A general framework for second-order blind separation of stationary colored sources

Abdeldjalil Aissa El Bey, Karim Abed-Meraim, Yves Grenier, Yingbo Hua

A General Framework for Second Order Blind Separation of Stationary Colored Sources

Abdeldjalil Aïssa-El-Bey [a,*], Karim Abed-Meraim [b,c], Yves Grenier [b] and Yingbo Hua [d]

[a] *Institut TELECOM; TELECOM Bretagne, Signal & Communications Department, Technopôle Brest-Iroise, BP 832, 29285 Brest, France*

[b] *Institut TELECOM; TELECOM ParisTech, TSI Department, 37-39 rue Dareau 75014, Paris, France*

[c] *University of Sharjah, ECUOS/ECE Department, 27272 Sharjah, UAE.*

[d] *University of California, College of Engineering, Department of Electrical Engineering, CA, 92521, USA*

Abstract

This paper focuses on the blind separation of stationary colored sources using the second order statistics of their instantaneous mixtures. We first start by presenting a brief overview of existing contributions in that field. Then, we present necessary and sufficient conditions for the identifiability and partial identifiability using a finite set of correlation matrices. These conditions depend on the autocorrelation function of the unknown sources. However, it is shown here that they can be tested directly from the observation through the decorrelator output. This issue is of prime importance to decide whether the sources have been well separated. If that's not the case then, further treatments will be needed. We then propose an identifiability testing based on resampling (jackknife) technique that is validated by simulation

results. Secondly, we present an iterative blind source separation method using second order statistics (SOS) and natural gradient technique. This algorithm has a number of attractive properties including its simplicity and 'easy' generalization to adaptive or convolutive schemes. Asymptotic performance analysis of this method is performed. Several numerical simulations are presented, to assess the theoretical results w.r.t the 'separability' testing, to demonstrate the effectiveness of the gradient-type decorrelation method and to validate the theoretical expression of the asymptotic performance index.

1 Introduction

Source separation aims to recover multiple sources from multiple observations (mixtures) received by a set of linear sensors. The problem is said to be 'blind' when the observations have been linearly mixed by the transfer medium, while having no *a priori* knowledge of the transfer medium or the sources. Blind source separation (BSS) has applications in several areas, such as communication, speech and audio processing, biomedical engineering, geophysical data processing, etc [1]. BSS of instantaneous mixtures has attracted so far a lot of attention due to its many potential applications and its mathematical tractability that leads to several nice and simple BSS solutions [1–5]. Assume that m narrow band signals impinge on an array of $n \geq m$ sensors. The

* Corresponding author.
Email addresses: abdeldjalil.aissaelbey@telecom-bretagne.eu (Abdeldjalil Aïssa-El-Bey), karim.abed@telecom-paristech.fr (Karim Abed-Meraim), yves.grenier@telecom-paristech.fr (Yves Grenier), yhua@ee.ucr.edu (Yingbo Hua).

measured array output is a weighted superposition of the signals, corrupted by additive noise, i.e.

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

where $\boldsymbol{s}(t) = [s_1(t), \cdots, s_m(t)]^T$ is the $m \times 1$ complex *source vector*, $\boldsymbol{\eta}(t) = [\eta_1(t), \cdots, \eta_n(t)]^T$ is the $n \times 1$ complex *noise vector*, $\boldsymbol{A}$ is the $n \times m$ full column rank *mixing matrix*, and the superscript T denotes the transpose operator. The source signal vector $\boldsymbol{s}(t)$ is assumed to be a multivariate stationary complex stochastic process.

In this paper, we only consider the second order BSS methods. Hence, hence the component processes $s_i(t)$, $1 \leq i \leq m$ are assumed to be temporally coherent and mutually uncorrelated, with zero mean and second order moments:

$$\boldsymbol{S}(\tau) \stackrel{\text{def}}{=} E\left(\boldsymbol{s}(t+\tau)\boldsymbol{s}^{\star}(t)\right) = \text{diag}[\rho_1(\tau), \cdots, \rho_m(\tau)] \ , \tag{2}$$

where $\rho_i(\tau) \stackrel{\text{def}}{=} E(s_i(t+\tau)s_i^*(t))$, the expectation operator is E, and superscripts $*$ and $\star$ denote the conjugate of a complex and the complex conjugate transpose of a vector, respectively. The additive noise $\boldsymbol{\eta}(t)$ is modeled as a white stationary zero-mean complex random process of covariance $E(\boldsymbol{\eta}(t)\boldsymbol{\eta}^{\star}(t)) = \sigma^2\boldsymbol{Q}$. The latter matrix is proportional to identity, (i.e. $E(\boldsymbol{\eta}(t)\boldsymbol{\eta}^{\star}(t)) = \sigma^2\boldsymbol{I}$) when the noise is spatially white. Under these assumptions, the observed data correlation matrices are given by :

$$\boldsymbol{R}_x(\tau) = \boldsymbol{A}\boldsymbol{S}(\tau)\boldsymbol{A}^{\star} + \delta(\tau)\sigma^2\boldsymbol{Q} \ .$$

From this expression, one can observe that the noise free correlation matrices are 'diagonalizable' under the linear transformation $\boldsymbol{B} = \boldsymbol{A}^{\#}$ where the superscript $(\cdot)^{\#}$ is the Moore-Penrose's pseudo-inversion operator, i.e. $\boldsymbol{B}\boldsymbol{R}_x(\tau)\boldsymbol{B}^H$ are diagonal $\forall\tau \neq 0$. Hence, the source separation is achieved by decorrelating

the signals at different time lags.

Before all, note that complete blind identification of separating (demixing) matrix $\boldsymbol{B}$ (or the equivalently mixing matrix $\boldsymbol{A}$) is impossible in the blind context, because the exchange of a fixed scalar between the source signal and the corresponding column of $\boldsymbol{A}$ leaves the observations unaffected. Also note that the signals *numbering* is immaterial. It follows that the best that can be done is to determine $\boldsymbol{B}$ up to a permutation and scalar shifts of its columns [3], i.e., $\boldsymbol{B}$ is a separating matrix if and only if :

$$\boldsymbol{B}\boldsymbol{y}(t) = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{s}(t) \tag{3}$$

where $\boldsymbol{P}$ is a permutation matrix and $\boldsymbol{\Lambda}$ a non-singular diagonal matrix.

Under the above assumptions, we provide a general framework for the BSS including the study of the identifiability and its testing as well as the introduction of a simple but efficient decorrelation method and its performance analysis. The paper is organized as follows. Section 2 reviews the principal contributions to the BSS problem using the second order statistics. Section 3 states the necessary and sufficient second order identifiability conditions. We then propose an identifiability testing based on resampling technique in section 4. Section 5 proposes a blind source separation algorithm using relative gradient technique. Section 6 is devoted to the performance analysis of the considered decorrelation method and the validation of the identifiability testing technique. Section 7 is for concluding remarks.

2 Overview of related works

The work using the second order statistics to achieve the blind source separation was initiated by L. Féty et al. [6]. Féty's method is based on the simultaneous diagonalization of the correlation matrices $\boldsymbol{R}_x(0)$ and $\boldsymbol{R}_x(1)$. Independently, L. Tong et al. proposed a similar technique, namely the AMUSE technique (Algorithm for Multiple Unknown Signals Extraction) [7, 8], that achieves the BSS by the simultaneous diagonalization of two symmetric matrices $\boldsymbol{R}_x(0)$ and $(\boldsymbol{R}_x(\tau_k) + \boldsymbol{R}_x^{\star}(\tau_k))/2$ with $\tau_k \neq 0$. This method has been extended in [9] where a generalized eigenvalue decomposition of a matrix pencil $(\boldsymbol{R}_x(\tau_1), \boldsymbol{R}_x(\tau_2))$ is considered.

Later on, A. Belouchrani et al. proposed the SOBI (Second-Order Blind Identification) algorithm [3] that generalizes the previous methods to the case where more than two correlation matrices are used. In the SOBI algorithm, the separation is achieved in two steps; the first step is the whitening of the observed signal vector by linear transformation. The second step consists of applying a joint approximate diagonalization algorithm to a set of different time-lag correlation matrices of the whitened signal vector. A variant of SOBI has been presented by D. Nuzillard et al. in [10] allowing direct signal separation from frequency domain data by exploiting the source correlation properties expressed in the time domain. This algorithm is referred to as f-SOBI, standing for "frequency domain SOBI".

Numerous approaches have been proposed in recent years both for the formulation of the diagonalization criterion and for the algorithms considered for its minimization. One of the most popular and computationally appealing approach for the joint diagonalization of a set of matrices $\boldsymbol{M}_1, \cdots, \boldsymbol{M}_K$ is the

unitary Approximate Joint Diagonalization (J.F. Cardoso et al. [11]), which minimizes the criterion

$$\sum_{k=1}^{K} \text{off}(\boldsymbol{B}\boldsymbol{M}_k\boldsymbol{B}^\star) \tag{4}$$

with respect to $\boldsymbol{B}$, subject to the unitary constraint $\boldsymbol{B}^\star\boldsymbol{B} = \boldsymbol{I}$, where

$$\text{off}(\boldsymbol{P}) = \sum_{i \neq j} |P_{ij}|^2 \ . \tag{5}$$

The unitary constraint implies the assumption of a unitary mixing matrix. Hence, in the general case, a pre-processing "spatial hard-whitening" stage is required, in which the non-unitary factor of the overall demixing matrix is found and applied to the data.

In [5, 12] an iterative algorithm using relative gradient technique has been considered for the minimization of (4) without unitary constraint. An alternative approach for non-unitary Approximate Joint Diagonalization has been proposed by Yeredor (the "AC-DC" algorithm [13]), which minimizes

$$\sum_{k=1}^{K} \|\boldsymbol{M}_k - \boldsymbol{A}\boldsymbol{D}_k\boldsymbol{A}^\star\|^2 \tag{6}$$

without constraining $\boldsymbol{A}$ to be unitary. In (6), $\boldsymbol{D}_k,\ k = 1, \cdots, K$ represent diagonal matrices. While computationally efficient in small-scale problems, this algorithm has been observed to exhibit extremely slow convergence in large-scale problems. This criterion is also considered in [14] where a gradient descent technique is used for its minimization and in [15] where quadratic optimization is used.

A computationally efficient unconstrained minimization algorithm was proposed by D.T. Pham et al. [16], whose target criterion is the Kullback-Leibler divergence between the $n \times n$ operand and the diagonal matrix with the same

diagonal as the operand :

$$\sum_{k=1}^{K} \lambda_k [\log\det \operatorname{diag}(\boldsymbol{B}\boldsymbol{M}_k\boldsymbol{B}^\star) - \log\det(\boldsymbol{B}\boldsymbol{M}_k\boldsymbol{B}^\star)] \tag{7}$$

where $\lambda_k,\ k = 1, \cdots, K$ are positive scalar factors. This approach requires all the target matrices to be positive definite, which limits its applicability as a generic BSS tool. Another class of BSS techniques based on the second order statistics is the one using the maximum likelihood principle. This method uses the Gaussian asymptotic property of the discrete Fourier transform of the second order stationary observations [17–19].

There are multiple potential applications of blind source separation using second order statistics. Among others, one may cite the work of G. Chabriel et al. who suggest in [20] a method for the modeling and the identification of mixtures of multiple propagating waves recorded by a compact set of sensors. These mixtures, depending on attenuation coefficients and propagating delays, are represented as instantaneous mixtures of different temporal derivatives of sources generating the waves. It is shown that separation can be achieved by a second-order statistical analysis of the recordings, when a sufficient number of sensors is available. A. Ziehe et al. in [21] use second order BSS technique for cleaning biomagnetic measurements of evoked responses in the peripheral nervous system. M. Valkama et al. [22] present an application to BSS using second order statistics in radio communications where antenna arrays receive mixtures of different communication signals.

Besides the methodology studies, the algorithmic derivations and the different BSS applications, many authors focused on the second order identifiability problem. L. Tong et al. have proved in [8] that the sources are blindly separable using all correlation matrices, i.e. $\{\boldsymbol{R}_x(\tau)|\tau \in \mathbb{Z}\}$ only if they have different spectral shapes. Now, in practice, one uses only few (finite set) cor-

relation matrices for the source separation. For the method in [6] (resp. [7,8]) the source separation is possible only if the non-zero generalized eigenvalues of $\{\boldsymbol{R}_x(0),\ \boldsymbol{R}_x(1)\}$ (resp. $\{\boldsymbol{R}_x(0),\ (\boldsymbol{R}_x(\tau_k) + \boldsymbol{R}_x^\star(\tau_k))/2\}$) are all distinct.
In [3], the authors have shown that BSS using SOBI algorithm with the correlation matrix set $\{\boldsymbol{R}_x(\tau_1), \cdots, \boldsymbol{R}_x(\tau_K)\}$ is possible only if the vectors $\boldsymbol{\rho}_i = [\rho_i(\tau_1), \cdots, \rho_i(\tau_K)]$, $i = 1, \cdots, m$ are pairwise linearly independent. This result has been generalized in [23] to establish a necessary and sufficient identifiability condition with a finite set of correlation matrices. It is recalled and further developed in section 3.

