldsymbol{s}}^T \right\}, \tag{5}$$

where $\eta > 0$ is a learning rate and $\hat{\boldsymbol{s}} = \boldsymbol{A}^{-1}\boldsymbol{x}$. The element-wise nonlinear function $\varphi(\cdot)$ is identical to the negative score function in ML estimation.

In the conventional ICA algorithms, one important thing lies in how one selects the nonlinear function $\varphi(\cdot)$ whose optimal form depends on the probability distribution of source which is unknown in advance. It is necessary to employ the hypothesized density in a smart way [10].

3. CORRELATION MATCHING APPROACH

For nonstationary sources, their variances are slowly time varying. Thus only multiple correlation matrices instead of probability density function allows us to perform the BSS task. In this section we describe two different algorithms.

3.1. Mixing Matrix Estimation

Let us denote by $\boldsymbol{R}_x^\circ(k)$ the correlation matrix of observation vector $\boldsymbol{x}(t)$ calculated using the samples in the kth time-windowed data frame. In the same manner we define the model correlation matrix by $\boldsymbol{R}_x(k) = \boldsymbol{A}\boldsymbol{R}_s(k)\boldsymbol{A}^T$. Note that the correlation matrix of source vector, $\boldsymbol{R}_s(k)$ is a diagonal matrix for all $k = 1, \dots, K$ where K is the number of frames.

Then we define the error between the correlation matrix of observed signals and the model by

$$\boldsymbol{E}(k) = \boldsymbol{R}_x^\circ(k) - \boldsymbol{R}_x(k), \tag{6}$$

for $k = 1, \dots, K$.

When $\boldsymbol{E}(k) = 0$ (i.e. correlation matching is achieved) for $k = i, j$ $(i \neq j)$, we have

$$\boldsymbol{R}_x^\circ(i) = \boldsymbol{A}\boldsymbol{R}_s(i)\boldsymbol{A}^T, \tag{7}$$

$$\boldsymbol{R}_x^\circ(j) = \boldsymbol{A}\boldsymbol{R}_s(j)\boldsymbol{A}^T. \tag{8}$$

There exists a closed-form solution for $\boldsymbol{A}$ which satisfies (7) and (8). In such a case, the mixing matrix $\boldsymbol{A}$ can be estimated by solving the generalized eigenvalue problem

$$\boldsymbol{R}_x^\circ(i)\left[\boldsymbol{R}_x^\circ(j)\right]^{-1}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{R}_s(i)\left[\boldsymbol{R}_s(j)\right]^{-1}. \tag{9}$$

The $\boldsymbol{A}$ that satisfies (9) is a solution to the task of BSS, if all the diagonal elements of $\boldsymbol{R}_s(i)$ and $\boldsymbol{R}_s(j)$ are distinct [9]. In practice, however, it is not clear which i and j guarantee the condition that $\boldsymbol{R}_s(i)$ and $\boldsymbol{R}_s(j)$ have distinctive diagonal elements.

In order to overcome this drawback, we consider multiple data frames, i.e., $K \geq 2$. The cost function that we consider here is

$$\mathcal{J} = \sum_{k=1}^{K} \operatorname{tr}\left\{\boldsymbol{E}(k)\boldsymbol{E}^T(k)\right\}, \tag{10}$$

where $\operatorname{tr}\{\cdot\}$ represents the trace operator. In order to avoid degenerate solutions, the optimization of the cost function (10) should be carried out under some constraints. One simple constraint is to restrict all the diagonal elements of the estimate of $\boldsymbol{A}$ to be unity [15].

The LS estimate of the mixing matrix is obtained by minimizing the cost function (10). In order to find the minima of the cost function (10), we compute the gradients with respect to the corresponding parameters which are given by

$$\frac{\partial \mathcal{J}}{\partial \boldsymbol{A}} = -4\sum_{k=1}^{K} \boldsymbol{E}(k)\boldsymbol{A}\boldsymbol{R}_s(k), \tag{11}$$

$$\frac{\partial \mathcal{J}}{\partial \boldsymbol{R}_s(k)} = -2\operatorname{diag}\left\{\boldsymbol{A}^T\boldsymbol{E}(k)\boldsymbol{A}\right\}. \tag{12}$$

The LS estimate of the mixing matrix $\boldsymbol{A}$ and source correlation matrix $\boldsymbol{R}_s(k)$ are computed iteratively by gradient descent method.

