Orthogonal Least Square Approach to Nonstationary Source Separation

최희열[0], 최승진
포항공과대학교 컴퓨터공학과

Heeyoul Choi[0] and Seungjin Choi
Dept. Computer Science and Engineering, POSTECH
(hychoi[0], seungjin)@postech.ac.kr

Abstract: Blind source separation (BSS) is a fundamental problem that is encountered in many practical applications. In most existing methods, stationary sources are considered higher-order statistics is necessary either explicitly or implicitly. But, many natural signals are nonstationary, and it is possible to perform BSS using only second-order statistics. Our method is based on only second order statistics. The algorithms are developed using the gradient descent method in orthogonality constraint and their performance is confirmed by numerical experiments.

1. Introduction

Blind source separation (BSS) problem has been a challenging field in many practical applications. To solve this problem it is required to recover the original information-bearing signal when transmitted through a corrupted environment without resorting to any prior knowledge except for statistical independence of sources. One popular approach to BSS might be independent component analysis (ICA) which decompose the multivariate observations into a linear sum of statistically independent components. Most of source separation methods have focused on stationary sources, so higher-order statistics (HOS) is necessary for successful separation. But, many natural signals are inherently nonstationary stochastic processes. It was shown in [1] that source separation could be achieved by decorrelation if sources are independent second-order nonstationary stochastic processes.

In this paper we formulate the BSS task problem as a correlation matching problem and develop efficient iterative algorithms.

2. Blind Source Separation

In this section, we describe the data model and compare correlation matching method with probability density matching method.

2.1 Data Model

In the context of source separation, let us assume that the m dimensional vector of measurement signals, $x(t)$, is generated by a linear data model described by

$$x(t) = As(t), \tag{1}$$

where $s(t)$ is the n dimensional vector whose elements are called sources. The matrix $A \in R^{m \times n}$ is called a mixing matrix. The task of source separation is to estimate the mixing matrix A (or its inverse), given only a finite number of measurement signals, $\{x(t)\}$, $t = 1, \cdots, N$. Source vector $s(t)$ is assumed to be statistically independent.

2.2 Probability Density Matching

BSS or ICA can be illustrated as a probability density matching problem [2]. Let us denote the observed density and model density by $p^o(x)$ and $p(x)$, respectively. As an optimization function to find A which best match $p^o(x)$ and $p(x)$, the Kullback-Leibler divergence is considered [3]. This gives the risk R that has the form

$$R = KL\left[p^0(x) \| p(x)\right] = \int p^0(x) \log \frac{p^o(x)}{p(x)} dx \quad (2)$$

And, the loss function L is

$$L = \log\left|\det A\right| - \sum_{i=1}^{n} \log p_i(s_i) \quad (3)$$

where $\log p^0(x)$ was neglected since it does not depend on A. Popular ICA algorithms were derived from the minimization of the loss function (3) using the natural gradient [3]. The adaptation algorithm for the mixing matrix A (see [3] for more details) has the form

$$\Delta A = -\eta A\{I - \varphi(\hat{s})\hat{s}^T\} \quad (4)$$

where $\eta > 0$ is a learning rate and $\hat{s} = A^{-1}x$. In the conventional ICA algorithms, one important thing lies in how one selects the nonlinear function 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 [4].

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 $R_x^0(k)$ the correlation matrix of observation vector $x(t)$ calculated using the samples in the kth time-windowed data frame. In the same manner we define the model correlation matrix by $R_x(k) = AR_s(k)A^T$. Note that the correlation matrix of source vector, $R_s(k)$ is a diagonal matrix for all $k=1,\dots,K$ where K is the number of frames.

Then the error between the correlation matrix of observed signals and the model for all k is

$$E(k) = R_x^0(k) - R_x(k) \quad (5)$$

From $E(k) = 0$, we have $(i \neq j)$

$$R_x(i) = AR_s(i)A^T \quad (6)$$

$$R_x(j) = AR_s(j)A^T \quad (7)$$

There exists a closed-form solution for A which satisfies (6), (7). In such a case, the mixing matrix A can be estimated by solving the generalized eigenvalue problem. In practice, however, it is not clear which i and j guarantee the condition that $R_s(i)$ and $R_s(j)$ 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

$$J = \sum_{k=1}^{K} tr\left\{E(k)E^T(k)\right\} \quad (8)$$

In order to avoid degenerate solutions, the optimization of the cost function (8) should be carried out under some constraints. One simple constraint is to restrict all the diagonal elements of the estimate of A to be unity [5].