3 Second Order Identifiability

3.1 Necessary and sufficient conditions of identifiability

In [8], Tong *et al.* showed that sources are blindly separable based on (the whole set) of second order statistics only if they have different spectral density functions. In practice we achieve the BSS using only a finite set of correlation matrices. Therefore, the previous identifiability result was generalized to that case in [3, 23] leading to the necessary and sufficient identifiability conditions given by the following theorem :

Theorem 1 *Let $\tau_1 < \tau_2 < \cdots < \tau_K$ be $K \geq 1$ time lags, and define $\boldsymbol{\rho}_i = [\rho_i(\tau_1), \rho_i(\tau_2), \cdots, \rho_i(\tau_K)]$ and $\tilde{\boldsymbol{\rho}}_i = [\Re e(\boldsymbol{\rho}_i), \Im m(\boldsymbol{\rho}_i)]$ where $\Re e(x)$ and $\Im m(x)$ denote the real part and imaginary part of x, respectively. Taking advantage of the indetermination, we assume without loss of generality that the sources are scaled such that $\|\boldsymbol{\rho}_i\| = \|\tilde{\boldsymbol{\rho}}_i\| = 1$, for all i* [1] *. Then, BSS can be achieved*

[1] We implicitly assume here that $\boldsymbol{\rho}_i \neq 0$, otherwise the source signal could not

using the output correlation matrices at time lags $\tau_1, \tau_2, \cdots, \tau_K$ *if and only if for all* $1 \leq i \neq j \leq m$ *:*

$$\tilde{\boldsymbol{\rho}}_i \text{ and } \tilde{\boldsymbol{\rho}}_j \text{ are (pairwise) linearly independent} \tag{8}$$

Proof: The proof of sufficiency is given in Appendix A. Here we here only prove that (8) is necessary to achieve BSS using correlation functions $\boldsymbol{R}_x(\tau_k) \stackrel{\text{def}}{=} E(\boldsymbol{x}(t+\tau_k)\boldsymbol{x}^\star(t))$, $k = 1, \cdots, K$. In fact, if two sources, say s_1 and s_2, have correlation coefficients such that $\tilde{\boldsymbol{\rho}}_1 = \epsilon\tilde{\boldsymbol{\rho}}_2$ where $\epsilon = \pm 1$, then any 'virtual' signal of the form $\tilde{\boldsymbol{x}}(t) = \tilde{\boldsymbol{A}}\tilde{\boldsymbol{s}}(t) + \boldsymbol{w}(t)$ where $\tilde{\boldsymbol{A}} = [\tilde{\boldsymbol{a}}_1, \tilde{\boldsymbol{a}}_2, \boldsymbol{a}_3, \cdots, \boldsymbol{a}_m]$ and $\tilde{\boldsymbol{s}}(t) = [\tilde{s}_1(t), \tilde{s}_2(t), s_3(t), \cdots, s_m(t)]^T$ with

$$[\tilde{\boldsymbol{a}}_1, \tilde{\boldsymbol{a}}_2] = [\boldsymbol{a}_1, \boldsymbol{a}_2]\boldsymbol{T}$$

and

$$\begin{bmatrix} \tilde{s}_1(t) \\ \\ \tilde{s}_2(t) \end{bmatrix} = \boldsymbol{T}^{-1} \begin{bmatrix} s_1(t) \\ \\ s_2(t) \end{bmatrix} \tag{9}$$

with

$$\boldsymbol{T} = \begin{bmatrix} \cos\theta & \sin\theta \\ \\ -\sin\theta & \cos\theta \end{bmatrix}, \quad \text{if} \quad \epsilon = +1$$

$$\boldsymbol{T} = \begin{bmatrix} \cosh\theta & \sinh\theta \\ \\ \sinh\theta & \cosh\theta \end{bmatrix}, \quad \text{if} \quad \epsilon = -1$$

---

be detected (and a fortiori could not be estimated) from the considered set of correlation matrices. This hypothesis will be held in the sequel.

verifies $\boldsymbol{R}_{\tilde{x}}(\tau_k) = \boldsymbol{R}_x(\tau_k)$ and $\boldsymbol{S}(\tau_k) = \tilde{\boldsymbol{S}}(\tau_k)$ for $k = 1, \cdots, K$, where $\tilde{\boldsymbol{S}}(\tau_k) \stackrel{\text{def}}{=} E(\tilde{\boldsymbol{s}}(t+\tau_k)\tilde{\boldsymbol{s}}^{\star}(t))$. $\square$

Interestingly, we can see from condition (8) that BSS can be achieved from only one correlation matrix $\boldsymbol{R}_x(\tau_k)$ provided that the vectors $[\Re e(\rho_i(\tau_k)), \Im m(\rho_i(\tau_k))]$ and $[\Re e(\rho_j(\tau_k)), \Im m(\rho_j(\tau_k))]$ are pairwise linearly independent for all $i \neq j$.

Note also that from (8), BSS can be achieved if at most one temporally white source signal exists. Similarly, recall that when using higher order statistics, BSS can only be achieved if at most one Gaussian source signal exists.

Under the condition of Theorem 1, the BSS can be achieved by decorrelation according to the following result:

Theorem 2 *Let $\tau_1, \tau_2, \cdots, \tau_K$ be K time lags and $\boldsymbol{z}(t) = [z_1(t), \cdots, z_m(t)]^T$ be an $m \times 1$ vector given by $\boldsymbol{z}(t) = \boldsymbol{B}\boldsymbol{x}(t)$. Define $r_{ij}(\tau_k) \stackrel{\text{def}}{=} E(z_i(t+\tau_k)z_j^*(t))$. If the identifiability condition holds, then $\boldsymbol{B}$ is a separating matrix (i.e. $\boldsymbol{B}\boldsymbol{y}(t) = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{s}(t)$ for a given permutation matrix $\boldsymbol{P}$ and a non-singular diagonal matrix $\boldsymbol{\Lambda}$) if and only if*

$$r_{ij}(\tau_k) = 0 \quad \textit{and} \quad \sum_{k=1}^{K} |r_{ii}(\tau_k)| > 0 \tag{10}$$

for all $1 \leq i \neq j \leq m$ and $k = 1, 2, \cdots, K$.

Proof: see the proof of Theorem 3 in Appendix A.

Note that, if one of the time lags is zero, the result of Theorem 2 holds only under the noiseless assumption. In that case, we can replace the condition $\sum_{k=1}^{K} |r_{ii}(\tau_k)| > 0$ by $r_{ii}(0) > 0$, for $i = 1, \cdots, m$. On the other hand, if all the time lags are non-zero and if the noise is temporally white (but can be spatially colored with unknown spatial covariance matrix) then the above result holds

without the noiseless assumption.

3.2 Partial Identifiability

It is generally believed that when the identifiability conditions are not met, the BSS cannot be achieved. This is only half the truth as it is possible to partially separate the sources in the sense that we can extract those which satisfy the identifiability conditions. More precisely, the sources can be separated in blocks each of them containing a mixture of sources that are not separable using the considered set of statistics. For example, consider a mixture of 3 sources such that $\tilde{\boldsymbol{\rho}}_1 = \tilde{\boldsymbol{\rho}}_2$ while $\tilde{\boldsymbol{\rho}}_1$ and $\tilde{\boldsymbol{\rho}}_3$ are linearly independent. In that case, source s_3 can be extracted while sources s_1 and s_2 cannot. In other words, by decorrelating the observed signal at the considered time lags, one obtain 3 signals one of them being s_3 (up to a scalar constant) and the two others are linear mixtures of s_1 and s_2.

This result can be mathematically formulated as follows: assume there are d distinct groups of sources each of them containing d_i source signals with same (up to a sign) correlation vector $\tilde{\boldsymbol{\rho}}_i$, $i = 1, \cdots, d$ (clearly, $m = d_1 + \cdots + d_d$). The correlation vectors $\tilde{\boldsymbol{\rho}}_1, \cdots, \tilde{\boldsymbol{\rho}}_d$ are pairwise linearly independent. We write $\boldsymbol{s}(t) = [\boldsymbol{s}_1^T(t), \cdots, \boldsymbol{s}_d^T(t)]^T$ where each sub-vector $\boldsymbol{s}_i(t)$ contains the d_i source signals with correlation vector $\tilde{\boldsymbol{\rho}}_i$.

Theorem 3 *Let $\boldsymbol{z}(t) = \boldsymbol{B}\boldsymbol{x}(t)$ be an $m \times 1$ random vector satisfying equation (10) for all $1 \leq i \neq j \leq m$ and $k = 1, \cdots, K$. Then, there exists a permutation matrix $\boldsymbol{P}$ such that $\overline{\boldsymbol{z}}(t) \stackrel{\text{def}}{=} \boldsymbol{P}\boldsymbol{z}(t) = [\overline{\boldsymbol{z}}_1^T(t), \cdots, \overline{\boldsymbol{z}}_d^T(t)]^T$ where $\overline{\boldsymbol{z}}_i(t) = \boldsymbol{U}_i \boldsymbol{s}_i(t)$, $\boldsymbol{U}_i$ being a $d_i \times d_i$ non-singular matrix. Moreover, sources belonging to the same group, i.e., having same (up to a sign) correlation vector $\tilde{\boldsymbol{\rho}}_i$ can not be*

separated using only the correlation matrices $\boldsymbol{R}_x(\tau_k),\ k = 1, \cdots, K$.

This result shows that when some of the sources have same (up to a sign) correlation vectors then the best that can be done is to separate them per blocks and this can be achieved by decorrelation. However, this result would be useless if we cannot check the linear dependency of the correlation vectors $\tilde{\boldsymbol{\rho}}_i$ and partition the signals per groups (as shown above) according to their correlation vectors. This leads us to the important problem of testing the identifiability condition that is discussed next.

4 Identifiability testing

4.1 Theoretical result

The necessary and sufficient identifiability condition (8) depends on the correlation coefficients of the source signals. The latter being unknown, it is therefore impossible to *a priori* check whether the sources are 'separable' or not from a given set of output correlation matrices. However, it is possible to check *a posteriori* whether the sources have been 'separated' or not. We have the following result:

Theorem 4 *Let* $\tau_1 < \tau_2 < \cdots < \tau_K$ *be* K *distinct time lags and* $\boldsymbol{z}(t) = \boldsymbol{B}\boldsymbol{x}(t)$. *Assume that* $\boldsymbol{B}$ *is a matrix such that* $\boldsymbol{z}(t)$ *satisfies*[2] *equation* (10) *for all* $1 \leq i \neq j \leq m$ *and* $k = 1, \cdots, K$. *Then there exists a permutation matrix* $\boldsymbol{P}$

[2] Because of the inherent indetermination of the BSS problem, we assume without loss of generality that the exact and estimated sources are similarly scaled, i.e., $\|\tilde{\boldsymbol{\rho}}_i\| = 1$.

such that for $k = 1, \cdots, K$.

$$E(\boldsymbol{z}(t+\tau_k)\boldsymbol{z}^{\star}(t)) = \boldsymbol{P}^T \boldsymbol{S}(k) \boldsymbol{P} \tag{11}$$

In other words the entries of $\overline{\boldsymbol{z}}(t) \stackrel{\text{def}}{=} \boldsymbol{P}\boldsymbol{z}(t)$ *have the same correlation coefficients as those of* $\boldsymbol{s}(t)$ *at time lags* $\tau_1, \cdots, \tau_K$, *i.e.* $E(\overline{z}_i(t+\tau_k)\overline{z}_i^*(t)) = \rho_i(\tau_k)$ *for* $k = 1, \cdots, K$ *and* $i = 1, \cdots, m$.

Proof: see Appendix B.

From Theorem 4, the existence of condition (8) can be checked by using the approximate correlation coefficients $r_{ii}(\tau_k) \stackrel{\text{def}}{=} E(z_i(t+\tau_k)z_i^*(t))$. It is important to point out that even if equation (10) holds, it does not mean that the source signals have been separated. Three situations may happen:

(1) For all pairs (i, j), $\tilde{\boldsymbol{\rho}}_i$ and $\tilde{\boldsymbol{\rho}}_j$ (computed from $\boldsymbol{z}(t)$) are pairwise linearly independent. Then we are sure that the sources have been separated and that $\boldsymbol{z}(t) = \boldsymbol{s}(t)$ up to the inherent indeterminacies of the BSS problem. In fact, testing the identifiability condition is equivalent to pairwise testing the angles between $\tilde{\boldsymbol{\rho}}_i$ and $\tilde{\boldsymbol{\rho}}_j$ for all $1 \leq i \neq j \leq m$. The larger the angle between $\tilde{\boldsymbol{\rho}}_i$ and $\tilde{\boldsymbol{\rho}}_j$, the better the quality of source separation (see performance analysis in [3]).

(2) For all pairs (i, j), $\tilde{\boldsymbol{\rho}}_i$ and $\tilde{\boldsymbol{\rho}}_j$ are linearly dependent. Thus, the sources haven't been separated and $\boldsymbol{z}(t)$ is still a linear combination of $\boldsymbol{s}(t)$.

