Algebraic Differential Decorrelation for Nonstationary Source Separation

Seungjin Choi [§1], Andrzej Cichocki [†]

[§] Department of Computer Science and Eingeering
Pohang University of Science and Technology, Korea
seungjin@postech.ac.kr

[†] Laboratory for Advanced Brain Signal Processing
Brain Science Institute, RIKEN, Japan
cia@brain.riken.go.jp

December 17, 2001

Abstract

In this letter we introduce a *differential correlation* which is able to capture the time-varying statistics of nonstationary signals and show that the minimization of differential cross-correlations between observation signals can achieve the nonstationary source separation. As an implementation, we employ an algebraic method, the joint approximate diagonalization, so the resulting method is referred to as *Algebraic Differential DEcorrelation* (ADDE). The useful behavior of the method is confirmed by computer simulations.

Indexing terms: Blind source separation, differential correlation, joint approximate diagonalization, nonstationarity.

Appeared in
Electronics Letters, vol. 37, no. 23, pp. 1414-1415, Nov. 18, 2001

[1]Please address correspondence to Prof. Seungjin Choi, Department of Computer Science and Engineering, Pohang University of Science and Technology, San 31 Hyoja-dong, Nam-gu, Pohang, Kyungbuk 790-784, Korea, Tel: +82-54-279-2259, Fax: +82-54-279-2299, Email: seungjin@postech.ac.kr

1 Introduction

Source separation is a fundamental problem that is encountered in many practical applications such as telecommunications, image/speech processing, and biomedical signal analysis where multiple sensors are involved. In its simplest form, the m-dimensional observation vector $\boldsymbol{x}(t) \in \mathbb{R}^m$ is assumed to be generated by

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

where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is the unknown mixing matrix, $\boldsymbol{s}(t)$ is the n-dimensional source vector (which is also unknown and $n \leq m$).

In this letter, we consider the case where sources are nonstationary (especially second-order nonstationary in the sense that their variances are time-varying) [4]. In such a case, it might be useful to exploit the information on how fast the correlations between signals are changing. To this end, we introduce a concept of *differential correlation* and present an Algebraic Differential DEcorrelation (ADDE) method for nonstationary source separation. For the sake of simplicity we assume that sources have zero mean, but this is not necessary for our algorithm.

2 Differential Correlation

A differential statistic is defined by the derivative of statistic with respect to time (or its discrete-time counterpart is defined by the difference between statistics) and is closely related to the differential learning [3, 6]. We define the time-delayed correlation matrix of the observation vector $\boldsymbol{x}(t)$ by

$$\boldsymbol{R}_x(t_k, \tau) = E\left\{\boldsymbol{x}(t_k)\boldsymbol{x}^T(t_k - \tau)\right\}. \tag{2}$$

In practice, the sample correlation matrix $\widehat{\boldsymbol{R}}_x(t_k, \tau)$ is computed using the samples in the kth time-windowed data frame. Here we use the notation $\boldsymbol{R}_x(t_k, \tau)$ for both ensemble correlation and sample correlation.

The differential correlation matrix is defined by

$$\delta \boldsymbol{R}_x(t, \tau) = \frac{\partial \boldsymbol{R}_x(t, \tau)}{\partial t}. \tag{3}$$

Or its discrete-time counterpart is defined by

$$\delta \boldsymbol{R}_x(t_k, t_l, \tau) = \boldsymbol{R}_x(t_k, \tau) - \boldsymbol{R}_x(t_l, \tau). \tag{4}$$

The definition in (4) will be used hereafter.

One can easily see that the linear data model (1) has the following decomposition

$$\delta \boldsymbol{R}_x(t_k, t_l, \tau) = \boldsymbol{A} \delta \boldsymbol{R}_s(t_k, t_l, \tau) \boldsymbol{A}^T, \text{ for } t_k \neq t_l, \tag{5}$$

where $\delta \boldsymbol{R}_s(t_k, t_l, \tau)$ is the differential correlation matrix of source vector $\boldsymbol{s}(t)$ that is assumed to be a diagonal matrix. It follows from (5) that the mixing matrix $\boldsymbol{A}$ can be estimated by solving a generalized eigenvalue problem

$$\delta \boldsymbol{R}_x(t_2, t_3, 0) \boldsymbol{U} = \delta \boldsymbol{R}_x(t_1, t_2, 0) \boldsymbol{U} \boldsymbol{\Lambda}, \tag{6}$$

where $\boldsymbol{U}$ and $\boldsymbol{\Lambda}$ correspond to the eigenvector and eigenvalue matrices, respectively. In such a case, the mixing matrix is given by $\boldsymbol{A} = \boldsymbol{U}^{-T}$. However, this is a valid solution only if all the diagonal elements of $\boldsymbol{\Lambda}$ are distinct. In order to avoid this difficulty, we exploit multiple differential correlation matrices and apply the joint approximate diagonalization method to estimate the mixing matrix. It is described in next section.

3 The Algorithm: ADDE

First we whiten the data $\boldsymbol{x}(t)$ to obtain $\boldsymbol{z}(t) = \boldsymbol{Q}\boldsymbol{x}(t)$ where $\boldsymbol{Q}$ is a whitening transformation. Then we have

$$\boldsymbol{z}(t) = \boldsymbol{Q}\boldsymbol{A}\boldsymbol{s}(t) = \boldsymbol{B}\boldsymbol{s}(t), \tag{7}$$

where $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ is an orthogonal matrix, i.e., $\boldsymbol{B}\boldsymbol{B}^T = \boldsymbol{I}$. Then, the whitened vector $\boldsymbol{z}(t)$ satisfies

$$\delta \boldsymbol{R}_z(t_i, t_{i+1}, \tau_j) = \boldsymbol{B} \delta \boldsymbol{R}_s(t_i, t_{i+1}, \tau_j) \boldsymbol{B}^T. \tag{8}$$

Note that it is not necessary to exploit two adjacent data frames to compute the differential correlation matrix, but here we show just one exemplary case.