We can avoid the constraint that $[\boldsymbol{A}]_{ii} = 1$ for $i = 1, \dots, n$ by pre-whiten the observation data. Using the samples at whole frames we compute the sample correlation matrix $\boldsymbol{R}_x^\circ = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^T$ where $\boldsymbol{U}$ and $\boldsymbol{\Lambda}$ are eigenvector and eigenvalue matrices. The whitening transformation matrix $\boldsymbol{Q}$ is $\boldsymbol{Q} = \boldsymbol{\Lambda}^{-\frac{1}{2}}\boldsymbol{U}^T$. For the sake of simplicity we assume that the observation data is already whitened by a transformation $\boldsymbol{Q}$. In such a case the problem of BSS is to find a orthogonal mixing matrix. This can be done using the method of gradient in orthogonality [2]. **Algorithm Outline**

(1) Here we assume that the observation data $\boldsymbol{x}$ is already pre-whitened. Thus the mixing matrix $\boldsymbol{A}$ is an orthogonal matrix.

(2) The $\boldsymbol{A}$ is adapted by the gradient descent method in orthogonality constraint

$$\Delta \boldsymbol{A} = -\eta \left\{ \frac{\partial \mathcal{J}}{\partial \boldsymbol{A}} - \boldsymbol{A} \left(\frac{\partial \mathcal{J}}{\partial \boldsymbol{A}} \right)^T \boldsymbol{A} \right\}. \tag{13}$$

(3) The model source correlation matrix is updated by the conventional gradient method that has the form

$$\Delta \boldsymbol{R}_s(k) = \eta \mathrm{diag} \left\{ \boldsymbol{A}^T \boldsymbol{E}(k) \boldsymbol{A} \right\}, \tag{14}$$

for $k = 1, \dots, n$.

3.2. Demixing Matrix Estimation

Now we consider a demixing model that is described by

$$\boldsymbol{y}(t) = \boldsymbol{W}\boldsymbol{x}(t), \tag{15}$$

where $\boldsymbol{W}$ is the demixing matrix. Then we have

$$\boldsymbol{R}_y(k) = \boldsymbol{W}\boldsymbol{R}_x(k)\boldsymbol{W}^T. \tag{16}$$

We define the error between the correlation matrices of the estimated source vector $\boldsymbol{y}$ and model source vector $\boldsymbol{s}$,

$$\boldsymbol{E}(k) = \boldsymbol{W}\boldsymbol{R}_x(k)\boldsymbol{W}^T - \boldsymbol{R}_s(k). \tag{17}$$

Then the correlation matching principle leads to the following optimization function

$$\mathcal{J} = \sum_{k=1}^{K} \mathrm{tr} \left\{ \boldsymbol{E}(k) \boldsymbol{E}(k)^T \right\}. \tag{18}$$

In fact the correlation matching method seeks for $\boldsymbol{W}$ that jointly diagonalizes $\boldsymbol{R}_x(k)$ for K different frames The gradients are

$$\frac{\partial \mathcal{J}}{\partial \boldsymbol{W}} = 4 \sum_{k=1}^{K} \boldsymbol{E}(k) \boldsymbol{W} \boldsymbol{R}_x(k), \tag{19}$$

$$\frac{\partial \mathcal{J}}{\partial \boldsymbol{R}_s(k)} = -2\mathrm{diag} \left\{ \boldsymbol{E}(k) \right\}. \tag{20}$$

We can find the LS estimate of the demixing matrix $\boldsymbol{W}$ using the same method as the one described in Section 3.1.

4. NUMERICAL EXAMPLE

A simple numerical example is presented. For sources, we have used two digitized speech signals sampled at 8 kHz (see Figure 4). Two mixture signals (see Figure 4) were generated using the mixing matrix $\boldsymbol{A}$ given by

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0.9 \\ 0.7 & 1 \end{bmatrix}. \tag{21}$$

The methods described in previous section were applied to estimate the mixing matrix and recover two original speech signals. The observation data was partitioned into 10 nonoverlapping frames. The learning rate $\eta = 0.001$ was used. The recovered speech signals were shown in Figure 4. In contrast to most methods of ICA, here we used only multiple correlation matrices to estimate the mixing matrix and were able to successfully recover the source signals without knowing the mixing matrix nor sources.