(3) A few pairs (i, j) out of all pairs satisfy $\tilde{\boldsymbol{\rho}}_i$ and $\tilde{\boldsymbol{\rho}}_j$ linearly dependent. Therefore the sources have been separated in blocks.

Now, having only one signal realization at hand, we propose to use a resampling technique to evaluate the statistics needed for the testing.

Note that in practice the source correlation coefficients are calculated from noisy finite sample data. Due to the joint effects of noise and finite sample size, it is impossible to obtain the exact source correlation coefficients to test the identifiability condition. The identifiability condition should be tested using a certain threshold α, i.e., decide that $\tilde{\boldsymbol{\rho}}_i$ and $\tilde{\boldsymbol{\rho}}_j$ are linearly independent if $\left||\tilde{\boldsymbol{\rho}}_i\tilde{\boldsymbol{\rho}}_j^T| - 1\right| > \alpha$.

To find α we use the fact that the estimation error of $\tilde{\boldsymbol{\rho}}_i\tilde{\boldsymbol{\rho}}_j^T$ is asymptotically Gaussian [3] and hence one can build the confidence interval of such a variable according to its variance. This algorithm can be summarized as follows:

(1) Estimate a demixing matrix $\boldsymbol{B}$ and $\boldsymbol{z}(t) \stackrel{\text{def}}{=} \boldsymbol{B}\boldsymbol{x}(t)$ using an existing second order decorrelation algorithm (e.g. SOBI [3]).

(2) For each component $z_i(t)$, estimate the corresponding normalized vector $\widetilde{\boldsymbol{\rho}}_i$.

(3) Calculate the scalar product $\hat{\mathcal{R}}(i,j) = |\tilde{\boldsymbol{\rho}}_i\tilde{\boldsymbol{\rho}}_j^T|$ for each pair (i,j).

(4) Estimate $\hat{\sigma}_{(i,j)}$ the standard deviation of $\hat{\mathcal{R}}(i,j)$ using resampling technique (see section 4.3).

(5) Choose $\alpha_{(i,j)}$ according to the confidence interval. e.g. to have a confidence interval equal to 99.7% we choose $\alpha_{(i,j)} = 3\hat{\sigma}_{(i,j)}$, and compare $|\hat{\mathcal{R}}(i,j)-1|$ to $\alpha_{(i,j)}$ to test whether sources i and j have been separated or not.

[3] More precisely, one can prove that the estimation error $\sqrt{T}\delta(\tilde{\boldsymbol{\rho}}_i\tilde{\boldsymbol{\rho}}_j^T)$ is asymptotically, i.e. for large sample size T, Gaussian with zero mean and finite variance.

In many signal processing applications one is interested in forming estimates of a certain number of unknown parameters of a random process, using a set of sample values. Further, one is interested in finding the sampling distribution of the estimators, so that the respective means, variances, and cumulants can be calculated, or in making some kind of probability statements with respect to the unknown true values of the parameters.

The bootstrap [24] was introduced by Efron [25] as an approach to calculate confidence intervals for parameters in circumstances where standard methods cannot be applied. The bootstrap has subsequently been used to solve many other problems that would be too complicated for traditional statistical analysis.

In simple words, the bootstrap does with a computer what the experimenter would do in practice, i.e. if it were possible: he or she would repeat the experiment. With the bootstrap, the observations are randomly reassigned, and the estimates recomputed. These assignments and recomputations are done hundreds or thousands of times and treated as repeated experiments.

The jackknife [26] is another resampling technique for estimating the standard deviation. As an alternative to the bootstrap, the jackknife method can be thought of as drawing T samples of size $(T-1)$ each without replacement from the original sample of size T [26].

Suppose we are given the sample $\mathcal{X} = \{X_1, X_2, \ldots, X_T\}$ and an estimate, $\hat{\vartheta}$, from $\mathcal{X}$. The jackknife method is based on the sample delete-one observation at a time,

$$\mathcal{X}^{(i)} = \{X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_T\}$$

for $i = 1, 2, \ldots, T$, called the jackknife sample. This i^{th} jackknife sample consists of the data set with the i^{th} observation removed. For each i^{th} jackknife sample, we calculate the i^{th} jackknife estimate, $\hat{\vartheta}^{(i)}$ of ϑ, $i = 1, 2, \ldots, T$. The jackknife estimate of the standard deviation of $\hat{\vartheta}$ is defined by

$$\hat{\sigma} = \sqrt{\frac{T-1}{T} \sum_{i=1}^{T} \left(\hat{\vartheta}^{(i)} - \frac{1}{T} \sum_{j=1}^{T} \hat{\vartheta}^{(j)} \right)^2} \tag{12}$$

The jackknife is computationally less expensive if T is less than the number of replicates used by the bootstrap for standard deviation estimation because it requires computation of $\hat{\vartheta}$ only for the n jackknife data sets. For example, if $L = 25$ resamples are necessary for standard deviation estimation with the bootstrap, and the sample size is $T = 10$, then clearly the jackknife would be computationally less expensive than the bootstrap. In order to test the separability of the estimated signals, we have used a jackknife method to estimate the variance of the scalar product quantities $\mathcal{R}(i, j)$ for $i, j = 1, 2, \ldots, m$. This is done according to the following steps:

(1) From each signal $\boldsymbol{z}_i = [z_i(0), \ldots, z_i(T-1)]^T$, generate T vectors such as $\boldsymbol{z}_i^{(j)} = [z_i(0), \ldots, z_i(j-1), z_i(j+1), \ldots, z_i(T-1)]^T$ and $j = 0, 1, \ldots, T-1$.

(2) For each vector $\boldsymbol{z}_i^{(j)}$, estimate the corresponding vector $\widetilde{\boldsymbol{\rho}}_i^{(j)}$.

(3) Estimate $\hat{\boldsymbol{\mathcal{R}}}$ such as its $(i, j)^{th}$ entry is

$$\hat{\mathcal{R}}(i, j) = \frac{1}{T} \sum_{k=0}^{T-1} \frac{\langle \widetilde{\boldsymbol{\rho}}_i^{(k)}, \widetilde{\boldsymbol{\rho}}_j^{(k)} \rangle}{\| \widetilde{\boldsymbol{\rho}}_i^{(k)} \| \| \widetilde{\boldsymbol{\rho}}_j^{(k)} \|}$$

where $\langle \cdot, \cdot \rangle$ denotes the scalar product and $\| \cdot \|$ is the euclidian norm.

(4) Estimate the standard deviation of $\hat{\mathcal{R}}(i, j)$ by:

$$\hat{\sigma}_{(i,j)} = \sqrt{\frac{T-1}{T} \sum_{k=0}^{T-1} \left(\frac{\langle \widetilde{\boldsymbol{\rho}}_i^{(k)}, \widetilde{\boldsymbol{\rho}}_j^{(k)} \rangle}{\| \widetilde{\boldsymbol{\rho}}_i^{(k)} \| \| \widetilde{\boldsymbol{\rho}}_j^{(k)} \|} - \frac{1}{T} \sum_{l=0}^{T-1} \frac{\langle \widetilde{\boldsymbol{\rho}}_i^{(l)}, \widetilde{\boldsymbol{\rho}}_j^{(l)} \rangle}{\| \widetilde{\boldsymbol{\rho}}_i^{(l)} \| \| \widetilde{\boldsymbol{\rho}}_j^{(l)} \|} \right)^2}$$

4.4 Discussion

Some useful comments are provided here to get more insight onto the considered testing method and its potential applications and extensions.

- The asymptotic performance analysis of SOBI derived in [3], shows that the separation performance of two sources s_i and s_j depends on the angle between their respective correlation vectors $\widetilde{\boldsymbol{\rho}}_i$ and $\widetilde{\boldsymbol{\rho}}_j$. Hence, measuring this angle gives a hint on the interference rejection level of the two considered sources.

  As a consequence, one can use the measure of this angle not only to test the separability of the two sources but also to guarantee a target (minimum) separation quality. Choosing the threshold $\alpha_{(i,j)}$ accordingly is an important issue that deserves further investigation.
- The testing method can be incorporated into a two stage separation procedure where the first stage consists in a second order decorrelation method (e.g. SOBI). The second stage would be an HOS-based separation method applied only when the testing indicates a failure of separation at the first step.
- In many practical situations, one might be interested in only one or few source signals. This is the case for example in the interference mitigation problem in [27] or in the power plants monitoring applications [28]. In this situation, the partial identifiability result is of high interest as it proves that the desired source signal can still be extracted even if a complete source separation cannot be achieved.

5 Separation algorithm using relative gradient

5.1 SOS-based separation criteria

Based on Theorem 2, we can define different objective functions for signal decorrelation. In [29], the following criterion [4] was used

$$G(\boldsymbol{z}) = \sum_{k=1}^{K} \log\det\left(\mathrm{diag}(\boldsymbol{R}_z(\tau_k))\right) - \log\det\left(\boldsymbol{R}_z(\tau_k)\right) \tag{13}$$

where $\mathrm{diag}(\boldsymbol{A})$ is the diagonal matrix obtained by zeroing the off diagonal entries of $\boldsymbol{A}$. Another criterion used in [5] is

$$G(\boldsymbol{z}) = \sum_{k=1}^{K} \sum_{1\le i<j\le m} [|r_{ij}(\tau_k)+r_{ji}(\tau_k)|^2+|r_{ij}(\tau_k)-r_{ji}(\tau_k)|^2]+\sum_{i=1}^{m}|\sum_{k=1}^{K}|r_{ii}(\tau_k)|-1|^2. \tag{14}$$

The last term in (14) is introduced to avoid trivial (zero-valued) solutions. Equations (13) and (14) are non-negative functions which are zero if and only if $\boldsymbol{R}_z(\tau_k) = E(\boldsymbol{z}(t+\tau_k)\boldsymbol{z}^\star(t))$ are diagonal for $k = 1, \cdots, K$ or equivalently if (10) is met. Hence, one can achieve the BSS through signal decorrelation by minimizing one of the previous cost functions.

5.2 Iterative Decorrelation Algorithm (IDA)

The separation criteria we have presented take the form:

$$\boldsymbol{B} \text{ is a separating matrix} \Longleftrightarrow G(\boldsymbol{z}(t)) = 0 \tag{15}$$

where $\boldsymbol{z}(t) = \boldsymbol{B}\boldsymbol{x}(t)$ and G is a given objective function. An efficient approach to solve (15) is the one proposed in [5, 18]. It is a block technique based on

[4] In that paper, only the case where $\tau_1 = 0$ was considered.

the processing of T received samples and consists of searching the zeros of the sample version of (15). Solutions are obtained iteratively in the form:

$$\boldsymbol{B}^{(p+1)} = (\boldsymbol{I} + \boldsymbol{\epsilon}^{(p)})\boldsymbol{B}^{(p)} \tag{16}$$

$$\boldsymbol{z}^{(p+1)}(t) = (\boldsymbol{I} + \boldsymbol{\epsilon}^{(p)})\boldsymbol{z}^{(p)}(t) \tag{17}$$

At iteration p, a matrix $\boldsymbol{\epsilon}^{(p)}$ is determined from a local linearization of $G(\boldsymbol{B}\boldsymbol{x}(t))$. It is an approximate Newton technique with the benefit that $\boldsymbol{\epsilon}^{(p)}$ can be very simply computed (no Hessian inversion) under the additional assumption that $\boldsymbol{B}^{(p)}$ is close to a separating matrix. This procedure is illustrated as follows:

Using (17), we have:

$$r_{ij}^{(p+1)}(\tau_k) = r_{ij}^{(p)}(\tau_k) + \sum_{q=1}^{m} \epsilon_{jq}^{*(p)} r_{iq}^{(p)}(\tau_k) + \sum_{l=1}^{m} \epsilon_{il}^{(p)} r_{lj}^{(p)}(\tau_k) + \sum_{l,q=1}^{m} \epsilon_{il}^{(p)} \epsilon_{jq}^{*(p)} r_{lq}^{(p)}(\tau_k) \tag{18}$$

where

$$r_{ij}^{(p)}(\tau_k) \stackrel{\text{def}}{=} E\left(z_i^{(p)}(t+\tau_k) z_j^{*(p)}(t)\right) \tag{19}$$

$$\approx \frac{1}{T-\tau_k} \sum_{t=1}^{T-\tau_k} z_i^{(p)}(t+\tau_k) z_j^{*(p)}(t) \tag{20}$$

Under the assumption that $\boldsymbol{B}^{(p)}$ is close to a separating matrix, it follows that

$$|\epsilon_{ij}^{(p)}| \ll 1$$

and

$$|r_{ij}^{(p)}(\tau_k)| \ll 1 \quad \text{for} \quad i \neq j$$

and thus, a first order approximation of $r_{ij}^{(p+1)}(\tau_k)$ is given by:

$$r_{ij}^{(p+1)}(\tau_k) \approx r_{ij}^{(p)}(\tau_k) + \epsilon_{ji}^{*(p)} r_{ii}^{(p)}(\tau_k) + \epsilon_{ij}^{(p)} r_{jj}^{(p)}(\tau_k) \tag{21}$$

similarly, we have:

$$r_{ji}^{(p+1)}(\tau_k) \approx r_{ji}^{(p)}(\tau_k) + \epsilon_{ij}^{*(p)} r_{jj}^{(p)}(\tau_k) + \epsilon_{ji}^{(p)} r_{ii}^{(p)}(\tau_k) \tag{22}$$

