Journal of VLSI Signal Processing, ?, 1–13 (2001)

Second Order Nonstationary Source Separation

SEUNGJIN CHOI
Department of Computer Science and Engineering, POSTECH, Korea

ANDRZEJ CICHOCKI
Lab for Advanced Brain Signal Processing, Brain Science Institute, RIKEN, Japan

ADEL BELOUCHARNI
Department of Electrical Engineering, Ecole Nationale Polytechnique, Algeria

Received ?; Revised ?

Editors: ?

Abstract. This paper addresses a method of blind source separation that jointly exploits the nonstationarity and temporal structure of sources. The method needs only multiple time-delayed correlation matrices of the observation data, each of which is evaluated at different time-windowed data frame, to estimate the demixing matrix. The method is insensitive to the temporally white noise since it is based on only time-delayed correlation matrices (with non-zero time-lags) and is applicable to the case of either nonstationary sources or temporally correlated sources. We also discuss the extension of some existing methods with the overview of second-order blind source separation methods. Extensive numerical experiments confirm the validity and high performance of the proposed method.

Keywords: Blind source separation, joint approximate diagonalization, noisy mixtures, nonstationarity, simultaneous diagonalization, temporal correlations

1. Introduction

Blind source separation (BSS) is a fundamental problem that is encountered in many practical applications such as telecommunications, array signal processing, image processing, speech processing (cocktail party problem), 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 $\boldsymbol{s}(t)$.

The task of BSS is to estimate the mixing matrix $\boldsymbol{A}$ (or its pseudo-inverse, $\boldsymbol{W} = \boldsymbol{A}^{\#}$ that is referred to as the demixing matrix), given only a finite number of observation data $\{\boldsymbol{x}(t)\}$, $t = 1, \ldots, N$. Two indeterminacies cannot be resolved in BSS without any prior knowledge. They include scaling and permutation ambiguities. Thus if the estimate of the mixing matrix, $\widehat{\boldsymbol{A}}$ satisfies $\boldsymbol{G} = \boldsymbol{W}\boldsymbol{A} = \widehat{\boldsymbol{A}}^{\#}\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}$ where $\boldsymbol{G}$ is the global transformation which combines the mixing and demixing system, $\boldsymbol{P}$ is some permutation matrix, and

$\boldsymbol{\Lambda}$ is some nonsingular diagonal matrix, then $(\widehat{\boldsymbol{A}}, \widehat{\boldsymbol{s}})$ and $(\boldsymbol{A}, \boldsymbol{s})$ are said to be related by a waveform-preserving relation [24].

A variety of methods/algorithms for BSS have been developed for last decade (for example, see [16] and references therein). Although many different BSS algorithms are available, their principles can be summarized by three distinctive approaches.

- The most popular approach to BSS exploits the marginal non-Gaussianity. When stationary sources are assumed to be i.i.d. with no temporal structure, higher-order statistics (HOS) is essential (implicitly or explicitly) to solve the BSS problem. In such a case, the method does not allow more than one Gaussian source. Exemplary algorithms can be found in [1, 7].
- If sources are spatially uncorrelated (less restrictive condition than the statistical independence) but temporally correlated (i.e., each source has non-vanishing temporal correlations), then second-order statistics (SOS) is sufficient to estimate the mixing matrix. Along this line, several methods have been developed [25, 18, 3]. Note that these SOS methods do not allow the separation of sources with identical spectra shape.
- The approaches described above inherently assume stationary sources. The property that we pay attention to in this paper, is the nonstationarity of sources. Mainly we are interested in the second-order nonstationarity (in the sense that sources have time-varying variances) which leads to SOS BSS algorithms. The nonstationarity was first taken into account by Matsuoka *et al.* [17] and was elaborated by Choi *et al.* using the natural gradient [13]. Decorrelation learning methods were shown to be sufficient to perform the BSS task. In contrast to other approaches, the methods based on nonstationarity allow the separation of Gaussian sources with identical spectra. However, they do not allow the separation of sources with identical nonstationary properties. One can also find some recent work on nonstationary source separation [12, 21, 11, 4].

In this paper we present the methods which jointly exploit the nonstationarity and temporal structure of sources to estimate the mixing matrix (or the demixing matrix) in the presence of spatially correlated but temporally white noise (which is not necessarily Gaussian). Thus our methods works even for the case where multiple Gaussian sources with no temporal correlations exit, whereas the SOS methods based on temporal correlations fail to separate such sources. Moreover, we show that if we use just time-delayed correlations of the observation data, we can find an estimate of the demixing matrix which is not sensitive to the white noise. To this end, we introduce a new whitening method and present the Second-Order Nonstationary source Separation (SEONS) method. Extensive computer simulations confirm the high performance of our methods.

Throughout this paper, the following assumptions are made:

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

(AS2) Sources are spatially uncorrelated but are temporally correlated (colored) stochastic signals with zero mean and time-varying variances, i.e.,

$$E\{s_i(t)s_j(t-\tau)\} = \delta_{ij}\gamma_i(t,\tau), \tag{2}$$

where δ_{ij} is the Kronecker delta. Source variances, $\{\gamma_i(t,\tau)\}$, are not constant with time t and are not zeros for some time-lags $\{\tau = 1, 2, \ldots\}$.

(AS3) Additive noises $\{v_i(t)\}$ are spatially correlated but temporally white, i.e.,

$$E\{\boldsymbol{v}(t)\boldsymbol{v}^T(t-\tau)\} = \delta_\tau \boldsymbol{\Gamma}, \tag{3}$$

where $\boldsymbol{\Gamma}$ is an arbitrary $m \times m$ matrix.

The rest of this paper is organized as follows. Next section briefly reviews some fundamental principles and SOS BSS methods. It is described mainly for the case of temporally correlated sources. In Section 3, we discuss the symmetric-definite pencils and present an extension of the matrix pencil method [10]. In Section 4, we introduce a new whitening method which is not sensitive to the white noise. Section 5 explains how we jointly exploit the nonstationarity and temporal structure of sources in the task of BSS and present the Second-Order Nonstationary source Separation (SEONS) method. The extension of the Pham-Cardoso algorithm [21] is also discussed. In Section 6, three numerical experimental results are presented to show the validity and high performance of our methods. Finally conclusion is drawn in Section 7

2. Second-order BSS Methods

2.1. A Fundamental Principle

For the moment, we consider stationary sources which are spatially uncorrelated but temporally correlated and assume the noise is isotropic Gaussian process, i.e., the covariance matrix of noise vector $\boldsymbol{v}(t)$ has the form

$$\boldsymbol{R}_v(0) = E\{\boldsymbol{v}(t)\boldsymbol{v}^T(t)\} = \sigma_v^2 \boldsymbol{I}_m, \quad (4)$$

where E denotes the statistical expectation operator, $\boldsymbol{I}_m$ is the $m \times m$ identity matrix, and σ_v^2 is the noise variance. The principle described here can be also applied to the case of nonstationary sources and its details are deferred to Section 5.

