BLIND SEPARATION OF JOINTLY STATIONARY CORRELATED SOURCES

M. R. Aghabozorgi and A. M. Doost-Hoseini

Department of Electrical and Computer Engineering, Isfahan University of Technology
Isfahan, Iran, P7530921@sepahan.iut.ac.ir

Abstract The separation of unobserved sources from mixed observed data is a fundamental signal processing problem. Most of the proposed techniques for solving this problem rely on independence or at least uncorrelation assumption for source signals. This paper introduces a technique for cases that source signals are correlated with each other. The method uses Wold decomposition principle for extracting desired and proper information from the predictable part of the observed data, and exploits approaches based on second-order statistics to estimate the mixing matrix and source signals. Simulation results are provided to illustrate the effectiveness of the method.

Key Words Blind Source Separation, Wold Decomposition, Second-Order Statistics

چکـیده   مسـاله جداسازی کور منابع سیگنال نامعلوم، که فقط مشاهداتی از ترکیبهای خطی آنها (به واسطه ترکیـب کنـنده ای نـامعلوم) توسط چند گیرنده دریافت شده است، یکی از مسائل اساسی در حوزه پردازش سیگنال میباشد. اغلب تکنیکهای حل این مساله بر اساس فرض استقلال و یا حداقل فرض نا هم بسته بودن برای منابع، عمل میکنند. این مقاله روشی را برای منابع هم بسته معرفی میکند. روش پیشنهادی ابتدا بر پایه قضـیه تجـزیه Wold اطلاعـات مناسـب و مطلوبـی را از بخش قابل پیشگویی مشاهدات استخراج میکند و سپس با رهیافت های مبتنی بر اساس آمارگان های مرتبه دوم (توابع هم بستگی سیگنال ها)، ماتریس ترکیب کننده و در نتیجه سیگنال های منابع را تخمین میزند. نتایج ناشی از شبیه سازی عملکرد مؤثر روش را نشان میدهد.

1. INTRODUCTION

In this section, blind source separation (BSS) is generally introduced and a brief review of existing algorithms is given.

Blind source separation is a basic and challenging research problem in signal processing, which has received a great deal of attention in recent years. It has a broad range of applications such as: array signal processing [1,2], speech processing [3], image reconstruction [4], communication systems [5,6], biomedical signal processing [7], and semiconductor manufacturing [8].

BSS consists of recovering source signals from several observed noisy mixtures of them. The observations are obtained from a set of sensors, each receiving different combinations of the source signals. The problem is called "blind" because no information is available about the mixture, i.e. recovering of source signals is achieved without the knowledge of the characteristics of the transmission channel. The lack of prior information causes the difficulty of the problem, but it is precisely the strength of the BSS model, making it a versatile tool for signal processing problems.

Thus far, the problem of the BSS has been solved using various techniques and algorithms based on different assumptions and models. In the following a brief review of some of these approaches follows.

A common model used by many researchers is the combination system, i.e. the transformation from the source signals to the sensors is linear time-invariant (LTI) and instantaneous. This is the model that we use in this paper. A complicated model for the combination system is convolutive form that has been investigated by some researchers [9,10].

The lack of prior information must be compensated by some assumptions that can be

divided into the following categories:

(a) Particular Source Statistics The most popular condition used by BSS techniques is the statistical independence assumption between the source signals. These techniques assume that the primary sources are statistically independent, and therefore the goal in these techniques is to achieve a separation process that produces outputs as independent as possible [11-13]. A less stringent condition is uncorrelation of sources. Techniques, that use this condition, exploit temporal correlation of each source signal (second-order blind identification), and use a joint diagonalization method of several correlation matrices [14], [15]. Other various assumptions in this category have been used such as cyclostationary condition that used by SCORE algorithm, which relies on the assumption that the different sources have different cyclostationary features [16,17].