From (21) and (22), we have:

$$r_{ij}^{(p+1)}(\tau_k)+r_{ji}^{(p+1)}(\tau_k) \;\approx\; 2r_{jj}^{(p)}(\tau_k)\Re e(\epsilon_{ij}^{(p)})+2r_{ii}^{(p)}(\tau_k)\Re e(\epsilon_{ji}^{(p)})+(r_{ij}^{(p)}(\tau_k)+r_{ji}^{(p)}(\tau_k))$$

$$r_{ij}^{(p+1)}(\tau_k)-r_{ji}^{(p+1)}(\tau_k) \;\approx\; 2\jmath r_{jj}^{(p)}(\tau_k)\Im m(\epsilon_{ij}^{(p)})-2\jmath r_{ii}^{(p)}(\tau_k)\Im m(\epsilon_{ji}^{(p)})+(r_{ij}^{(p)}(\tau_k)-r_{ji}^{(p)}(\tau_k))$$

with $\jmath = \sqrt{-1}$. By replacing the previous equation into (14), we obtain the following least squares (LS) minimization problem

$$\min \left\| \left[\boldsymbol{r}_{jj}^{(p)},\ \boldsymbol{r}_{ii}^{(p)}\right] \boldsymbol{E}_{ij}^{(p)} + \left[\frac{1}{2}(\boldsymbol{r}_{ij}^{(p)} + \boldsymbol{r}_{ji}^{(p)}),\ \frac{1}{2\jmath}(\boldsymbol{r}_{ij}^{(p)} - \boldsymbol{r}_{ji}^{(p)})\right] \right\|$$

where

$$\boldsymbol{E}_{ij}^{(p)} \stackrel{\text{def}}{=} \begin{bmatrix} \Re e(\epsilon_{ij}^{(p)}) & \Im m(\epsilon_{ij}^{(p)}) \\ \Re e(\epsilon_{ji}^{(p)}) & -\Im m(\epsilon_{ji}^{(p)}) \end{bmatrix} \tag{23}$$

$$\boldsymbol{r}_{ij}^{(p)} = [r_{ij}^{(p)}(\tau_1), \cdots, r_{ij}^{(p)}(\tau_K)]^T \tag{24}$$

A solution to the LS minimization problem is given by:

$$\boldsymbol{E}_{ij}^{(p)} = -\left[\boldsymbol{r}_{jj}^{(p)}, \boldsymbol{r}_{ii}^{(p)}\right]^{\#} \left[\frac{1}{2}(\boldsymbol{r}_{ij}^{(p)} + \boldsymbol{r}_{ji}^{(p)}), \frac{1}{2\jmath}(\boldsymbol{r}_{ij}^{(p)} - \boldsymbol{r}_{ji}^{(p)})\right] \tag{25}$$

where $\boldsymbol{A}^{\#}$ denotes the pseudo-inverse of matrix $\boldsymbol{A}$. Equations (23) and (25) provide the explicit expression of $\epsilon_{ij}^{(p)}$ for $i \neq j$. For $i = j$, the minimization of (14) using the first order approximation leads to :

$$\left| \sum_{k=1}^{K} r_{ii}^{(p)}(\tau_k) \left(1 + 2\Re e(\epsilon_{ii}^{(p)})\right) \right| - 1 = 0. \tag{26}$$

Without loss of generality, we take advantage of the phase indeterminacy to assume that ϵ_{ii} are real-valued and hence $\Re e(\epsilon_{ii}) = \epsilon_{ii}$. Consequently, we obtain :

$$\epsilon_{ii}^{(p)} = \frac{1 - \sum_{k=1}^{K} |r_{ii}^{(p)}(\tau_k)|}{2\sum_{k=1}^{K} |r_{ii}^{(p)}(\tau_k)|} \tag{27}$$

In the case of real-valued signals, the LS minimization becomes :

$$\min \left\| \boldsymbol{H}_{ij}^{(p)} \boldsymbol{e}_{ij}^{(p)} + \boldsymbol{\psi}_{ij}^{(p)} \right\|$$

where

$$\boldsymbol{H}_{ij}^{(p)} = \begin{bmatrix} 1 \\ \\ 1 \end{bmatrix} \otimes \left[\boldsymbol{r}_{jj}^{(p)},\ \boldsymbol{r}_{ii}^{(p)} \right] \tag{28}$$

$$\boldsymbol{e}_{ij}^{(p)} = \left[\epsilon_{ij}^{(p)}, \epsilon_{ji}^{(p)} \right]^T \tag{29}$$

$$\boldsymbol{\psi}_{ij}^{(p)} = \begin{bmatrix} \boldsymbol{r}_{ij}^{(p)} \\ \\ \boldsymbol{r}_{ji}^{(p)} \end{bmatrix} \tag{30}$$

and $\otimes$ denotes the Kronecker product. A solution to the LS minimization problem is given by:

$$\boldsymbol{e}_{ij}^{(p)} = -\boldsymbol{H}_{ij}^{(p)\#} \boldsymbol{\psi}_{ij}^{(p)} \tag{31}$$

Remark: A main advantage of the above algorithm is its flexibility and easy implementation in the adaptive case. This is the focus of the next subsection. This algorithm can Also be extended to deal with BSS of convolutive mixtures as shown in [30].

5.3 Adaptive implementation

To derive an adaptive version of the above batch algorithm, we replace in the above formulae the iteration index p by the time index t and estimate adaptively the correlation coefficients $r_{ij}^{(t)}(\tau_k)$. The adaptive algorithm can be summarized as follows : At time instant t

- Update the correlation matrices, i.e., $\boldsymbol{R}_z(\tau_k)$, $k = 1, \ldots, K$, using the fol-

lowing averaging technique :

$$\boldsymbol{z}(t) = \boldsymbol{B}^{(t-1)}\boldsymbol{x}(t)$$

$$\boldsymbol{R}_z^{(t)}(\tau_k) = \lambda \boldsymbol{R}_z^{(t-1)}(\tau_k) + (1-\lambda)\boldsymbol{z}(t)\boldsymbol{z}^\star(t-\tau_k)$$

where $0 < \lambda < 1$ is a positive forgetting factor. Note that $r_{ij}^{(t)}(\tau_k)$ is the (i,j)-th entry of $\boldsymbol{R}_z^{(t)}(\tau_k)$.

- Estimate $\boldsymbol{\epsilon}^{(t)}$ using equations (25) and (27) and the updated correlation coefficients $r_{ij}^{(t)}(\tau_k)$.
- Update the value of the separating matrix, the correlation matrices $\boldsymbol{R}_z(\tau_k),\ k = 1, \ldots, K$, and the estimated sources $\boldsymbol{z}(t+1-\tau_k),\ k = 1, \ldots, K$:

$$\boldsymbol{B}^{(t)} = (\boldsymbol{I} + \boldsymbol{\epsilon}^{(t)})\boldsymbol{B}^{(t-1)}$$

$$\boldsymbol{R}_z^{(t)}(\tau_k) = (\boldsymbol{I} + \boldsymbol{\epsilon}^{(t)})\boldsymbol{R}_z^{(t)}(\tau_k)(\boldsymbol{I} + \boldsymbol{\epsilon}^{(t)})^\star$$

$$\boldsymbol{z}(t+1-\tau_k) = (\boldsymbol{I} + \boldsymbol{\epsilon}^{(t)})\boldsymbol{z}(t+1-\tau_k).$$

Besides its computational simplicity, this algorithm has the advantage of uniform performance (i.e. it has, in noiseless case, the same asymptotic performance whatever the mixing matrix is) and stability [31].

6 Performance analysis

6.1 Theoretical performance analysis

In this section, asymptotic (i.e. for large sample sizes) performance analysis results of the previous separation method is given. We consider the case of instantaneous mixture with i.i.d complex-valued sources satisfying, in addition

to the identifiability condition $\sum_{k\in\mathbb{Z}} |\rho_i(\tau_k)| < +\infty$ for $i = 1, \ldots, m$. The noise is assumed Gaussian with variance $\sigma^2 \boldsymbol{I}$. Assuming that the permutation indeterminacy is $\boldsymbol{P} = \boldsymbol{I}$, one can write:

$$\boldsymbol{B}\boldsymbol{A} = \boldsymbol{I} + \boldsymbol{\delta} \tag{32}$$

and hence, the separation quality is measured in our simulations from the mixing matrix $\boldsymbol{A}$ and the decorrelation matrix $\boldsymbol{B}$ using the mean rejection level criterion [3] defined as:

$$\mathcal{I}perf \stackrel{\text{def}}{=} \sum_{1\leq p\neq q\leq m} \frac{E\left(|(BA)_{pq}|^2\right)\rho_q(0)}{E\left(|(BA)_{pp}|^2\right)\rho_p(0)} \tag{33}$$

$$= \sum_{1\leq p\neq q\leq m} E\left(|\delta_{pq}|^2\right)\frac{\rho_q(0)}{\rho_p(0)} \tag{34}$$

Our performance analysis consists in deriving the closed-form expression of the asymptotical variance errors:

$$\lim_{T\to+\infty} T\ E\left(|\delta_{pq}|^2\right) \tag{35}$$

By using a similar approach to that in [18] based on the central-limit and continuity theorems, one obtains the following result :

Theorem 5 *Vector* $\boldsymbol{\delta}_{ij} \stackrel{\text{def}}{=} [\Re e(\delta_{ij})\ \Re e(\delta_{ji})\ \Im m(\delta_{ij})\ -\Im m(\delta_{ji})]^T$ *is asymptotically Gaussian distributed with asymptotic covariance matrix*

$$\boldsymbol{\Delta}_{ij} \stackrel{\text{def}}{=} \lim_{T\to+\infty} T\ E\left(\boldsymbol{\delta}_{ij}\boldsymbol{\delta}_{ij}^T\right) \tag{36}$$

$$= \frac{1}{4}\boldsymbol{\mathcal{H}}_{ij}^{\#}\boldsymbol{\Psi}_{ij}\boldsymbol{\mathcal{H}}_{ij}^{\#\star} \tag{37}$$

where

$$\boldsymbol{\mathcal{H}}_{ij} = \boldsymbol{I}_2 \otimes \left[\boldsymbol{\rho}_i^T\ \boldsymbol{\rho}_j^T\right] \tag{38}$$

$$\boldsymbol{\Psi}_{ij} = \begin{bmatrix} \boldsymbol{\Xi}_{11}^{(ij)} & \boldsymbol{\Xi}_{12}^{(ij)} \\ \\ \boldsymbol{\Xi}_{21}^{(ij)} & \boldsymbol{\Xi}_{22}^{(ij)} \end{bmatrix} \tag{39}$$

with

$$\boldsymbol{\Xi}_{11}^{(ij)} = \boldsymbol{\Gamma}_{11}^{(ij)} + \boldsymbol{\Gamma}_{12}^{(ij)} + \boldsymbol{\Gamma}_{21}^{(ij)} + \boldsymbol{\Gamma}_{22}^{(ij)} \tag{40}$$

$$\boldsymbol{\Xi}_{12}^{(ij)} = \jmath \left[\boldsymbol{\Gamma}_{11}^{(ij)} - \boldsymbol{\Gamma}_{12}^{(ij)} + \boldsymbol{\Gamma}_{21}^{(ij)} + \boldsymbol{\Gamma}_{22}^{(ij)} \right] \tag{41}$$

$$\boldsymbol{\Xi}_{21}^{(ij)} = -\jmath \left[\boldsymbol{\Gamma}_{11}^{(ij)} + \boldsymbol{\Gamma}_{12}^{(ij)} - \boldsymbol{\Gamma}_{21}^{(ij)} + \boldsymbol{\Gamma}_{22}^{(ij)} \right] \tag{42}$$

$$\boldsymbol{\Xi}_{22}^{(ij)} = \boldsymbol{\Gamma}_{11}^{(ij)} - \boldsymbol{\Gamma}_{12}^{(ij)} - \boldsymbol{\Gamma}_{21}^{(ij)} + \boldsymbol{\Gamma}_{22}^{(ij)} \tag{43}$$

and

$$\boldsymbol{\Gamma}_{11}^{(ij)}(k,k') = \sum_{\tau \in \mathbb{Z}} r_{ii}(\tau_k + \tau) r_{jj}(\tau_{k'} + \tau) \tag{44}$$

$$\boldsymbol{\Gamma}_{22}^{(ij)}(k,k') = \sum_{\tau \in \mathbb{Z}} r_{ii}(\tau_{k'} + \tau) r_{jj}(\tau_k + \tau) \tag{45}$$

$$\boldsymbol{\Gamma}_{12}^{(ij)}(k,k') = \sum_{\tau \in \mathbb{Z}} r_{ii}(\tau_k + \tau) r_{jj}(\tau_{k'} - \tau) \tag{46}$$

$$r_{ii}(\tau_k) = \rho_i(\tau_k) + \delta(\tau_k)\sigma^2 \boldsymbol{b}_i \boldsymbol{b}_i^T \ . \tag{47}$$

Note that $\boldsymbol{\Gamma}_{21}^{(ij)} = \boldsymbol{\Gamma}_{12}^{(ij)T}$ *and* $\boldsymbol{b}_i$ *represents the* i^{th} *row of* $\boldsymbol{B} = \boldsymbol{A}^{\#}$.

