DIFFERENTIAL DECORRELATION FOR NONSTATIONARY SOURCE SEPARATION

Seungjin Choi [§], *Andrzej Cichocki* [†], *Yannick Deville* [‡]

[§] Department of Computer Science & Engineering, POSTECH, Korea
seungjin@postech.ac.kr
[†] Laboratory for Advanced Brain Signal Processing, Brain Science Institute, RIKEN, Japan
cia@brain.riken.go.jp
[‡] LAMI, Université Paul Sabatier, France
ydeville@cict.fr

ABSTRACT

In this paper we consider the problem of source separation for the case that sources are (second-order) nonstationary, especially their variances are slowly time varying. The differential correlation is exploited in order to capture the time-varying statistics of signals. We show that nonstationary source separation can be achieved by differential decorrelation. Algebraic methods are presented and discussed.

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

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

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 $n \leq m$), and $\boldsymbol{v}(t)$ is the additive noise vector that is statistically independent of $\boldsymbol{s}(t)$.

A variety of methods and algorithms for source separation have been developed for last decade (for example, see [1] and references therein). Most of source separation algorithms focused on stationary signals and exploited the non-Gaussianity [2] or the temporal structure of source [3]. However, many natural signals are inherently nonstationary, so the nonstationarity is one important property that can be used for the task of source separation. Earlier work on nonstationary source separation can be found in [4] and some recent developments are in [5, 6, 7, 8].

In this paper we focus on the problem of nonstationary source separation In order to capture the time-varying statistics of signals, we employ the concept of differential learning [9, 10] which will be described in Sec. 2. In our previous work [7, 11], we developed algebraic methods for nonstationary source separation which exploited multiple time-delayed correlation matrices of observation data evaluated in several different time-windowed data frame. The simultaneous diagonalization (or generalized eigenvalue problem) and the joint approximate diagonalization method [12] were used in the framework of nonstationary source separation. Here we adopt the differential correlation and generalize our previous work. We discuss the benefits and drawbacks of methods based on differential correlation.

Throughout this paper, the following assumptions are made:

(AS1) The mixing matrix $\boldsymbol{A}$ is of full column rank.

(AS2) Sources are spatially uncorrelated but are temporally correlated (colored) stochastic signals with zero mean.

(AS3) Sources are second-order nonstationary signals in the sense that their variances are time varying.

(AS4) Additive noises $\{v_i(t)\}$ are stationary stochastic processes.

2. DIFFERENTIAL CORRELATION

Since the statistics of nonstationary signals are time-varying, how fast or how slow their statistics are changing, might have important information. In order to capture the time-varying statistics of nonstationary signals we exploit the differential statistics which is defined by the derivative of statistics with respect to time (or its discrete-time counterpart is defined by the difference between statistics). Recent work on differential learning for source separation can be found in [9, 10, 13].

The time-delayed correlation matrix of the observation vector $\boldsymbol{x}(t)$ is defined 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 $\hat{\boldsymbol{R}}_x(t_k, \tau)$ 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

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

Here we adopt the definition in (4).

One can easily see that under the assumptions (AS1)-(AS4), we have the following decomposition

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

where $\delta \boldsymbol{R}_s(t_k, t_l, \tau)$ and $\delta \boldsymbol{R}_v(t_k, t_l, \tau)$ are the differential correlation matrices of source vector $\boldsymbol{s}(t)$ and noise vector $\boldsymbol{v}(t)$ defined in the same way as in (4), respectively.

Remarks:

- $\delta \boldsymbol{R}_s(t_k, t_l, \tau)$ is a diagonal matrix due to the assumptions (AS2) and (AS3).
- $\delta \boldsymbol{R}_v(t_k, t_l, \tau)$ disappears when: (1) the noise is a stationary stochastic process; (2) the noise is temporally white regardless of its spatial dependence. This observation enables us to find a method for robust estimation of the mixing matrix $\boldsymbol{A}$.
- When $\delta \boldsymbol{R}_v(t_k, t_l, \tau)$ disappears, the mixing matrix $\boldsymbol{A}$ or its pseudo-inverse (demixing matrix $\boldsymbol{W}$) can be estimated by diagonalizing a set of matrices, $\{\delta \boldsymbol{R}_x(t_k, t_l, \tau)\}$. Algebraic methods are discussed in Sections 3 and 4.