The correlation matrices of the observation vector $\boldsymbol{x}(t)$ satisfy

$$\begin{aligned} \boldsymbol{R}_x(0) - \sigma_v^2 \boldsymbol{I}_m &= \boldsymbol{A}\boldsymbol{R}_s(0)\boldsymbol{A}^T, & (5) \\ \boldsymbol{R}_x(\tau) &= \boldsymbol{A}\boldsymbol{R}_s(\tau)\boldsymbol{A}^T, & (6) \end{aligned}$$

for some non-zero time-lag τ and both $\boldsymbol{R}_s(0)$ and $\boldsymbol{R}_s(\tau)$ are diagonal matrices since sources are assumed to be spatially uncorrelated. In the case of overdetermined mixtures ($m > n$), the noise variance σ_v^2 can be estimated from the least singular value of $\boldsymbol{R}_x(0)$ (or the average of minor $m-n$ singular values of $\boldsymbol{R}_x(0)$). However such an estimate of noise variance is not reliable or is difficult to calculate when signal to noise ratio (SNR) is low or the variance of each sensor noise is different.

Denote $\widetilde{\boldsymbol{R}}_x(0) = \boldsymbol{R}_x(0) - \sigma_v^2 \boldsymbol{I}_m$. Then the pseudo-inverse of the mixing matrix, $\boldsymbol{A}^{\#}$ can be identified up to its re-scaled and permuted version by the simultaneous diagonalization of $\widetilde{\boldsymbol{R}}_x(0)$ and $\boldsymbol{R}_x(\tau)$, provided that $\boldsymbol{R}_s^{-1}(0)\boldsymbol{R}_s(\tau)$ has distinct diagonal elements. In fact, this is the main idea of AMUSE [25] that was motivated by FOBI [6]. This fundamental result is described in the following theorem (that is already exploited in literature).

Theorem 1. *Let $\boldsymbol{\Lambda}_1, \boldsymbol{D}_1 \in \mathbb{R}^{n\times n}$ be diagonal matrices with positive diagonal entries and $\boldsymbol{\Lambda}_2, \boldsymbol{D}_2 \in \mathbb{R}^{n\times n}$ be diagonal matrices with non-zero diagonal entries. Suppose that $\boldsymbol{G} \in \mathbb{R}^{n\times n}$ satisfies the following decompositions:*

$$\begin{aligned} \boldsymbol{D}_1 &= \boldsymbol{G}\boldsymbol{\Lambda}_1\boldsymbol{G}^T, & (7) \\ \boldsymbol{D}_2 &= \boldsymbol{G}\boldsymbol{\Lambda}_2\boldsymbol{G}^T. & (8) \end{aligned}$$

Then the matrix $\boldsymbol{G}$ is the generalized permutation matrix, i.e., $\boldsymbol{G} = \boldsymbol{P}\boldsymbol{\Lambda}$ if $\boldsymbol{D}_1^{-1}\boldsymbol{D}_2$ and $\boldsymbol{\Lambda}_1^{-1}\boldsymbol{\Lambda}_2$ have distinct diagonal entries.

Proof: See Appendix

2.2. Simultaneous Diagonalization

In general we can find a linear transformation which simultaneously diagonalizes two symmetric matrices. For the sake of simplicity, the simultaneous diagonalization is explained in the case of $m = n$ and noise-free mixtures. Thus we deal with $\boldsymbol{R}_x(0)$ and $\boldsymbol{R}_x(\tau)$. The simultaneous diagonalization consists of two steps (whitening followed by an unitary transformation):

(1) First, the matrix $\boldsymbol{R}_x(0)$ is whitened by

$$\boldsymbol{z}(t) = \boldsymbol{D}_1^{-\frac{1}{2}}\boldsymbol{U}_1^T\boldsymbol{x}(t), \quad (9)$$

where $\boldsymbol{D}_1$ and $\boldsymbol{U}_1$ are the eigenvalue and eigenvector matrices of $\boldsymbol{R}_x(0)$ as

$$\boldsymbol{R}_x(0) = \boldsymbol{U}_1\boldsymbol{D}_1\boldsymbol{U}_1^T. \quad (10)$$

Then we have

$$\begin{aligned} \boldsymbol{R}_z(0) &= \boldsymbol{D}_1^{-\frac{1}{2}}\boldsymbol{U}_1^T\boldsymbol{R}_x(0)\boldsymbol{U}_1\boldsymbol{D}_1^{-\frac{1}{2}} = \boldsymbol{I}_m, \\ \boldsymbol{R}_z(\tau) &= \boldsymbol{D}_1^{-\frac{1}{2}}\boldsymbol{U}_1^T\boldsymbol{R}_x(\tau)\boldsymbol{U}_1\boldsymbol{D}_1^{-\frac{1}{2}}. \end{aligned}$$

(2) Second, a unitary transformation is applied to diagonalize the matrix $\boldsymbol{R}_z(\tau)$. The eigen-decomposition of $\boldsymbol{R}_z(\tau)$ has the form

$$\boldsymbol{R}_z(\tau) = \boldsymbol{U}_2\boldsymbol{D}_2\boldsymbol{U}_2^T. \quad (11)$$

Then $\boldsymbol{y}(t) = \boldsymbol{U}_2^T\boldsymbol{z}(t)$ satisfies

$$\begin{aligned} \boldsymbol{R}_y(0) &= \boldsymbol{U}_2^T\boldsymbol{R}_z(0)\boldsymbol{U}_2 = \boldsymbol{I}_m, \\ \boldsymbol{R}_y(\tau) &= \boldsymbol{U}_2^T\boldsymbol{R}_z(\tau)\boldsymbol{U}_2 = \boldsymbol{D}_2. \end{aligned}$$

Thus both matrices $\boldsymbol{R}_x(0)$ and $\boldsymbol{R}_x(\tau)$ are simultaneously diagonalized by a linear transform $\boldsymbol{W} = \boldsymbol{U}_2^T\boldsymbol{D}_1^{-\frac{1}{2}}\boldsymbol{U}_1^T$. It follows from Theorem 1 that $\boldsymbol{W} = \boldsymbol{U}_2^T\boldsymbol{D}_1^{-\frac{1}{2}}\boldsymbol{U}_1^T$ is a valid demixing matrix if all the diagonal elements of $\boldsymbol{D}_2$ are distinct.

2.3. Generalized Eigenvalue Problem

The simultaneous diagonalization of two symmetric matrices can be carried out without going through two-step procedures. From the discussion in Section 2.2, we have

$$\boldsymbol{W}\boldsymbol{R}_x(0)\boldsymbol{W}^T = \boldsymbol{I}_m, \quad (12)$$
$$\boldsymbol{W}\boldsymbol{R}_x(\tau)\boldsymbol{W}^T = \boldsymbol{D}_2. \quad (13)$$

The linear transformation $\boldsymbol{W}$ which satisfies (12) and (13) is the eigenvector matrix of $\boldsymbol{R}_x^{-1}(0)\boldsymbol{R}_x(\tau)$ [14]. In other words, the matrix $\boldsymbol{W}$ is the generalized eigenvector matrix of the pencil $\boldsymbol{R}_x(\tau) - \lambda\boldsymbol{R}_x(0)$ [18].

Recently Chang *et al.* proposed the matrix pencil method for BSS [10] where they exploited $\boldsymbol{R}_x(\tau_1)$ and $\boldsymbol{R}_x(\tau_2)$ for $\tau_1 \neq \tau_2 \neq 0$. Since the noise vector was assumed to be temporally white, two matrices $\boldsymbol{R}_x(\tau_1)$ and $\boldsymbol{R}_x(\tau_2)$ are not theoretically affected by the noise vector, i.e.,

$$\boldsymbol{R}_x(\tau_1) = \boldsymbol{A}\boldsymbol{R}_s(\tau_1)\boldsymbol{A}^T, \quad (14)$$
$$\boldsymbol{R}_x(\tau_2) = \boldsymbol{A}\boldsymbol{R}_s(\tau_2)\boldsymbol{A}^T. \quad (15)$$

Thus it is obvious to see that we can find an estimate of demixing matrix that is not sensitive to the white noise. A similar idea was also exploited in [12, 11].