In the case of real-valued signals, the preceding result becomes :

$$\boldsymbol{\Delta}_{ij} = \boldsymbol{\mathcal{H}}_{ij}^{\#} \boldsymbol{\Psi}_{ij} \boldsymbol{\mathcal{H}}_{ij}^{\#T} \tag{48}$$

where

$$\boldsymbol{\mathcal{H}}_{ij} = \begin{bmatrix} 1 & \\ & \\ & 1 \end{bmatrix} \otimes \begin{bmatrix} \boldsymbol{\rho}_i^T & \boldsymbol{\rho}_j^T \end{bmatrix} \tag{49}$$

$$\boldsymbol{\Psi}_{ij} = \begin{bmatrix} \boldsymbol{\Gamma}_{11}^{(ij)} & \boldsymbol{\Gamma}_{12}^{(ij)} \\ \\ \boldsymbol{\Gamma}_{21}^{(ij)} & \boldsymbol{\Gamma}_{22}^{(ij)} \end{bmatrix} \tag{50}$$

Using the inverse Fourier transform relation (i.e. $\rho_i(\tau) = \int_{-\frac{1}{2}}^{\frac{1}{2}} S_i(f) e^{\jmath 2\pi\tau f} df$), the previous result can be rewritten by using the normalization assumption of Theorem 1 ($\|\boldsymbol{\rho}_i\| = 1$ for all i) as :

$$\boldsymbol{\Delta}_{ij} = \frac{1}{\left(\rho_{ij}^2 - 1\right)^2} \begin{bmatrix} 1 & -\rho_{ij} \\ -\rho_{ij} & 1 \end{bmatrix} \boldsymbol{\mathcal{D}} \begin{bmatrix} 1 & -\rho_{ij} \\ -\rho_{ij} & 1 \end{bmatrix} \tag{51}$$

where

$$\rho_{ij} = \boldsymbol{\rho}_i \boldsymbol{\rho}_j^T \tag{52}$$

and

$$\boldsymbol{\mathcal{D}} = \int_{-\frac{1}{2}}^{\frac{1}{2}} S_i(f) S_j(f) \boldsymbol{\mathcal{V}}_{ij} \boldsymbol{\mathcal{V}}_{ij}^T df \ . \tag{53}$$

Note that

$$\boldsymbol{\mathcal{V}}_{ij} = \left[\Re e\left(\varrho_i(f)\right) \quad \Re e\left(\varrho_j(f)\right)\right]^T \ , \tag{54}$$

with

$$\varrho_l(f) = \sum_{k=1}^{K} \rho_l(\tau_k) \exp(-\jmath 2\pi\tau_k f) \quad \text{for } l \in \{i, j\} \tag{55}$$

and

$$S_l(f) = S_{s_l}(f) + \sigma^2 \|\boldsymbol{b}_l\|^2, \quad \text{for } l \in \{i, j\} \tag{56}$$

where $S_{s_l}(f)$ is the power spectral density of the l^{th} source. By replacing (56) in (53), we obtain :

$$\begin{aligned} \boldsymbol{\mathcal{D}} = & \int_{-\frac{1}{2}}^{\frac{1}{2}} S_{s_i}(f) S_{s_j}(f) \boldsymbol{\mathcal{V}}_{ij} \boldsymbol{\mathcal{V}}_{ij}^T df + \sigma^4 \|\boldsymbol{b}_i\|^2 \|\boldsymbol{b}_j\|^2 \int_{-\frac{1}{2}}^{\frac{1}{2}} \boldsymbol{\mathcal{V}}_{ij} \boldsymbol{\mathcal{V}}_{ij}^T df \\ & + \sigma^2 \|\boldsymbol{b}_i\|^2 \int_{-\frac{1}{2}}^{\frac{1}{2}} S_{s_j}(f) \boldsymbol{\mathcal{V}}_{ij} \boldsymbol{\mathcal{V}}_{ij}^T df + \sigma^2 \|\boldsymbol{b}_j\|^2 \int_{-\frac{1}{2}}^{\frac{1}{2}} S_{s_i}(f) \boldsymbol{\mathcal{V}}_{ij} \boldsymbol{\mathcal{V}}_{ij}^T df \end{aligned}$$

From the above expression, we notice that in the noiseless case the performance of the considered BSS method is independent from the mixture matrix (i.e. equivariance property). In that case, the performance limit is essentially

function of the "non-collinearity" of vectors $\boldsymbol{\rho}_i$ and $\boldsymbol{\rho}_j$ (one can see it mainly from the term $\left(\rho_{ij}^2 - 1\right)^2$ that appears in the denominator of equation (51)).

6.2 Simulation-based performance analysis

We present in this section some numerical simulations to evaluate the performance of the previous separation algorithm. We consider in our simulation an array of $n = 4$ sensors with half wavelength spacing, receiving two signals in the presence of stationary real temporally white noise. The two source signals are generated by filtering real white Gaussian processes by an autoregressive (AR) model of order 1 with coefficient $a_1 = 0.95e^{\jmath 0.5}$ and $a_2 = 0.5e^{\jmath 0.7}$ (except for Figure 6). The sources have direction of arrivals (DOA) $\phi_1 = 30$ and $\phi_2 = 45$ degrees respectively. Note that the simulation results shown here do not depend on this explicit choice of the source DOAs but rather on the angle difference $\phi_2 - \phi_1$, as illustrated by Figure 7. The number of time lags is $K = 5$ (except for Figure 8). The signal to noise ratio is defined as $\mathrm{SNR} = 10 \log_{10} \frac{\sigma_s^2}{\sigma_n^2}$, where σ_n^2 and σ_s^2 are the noise variance and signal variance respectively. The mean rejection level is estimated over 1000 Monte-Carlo runs.

In Figure 1, we compare the separation performance obtained by the decorrelation algorithm with SOBI algorithm.

Figure 2 shows the mean rejection levels against the signal to noise ratio SNR. We compare the IDA algorithm with the SOBI algorithm which is based on a joint diagonalization of a set of covariance matrices [3]. The additive noise is temporally white but spatially colored. The noise covariance is assumed to be of the form $E(\boldsymbol{\eta}(t)\boldsymbol{\eta}^\star(t)) = n\sigma_n^2 \boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{Q}}^H / \|\boldsymbol{\mathcal{Q}}\|^2$, where $\boldsymbol{\mathcal{Q}}$ is given by $\boldsymbol{\mathcal{Q}}_{ij} =$

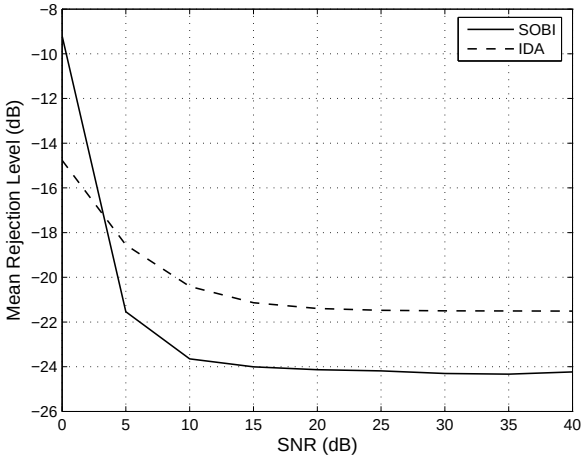


Fig. 1. Comparison between IDA and SOBI for spatially white noise: Mean Rejection Level in dB versus SNR for 2 autoregressive sources 4 sensors and $T = 1000$.

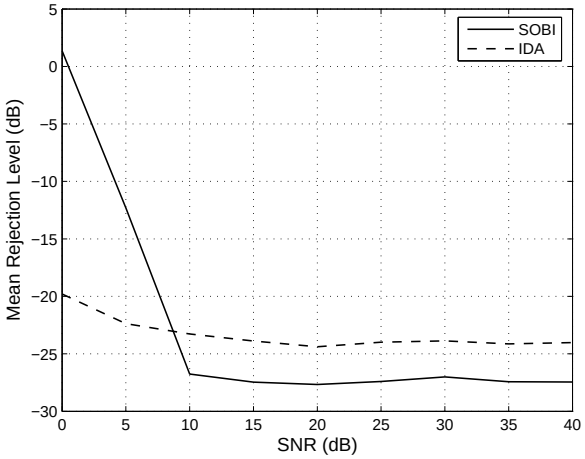


Fig. 2. Comparison between IDA and SOBI for spatially colored noise: Mean Rejection Level in dB versus SNR for 2 autoregressive sources $T = 1000$.
$0.9^{|i-j|}$ (i.e. $\boldsymbol{\eta}(t) = \sqrt{n}\sigma_n \frac{\mathcal{Q}}{\|\mathcal{Q}\|}\overline{\boldsymbol{\eta}}(t)$ where $\overline{\boldsymbol{\eta}}(t)$ is a unit norm white Gaussian noise). In this case, IDA performs much better than SOBI at low SNR. This is an important advantage of the IDA method over existing SOS methods that often assume the noise covariance matrix known up to a scalar constant.

In Figure 3, The mean rejection level $\mathcal{I}perf$ is plotted in dB against the sample size. The figure is for SNR = 40dB. This figure shows that the asymp-

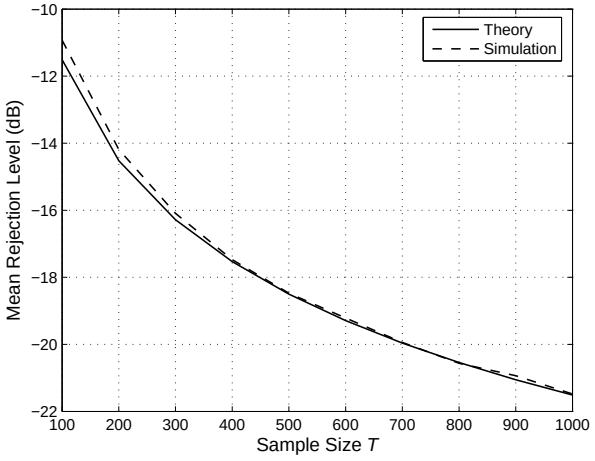
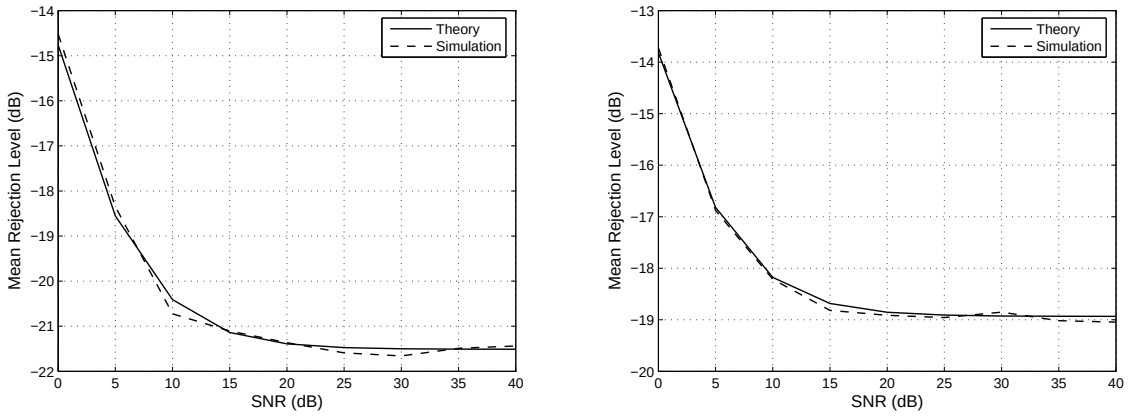


Fig. 3. Mean Rejection Level in dB versus the sample size T for 2 autoregressive sources 4 sensors and SNR = 40dB.

totic closed form expressions of the rejection level are pertinent from snapshot length of about 100 samples. In the plots $E\left(|\delta_{ij}|^2\right)$ and $E\left(|\delta_{ji}|^2\right)$ are replaced by $\frac{\mathbf{\Delta}_{ij}(1,1)+\mathbf{\Delta}_{ij}(3,3)}{T}$ and $\frac{\mathbf{\Delta}_{ij}(2,2)+\mathbf{\Delta}_{ij}(4,4)}{T}$ respectively. This means that asymptotic conditions are reached for short data block size.



(a) (b)

Fig. 4. Mean Rejection Level in dB versus the SNR for $T = 1000$: (a) 4 sensors and 2 AR sources with DOAs $\phi_1 = 30$ and $\phi_2 = 45$ degrees; (b) 7 sensors and 5 AR sources with DOAs $\phi_1 = 15$, $\phi_2 = 30$, $\phi_3 = 45$, $\phi_4 = 60$ and $\phi_5 = 120$ degrees respectively.

Figure 4 shows the mean rejection level against the signal to noise ratio SNR. We compare the empirical performance with theoretical performance for $T = 1000$ sample size.