3. SYMMETRIC-DEFINITE PENCIL

Let us consider two different differential correlation matrices

$$\begin{aligned} \delta \boldsymbol{R}_1 &= \delta \boldsymbol{R}_x(t_1, t_2, \tau) = \boldsymbol{A}\boldsymbol{\Lambda}_1 \boldsymbol{A}^T, \\ \delta \boldsymbol{R}_2 &= \delta \boldsymbol{R}_x(t_3, t_4, \tau) = \boldsymbol{A}\boldsymbol{\Lambda}_2 \boldsymbol{A}^T, \end{aligned} \tag{6}$$

where

$$\begin{aligned} \boldsymbol{\Lambda}_1 &= \delta \boldsymbol{R}_s(t_1, t_2, \tau) = \text{diag}\{\gamma_1, \ldots, \gamma_n\}, \\ \boldsymbol{\Lambda}_2 &= \delta \boldsymbol{R}_s(t_3, t_4, \tau) = \text{diag}\{\beta_1, \ldots, \beta_n\}. \end{aligned} \tag{7}$$

The simultaneous diagonalization of two matrices $\delta \boldsymbol{R}_1$ and $\delta \boldsymbol{R}_2$ allows us to estimate the mixing matrix $\boldsymbol{A}$ or the demixing matrix $\boldsymbol{W} = \boldsymbol{A}^{-1}$, provided that $\left\{\lambda_i = \frac{\beta_i}{\gamma_i}\right\}$ are distinct. Typically the simultaneous diagonalization consists of steps: (1) whitening and (2) unitary transform. Alternatively the simultaneous diagonalization can be carried out by solving the generalized eigenvalue problem,

$$\delta \boldsymbol{R}_2 \boldsymbol{U} = \delta \boldsymbol{R}_1 \boldsymbol{U} \text{diag}\{\lambda_1, \ldots, \lambda_n\}. \tag{8}$$

Then the mixing matrix $\boldsymbol{A}$ corresponds to $\boldsymbol{U}^{-T}$, provided that $\left\{\lambda_i = \frac{\beta_i}{\gamma_i}\right\}$ are distinct. Note that (8) is identical to the problem

$$\delta \boldsymbol{R}_1^{-1} \delta \boldsymbol{R}_2 \boldsymbol{U} = \boldsymbol{U} \text{diag}\{\lambda_1, \ldots, \lambda_n\}. \tag{9}$$

Remarks:

- The method based on the generalized eigenvalue problem gives an closed-form solution.
- It is expected to the method here is less sensitive to additive noise, compared to many existing source separation methods. In the presence of spatially correlated but temporally white noise, the differential correlation with non-zero time-lag ($\tau \neq 0$) of noise vector becomes zero matrix. This is also true for regular time-delayed correlation matrix. Some related work can be found in [7, 11]. If the noise is spatially and temporally correlated, but is stationary, then the differential correlation of noise vector is zero matrix (at least theoretically).
- A numerical instability might happen because of two reasons: (1) $\delta \boldsymbol{R}_1$ and $\delta \boldsymbol{R}_2$ might not be symmetric; (2) $\delta \boldsymbol{R}_1$ and $\delta \boldsymbol{R}_2$ are not always positive definite even when $\tau = 0$.

The set of all matrices of the form $\delta \boldsymbol{R}_2 - \lambda \delta \boldsymbol{R}_1$ with $\lambda \in \mathbb{R}$ is said to be a *pencil*. Frequently we encounter into the case where $\boldsymbol{R}_2$ is symmetric and $\delta \boldsymbol{R}_1$ is symmetric and positive definite. Pencils of this variety are referred to as *symmetric-definite pencils* [14].