In general, the generalized eigenvalue decomposition requires the symmetric-definite pencil (one matrix is symmetric and the other is symmetric and positive definite). However $\boldsymbol{R}_x(\tau_2) - \lambda\boldsymbol{R}_x(\tau_1)$ is not symmetric-definite, which might cause a numerical instability problem which results in complex-valued eigenvectors. Section 3 describes how we can construct a symmetric-definite pencil and discusses the extension of the matrix pencil method.

2.4. *SOBI*

The AMUSE [25] and the matrix pencil method [10] exploits only two different correlation matrices of the observation vector, so their performance is degraded if some eigenvalues of $\boldsymbol{R}_s(\tau)$ are close each other. In order to avoid this drawback, Belouchrani *et al.* [3] proposed the SOBI algorithm based on the joint approximate diagonalization of multiple time-delayed correlation matrices.

Let us recall the whitened vector $\boldsymbol{z}(t)$ in (9). Let us consider a set of time-delayed correlation matrices of $\boldsymbol{z}(t)$, $\{\boldsymbol{R}_z(\tau_i)\}$, $i = 1, \ldots, K$. The SOBI finds an unitary transformation $\boldsymbol{V}$ such that

$$\boldsymbol{V}^T\boldsymbol{R}_z(\tau_i)\boldsymbol{V} = \boldsymbol{\Lambda}_i, \text{ for } i = 1, \ldots, K, \quad (16)$$

where $\{\boldsymbol{\Lambda}_i\}$ is a set of diagonal matrices. Since the SOBI exploits several correlation matrices, it reduces the probability that some bad choice of time-lag τ_i results in the case where $\boldsymbol{R}_z(\tau_i)$ has the same diagonal elements. In addition, the exploitation of several correlation matrices increases the statistical efficiency [3].

3. Symmetric-Definite Pencils

The set of all matrices of the form $\boldsymbol{R}_1 - \lambda\boldsymbol{R}_2$ with $\lambda \in \mathbb{R}$ is said to be a *pencil*. Frequently we encounter into the case where $\boldsymbol{R}_1$ is symmetric and $\boldsymbol{R}_2$ is symmetric and positive definite. Pencils of this variety are referred to as *symmetric-definite pencils* [15].

Theorem 2. (pp. 468 in [15]) *If $\boldsymbol{R}_1 - \lambda\boldsymbol{R}_2$ is symmetric-definite, then there exists a nonsingular matrix $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n]$ such that*

$$\boldsymbol{U}^T\boldsymbol{R}_1\boldsymbol{U} = \textit{diag}\{\gamma_1(\tau_1), \ldots, \gamma_n(\tau_1)\}, \quad (17)$$
$$\boldsymbol{U}^T\boldsymbol{R}_2\boldsymbol{U} = \textit{diag}\{\gamma_1(\tau_2), \ldots, \gamma_n(\tau_2)\}. \quad (18)$$

Moreover $\boldsymbol{R}_1\boldsymbol{u}_i = \lambda_i\boldsymbol{R}_2\boldsymbol{u}_i$ for $i = 1, \ldots, n$, and $\lambda_i = \frac{\gamma_i(\tau_1)}{\gamma_i(\tau_2)}$.

It is apparent from Theorem 2 that $\boldsymbol{R}_1$ should be symmetric and $\boldsymbol{R}_2$ should be symmetric and positive definite so that the generalized eigenvector $\boldsymbol{U}$ can be a valid solution if $\{\lambda_i\}$ are distinct. Unfortunately in [10], the symmetric-definite pencil was not considered, so we might have a numerical instability problem in the calculation of the generalized eigenvectors.

Now we explain how we construct a symmetric-definite pencil. Let us consider two time-delayed correlation matrices $\boldsymbol{R}_x(\tau_1)$ and $\boldsymbol{R}_x(\tau_2)$ for non-zero time-lags τ_1 and τ_2. For the requirement of symmetry, we replace $\boldsymbol{R}_x(\tau_1)$ and $\boldsymbol{R}_x(\tau_2)$ by $\boldsymbol{M}_x(\tau_1)$ and $\boldsymbol{M}_x(\tau_2)$ that are defined by

$$\boldsymbol{M}_x(\tau_1) = \frac{1}{2}\left\{\boldsymbol{R}_x(\tau_1) + \boldsymbol{R}_x^T(\tau_1)\right\}, \quad (19)$$
$$\boldsymbol{M}_x(\tau_2) = \frac{1}{2}\left\{\boldsymbol{R}_x(\tau_2) + \boldsymbol{R}_x^T(\tau_2)\right\}. \quad (20)$$

Then the pencil $\boldsymbol{M}_x(\tau_2) - \lambda\boldsymbol{M}_x(\tau_1)$ is a symmetric pencil. In general, the matrix $\boldsymbol{M}_x(\tau_1)$ is not positive definite for $\tau_1 \neq 0$. Thus instead of $\boldsymbol{M}_x(\tau_1)$, we consider a linear combination of several time-delayed correlation matrices, i.e.,

$$\boldsymbol{C}_1 = \sum_{i=1}^{J} \alpha_i \boldsymbol{M}_x(\tau_i). \quad (21)$$

The set of coefficients, $\{\alpha_i\}$, is chosen in such a way that the symmetric matrix $\boldsymbol{C}_1$ is positive definite. One simple way to do this is to use the finite step global convergence (FSGC) algorithm [23]. This method is referred to as *Extended Matrix Pencil Method* that is summarized below.

Algorithm Outline: Extended Matrix Pencil Method

1. Compute $\boldsymbol{M}_x(\tau_2)$ for some time-lag $\tau_2 \neq 0$ and calculate the matrix $\boldsymbol{C}_1 = \sum_{i=1}^{J} \alpha_i \boldsymbol{M}_x(\tau_i)$ by the FSGC method.
2. Find the generalized eigenvector matrix $\boldsymbol{V}$ of the pencil $\boldsymbol{M}_x(\tau_2) - \lambda \boldsymbol{C}_1$ which satisfies
$$\boldsymbol{M}_x(\tau_2)\boldsymbol{V} = \boldsymbol{C}_1 \boldsymbol{V} \boldsymbol{\Lambda}. \tag{22}$$
3. The demixing matrix is given by $\boldsymbol{W} = \boldsymbol{V}^T$.

4. Robust Whitening

The whitening (or data sphering) is an important preprocessing step in a variety of BSS methods. The conventional whitening exploits the equal-time correlation matrix of the data $\boldsymbol{x}(t)$, so that the effect of additive noise can not be removed. The idea of a new whitening method lies in utilizing the time-delayed correlation matrices that are not sensitive to the white noise. A new whitening method is named as a *robust whitening*, motivated by the fact that it is not sensitive to the white noise. However, it is somewhat different from (Huber's) robust statistics.

The time-delayed correlation matrix of the observation data $\boldsymbol{x}(t)$ has the form

$$\begin{aligned} \boldsymbol{R}_x(\tau) &= E\{\boldsymbol{x}(t)\boldsymbol{x}^T(t-\tau)\} \\ &= \boldsymbol{A}\boldsymbol{R}_s(\tau)\boldsymbol{A}^T, \end{aligned} \tag{23}$$

for $\tau \neq 0$. One can easily see that the transformation $\boldsymbol{R}_x^{-\frac{1}{2}}(\tau)$ whiten the data $\boldsymbol{x}(t)$ without the effect of the noise vector $\boldsymbol{v}(t)$. It reduces the noise effect and project the data onto the signal subspace, in contrast to the conventional whitening transformation $\boldsymbol{R}_x^{-\frac{1}{2}}(0)$. Some source separation methods already employed this robust whitening transformation [19, 12, 11, 2, 5].