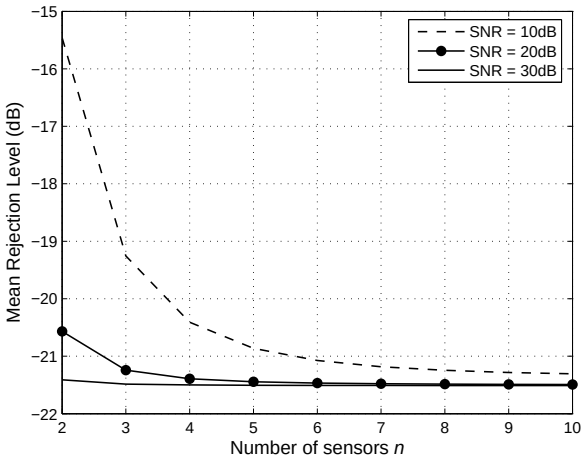


Fig. 5. Mean Rejection Level in dB versus the number of sensors n for 2 autoregressive sources and T=1000.

Figure 5 shows the mean rejection level versus the number of sensors using the theoretical formulation for $T = 1000$ sample size. We observe that, the more the number of sensors, the lower the rejection level is in the low SNR case. For high SNRs the number of sensors has negligible effect on the separation performance (in accordance with the uniform performance property).

Figure 6 shows $\mathcal{I}perf$ versus the spectral shift $\delta\theta$. The spectral shift $\delta\theta$ represents the spectral overlap of the two sources. In this figure, the noise is assumed to be spatially white and its level is kept constant at 10dB and 30dB. We let $a_1 = 0.7e^{\jmath 0.5}$ and $a_2 = 0.5e^{\jmath(0.5+\delta\theta)}$. The plot evidences a significant increase in rejection performance by increasing $\delta\theta$.

Figure 7 assesses the performance of IDA versus the angle difference $\delta\phi = \phi_2 - \phi_1$ for different values of the SNR. The number of sensors is $n = 4$, the

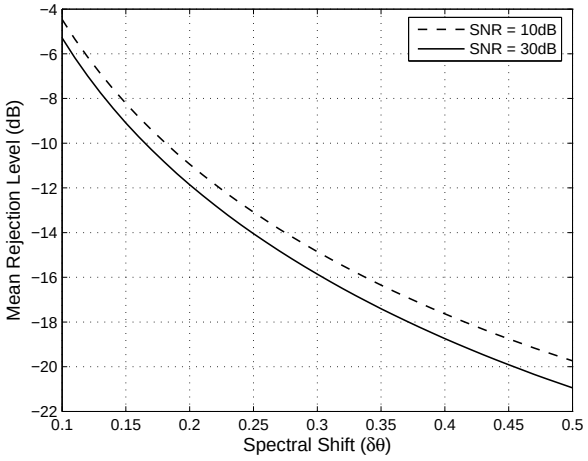


Fig. 6. Mean Rejection Level in dB versus the spectral shift $\delta\theta$ for 2 autoregressive sources, 4 sensors and T=1000.

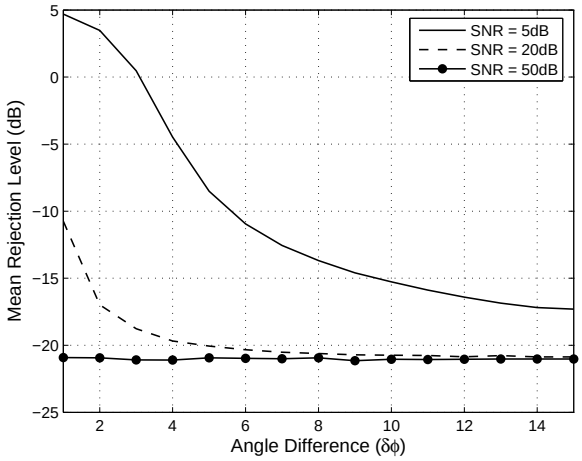


Fig. 7. Mean Rejection Level in dB versus the angle difference $\delta\phi$ for 2 autoregressive sources, 4 sensors and T=1000.

sample size $T = 1000$, two AR sources with coefficient $a_1 = 0.95e^{j0.5}$ and $a_2 = 0.5e^{j0.7}$ are considered. Their respective DOAs are $\phi_1 = 30$ degrees and $\phi_2 = \phi_1 + \delta\phi$. From the plots, we observe a significant performance degradation when $\delta\phi$ is close to zero. Indeed in that case, the identifiability condition w.r.t. the mixture matrix $\boldsymbol{A}$ is ill-satisfied. Also, we can observe that in the absence of noise or equivalently when the noise is negligible (e.g. for SNR=50dB), the

method has uniform performance independent from the mixture matrix $\boldsymbol{A}$ and its numerical conditioning.

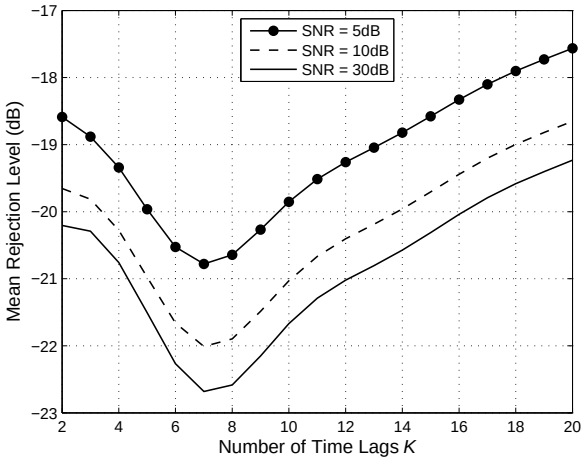


Fig. 8. Mean Rejection Level in dB versus the number of time lags K for 2 autoregressive sources, 4 sensors and T=1000.

The plots in Figure 8 illustrate the effect of the number of time lags K for different SNRs. In this simulation the sources arrive from the directions $\phi_1 = 10$ and $\phi_2 = 13$ degrees. This choice is to have an ill-conditioned mixture matrix and hence a difficult separation context. In that case, the increase of the number of correlation matrices is needed to improve the separation quality. Otherwise, in a 'good' context, increasing the number of time lags would not significantly affect the performance of the considered algorithm. From this figure, we also observe that a large increase of the number of correlation matrices leads to a degradation of the separation performance. The reason is that, the correlation coefficients of considered sources decrease exponentially towards zero and consequently the signal term in the large time lags correlation matrices is negligible and their estimations are too noisy.

6.3 Performance assessment of the testing technique

We present in this section some simulation results to illustrate the performance of the separability testing method. In the simulated environment we consider uniform linear array with $n = 2$ sensors receiving the signals from $m = 2$ unit-power first order autoregressive sources (with coefficients $a_1 = 0.95e^{\jmath 0.5}$ and $a_2 = 0.5e^{\jmath 0.7}$) in the presence of stationary complex temporally white noise. The considered sources are separable according to the identifiability result, i.e. their respective correlation vectors $\widetilde{\boldsymbol{\rho}}_1$ and $\widetilde{\boldsymbol{\rho}}_2$ are linearly independent. The time lags (delays) implicitly involved are $\tau_0, \cdots, \tau_9$ (i.e., $K = 10$). We use SOBI algorithm [3] to obtain the decorrelated sources. The statistics in the curves are evaluated over 2000 Monte-Carlo runs.

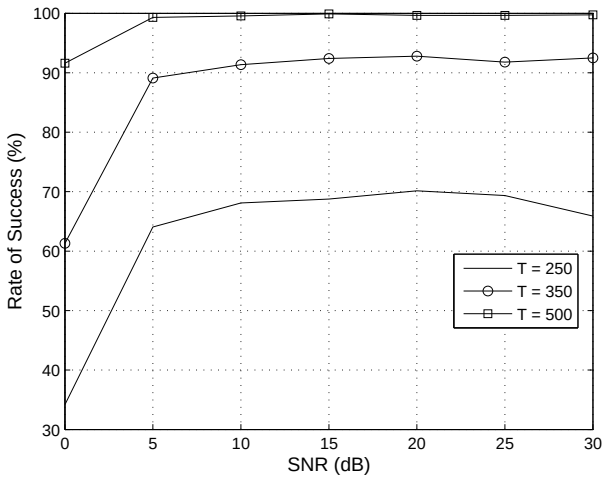


Fig. 9. Rate of success versus SNR for 2 autoregressive sources and 2 sensors and $\beta = 99.7\%$: comparison of the performance of our testing algorithm for different sample sizes T.

We first present in Figure 9 a simulation example where we compare the rate of success of the testing procedure (success means that we decide the 2 sources

have been separated) to detect the sources separability for different sample sizes versus the SNR in dB. The confidence interval is fixed to $\beta = 99.7\%$. One can observe from this figure that the performance of the testing procedure degrades significantly for a small sample size due to the increased estimation errors and the fact that we use the asymptotic normality of considered statistics.

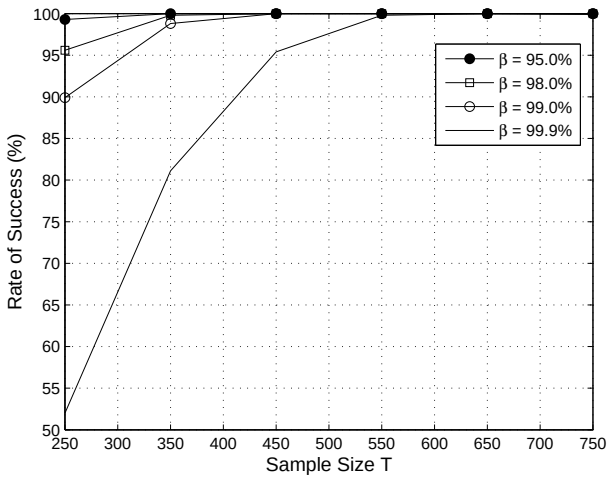


Fig. 10. Rate of success versus sample size T for 2 autoregressive sources and 2 sensors and SNR = 25dB: comparison of the performance of our algorithm for different confidence interval β.

In Figure 10, we present a simulation example where we compare the rate of success according to the sample size for different confidence intervals. The SNR is set to 25dB. Clearly, the lower the confidence interval is, the higher is the rate of success of the testing procedure. Also, as observed in Figure 9, the rate of success increases rapidly when increasing the sample size.

In Figure 11, we present a simulation example where we plot the rate of success versus the confidence interval β for different sample sizes and for SNR = 25dB. This plot shows somehow the evolution of the rate of success w.r.t. the 'false

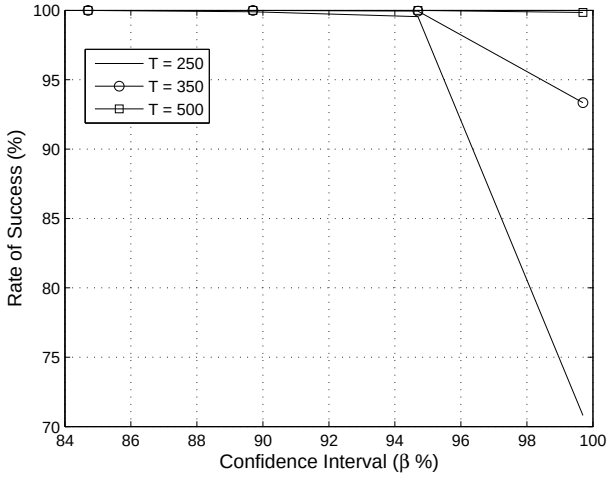


Fig. 11. Rate of false alarm versus confidence interval β for 2 autoregressive sources and 2 sensors and SNR = 25dB: comparison of the performance of our algorithm for different sample size T.

alarm rate' and confirms the results of the two previous figures.

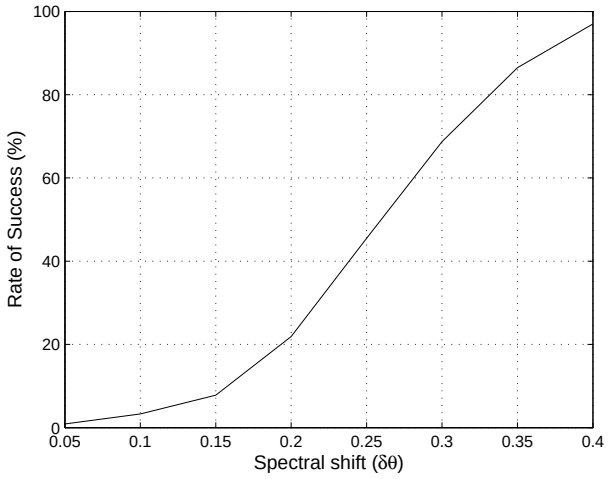


Fig. 12. Rate of success versus spectral shift $\delta\theta$ for 2 autoregressive sources and 5 sensors and SNR = 25dB.