Theorem 1 (pp. 468 in [14]) *If $\delta \boldsymbol{R}_2 - \lambda \delta \boldsymbol{R}_1$ is symmetric-definite, then there exists a nonsingular matrix $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n]$ such that*

$$\boldsymbol{U}^T \delta \boldsymbol{R}_1 \boldsymbol{U} = \textit{diag}\{\gamma_1, \ldots, \gamma_n\}, \tag{10}$$

$$\boldsymbol{U}^T \delta \boldsymbol{R}_2 \boldsymbol{U} = \textit{diag}\{\beta_1, \ldots, \beta_n\}. \tag{11}$$

Moreover $\delta \boldsymbol{R}_2 \boldsymbol{u}_i = \lambda_i \delta \boldsymbol{R}_1 \boldsymbol{u}_i$ for $i = 1, \ldots, n$, and $\lambda_i = \frac{\beta_i}{\gamma_i}$.

For the requirement of symmetry, we replace $\delta \boldsymbol{R}_1$ and $\delta \boldsymbol{R}_2$ by $\delta \boldsymbol{M}_1$ and $\delta \boldsymbol{M}_2$ that are defined by

$$\delta \boldsymbol{M}_1 = \frac{1}{2}\left\{\delta \boldsymbol{R}_1 + \delta \boldsymbol{R}_1^T\right\}, \tag{12}$$

$$\delta \boldsymbol{M}_2 = \frac{1}{2}\left\{\delta \boldsymbol{R}_2 + \delta \boldsymbol{R}_2^T\right\}. \tag{13}$$

In general $\delta \boldsymbol{M}_1$ is not always positive definite, regardless of the value of τ. Thus we consider a linear combination

$$\delta \boldsymbol{C}_1 = \sum_{i=1}^{J} \alpha_i \delta \boldsymbol{M}_x(t_1, t_2, \tau_i) \tag{14}$$

for $J \geq 2$. The coefficients α_i are selected in such as way that $\delta \boldsymbol{C}_1$ becomes positive definite. For example, we can use the same method as employed in the finite-step global convergence (FSGC) algorithm [15], in order to find $\{\alpha_i\}$ such that the matrix $\delta \boldsymbol{C}_1$ is positive definite. Thus, the pencil $\delta \boldsymbol{M}_2 - \lambda \delta \boldsymbol{C}_1$ is symmetric-definite, so the generalized eigenvector matrix $\boldsymbol{U}$ that solves

$$\delta \boldsymbol{M}_2 \boldsymbol{U} = \delta \boldsymbol{C}_1 \boldsymbol{U} \text{diag}\{\lambda_1, \ldots, \lambda_n\}, \tag{15}$$

can be computed without any numerical instability problem. The method of choosing a set of coefficients, $\{\alpha_i\}$ such that the matrix $\delta \boldsymbol{C}_1$ is positive definite, is summarized below.

Algorithm Outline: Selection of $\{\alpha_i\}$

1. Estimate differential correlation matrices and construct an $m \times mJ$ matrix

$$\mathcal{M} = [\delta \boldsymbol{M}_x(t_1, t_2, \tau_1) \cdots \delta \boldsymbol{M}_x(t_1, t_2, \tau_J)]. \tag{16}$$

Then compute the singular value decomposition (SVD) of $\mathcal{M}$, i.e.,

$$\mathcal{M} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T, \tag{17}$$

where $\boldsymbol{U} \in \mathbb{R}^{m \times m}$ and $\boldsymbol{V} \in \mathbb{R}^{mJ \times mJ}$ are orthogonal matrices, and $\boldsymbol{\Sigma}$ has nonzero entries at (i, i) position ($i = 1, \ldots, n$) and zeros elsewhere. The number of sources, n can be detected by inspecting the singular values. Define $\boldsymbol{U}_s$ by

$$\boldsymbol{U}_s = [\boldsymbol{u}_1 \cdots \boldsymbol{u}_n], \tag{18}$$

where $\boldsymbol{u}_i$ is the ith column vector of the matrix $\boldsymbol{U}$ and $n \leq m$.