In general, however, the matrix $\boldsymbol{R}_x(\tau)$ is not always positive definite, so the whitening transformation $\boldsymbol{R}_x^{-\frac{1}{2}}(\tau)$ may not be valid for some time-lag τ. The idea of the robust whitening is to consider a linear combination of several time-delayed correlation matrices, i.e.,

$$\boldsymbol{C}_x = \sum_{i=1}^{K} \alpha_i \boldsymbol{M}_x(\tau_i). \tag{24}$$

A proper choice of $\{\alpha_i\}$ results in the positive definite matrix $\boldsymbol{C}_x$, as in the extended matrix pencil method. Once again, the FSGC method [23] can be used to find a set of coefficients $\{\alpha_i\}$ such that the matrix $\boldsymbol{C}_x$ is positive definite.

The matrix $\boldsymbol{C}_x$ has the eigen-decomposition

$$\boldsymbol{C}_x = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^T, \tag{25}$$

where $\boldsymbol{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m]$ and

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{D}_1 & \\ & \boldsymbol{0} \end{bmatrix}, \tag{26}$$

where $\boldsymbol{D}_1 \in \mathbb{R}^{n \times n}$ is a diagonal matrix whose diagonal elements are n principal eigenvalues of $\boldsymbol{C}_x$. Let $\boldsymbol{U}_1 = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n]$. Then the robust whitening transformation matrix is given by $\boldsymbol{Q} = \boldsymbol{D}_1^{-\frac{1}{2}} \boldsymbol{U}_1^T$. The transformation $\boldsymbol{Q}$ project the data onto n-dimensional signal subspace as well as carrying out whitening.

Let us denote the whitened n-dimensional data by $\boldsymbol{z}(t)$

$$\begin{aligned} \boldsymbol{z}(t) &= \boldsymbol{Q}\boldsymbol{x}(t) \\ &= \boldsymbol{B}\boldsymbol{s}(t) + \boldsymbol{Q}\boldsymbol{v}(t), \end{aligned} \tag{27}$$

where $\boldsymbol{B} \in \mathbb{R}^{n \times n}$. The whitened data $\boldsymbol{z}(t)$ (in the sense that $\sum_{i=1}^{K} \alpha_i \boldsymbol{M}_z(\tau_i) = \boldsymbol{I}$) is a unitary mixture of sources with additive noise, i.e., $\boldsymbol{B}\boldsymbol{B}^T = \boldsymbol{I}$.

Algorithm Outline: Robust whitening

1. Estimate time-delayed correlation matrices and construct an $m \times mJ$ matrix
$$\mathcal{M} = [\boldsymbol{M}_x(\tau_1) \cdots \boldsymbol{M}_x(\tau_J)]. \tag{28}$$
Then compute the singular value decomposition (SVD) of $\mathcal{M}$, i.e.,
$$\mathcal{M} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T, \tag{29}$$
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
$$\boldsymbol{U}_s = [\boldsymbol{u}_1 \cdots \boldsymbol{u}_n], \tag{30}$$
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
$$\boldsymbol{F}_i = \boldsymbol{U}_s^T \boldsymbol{M}_x(\tau_i)\boldsymbol{U}_s. \tag{31}$$
3. Choose any initial $\boldsymbol{\alpha} = [\alpha_1 \cdots \alpha_J]^T$.
4. Compute
$$\boldsymbol{F} = \sum_{i=1}^{J} \alpha_i \boldsymbol{F}_i. \tag{32}$$
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
$$\boldsymbol{\delta} = \frac{\left[\boldsymbol{u}^T \boldsymbol{F}_1 \boldsymbol{u} \cdots \boldsymbol{u}^T \boldsymbol{F}_J \boldsymbol{u}\right]^T}{\|[\boldsymbol{u}^T \boldsymbol{F}_1 \boldsymbol{u} \cdots \boldsymbol{u}^T \boldsymbol{F}_J \boldsymbol{u}]\|}. \tag{33}$$
Go to step 4. This loop is terminated in a finite number of steps (see [23] for proof).
7. Compute
$$\boldsymbol{C} = \sum_{i=1}^{J} \alpha_i \boldsymbol{M}_x(\tau_i), \tag{34}$$
and perform an eigenvalue-decomposition of $\boldsymbol{C}$,
$$\boldsymbol{C} = [\boldsymbol{U}_{c1}, \boldsymbol{U}_{c2}] \begin{bmatrix} \boldsymbol{D}_1 & \\ & \boldsymbol{0} \end{bmatrix} [\boldsymbol{U}_{c1}, \boldsymbol{U}_{c2}]^T \tag{35}$$
where $\boldsymbol{U}_{c1}$ contains the eigenvectors associated with n principal singular values of $\boldsymbol{D}_1$.
8. The robust whitening transformation is performed by
$$\boldsymbol{z}(t) = \boldsymbol{Q}\boldsymbol{x}(t), \tag{36}$$
where $\boldsymbol{Q} = \boldsymbol{D}_1^{-\frac{1}{2}} \boldsymbol{U}_{c1}^T$.

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

5. Second-order Nonstationary Source Separation

This section describes BSS methods which jointly exploit the nonstationarity and temporal structure of sources. They include: (1) the extended matrix pencil method (nonstationary case); (2) SEONS algorithm; (3) the extended Pham-Cardoso method.

Now we consider the case where sources are second-order nonstationary and have non-vanishing temporal correlations. It follows from the assumptions (AS1)-(AS3) that we have
$$\boldsymbol{M}_x(t_k, \tau_i) = \boldsymbol{A}\boldsymbol{M}_s(t_k, \tau_i)\boldsymbol{A}^T, \tag{37}$$
for $\tau_i \neq 0$. In practice $\boldsymbol{M}_x(t_k, \tau_i)$ is computed using the samples in the kth time-windowed data frame, i.e.,
$$\begin{aligned} \boldsymbol{R}_x(t_k, \tau_i) &= \frac{1}{N_k} \sum_{t \in \mathcal{N}_k} \boldsymbol{x}(t)\boldsymbol{x}^T(t - \tau_i), \\ \boldsymbol{M}_x(t_k, \tau_i) &= \frac{1}{2}\left\{\boldsymbol{R}_x(t_k, \tau_i) + \boldsymbol{R}_x^T(t_k, \tau_i)\right\}, \end{aligned}$$
where $\mathcal{N}_k$ is a set of data points in the kth time-windowed frame and N_k is the number of data points in $\mathcal{N}_k$.

It is straightforward to see that the extended matrix pencil method can be also applied to the case of nonstationary sources.