The simulation example presented in Figure 12 assumes two source signals with parameters $a_1 = 0.5e^{\jmath 0.5}$ and $a_2 = 0.5e^{\jmath(0.5+\delta\theta)}$, where $\delta\theta$ represents the spectral overlap of the two sources. The number of sensors is $n = 5$, the sample

size is $T = 1000$ and the SNR = 25dB. Figure 12 shows the rate of success versus the spectral shift $\delta\theta$. As we can see, small values of $\delta\theta$ lead to high rates of 'non-separability' decision by our testing procedure. Indeed, when $\delta\theta$ is close to zero the two vectors $\widetilde{\boldsymbol{\rho}}_1$ and $\widetilde{\boldsymbol{\rho}}_2$ are close to 'linear dependency'. That means that the separation quality of the two sources is poor in that case which explains the observed testing results.

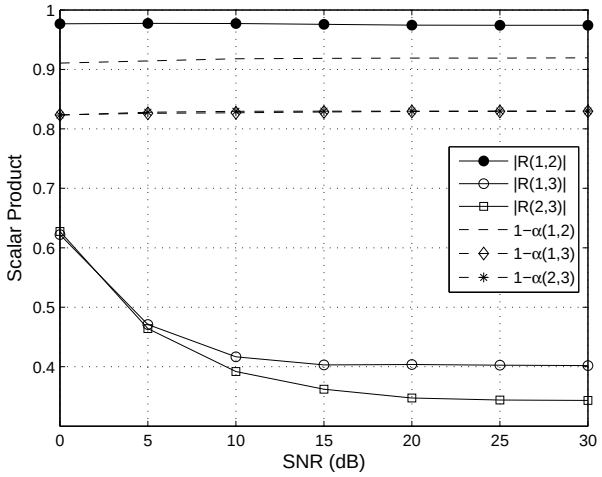


Fig. 13. Average values of the $|\mathcal{R}(i,j)|$ and thresholds $1-\alpha_{(i,j)}$ versus SNR for 3 sources and 3 sensors : 2 sources are complex white Gaussian processes and the third one is an autoregressive signal.

In the last figure, we assume there exist three sources. The first two sources are complex white Gaussian processes (hence $\widetilde{\boldsymbol{\rho}}_1 = \widetilde{\boldsymbol{\rho}}_2$) and the third one is an autoregressive signal with coefficient $a_3 = 0.95e^{j0.5}$. The plots in Figure 13 compares the average values of scalar products for $\widetilde{\boldsymbol{\rho}}_i$ and $\widetilde{\boldsymbol{\rho}}_j$ $(i,j = 1,2,3)$ with their corresponding threshold values $1-\alpha_{(i,j)}$ versus the SNR. The sample size is fixed to $T = 500$ and the number of sensors is $n = 3$. This example illustrates the situation where two of the sources (here sources 1 and 2) cannot be separated (this is confirmed by the testing result) while the third one

is extracted correctly (the plots show clearly that $\mathcal{R}(1,3) < 1 - \alpha_{(1,3)}$ and $\mathcal{R}(2,3) < 1 - \alpha_{(2,3)}$).

7 Conclusion

The second order blind source separation of stationary colored sources is a simple decorrelation problem. The signal decorrelation for a finite set of time lags leads to source separation under certain conditions that are fully detailed in this paper. The separability conditions depend on the unknown sources and hence cannot be verified (tested) directly. However, we present in this paper a testing procedure that uses the output signals of the decorrelator to verify a posteriori whether the sources have been correctly separated or not. The signal decorrelation can be achieved in many different ways. We have presented in this work, one of them that has the advantage of simplicity and efficiency both in block and adaptive processing. Performance analysis using computer simulations and validation of the different theoretical results are provided at the end of the paper. Testing IDA in real world data is one of the perspectives of this work.

A Proof of Theorem 3

Before proceeding, note that Theorem 3 is a generalization of Theorem 2 and hence its proof is implicitly a proof of Theorem 2. Also, note that the result of Theorem 2 is a proof for the sufficiency of the identifiability condition of Theorem 1. Indeed, since $\boldsymbol{A}$ is full column rank, we know that a decorrelation matrix exist (e.g. $\boldsymbol{B} = \boldsymbol{A}^{\#}$). Now, Theorem 2 result demonstrates that under

condition (8), a decorrelation matrix at the considered time lags achieves the desired blind source separation.

For proof of Theorem 3, let us notice that equation (10) is equivalent to [5]

$$\boldsymbol{C}\boldsymbol{S}(\tau_k)\boldsymbol{C}^H \textit{ diagonal for } k = 1, \cdots, K$$

$$\boldsymbol{C}(\sum_{k=1}^{K} \boldsymbol{S}(\tau_k)^H \boldsymbol{C}^H \boldsymbol{C}\boldsymbol{S}(\tau_k))\boldsymbol{C}^H > 0$$

where $\boldsymbol{C} = \boldsymbol{B}\boldsymbol{A}$, ^{H} denotes the matrix conjugate transpose, and the inequality $\boldsymbol{A} > 0$ stands for $\boldsymbol{A}$ positive definite. This latter inequality implies in particular that $\boldsymbol{C}$ is full rank. Note that $\boldsymbol{C}\boldsymbol{S}(\tau_k)\boldsymbol{C}^H$ is diagonal if and only if both $\boldsymbol{C}\Re e(\boldsymbol{S}(\tau_k))\boldsymbol{C}^H$ and $\boldsymbol{C}\Im m(\boldsymbol{S}(\tau_k))\boldsymbol{C}^H$ are diagonal.

Then we shall show that there exist linear combinations of $\Re e(\boldsymbol{S}(k))$ and $\Im m(\boldsymbol{S}(k))$, i.e., $\boldsymbol{S}_1 = \sum_{k=1}^{K} \alpha_k \Re e(\boldsymbol{S}(\tau_k)) + \alpha'_k \Im m(\boldsymbol{S}(\tau_k))$ and $\boldsymbol{S}_2 = \sum_{k=1}^{K} \beta_k \Re e(\boldsymbol{S}(\tau_k)) + \beta'_k \Im m(\boldsymbol{S}(\tau_k))$, such that **(i)** $\boldsymbol{S}_2$ is non-singular and **(ii)** the diagonal entries of $\boldsymbol{S}_1\boldsymbol{S}_2^{-1}$ take d distinct values of multiplicity $d_1, \cdots, d_d$, respectively. A simple way to prove it is to consider linear combinations of the form $\alpha_k = x^k, \alpha'_k = x^{k+K}$ and $\beta_k = y^k, \beta'_k = y^{k+K}$ (x and y being real scalars). The diagonal entries of $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are $P_i(x) = \sum_{k=1}^{K} \Re e(\rho_i(\tau_k))x^k + x^K \sum_{k=1}^{K} \Im m(\rho_i(\tau_k))x^k$ and $P_i(y)$, $i = 1, \cdots, m$, respectively. Under the assumption of Theorem 2, we have [6]

$$\begin{array}{rl} P_i(y) \not\equiv 0 & \text{for all } i \\ P_i(x)P_j(y) \not\equiv P_i(y)P_j(x) & \text{if } i \text{ and } j \text{ belong to two distinct groups} \\ P_i(x)P_j(y) \equiv P_i(y)P_j(x) & \text{if } i \text{ and } j \text{ belong to the same group} \end{array}$$

[5] The noise term is neglected. However, this term equals zero if all the time lags are non-zero and if the noise is temporally white.

[6] The equalities below are understood in terms of equalities of polynomials.

Therefore, $P(x,y) \stackrel{\text{def}}{=} \prod_{i=1}^{d} P_i(y) \prod_{1\leq i<j\leq d}(P_i(x)P_j(y) - P_j(x)P_i(y)) \not\equiv 0$ (here i and j denote the group index not the source index) and thus there exist infinite values of (x,y) such that $P(x,y) \neq 0$. Let (x_0,y_0) be such a value and let $\boldsymbol{S}_1 = \sum_{k=1}^{K} x_0^k \Re e(\boldsymbol{S}(\tau_k)) + x_0^{k+K} \Im m(\boldsymbol{S}(\tau_k))$ and $\boldsymbol{S}_2 = \sum_{k=1}^{K} y_0^k \Re e(\boldsymbol{S}(\tau_k)) + y_0^{k+K} \Im m(\boldsymbol{S}(\tau_k))$. $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ satisfy conditions **(i)** and **(ii)**.

To complete the proof, note that

$$\boldsymbol{M}_1 \stackrel{\text{def}}{=} \sum_{k=1}^{K} x_0^k \boldsymbol{C} \Re e(\boldsymbol{S}(\tau_k)) \boldsymbol{C}^H + x_0^{k+K} \boldsymbol{C} \Im m(\boldsymbol{S}(\tau_k)) \boldsymbol{C}^H = \boldsymbol{C}\boldsymbol{S}_1\boldsymbol{C}^H$$

is diagonal and

$$\boldsymbol{M}_2 \stackrel{\text{def}}{=} \sum_{k=1}^{K} y_0^k \boldsymbol{C} \Re e(\boldsymbol{S}(\tau_k)) \boldsymbol{C}^H + y_0^{k+K} \boldsymbol{C} \Im m(\boldsymbol{S}(\tau_k)) \boldsymbol{C}^H = \boldsymbol{C}\boldsymbol{S}_2\boldsymbol{C}^H$$

is diagonal and non-singular. It follows that, $\boldsymbol{M}_1\boldsymbol{M}_2^{-1} = \boldsymbol{C}\boldsymbol{S}_1\boldsymbol{S}_2^{-1}\boldsymbol{C}^{-1}$ is a diagonal matrix with d distinct eigenvalues (condition **(ii)**). Using standard spectral theory, e.g., [32], we conclude that $\boldsymbol{C} = \boldsymbol{P}\boldsymbol{\Lambda}$ for a given permutation matrix $\boldsymbol{P}$ and a given non-singular block-diagonal matrix $\boldsymbol{\Lambda}$, i.e., $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{U}_1, \cdots, \boldsymbol{U}_d)$ where $\boldsymbol{U}_i$ is a $d_i \times d_i$ non-singular matrix. Finally, the fact that sources belonging to the same group cannot be separated from the considered set of statistics is simply a direct consequence of the necessity of condition (8). □

B Proof of Theorem 4

According to the above development, we have $\boldsymbol{C} = \boldsymbol{P}\text{diag}(\boldsymbol{U}_1, \cdots, \boldsymbol{U}_d)$. On the other hand we have $\boldsymbol{C}\boldsymbol{S}(\tau_k)\boldsymbol{C}^H$ diagonal for $k = 1, \cdots, K$. This leads

to $\boldsymbol{U}_i\boldsymbol{S}_i(\tau_k)\boldsymbol{U}_i^H$ diagonal $\forall i = 1, \cdots, d$, $\forall k = 1, \cdots, K$. Now since we have grouped the sources in such a way each group corresponds to sources with (up to a sign) correlation vector, i.e. $\boldsymbol{S}_i(\tau_k) = \rho_i(\tau_k)\boldsymbol{\Sigma}_i$ where $\boldsymbol{\Sigma}_i$ are diagonal matrices with diagonal entries equal to ± 1, the previous equation becomes $\boldsymbol{U}_i\boldsymbol{\Sigma}_i\boldsymbol{U}_i^H$ diagonal for $i = 1, \cdots, d$. Moreover, since $\boldsymbol{\Sigma}_i$ is real diagonal and since we normalized to unity the correlation vectors (both estimated and exact), then $\boldsymbol{U}_i\boldsymbol{\Sigma}_i\boldsymbol{U}_i^H$ is real diagonal with diagonal entries equal to ± 1. Finally, since the signature (i.e., the number of positive eigenvalues and the number of negative eigenvalues) of a quadratic form $\boldsymbol{Q}$ is invariant by multiplication left and right by a non-singular matrix and its conjugate transpose, respectively, then $\boldsymbol{U}_i\boldsymbol{\Sigma}_i\boldsymbol{U}_i^H$ has the same signature (i.e., number of $+1$s and number of -1s) as $\boldsymbol{\Sigma}_i$. In other words, there exists a $d_i \times d_i$ permutation matrix $\boldsymbol{P}_i$ such that $\boldsymbol{U}_i\boldsymbol{\Sigma}_i\boldsymbol{U}_i^H = \boldsymbol{P}_i\boldsymbol{\Sigma}_i\boldsymbol{P}_i^T$. It follows then that, for $k = 1, \cdots, K$

$$E(\boldsymbol{z}(t+\tau_k)\boldsymbol{z}^\star(t)) = \boldsymbol{C}\boldsymbol{S}(\tau_k)\boldsymbol{C}^H = \boldsymbol{P}\boldsymbol{S}(\tau_k)\boldsymbol{P}^T \tag{B.1}$$

for a given permutation matrix $\boldsymbol{P}$. Equation (B.1) means that the entries $z_i(t)$ of $\boldsymbol{z}(t)$ have the same correlation coefficients as $s_{P(i)}(t)$ for $i = 1, \cdots, m$, where $P(1), \cdots, P(m)$ are the images of $1, \cdots, m$ by the permutation $\boldsymbol{P}$. This verifies the conclusion of Theorem 4. □

C Proof of Theorem 5

We propose here to derive the expression of the asymptotic covariance error. Note that the asymptotic Gaussianity of the error $\boldsymbol{\delta}$ comes from the central limit theorem as shown in [3, 18]. The expression of $\boldsymbol{\delta}$ can be obtained by replacing in equations (17) and (23) $\boldsymbol{z}^{(p)}(t)$ and $\boldsymbol{\epsilon}^{(p)}$ by $\boldsymbol{s}(t)$ and $\boldsymbol{\delta}$ respectively.