We apply the joint approximate diagonalization to estimate the unitary mixing matrix $\boldsymbol{B}$, as in JADE [2], SOBI [1], and SEONS [5]. The algorithm ADDE is summarized below.

Algorithm Outline: ADDE

1. Pre-whiten the observation data using whole data points so that the whitened data $\boldsymbol{z}(t) = \boldsymbol{Q}\boldsymbol{x}(t)$ is a unitary mixture of sources (where $\boldsymbol{Q}$ is a whitening transformation).

2. Divide the data $\{\boldsymbol{z}(t)\}$ into K non-overlapping blocks and calculate $(K-1)J$ differential correlation matrices, $\delta\boldsymbol{R}_z(t_i, t_{i+1}, \tau_j)$ for $i = 1, \ldots, K-1$ and $j = 1, \ldots, J$ (for example, $\tau_j = j - 1$).

3. Find a unitary joint diagonalizer $\boldsymbol{V}$ of $\{\delta\boldsymbol{R}_z(t_i, t_{i+1}, \tau_j)\}$ which satisfies

$$\boldsymbol{V}^T \delta\boldsymbol{R}_z(t_i, t_{i+1}, \tau_j)\boldsymbol{V} = \boldsymbol{\Lambda}_{i,j}, \qquad (9)$$

where $\{\boldsymbol{\Lambda}_{i,j}\}$ is a set of diagonal matrices.

4 The demixing matrix is given by $\boldsymbol{W} = \boldsymbol{V}^T\boldsymbol{Q}$.

4 A Numerical Example

In this simulation, we used 1 digitized voice signal, 1 digitized music signal (both of which were sampled at 8 kHz), and 3 Guassian signals with no temporal correlations but their variances being time-varying. The mixture vector $\boldsymbol{x}(t)$ (with 10000 data points) was generated by the mixing matrix $\boldsymbol{A} \in \mathbb{R}^{5\times5}$, all the elements of which were drawn from standardized Gaussian distribution (i.e., zero mean and unit variance).

In order to evaluate the performance of algorithms, we calculated the performance index (PI) defined by

$$\text{PI} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \left\{ \left(\sum_{k=1}^{n} \frac{|g_{ik}|}{\max_j |g_{ij}|} - 1 \right) + \left(\sum_{k=1}^{n} \frac{|g_{ki}|}{\max_j |g_{ji}|} - 1 \right) \right\}, \qquad (10)$$

where g_{ij} is the (i,j)-element of the matrix $\boldsymbol{G} = \boldsymbol{W}\boldsymbol{A}$.

We evaluated the performance of three algorithms: JADE [2], SOBI [1], and the proposed algorithm ADDE. All three algorithms are based on the joint approximate diagonalization but exploit different statistics of the data. For SOBI, 20 different time-delayed correlation matrices of whitened data were exploited. For ADDE, we divided the whitened data into 100 different non-overlapping blocks (the length of each data frame is 100) and calculated 99 differential

Algorithm	Performance index
JADE	0.0146
SOBI	0.0394
ADDE	$\mathbf{8.4573 \times 10^{-4}}$

Table 1: Performance of JADE, SOBI, and ADDE.

correlation matrices by considering two adjacent data frames. Although JADE and SOBI are good algorithms, they had difficulty in the presence of nonstationary Gaussian sources with no temporal correlations, as expected. In such a case, the ADDE outperformed JADE and SOBI (see Table 1).

5 Discussion

In this letter, we introduced the differential correlation and showed that it was useful to capture the time-varying statistics for the case of second-order nonstationary source separation. We presented an algebraic differential decorrelation method, ADDE, based on the joint approximate diagonalization. The useful behavior of the method was confirmed by computer simulations. The ADDE, in principle, could be extended to the case of noisy mixtures. For example, in the presence of temporally white noise, the robust orthogonalization method (as in [5]) could be applied in our current framework.

6 Acknowledgment

This work was supported by Korea Ministry of Science and Technology under International Co-operative Research Program and Brain Science Engineering Research Program and by POSTECH BK 21.

References

[1] A. Belouchrani, K. Abed-Merain, J. -F. Cardoso, and E. Moulines. A blind source separation technique using second order statistics. *IEEE Trans. Signal Processing*, 45:434–444, Feb. 1997.

[2] J. -F. Cardoso and A. Souloumiac. Blind beamforming for non Gaussian signals. *IEE Proceedings-F*, 140(6):362–370, 1993.

[3] S. Choi. Differential Hebbian-type learning algorithms for decorrelation and independent component analysis. *Electronics Letters*, 34(9):900–901, 1998.

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

[5] S. Choi, A. Cichocki, and A. Belouchrani. Blind separation of second-order nonstationary and temporally colored sources. In *Proc. IEEE Workshop on Statistical Signal Processing*, pages 444–447, Singapore, 2001.

[6] Y. Deville and M. Benali. Differential source separation: Concept and application to a criterion based on differential normalized kurtosis. In *Proc. EUSIPCO*, 2000.