Algorithm Outline: Extended Matrix Pencil Method (nonstationary case)

1. We partition the observation data into two non-overlapping blocks, $\{\mathcal{N}_1, \mathcal{N}_2\}$.
2. Compute $\boldsymbol{M}_x(t_2, \tau_2)$ for some time-lag $\tau_2 \neq 0$ using the data points in $\mathcal{N}_2$.
3. Calculate the matrix $\boldsymbol{C}_1(t_1) = \sum_{i=1}^{J} \alpha_i \boldsymbol{M}_x(t_1, \tau_i)$ by the FSGC method using the data points in $\mathcal{N}_1$.
4. Find the generalized eigenvector matrix $\boldsymbol{V}$ of the pencil $\boldsymbol{M}_x(t_2, \tau_2) - \lambda \boldsymbol{C}_1(t_1)$ which satisfies
$$\boldsymbol{M}_x(t_2, \tau_2)\boldsymbol{V} = \boldsymbol{C}_1(t_1)\boldsymbol{V}\boldsymbol{\Lambda}. \tag{38}$$
3. The demixing matrix is given by $\boldsymbol{W} = \boldsymbol{V}^T$.

Remarks: The method in [22] employed two matrices $\boldsymbol{R}_x(t_1, 0)$ and $\boldsymbol{R}(t_2, 0)$ to estimate the demixing matrix.

In order to improve the statistical efficiency, we can employ the joint approximate diagonalization method [9] in our case, as in the JADE and SOBI. The joint approximate diagonalization method in [9] finds an unitary transformation that jointly diagonalizes several matrices (which do not have to be symmetric nor positive definite). The method SEONS is based on this joint approximate diagonalization. In this sense the SEONS includes the SOBI as its special case (if sources are stationary). The algorithm is summarized below.

Algorithm Outline: SEONS

1. The robust whitening method (described in Section 4) is applied to obtain the whitened vector $\boldsymbol{z}(t) = \boldsymbol{Q}\boldsymbol{x}(t)$. In the robust whitening step, we used the whole available data points.
2. Divide the whitened data $\{\boldsymbol{z}(t)\}$ into K non-overlapping blocks and calculate $\boldsymbol{M}_z(t_k, \tau_j)$ for $k = 1, \ldots, K$ and $j = 1, \ldots, J$. In other words, at each time-windowed data frame, we compute J different time-delayed correlation matrices of $\boldsymbol{z}(t)$.
3. Find a unitary joint diagonalizer $\boldsymbol{V}$ of $\{\boldsymbol{M}_z(t_k, \tau_j)\}$ using the joint approximate diagonalization method in [9], which satisfies

$$\boldsymbol{V}^T \boldsymbol{M}_z(t_k, \tau_j) \boldsymbol{V} = \boldsymbol{\Lambda}_{k,j}, \tag{39}$$

where $\{\boldsymbol{\Lambda}_{k,j}\}$ is a set of diagonal matrices.
4 The demixing matrix is computed as $\boldsymbol{W} = \boldsymbol{V}^T \boldsymbol{Q}$.

Recently Pham [20] developed a joint approximate diagonalization method where non-unitary joint diagonalizer of several Hermitian positive matrices is computed by a way similar to the classical Jacobi method. Second-order nonstationarity was also exploited in [21], but only noise-free data was considered. The following extended Pham-Cardoso method generalizes the method in [21]. One advantage of the extended Pham-Cardoso is the fact that it does not require the whitening step because the joint approximate diagonalization method in [21] finds a non-unitary joint diagonalizer. However, it requires that the set of matrices to be diagonalized should be Hermitian and positive definite, so we need to find a linear combination of time-delayed correlation matrices that is positive definite at each data frame, which increase the computational complexity.

Algorithm Outline: Extended Pham-Cardoso

1. Divide the data $\{\boldsymbol{x}(t)\}$ into K non-overlapping blocks and calculate $\boldsymbol{M}_x(t_k, \tau_j)$ for $k = 1, \ldots, K$ and $j = 1, \ldots, J$.
2. At each data frame, we compute

$$\boldsymbol{C}_k = \sum_{i=1}^{J} \alpha_i^{(k)} \boldsymbol{M}_x(t_k, \tau_i) \tag{40}$$

by the FSGC method for $k = 1, \ldots, K$. Note that $\{\boldsymbol{C}_k\}$ is symmetric and positive definite.
3. Find a non-unitary joint diagonalizer $\boldsymbol{V}$ of $\{\boldsymbol{C}_k\}$ using the joint approximate diagonalization method in [20], which satisfies

$$\boldsymbol{V} \boldsymbol{C}_k \boldsymbol{V}^T = \boldsymbol{\Lambda}_k, \tag{41}$$

where $\{\boldsymbol{\Lambda}_k\}$ is a set of diagonal matrices.
4. The demixing matrix is computed as $\boldsymbol{W} = \boldsymbol{V}$.

6. Numerical Experiments

Several numerical experimental results are presented to evaluate the performance of our method (SEONS) and to compare it with some existing methods such as JADE [8], SOBI [3], the matrix pencil method [10], and Pham-Cardoso [21]. Through numerical experiments, we confirm the useful behavior of the proposed method, SEONS, in a variety of cases: (1) the case where several nonstationary Gaussian sources exist and each Gaussian source has no temporal correlation; (2) the case where additive noises are spatially correlated but temporally white Gaussian processes; (3) the case where measurement noises are i.i.d. with uniform distribution.

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

$$\begin{aligned} \text{PI} \;=\; & \frac{1}{n(n-1)} \sum_{i=1}^{n} \left\{ \left(\sum_{k=1}^{n} \frac{|g_{ik}|}{\max_j |g_{ij}|} - 1 \right) \right. \\ & \left. + \left(\sum_{k=1}^{n} \frac{|g_{ki}|}{\max_j |g_{ji}|} - 1 \right) \right\}, \end{aligned} \tag{42}$$

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.

6.1. Experiment 1

The first experiment was designed to evaluate the effectiveness of the proposed method in the presence of several nonstationary Gaussian signals. In this experiment, we used three speech signals that are sampled at 8 kHz and two Gaussian signals (with no temporal correlations) whose variances are slowly varying. These 5 sources were mixed using a randomly generated 5×5 mixing matrix to generate 5-dimensional observation

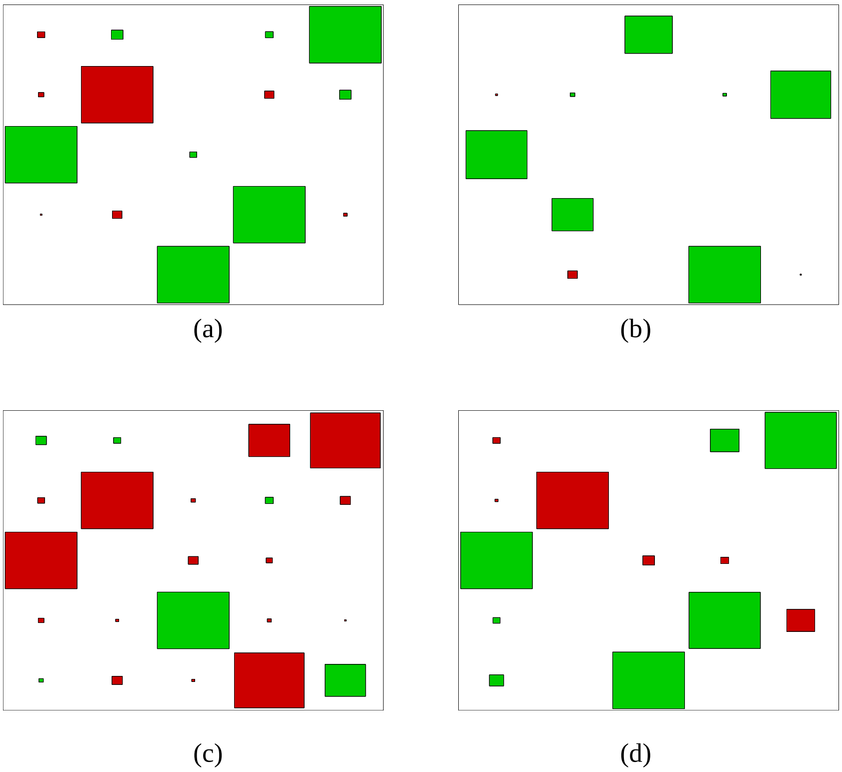

Fig. 1. Hinton diagrams of global system matrices in Experiment 1: (a) SEONS (PI=0.001); (b) Pham-Cardoso (PI=0.001); (c) JADE (PI=0.08); (d) SOBI (PI=0.01).

vector with 10000 data points. No measurement noise was added.