Hence, from the vectorized version of equation (25) the expression of matrix $\boldsymbol{\delta}_{ij}\boldsymbol{\delta}_{ij}^T$ can be written as :

$$\boldsymbol{\delta}_{ij}\boldsymbol{\delta}_{ij}^T = \frac{1}{4}\mathbb{H}_{ij}^{\#}\boldsymbol{\psi}_{ij}\mathbb{H}_{ij}^{\#\star} \tag{C.1}$$

where

$$\mathbb{H}_{ij} = \boldsymbol{I}_2 \otimes [\boldsymbol{r}_{ii}\ \boldsymbol{r}_{jj}] \tag{C.2}$$

$$\boldsymbol{\psi}_{ij} = \begin{bmatrix} \boldsymbol{\xi}_{11}^{(ij)} & \boldsymbol{\xi}_{12}^{(ij)} \\ \boldsymbol{\xi}_{21}^{(ij)} & \boldsymbol{\xi}_{22}^{(ij)} \end{bmatrix} \tag{C.3}$$

with

$$\boldsymbol{\xi}_{11}^{(ij)} = \boldsymbol{\gamma}_{11}^{(ij)} + \boldsymbol{\gamma}_{12}^{(ij)} + \boldsymbol{\gamma}_{21}^{(ij)} + \boldsymbol{\gamma}_{22}^{(ij)} \tag{C.4}$$

$$\boldsymbol{\xi}_{12}^{(ij)} = \jmath\left[\boldsymbol{\gamma}_{11}^{(ij)} - \boldsymbol{\gamma}_{12}^{(ij)} + \boldsymbol{\gamma}_{21}^{(ij)} + \boldsymbol{\gamma}_{22}^{(ij)}\right] \tag{C.5}$$

$$\boldsymbol{\xi}_{21}^{(ij)} = -\jmath\left[\boldsymbol{\gamma}_{11}^{(ij)} + \boldsymbol{\gamma}_{12}^{(ij)} - \boldsymbol{\gamma}_{21}^{(ij)} + \boldsymbol{\gamma}_{22}^{(ij)}\right] \tag{C.6}$$

$$\boldsymbol{\xi}_{22}^{(ij)} = \boldsymbol{\gamma}_{11}^{(ij)} - \boldsymbol{\gamma}_{12}^{(ij)} - \boldsymbol{\gamma}_{21}^{(ij)} + \boldsymbol{\gamma}_{22}^{(ij)} \tag{C.7}$$

and $\boldsymbol{\gamma}_{11}^{(ij)},\ \boldsymbol{\gamma}_{12}^{(ij)},\ \boldsymbol{\gamma}_{21}^{(ij)}$ and $\boldsymbol{\gamma}_{22}^{(ij)}$ are $K \times K$ matrices given by $\forall 1 \leq k, k' \leq K$:

$$\boldsymbol{\gamma}_{11}^{(ij)}(k,k') = \frac{1}{(T-\tau_k)(T-\tau_{k'})}\sum_{t=1}^{T-\tau_k}\sum_{t'=1}^{T-\tau_{k'}} z_i(t+\tau_k)z_j(t)z_i(t'+\tau_{k'})z_j(t') \tag{C.8}$$

$$\boldsymbol{\gamma}_{22}^{(ij)}(k,k') = \frac{1}{(T-\tau_k)(T-\tau_{k'})}\sum_{t=1}^{T-\tau_k}\sum_{t'=1}^{T-\tau_{k'}} z_j(t+\tau_k)z_i(t)z_j(t'+\tau_{k'})z_i(t') \tag{C.9}$$

$$\boldsymbol{\gamma}_{12}^{(ij)}(k,k') = \frac{1}{(T-\tau_k)(T-\tau_{k'})}\sum_{t=1}^{T-\tau_k}\sum_{t'=1}^{T-\tau_{k'}} z_i(t+\tau_k)z_j(t)z_j(t'+\tau_{k'})z_i(t') \tag{C.10}$$

$$\boldsymbol{\gamma}_{21}^{(ij)}(k,k') = \frac{1}{(T-\tau_k)(T-\tau_{k'})}\sum_{t=1}^{T-\tau_k}\sum_{t'=1}^{T-\tau_{k'}} z_j(t+\tau_k)z_i(t)z_i(t'+\tau_{k'})z_j(t') \tag{C.11}$$

Therefore, using the index $\tau = t - t'$, we obtain

$$\begin{aligned}\lim_{T\to+\infty} T\ E\left(\boldsymbol{\gamma}_{11}^{(ij)}(k,k')\right) &= \lim_{T\to+\infty}\frac{T(T-\tau_k-|\tau|)}{(T-\tau_k)(T-\tau_{k'})}\sum_{\tau=-T+\tau_{k'}}^{T-\tau_k} r_{ii}(\tau_k+\tau)r_{jj}(\tau_{k'}+\tau) \\ &= \boldsymbol{\Gamma}_{11}^{(ij)}(k,k')\end{aligned}$$

$$\lim_{T\to+\infty} T\ E\left(\boldsymbol{\gamma}_{22}^{(ij)}(k,k')\right) = \lim_{T\to+\infty} \frac{T(T-\tau_k-|\tau|)}{(T-\tau_k)(T-\tau_{k'})} \sum_{\tau=-T+\tau_{k'}}^{T-\tau_k} r_{ii}(\tau_{k'}+\tau) r_{jj}(\tau_k+\tau)$$

$$= \boldsymbol{\Gamma}_{22}^{(ij)}(k,k')$$

$$\lim_{T\to+\infty} T\ E\left(\boldsymbol{\gamma}_{12}^{(ij)}(k,k')\right) = \lim_{T\to+\infty} \frac{T(T-\tau_k-|\tau|)}{(T-\tau_k)(T-\tau_{k'})} \sum_{\tau=-T+\tau_{k'}}^{T-\tau_k} r_{ii}(\tau_k+\tau) r_{jj}(\tau_{k'}-\tau)$$

$$= \boldsymbol{\Gamma}_{12}^{(ij)}(k,k')$$

$$\lim_{T\to+\infty} T\ E\left(\boldsymbol{\gamma}_{21}^{(ij)}(k,k')\right) = \lim_{T\to+\infty} \frac{T(T-\tau_k-|\tau|)}{(T-\tau_k)(T-\tau_{k'})} \sum_{\tau=-T+\tau_{k'}}^{T-\tau_k} r_{ii}(\tau_{k'}-\tau) r_{jj}(\tau_k+\tau)$$

$$= \boldsymbol{\Gamma}_{21}^{(ij)}(k,k')$$

where $\boldsymbol{\Gamma}_{11}^{(ij)}(k,k')$, $\boldsymbol{\Gamma}_{22}^{(ij)}(k,k')$, $\boldsymbol{\Gamma}_{12}^{(ij)}(k,k')$ and $\boldsymbol{\Gamma}_{21}^{(ij)}(k,k')$ are given by equations (44), (45), (46) and (47). □

References

[1] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing.* UK: Wiley & Sons, Ltd., 2003.

[2] D. T. Pham and J. F. Cardoso, "Blind source separation of instantaneous mixtures of nonstationary sources," *IEEE Transactions on Signal Processing*, vol. 49, no. 9, pp. 1837–1848, September 2001.

[3] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, February 1997.

[4] J. Cardoso and A. Souloumiac, "A blind beamforming for non-gaussian signals," *IEE Proceedings F, Radar and Signal Processing*, vol. 140, no. 6, pp. 362–370, December 1993.

[5] K. Abed-Meraim, Y. Hua, and A. Belouchrani, "A general framework for blind source separation using second order statistics," in *8^{th} IEEE Digital Signal Processing Workshop*, Utah, USA, August 1998.

[6] L. Féty and J.-P. V. Uffelen, "News methods for signal separation," in *in Proc. 14th Conf. on HF radio systems and techniques*, London, U.K, April 1988, pp. 226–230.

[7] L. Tong, V. C. Soon, Y. F. Huang, and R. Li, "Amuse: A new blind identification algorithm," in *in Proc. IEEE ISCAS*, New Orleans, U.S.A, May 1990.

[8] L. Tong, R. Liu, V. Soon, and Y. Huang, "Indeterminacy and identifiability of blind identification," *IEEE Transactions on Circuits and Systems*, vol. 38, no. 5, pp. 499–509, May 1991.

[9] A. M. Tomé, "Blind source separation using a matrix pencil," in *in Proc. International Joint Conference on Neural Networks*, Como, Italy, 2000.

[10] D. Nuzillard and J. M. Nuzillard, "Second-order blind source separation in the fourier space of data," *Signal Processing*, vol. 83, no. 3, pp. 627–631, March 2003.

[11] J. F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM Journal on Matrix Analysis and Applications*, vol. 17, pp. 161–164, 1996.

[12] A. Ziehe, P. Laskov, G. Nolte, and K. R. Müller, "A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation," *Journal of Machin Learning Research*, vol. 5, pp. 777–800, 2004.

[13] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Transactions on Signal Processing*, vol. 50, pp. 1545–1553, 2002.

[14] H. C. Wu and J. C. Principe, "A unifying criterion for blind source separation and decorrelation : Simultaneous diagonalization of correlation matrices," in *in Proc. IEEE Workshop of Neural Networks for Signal Processing*, vol. 7, 1997, pp. 496–505.

[15] R. Vollgraf and K. Obermayer, "Quadratic optimization for simultaneous matrix diagonalization," *IEEE Transactions on Signal Processing*, vol. 54, no. 9, pp. 3270–3278, September 2006.

[16] D. T. Pham, "Joint approximate diagonalization of positive definite matrices," *SIAM Journal on Matrix Anal. and Appl*, vol. 22, pp. 1136–1152, 2001.

[17] D. T. Pham and P. Garat, "Séparation aveugle de sources temporellement corrélées," in *in In 14e colloque GRETSI*, Juanles-Pins, France, September 1993.

[18] ——, "Blind separation of mixture of independent sources through a quasi-maximum likelihood approach," *IEEE Transactions on Signal Processing*, vol. 45, no. 7, pp. 1712–1725, July 1997.

[19] S. Dégerine and A. Zaidi, "Separation of an instantaneous mixture of gaussian autoregressive sources by the exact maximum likelihood approach," *IEEE Transactions on Signal Processing*, vol. 52, no. 6, pp. 1499–1512, June 2004.

[20] G. Chabriel and J. Barrère, "An instantaneous formulation of mixtures for blind separation of propagating waves," *IEEE Transactions on Signal Processing*, vol. 54, no. 1, pp. 49–58, January 2006.

[21] A. Ziehe, K. R. Müller, G. Nolte, B. M. Mackert, and G. Curi, "Artifact reduction in magnetoneurography based on time-delayed second-order correlations," *IEEE Transactions on Biomedical Engineerin*, vol. 47, no. 1, pp. 75–87, January 2000.

[22] M. Valkama, M. Renfors, and V. Koivunen, "Blind signal estimation in conjugate signal models with application to i/q imbalance compensation," *IEEE Signal Processing Letter*, vol. 12, no. 11, pp. 733–736, November 2005.

[23] K. Abed-Meraim, Y. Xiang, and Y. Hua, "Generalized second order identifiability condition and relevant testing technique," in *Proc. ICASSP*, vol. 5, Istanbul, Turkey, June 2000, pp. 2989–2992.

[24] A. M. Zoubir and B. Boashash, "The bootstrap and its application in signal processing," *IEEE Signal Processing Magazine*, vol. 15, no. 1, pp. 56–76, January 1998.

[25] B. Efron, "The jackknife, the bootstrap and other resampling plans," in *CBMS Monograph 38, Society for Industrial and Applied Mathematics*, Philadelphia, USA, 1982.

[26] R. G. Miller, "The jackknife - a review," *Biometrika*, vol. 61, pp. 1–15, 1974.

[27] A. Belouchrani, M. G. Amin, and C. Wang, "Interference mitigation in spread spectrum communications using blind source separation," in *Proc. Asilomar Conference*, vol. 1, November 1996, pp. 718–722.

[28] G. D'Urso, P. Prieur, and C. Vincent, "Blind identification methods applied to electricite de france's civil works and power plants monitoring," in *Higher-Order Statistics, Proc. of the IEEE Signal Processing Workshop*, July 1997, pp. 82–86.

[29] M. Kawamoto, K. Matsuoka, and M. Oya, "Blind separation of sources using temporal correlation of the observed signal," *IEICE Transactions on Fundamentals of Electronics Communications and Computing*, vol. E80A, no. 4, pp. 695–704, April 1997.

[30] A. Aïssa-El-Bey, K. Abed-Meraim, and Y. Grenier, "Iterative blind source separation by decorrelation algorithm: algorithm and performance analysis,"

in *Proc. EUSIPCO*, Florence, Italy, September 2006.

[31] Y. Xiang, K. Abed-Meraim, and Y. Hua, "Adaptive blind source separation by second order statistics and natural gradient," in *Proc. ICASSP*, 1998.

[32] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge, UK: Cambridge Univ. Press, 1985.