2. For $i = 1, \ldots, J$, compute

$$\boldsymbol{F}_i = \boldsymbol{U}_s^T \delta \boldsymbol{M}_x(t_1, t_2, \tau_i) \boldsymbol{U}_s. \tag{19}$$

3. Choose any initial $\boldsymbol{\alpha} = [\alpha_1 \cdots \alpha_J]^T$.
4. Compute

$$\boldsymbol{F} = \sum_{i=1}^{J} \alpha_i \boldsymbol{F}_i. \tag{20}$$

5. Compute a Schur decomposition of $\boldsymbol{F}$ and check if $\boldsymbol{F}$ is positive definite or not. If $\boldsymbol{F}$ is positive definite, the algorithm is terminated. Otherwise, go to Step 6.
6. Choose an eigenvector $\boldsymbol{u}$ corresponding to the smallest eigenvalue of $\boldsymbol{F}$ and update $\boldsymbol{\alpha}$ via replacing $\boldsymbol{\alpha}$ by $\boldsymbol{\alpha} + \boldsymbol{\delta}$ where

$$\boldsymbol{\delta} = \frac{\left[\boldsymbol{u}^T \boldsymbol{F}_1 \boldsymbol{u} \cdots \boldsymbol{u}^T \boldsymbol{F}_J \boldsymbol{u}\right]^T}{\|[\boldsymbol{u}^T \boldsymbol{F}_1 \boldsymbol{u} \cdots \boldsymbol{u}^T \boldsymbol{F}_J \boldsymbol{u}]\|}. \tag{21}$$

   Go to step 4. This loop is terminated in a finite number of steps (see [15] for proof).

Note: In the case of $m = n$ (equal number of sources and sensors), step 1 and 2 are not necessary. Simply we let $\boldsymbol{F}_i = \delta \boldsymbol{M}_x(t_1, t_2, \tau_i)$.

Remark: The symmetric-definite pencil method described above in fact exploits both the nonstationarity and temporal structure of signals. Instead of using a linear combination of $\delta \boldsymbol{M}_x(t_1, t_2, \tau_i)$, we can consider a combination of $\delta \boldsymbol{M}_x(t_k, t_l, 0)$ (for $k = 1, 3, \ldots$ and $l = 2, 4, \ldots$) for the case where signals are nonstationary without temporal correlations.

4. JOINT APPROXIMATE DIAGONALIZATION

Let us consider a unitary mixture, $\boldsymbol{z}(t)$ described by

$$\boldsymbol{z}(t) = \boldsymbol{B}\boldsymbol{s}(t) + \boldsymbol{v}(t), \tag{22}$$

where $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ is an orthogonal matrix, i.e., $\boldsymbol{B}\boldsymbol{B}^T = \boldsymbol{I}$. The unitary mixture vector $\boldsymbol{z}(t)$ can be obtained by a whitening transformation $\boldsymbol{Q}$, i.e., $\boldsymbol{z}(t) = \boldsymbol{Q}\boldsymbol{x}(t)$. In order to reduce the effect of white noise, the robust whitening [16, 11] can be employed. It follows from the assumptions (AS2) and (AS3) that we have

$$\delta \boldsymbol{R}_z(t_1, t_2, \tau_i) = \boldsymbol{B} \delta \boldsymbol{R}_s(t_1, t_2, \tau_i) \boldsymbol{B}^T \tag{23}$$

for $\tau_i \neq 0$ ($i = 1, \ldots, K$).

We apply the joint approximate diagonalization (JAD) to estimate the unitary mixing matrix $\boldsymbol{B}$, as in JADE [17], SOBI [3], and SEONS [11]. The joint approximate diagonalization method finds a unitary joint diagonalizer $\boldsymbol{V}$ of $\{\delta \boldsymbol{R}_z(t_1, t_2, \tau_i)\}$ which satisfies

$$\boldsymbol{V}^T \delta \boldsymbol{R}_z(t_1, t_2, \tau_i) \boldsymbol{V} = \boldsymbol{\Lambda}_i, \tag{24}$$

where $\{\boldsymbol{\Lambda}_i\}$ is a set of diagonal matrices. The unitary mixing matrix $\boldsymbol{B} = \boldsymbol{V}^{-T}$.