In this experiment, we compared the SEONS with JADE, SOBI, and Pham-Cardoso [21]. The following parameters were used in this experiment:

- In SEONS and Pham-Cardoso, we partitioned the whole data (10000 data points) into 100 different frames (each frame contains 100 data points) to calculate 100 different equal-time correlation matrices. These matrices were used to estimate the demixing matrix.
- In SOBI, we used 20 different time-delayed correlation matrices to estimate the demixing matrix.

The result is shown in Fig. 1 in which the Hinton diagram of the global system matrix $\boldsymbol{G}$ is plotted. In Hinton diagram, each square's area represents the magnitude of the element of the matrix and each square's color represents the sign of the element (dark gray for negative value and light gray for positive value). For successful separation, each row and column has only one dominant square (regardless of its color). Small squares contribute performance degradation. One can observe that SEONS and Pham-Cardoso work well even in the presence of nonstationary Gaussian sources (see (a) and (b) in Fig. 1), compared to JADE and SOBI (see (c) and (d) in Fig. 1). For the case of JADE, one can observe the performance degradation in the first and last row of $\boldsymbol{G}$, which verifies that the two Gaussian sources are difficult to be separated out. The SOBI gives slightly better performance than JADE, but its performance is not comparable to SEONS (see the first and fourth row of $\boldsymbol{G}$, (d) in Fig. 1).

6.2. *Experiment 2*

The second experiment was designed to show the robustness of the SEONS in the presence of spatially correlated but temporally white noise. We used 3 digitized voice signals and 2 music signals, all of which were sampled at 8 kHz. All the elements of the mixing matrix $\boldsymbol{A} \in \mathbb{R}^{5\times 5}$ were drawn from standardized Gaus-

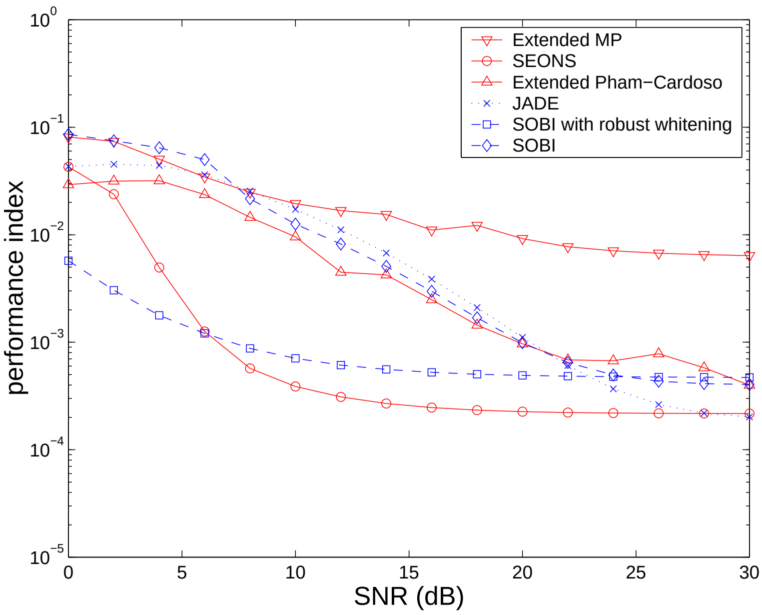

Fig. 2. Performance comparison for various BSS algorithms in Experiment 2.

sian distribution (i.e., zero mean and unit variance). As in the experiment 1, the whole data has 10000 samples.

The spatially correlated but temporally white Gaussian noise was generated in the following manner. We first generate white Gaussian noise with zero mean and unit variance which constitutes 5-dimensional noise vector $\widetilde{\boldsymbol{v}}$ (thus the covariance matrix is the identity matrix). Then it is transformed by an arbitrary matrix $\boldsymbol{H}$, each row of which is normalized. The transformed noise vector $\boldsymbol{v} = \boldsymbol{H}\widetilde{\boldsymbol{v}}$ has an covariance matrix $E\{\boldsymbol{v}(t)\boldsymbol{v}^T(t)\}$ whose diagonal elements are still one with non-zero off-diagonal elements. Its time-delayed correlation matrix is close to the zero matrix. The signal to noise ratio (SNR) was calculated by

$$\mathrm{SNR} = 10 \log_{10} \frac{\sigma_{s,i}^2}{\sigma_{v,i}^2}, \tag{43}$$

where

$$\sigma_{s,i}^2 = \frac{1}{N} \sum_{t=1}^{N} s_i^2(t), \tag{44}$$

and

$$\sigma_{v,i}^2 = \frac{c^2}{N} \sum_{t=1}^{N} v_i^2(t), \tag{45}$$

where the constant c^2 is chosen in such a way to generate the mixtures with pre-specified SNR level.

The algorithms that are tested in this experiment, include the extended matrix pencil method (Extended MP), SEONS, extended Pham-Cardoso, JADE, SOBI, and SOBI with robust whitening method [2, 5] (see Fig. 2). In the experiment, we used the following parameters:

- In the extended MP method, the observation data was partitioned into two non-overlapping blocks, $\{\mathcal{N}_1, \mathcal{N}_2\}$, so each data frame has 5000 samples. We found a linear combination of 5 time-delayed correlation matrices (with time-lags $\{1, 2, \ldots, 5\}$) in $\mathcal{N}_1$ such that the resulting matrix is positive definite. We also calculate the time-delayed correlation matrix with $\tau = 1$ using the samples in $\mathcal{N}_2$. The generalized eigenvectors of the pencil that consists of these two matrices (each of which calculated from $\mathcal{N}_1$ and $\mathcal{N}_2$, respectively), was computed to estimate the mixing matrix.
- In SEONS, we partitioned the data into 50 no overlapping blocks (each frame has 200 data points), $\{\mathcal{N}_1, \ldots, \mathcal{N}_{50}\}$. The robust whitening was performed using the combination of 5 time-delayed correlation matrices (with time-lags $\{1, 2, \ldots, 5\}$). In each data frame, we computed 5 time-delayed correlation matrices. A joint approx-

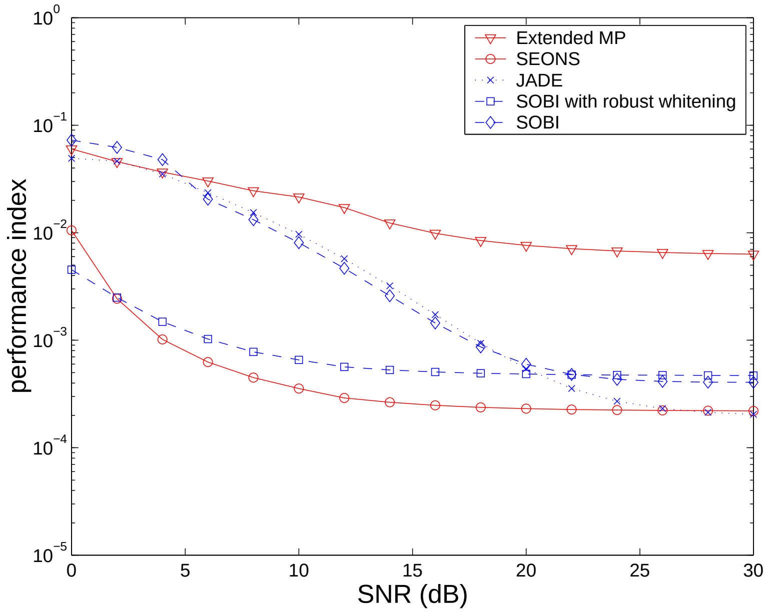

Fig. 3. Performance comparison for various BSS algorithms in Experiment 3.

imate diagonalizer of 250 correlation matrices (5 from each block = 5×50) was computed to estimate the demixing matrix.