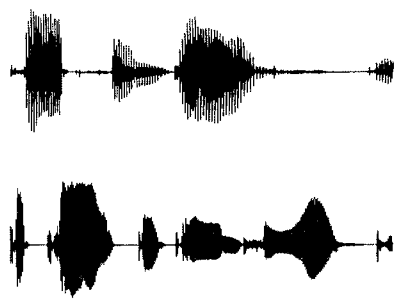

Figure 1: Original speech signals.

5. CONCLUSIONS

We have presented efficient iterative algorithms for blind separation of nonstationary sources. Our method requires only multiple correlation matrices in contrast to most existing BSS algorithms. In the framework of correlation matching we described how the mixing matrix

Figure 2: Mixture signals.

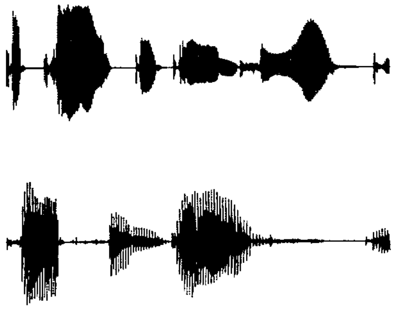

Figure 3: Recovered speech signals

or the demixing matrix could be estimated. Iterative algorithms were developed using the gradient descent method in orthogonality constraint and their performance were confirmed by numerical experiments.

Although we considered only noiseless mixtures, our method can be extended to the case of noisy mixtures. Currently we are working on noisy BSS problem in the framework of correlation matching.

6. ACKNOWLEDGMENT

This research was supported by Brain Science and Engineering Research Program sponsored by Korean Ministry of Science and Technology.

7. REFERENCES

[1] S. Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10(2):251–276, 1998.

[2] S. Amari. Natural gradient for over- and undercomplete bases in ICA. *Neural Computation*, 11(8):1875–1883, 1999.

[3] S. Amari and A. Cichocki. Adaptive blind signal processing - neural network approaches. *Proc. of the IEEE, Special Issue on Blind Identification and Estimation*, 86(10):2026–2048, Oct. 1998.

[4] H. Attias and C. E. Schreiner. Blind source separation and deconvolution: The dynamic component analysis algorithms. *Neural Computation*, 10:1373–1424, 1998.

[5] A. Bell and T. Sejnowski. An information maximisation approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.

[6] J. -F. Cardoso. Infomax and maximum likelihood for source separation. *IEEE Signal Processing Letters*, 4(4):112–114, Apr. 1997.

[7] J. -F. Cardoso. Blind signal separation: Statistical principles. *Proc. of the IEEE, Special Issue on Blind Identification and Estimation*, 86(10):2009–2025, Oct. 1998.

[8] S. Choi. Nonstationary source separation. *Journal of Electrical Engineering and Information Science*, 4(4):499–506, Aug. 1999.

[9] S. Choi and A. Cichocki. Blind separation of nonstationary sources in noisy mixtures. *Electronics Letters*, 36(9):848–849, Apr. 2000.

[10] S. Choi, A. Cichocki, and S. Amari. Flexible independent component analysis. In T. Constantinides, S. Y. Kung, M. Niranjan, and E. Wilson, editors, *Neural Networks for Signal Processing VIII*, pages 83–92, 1998.

[11] S. Choi and O. Lee. Sparse independent component analysis: Efficient coding of natural scenes? *Journal of Electrical Engineering and Information Science*, 5(2):84–92, Apr. 2000.

[12] A. Cichocki, I. Sabala, S. Choi, B. Orsier, and R. Szupiluk. Self-adaptive independent component analysis for sub-Gaussian and super-Gaussian mixtures with unknown number of source signals. In *Proc. Int. Symp. Nonlinear Theory and Applications*, pages 731–734, 1997.

[13] P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, 1994.

[14] K. Matsuoka, M. Ohya, and M. Kawamoto. A neural net for blind separation of nonstationary signals. *Neural Networks*, 8(3):411–419, 1995.

[15] L. C. Parra and C. Spence. Convolutive blind source separation of non-stationary sources. *IEEE Trans. Speech and Audio Processing*, to appear.