Algorithm Outline: Differential Decorrelation via JAD

1. Apply the robust orthogonalization [16, 11] to obtain the vector $\boldsymbol{z}(t)$.
2. Divide the data $\{\boldsymbol{z}(t)\}$ into two non-overlapping blocks and calculate $\delta \boldsymbol{R}_z(t_1, t_2, \tau_j)$ for $j = 1, \ldots, K$.
3. Find a unitary joint diagonalizer $\boldsymbol{V}$ of $\{\delta \boldsymbol{R}_z(t_1, t_2, \tau_j)\}$ which satisfies

$$\boldsymbol{V}^T \delta \boldsymbol{R}_z(t_1, t_2, \tau_j) \boldsymbol{V} = \boldsymbol{\Lambda}_j, \tag{25}$$

   where $\{\boldsymbol{\Lambda}_j\}$ is a set of diagonal matrices.
4. The unitary mixing matrix is given by $\boldsymbol{B} = \boldsymbol{V}^{-T}$.

Remark: Instead of exploiting a set of matrices, $\delta \boldsymbol{R}_z(t_1, t_2, \tau_j)$ for $j = 1, \ldots, K$, we can also employ a set of matrices $\delta \boldsymbol{R}_z(t_k, t_l, \tau_j)$. In other words, the data is divided into a bunch of small frames to calculate the differential correlations. At each data frame, we calculate several different time-delayed differential correlation matrices or single differential correlation matrix to find $\delta \boldsymbol{R}_z(t_k, t_l, \tau_j)$.

5. SIMULATIONS

In this simulation, we used 3 digitized voice signals and 2 music signals, all of which were sampled at 8 kHz. The mixture vector $\boldsymbol{x}(t)$ was generated by the mixing matrix $\boldsymbol{A} \in \mathbb{R}^{5 \times 5}$, all the elements of which were drawn from standardized Gaussian distribution (i.e., zero mean and unit variance). The length of whole data is 10000.

We evaluated the performance of our method (that is JAD-based differential decorrelation), JADE [17], and SOBI [3]. Note that all these three methods employ the same joint approximate diagonalization, although different statistics are exploited. In our proposed method, first the observation vector $\boldsymbol{x}(t)$ was transformed to $\boldsymbol{z}(t)$ by a linear transformation $\boldsymbol{Q}$ that is computed by the robust orthogonalization. The robust orthogonalization exploits a linear combination of multiple time-delayed correlation matrices of the observation vector [16, 11, 18]. Once the vector $\boldsymbol{z}(t)$ is computed, then we divide the whitened vector into two non-overlapping frames and compute $\delta \boldsymbol{R}_z(t_1, t_2, \tau_j)$ for $j = 1, \ldots, 20$. Then a unitary joint diagonalizer is computed.

In order to measure the performance of algorithms, we use the performance index (PI) defined by

$$\begin{aligned} \text{PI} \quad = \quad & \frac{1}{n(n-1)} \sum_{i=1}^{n} \left\{ \left(\sum_{k=1}^{n} \frac{|g_{ik}|}{\max_j |g_{ij}|} - 1 \right) \right. \\ & \left. + \left(\sum_{k=1}^{n} \frac{|g_{ki}|}{\max_j |g_{ji}|} - 1 \right) \right\}, \end{aligned} \tag{26}$$

where g_{ij} is the (i,j)-element of the global system matrix $\boldsymbol{G} = \boldsymbol{W}\boldsymbol{A}$ and $\max_j g_{ij}$ represents the maximum value among the elements in the ith row vector of $\boldsymbol{G}$, $\max_j g_{ji}$ does the maximum value among the elements in the ith column vector of $\boldsymbol{G}$. When the perfect separation is achieved, the performance index is zero. In practice, the value of performance index around 10^{-3} gives quite a good performance.

The result is shown in Fig. 1. In the range of low SNR, our method outperforms JADE and SOBI. In fact, this high gain results from exploiting the time-delayed differential correlation matrices with the robust orthogonalization. The method based on the generalized eigen-decomposition also works fine, but its performance is not comparable to the method based on the JAD.

6. DISCUSSION

In this paper, we introduced the differential decorrelation and showed how it could be used for the task of nonstationary source separation. This is our first step to investigate the differential correlation for nonstationary source separation.