- As in the SEONS, the extended Pham-Cardoso method used 50 no-overlapping blocks. At each block, we found a linear combination of 5 time-delayed correlation matrices (with time-lags $\{1, 2, \ldots, 5\}$), which is positive definite. Then a joint approximate diagonalizer of 50 correlation matrices was computed to estimate the demixing matrix.
- In SOBI, we used 20 time-delayed correlation matrices with time-lags $\{1, 2, \ldots, 20\}$.
- In SOBI with robust whitening, the same robust whitening as SEONS was carried out. Then 20 time-delayed correlation matrices with $\{1, 2, \ldots, 20\}$ was computed to estimate the mixing matrix.

At high SNR, most of algorithms worked very well, except for the extended MP method since it uses only two matrices. At low SNR, one can observe that the SOBI with robust whitening outperforms the SOBI (in which the whitening is based on the eigen-decomposition of the equal-time correlation matrix of the observation vector). The SEONS gives slightly better performance than the SOBI with robust whitening in most of ranges of SNR. In the range between 0 and 6 dB, the SEONS is worse than the SOBI with robust whitening. It might result from the fact that the SEONS takes only 200 data points to calculate the time-delayed correlation matrices, so the temporal whiteness of the noise vector is not really satisfied. One can use less number of blocks (so more data points for each block) to reduce this drawback. The advantage of SEONS over SOBI with robust whitening lies in the fact that the former works even for the case of nonstationary sources with identical spectra, whereas the latter does not (see the result of Experiment 1).

6.3. *Experiment 3*

The experimental setup is identical to the Experiment 2, except for the noise vector. Here we add the uniformly distributed noise vector. Since SEONS and SOBI with robust whitening is not affected by the noise covariance matrix, the performance is not degraded with respect the probability distribution of the noise. The result is shown in Fig. 3. Once again, SEONS and SOBI with robust whitening gave the best performance.

7. Conclusion

In this paper we have presented a method of BSS that jointly exploits the nonstationarity and temporal structure of sources. We introduced a new whitening method which exploited a linear combination of time-delayed correlation matrices and showed that it was not sensitive to the white noise. Based on this new whitening method, we have proposed the SEONS that needed only multiple time-delayed correlation matrices, each of which was evaluated at different time-windowed data frame, to estimate the demixing matrix. For the case of stationary sources (with temporal correlations), the correlation matrices evaluated at different time-windowed frames become identical. In such a case, the SEONS is identical to the SOBI. However, unlike SOBI, the SEONS could be also applied to the case where sources do not have temporal correlations but are nonstationary. Thus the SEONS includes the SOBI as its special case.

We also discussed the extension of several existing methods

- The matrix pencil method [10] was extended to the case of nonstationary sources. Moreover, we described how we chose a symmetric-definite pencil in order to avoid a numerical instability problem in the calculation of the generalized eigenvectors. It led to the extended matrix pencil method.
- The nonstationary source separation method proposed by Pham and Cardoso [21] was extended, in order to take the temporal structure of sources into account. At low SNR, SEONS gave better performance than the extended Pham-Cardoso.

8. Acknowledgment

This work was supported by Korea Ministry of Science and Technology under an International Cooperative Research Project and Brain Science and Engineering Research Program and by Korea Ministry of Information and Communication under Advanced backbone IT technology development project.

9. Appendix: Proof of Theorem 1

>From (7), there exists an orthogonal matrix $\boldsymbol{Q}$ such that

$$\left(\boldsymbol{G}\boldsymbol{\Lambda}_1^{\frac{1}{2}}\right) = \left(\boldsymbol{D}_1^{\frac{1}{2}}\right)\boldsymbol{Q}. \tag{46}$$

Hence,

$$\boldsymbol{G} = \boldsymbol{D}_1^{\frac{1}{2}}\boldsymbol{Q}\boldsymbol{\Lambda}_1^{-\frac{1}{2}}. \tag{47}$$

Substitute (47) into (8) to obtain

$$\boldsymbol{D}_1^{-1}\boldsymbol{D}_2 = \boldsymbol{Q}\boldsymbol{\Lambda}_1^{-1}\boldsymbol{\Lambda}_2\boldsymbol{Q}^T. \tag{48}$$

Since the right-hand side of (48) is the eigen-decomposition of the left-hand side of (48), the diagonal elements of $\boldsymbol{D}_1^{-1}\boldsymbol{D}_2$ and $\boldsymbol{\Lambda}_1^{-1}\boldsymbol{\Lambda}_2$ are the same. From the assumption that the diagonal elements of $\boldsymbol{D}_1^{-1}\boldsymbol{D}_2$ and $\boldsymbol{\Lambda}_1^{-1}\boldsymbol{\Lambda}_2$ are distinct, the orthogonal matrix $\boldsymbol{Q}$ must have the form $\boldsymbol{Q} = \boldsymbol{P}\boldsymbol{\Psi}$, where $\boldsymbol{\Psi}$ is an diagonal matrix whose diagonal elements are either $+1$ or -1. Hence, we have

$$\begin{aligned} \boldsymbol{G} &= \boldsymbol{D}_1^{\frac{1}{2}}\boldsymbol{P}\boldsymbol{\Psi}\boldsymbol{\Lambda}_1^{-\frac{1}{2}} \\ &= \boldsymbol{P}\boldsymbol{P}^T\boldsymbol{D}_1^{\frac{1}{2}}\boldsymbol{P}\boldsymbol{\Psi}\boldsymbol{\Lambda}_1^{-\frac{1}{2}} \\ &= \boldsymbol{P}\boldsymbol{\Lambda}, \end{aligned} \tag{49}$$

where

$$\boldsymbol{\Lambda} = \boldsymbol{P}^T\boldsymbol{D}_1^{\frac{1}{2}}\boldsymbol{P}\boldsymbol{\Psi}\boldsymbol{\Lambda}_1^{-\frac{1}{2}} \qquad \text{Q.E.D.} \tag{50}$$

References

1. 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.
2. A. Belouchrani. *Séparation Autodidacte de Sources: Algorithme, Performances et Application a des Signaux Expérimentaux*. PhD thesis, ENST, Telecom Paris, July 1995.
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*, 45:434–444, Feb. 1997.
4. A. Belouchrani and M. G. Amin. Blind source separation based on time-frequency signal representations. *IEEE Trans. Signal Processing*, 46:2888–2897, Nov. 1998.
5. A. Belouchrani and A. Cichocki. Robust whitening procedure in blind source separation context. *Electronics Letters*, 36(24):2050–2051, Nov. 2000.
6. J. -F. Cardoso. Source separation using higher-order moments. In *Proc. ICASSP*, 1989.
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. J. -F. Cardoso and A. Souloumiac. Blind beamforming for non Gaussian signals. *IEE Proceedings-F*, 140(6):362–370, 1993.
9. J. -F. Cardoso and A. Souloumiac. Jacobi angles for simultaneous diagonalization. *SIAM J. Mat. Anal. Appl.*, 17(1):161–164, Jan. 1996.
10. 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*, 48(3):900–907, Mar. 2000.
11. 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*, pages 405–414, Sidney, Austrailia, 2000.