The LS estimate of the mixing matrix is obtained by minimizing the cost function (8). In order to find the minima of the cost function (8), we compute the gradients with respect to the corresponding parameters which are given by

$$\frac{\partial J}{\partial A} = -4\sum_{k=1}^{K} E(k)AR_s(k) \quad (9)$$

$$\frac{\partial J}{\partial R_s(k)} = -2diag\{A^T E(k) A\} \tag{10}$$

The LS estimate of the mixing matrix A and source correlation matrix $R_s(k)$ are computed iteratively by gradient descent method.

We can avoid the constraint that $[A]_{ii}=1$ for $i=1,\dots n$ by pre-whitening the observation data. Using the samples at whole frames we compute the sample correlation matrix $R_x^0 = UDU^T$ where U and D are eigenvector and eigenvalue matrices. The whitening transformation matrix Q is $Q = D^{-\frac{1}{2}}U^T$. For the sake of simplicity we assume that the observation data is already whitened by a transformation 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 [6].

Algorithm Outline

(1) We assume that the observation data x is already pre-whitened. Thus the mixing matrix A is an orthogonal matrix.

(2) The A is adapted by the gradient descent method in orthogonality constraint.

$$\Delta A = -\eta \left\{ \frac{\partial J}{\partial A} - A \left(\frac{\partial J}{\partial A} \right)^T A \right\} \tag{11}$$

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

$$\Delta R_s(k) = \eta diag\{A^T E(k) A\} \tag{12}$$

3.2 Demixing Matrix Estimation

Now we consider a demixing model that is described by $y(t) = Wx(t)$ where W is the demixing matrix. Then we have $R_y(k) = WR_x(k)W^T$. We define the error between the correlation matrices of the estimated source vector y and model source vector s,

$$E(k) = WR_x(k)W^T - R_s(k) \tag{13}$$

Then the correlation matching principle leads to the following optimization function

$$J = \sum_{k=1}^{K} tr\{E(k)E^T(k)\}. \tag{14}$$

In fact the correlation matching method seeks for W that jointly diagonalizes $R_x(k)$ for K different frames. The gradients are

$$\frac{\partial J}{\partial W} = 4\sum_{k=1}^{K} E(k)WR_x(k) \tag{15}$$

$$\frac{\partial J}{\partial R_s(k)} = -2diag\{E(k)\} \tag{16}$$

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

4. Numerical Example

We confirm our method by simple experiment. For sources, we have used two digitized speech signals sampled at 8 kHz (see Figure 1.). Two mixture signals (see Figure 2) were generated using the mixing matrix A given by $\begin{bmatrix} 1 & 0.9 \\ 0.7 & 1 \end{bmatrix}$.

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 8 nonoverlapping frames and each frame size is 1500. The learning rate was 0.1. First signals are orthogonalized, and recovered (see Figure 3). 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.

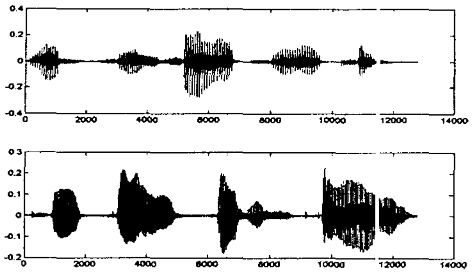

Figure 1 Original Speech Signals

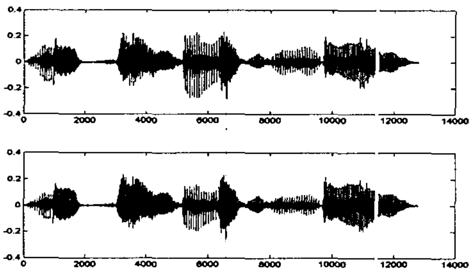

Figure 2 Observed Mixture signals

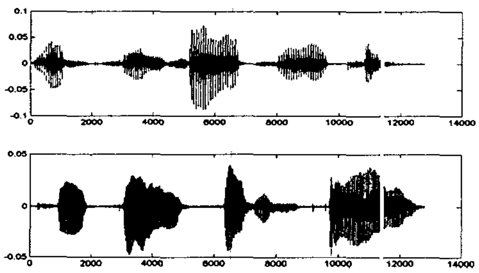

Figure 3 Recovered speech signals

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 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.

6. Acknowledgment

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

7. References

[1] K. Matsuoka, M. Ohya, and M. Kawamoto, "A neural net for blind separation of nonstationary signals," *Neural Networks*, vol. 8, no. 3, pp. 411-419, 1995.

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

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

[4] 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.

[5] L. C. Parra and C. Spence. Convolutive blind source separation of non-stationary sources. IEEE Trans. Speech and Audio Processing. pp. 320-327, May 2000.

[6] S. Amari. Natural gradient for over- and under complete bases in ICA. Neural Computation, 11(8):1875-1883, 1999.