We can also develop adaptive algorithms for differential decorrelation. For example, we can employ the iterative least squares

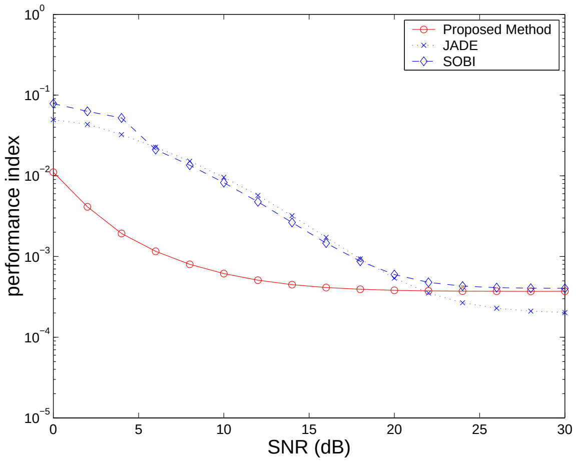

Figure 1: Performance comparison for the proposed method, JADE, and SOBI.

algorithm in [18] and replace the regular correlation by differential correlation. Many other efficient implementation will be also possible.

In simulations we only considered the case where additive noise is spatially correlated but temporally white. However, our method is potentially applicable to the case of any stationary noise. However, in practice, it is difficult to expect the differential correlation becomes zero when the sample correlation is used. The performance will depend on the number of data points that is are used to compute the sample correlation. Theoretical performance analysis will be our future work.

7. ACKNOWLEDGMENT

This work was supported by Korea Ministry of Science and Technology under Brain Science and Engineering Research Program and International Cooperative Research Program.

8. REFERENCES

[1] S. Haykin, *Unsupervised Adaptive Filtering: Blind Source Separation*. Prentice-Hall, 2000.

[2] S. Amari and A. Cichocki, "Adaptive blind signal processing - neural network approaches," *Proc. of the IEEE, Special Issue on Blind Identification and Estimation*, vol. 86, no. 10, pp. 2026–2048, Oct. 1998.

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

[4] 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.

[5] C. Chang, Z. Ding, S. F. Yau, and F. H. Y. Chan, "A matrix-pencil approach to blind separation of colored nonstationary signals," *IEEE Trans. Signal Processing*, vol. 48, pp. 900–907, Mar. 2000.

[6] S. Choi and A. Cichocki, "Blind separation of nonstationary sources in noisy mixtures," *Electronics Letters*, vol. 36, pp. 848–849, Apr. 2000.

[7] S. Choi and A. Cichocki, "Blind separation of nonstationary and temporally correlated sources from noisy mixtures," in *Proc. IEEE Workshop on Neural Networks for Signal Processing*, (Sidney, Austrailia), pp. 405–414, 2000.

[8] D. T. Pham and J. F. Cardoso, "Blind separation of instantaneous mixtures of nonstationary sources," in *Proc. ICA*, (Helsinki, Finland), pp. 187–192, 2000.

[9] S. Choi, "Differential Hebbian-type learning algorithms for decorrelation and independent component analysis," *Electronics Letters*, vol. 34, no. 9, pp. 900–901, 1998.

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

[11] 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*, (Singapore), 2001.

[12] J. F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM J. Mat. Anal. Appl.*, vol. 17, pp. 161–164, Jan. 1996.

[13] Y. Deville and S. Savoldelli, "A second-order differential approach for underdetermined convolutive source separation," in *Proc. ICASSP*, 2001.

[14] G. H. Golub and C. F. V. Loan, *Matrix Computations, 2nd edition*. Johns Hopkins, 1993.

[15] L. Tong, Y. Inouye, and R. Liu, "A finite-step global convergence algorithm for the parameter estimation of multichannel MA processes," *IEEE Trans. Signal Processing*, vol. 40, pp. 2547–2558, Oct. 1992.

[16] A. Belouchrani and A. Cichocki, "Robust whitening procedure in blind source separation context," *Electronics Letters*, vol. 36, pp. 2050–2051, Nov. 2000.

[17] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals," *IEE Proceedings-F*, vol. 140, no. 6, pp. 362–370, 1993.

[18] S. Choi and A. Cichocki, "Correlation matching approach to source sepration in the presence of spatially correlated noise," in *Proc. IEEE ISSPA*, (Kuala-Lumpur, Malaysia), 2001.