12. S. Choi and A. Cichocki. Blind separation of nonstationary sources in noisy mixtures. *Electronics Letters*, 36(9):848–849, Apr. 2000.
13. S. Choi, A. Cichocki, and S. Amari. Equivariant nonstationary source separation. *Neural Networks*, 2001, to appear.
14. K. Fukunaga. *An Introduction to Statistical Pattern Recognition*. Academic Press, New York, NY, 1990.
15. G. H. Golub and C. F. Van Loan. *Matrix Computations, 2nd edition*. Johns Hopkins, 1993.
16. S. Haykin. *Unsupervised Adaptive Filtering: Blind Source Separation*. Prentice-Hall, 2000.
17. K. Matsuoka, M. Ohya, and M. Kawamoto. A neural net for blind separation of nonstationary signals. *Neural Networks*, 8(3):411–419, 1995.
18. L. Molgedey and H. G. Schuster. Separation of a mixture of independent signals using time delayed correlations. *Physical Review Letters*, pages 3634–3637, 1994.
19. K. R. Müller, P. Philips, and A. Ziehe. JADE_{TD}: Combining higher-order statistics and temporal information for blind source separation (with noise). In *Proc. ICA'99*, pages 87–92, Aussois, France, 1999.
20. D. T. Pham. Joint approximate diagonalization of positive definite hermitian matrices. Technical Report LMC/IMAG, University of Grenoble, France, 2000.
21. D. T. Pham and J. -F. Cardoso. Blind separation of instantaneous mixtures of nonstationary sources. In *Proc. ICA*, pages 187–192, Helsinki, Finland, 2000.
22. A. Souloumiac. Blind source detection and separation using second order non-stationarity. In *Proc. ICASSP*, pages 1912–1915, 1995.
23. 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*, 40(10):2547–2558, Oct. 1992.
24. L. Tong, Y. Inouye, and R. Liu. Waveform-preserving blind estimation of multiple independent sources. *IEEE Trans. Signal Processing*, 41(7):2461–2470, July 1993.
25. L. Tong, V. C. Soon, Y. F. Huang, and R. Liu. AMUSE: a new blind identification alogrithm. In *Proc. ISCAS*, pages 1784–1787, 1990.

Seungjin Choi was born in Seoul, Korea, in 1964. He received the B.S. and M.S. degrees in electrical engineering from Seoul National University, Korea, in 1987 and 1989, respectively and the Ph.D degree in electrical engineering from the University of Notre Dame, Indiana, in 1996.

He was a Visiting Assistant Professor in the Department of Electrical Engineering at University of Notre Dame, Indiana during the Fall semester of 1996. He was with the Laboratory for Artificial Brain Systems, RIKEN, Japan in 1997 and was an Assistant Professor in the School of Electrical and Electronics Engineering, Chungbuk National University from 1997 to 2000. He is currently Assistant Professor of Computer Science and Engineering at Pohang University of Science and Technology, and is affiliated with the Intelligent Multimedia Lab as well as the Brain Research Center. He has also been an adjunct senior researcher at Laboratory for Advanced Brain Signal Processing, BSI, Japan since 1998. His primary research interests include probabilistic/statistical learning, statistical (blind) signal processing, independent component analysis, data analysis, pattern recognition, and multiuser communications.

Dr. Choi is a Technical Committee member of IEEE Neural Networks for Signal Processing (NNSP).

Andrzej Cichocki was born in Poland on August 1947. He received the M.Sc.(with honors), Ph.D., and Habilitate Doctorate (Dr.Sc.) degrees, all in electrical engineering and computer science, from Warsaw University of Technology (Poland) in 1972, 1975, and 1982, respectively.

Since 1972, he has been with the Institute of Theory of Electrical Engineering and Electrical Measurements at the Warsaw University of Technology, where he became a full Professor in 1995.

He is the co-author of two international books: MOS Switched-Capacitor and Continuous-Time Integrated Circuits and Systems (Springer-Verlag, 1989) and Neural Networks for Optimization and Signal Processing (J Wiley and Teubner Verlag,1993/94) and author or co-author of more than hundred fifty (150) scientific papers.

He spent at University Erlangen-Nuernberg (GERMANY) a few years as Alexander Humboldt Research Fellow and Guest Professor. In 1995-96 he has been working as a Team Leader of the Laboratory for Artificial Brain Systems, at the Frontier Research Program RIKEN (JAPAN), in the Brain Information Processing Group directed by professor Shunichi Amari. Currently he is head of the laboratory for Open Information Systems in the Brain Science Institute, Riken, Wako-schi, JAPAN.

He is reviewer of several international Journals, e.g. IEEE Trans. on Neural Networks, Signal Processing, Circuits and Systems, Biological Cybernetics, Electronics Letters, Neurocomputing, Neural Computation. He is also member of several international Scientific Committees and the associated Editor of IEEE Transaction on Neural Networks (since January 1998). His current research interests include signal and image processing (especially blind signal/image processing), neural networks and their electronic implementations, learning theory and algorithms, independent and principal component analysis, optimization problems, circuits and systems theory and their applications, artificial intelligence.

Adel Belouchrani was born in Algiers, ALGERIA, on 5 May 1967. He received the State Engineering degree in electronics from the Ecole Nationale Polytechnique (ENP) of Algiers, ALGERIA, in 1991, and the M.Sc. degree in signal processing from the Institut National Polytechnique de Grenoble (INPG) , FRANCE, in 1992, and the Ph. D. degree in the field of signal and image processing from Ecole Nationale Superieure des Telecommunications (ENST) , Paris, FRANCE, in 1995.

From 1995 to 1996, he has been a visiting scholar at the Electrical Engineering and Computer Sciences Department of University of California at Berkeley , USA. From 1996 to 1997, he was with the Department of Electrical and Computer Engineering of Villanova University , Pennsylvania, USA, as research associate. During this period, he served as consultant for Comcast Inc. During February 1997, he has been a visiting scientist at Laboratory for Artificial Brain System, Riken, JAPAN. From August 1997 to October 1997, he has been with Alcatel ETCA , BELGIUM. Dr Belouchrani is currently with the Electrical Engineering Department of the Ecole Nationale Polytechnique of Algiers, ALGERIA, as Associate Professor. He has been, from August to September 2000, a visiting researcher at Brain Signal Processing Laboratory , Brain Science Institute , Riken, JAPAN. He has been also, from May to June 2001, a visiting schalor at the Neuro-Dynamic systems Laboratory at University of California at San Diego , USA. He has been, from August to September 2001, a visiting researcher at Brain Signal Processing Laboratory , Brain Science Institute , Riken, JAPAN. His current research interest are in Statistical signal processing and (blind) array signal processing with applications to biomedical and communications problems, time frequency representations, exploitation of the spatial time frequency distributions in array signal processing et FPGA implementation of signal processing algorithms.