(B) Special Source Signals Structures BSS algorithms that are in this category achieve separation process by tracking a special character, assumed a priory for source signals, in the observed data. Constant modulus algorithm (CMA) is an important technique in this class. This algorithm exploits the constant modulus property of some signals, e.g. communication signals like FM, PM, FSK. The algorithm is derived by finding proper weight vectors such that weighted observations become CM signals [18,19]. Another property of source signals might be the finite alphabet of digital signals [20].

(C) Special Structure For Combination System Some approaches use properties of the mixer of source signals (mixing matrix), e.g. assumptions on the geometry of the antenna array that cause special structures for mixing matrix. In particular, the columns of the mixing matrix are assumed to be vectors on the array manifold, each associated with a certain direction of arrival (DOA) (like Vander monde structure based on DOA for uniform linear array). MUSIC [21,1], and ESPIRIT [22] algorithms, are in this class. It must be noted here that spatial smoothing (SS) [23] has been proposed to extend these algorithms to cases of correlated source signals.

In this paper, the aim is to propose a solution to BSS problem for correlated source signals without imposing special structures on signals or mixing matrix.

This paper is organized as follows: In section II, the problem of BSS is stated along with the related assumptions. Proposed pre-separation procedure is introduced in section III. Section IV expresses BSS algorithm, and simulation results are presented in section V. Concluding remarks are given in section VI.

2. PROBLEM FORMULATION

In this section a model for the problem and some notions of blind identification are presented.

2.1 The Model Assume that d signals $s_1(t), \ldots, s_d(t)$ are transmitted from d sources at different locations. By considering a narrowband time-invariant channel, what we receive at m sensors (antennas) will be an instantaneous linear combination of these signals that constructs observation data:

$$\mathbf{x}(t) = \mathbf{a}_1 . s_1(t) + \ldots + \mathbf{a}_d . s_d(t) + \mathbf{n}(t) \qquad (1)$$

Thus the model is as follows:
$$\mathbf{x}(t) = \mathbf{y}(t) + \mathbf{n}(t) = \mathbf{A} . \mathbf{s}(t) + \mathbf{n}(t) \qquad (2)$$

where

$\mathbf{x}(t) \in \Re^{m \times 1}$ is the observed data vector from m sensors,

$\mathbf{s}(t) \in \Re^{d \times 1}$

is the signal vector, composed of d unknown source signals,

$\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_d] \in \Re^{m \times d}$

Characterizes the unknown channel and is referred to as "mixing matrix",

$\mathbf{n}(t) \in \Re^{m \times 1}$

is the additive noise vector at the sensor array.

Following assumptions are considered in the model:

(A1) Each element of $\mathbf{s}(t)$ (source signals) is a zero-mean, stationary process.

(A2) the additive noise $\mathbf{n}(t)$ is assumed as a stationary, white zero-mean random process, independent of source signals.

$$E\left\{\mathbf{n}(t+\tau)\mathbf{n}^*(t)\right\} = N_0 . \delta(\tau). \mathbf{I} \qquad (3)$$

where $\delta(\tau)$ is Kronecker delta, N_0 is variance of noise, and $\mathbf{I}$ denotes the identity matrix.

(A3) Mixing matrix $\mathbf{A}$ has full column rank, i.e. rank($\mathbf{A}$) = d, this is the only condition on $\mathbf{A}$ and $\mathbf{A}$ can have any unknown form.

(A4) The number of sensors (m) must be greater or equal to d (the number of sources).

It must be emphasized here that we don't impose any assumption about independence or uncorrelation of source signals. In other words, the source signals can be correlated, and only the following assumption is considered:

(A5) Source signals are jointly stationary.

The aim of blind source separation (BSS) is to identify the mixing matrix $\mathbf{A}$ (and consequently recovering the source signals from the observations), or equivalently to find a matrix $\mathbf{B}$ such that (ideally): $\mathbf{B}.\mathbf{A} = \mathbf{I}$.

2.2 B. Blind Separability and Identifiability

An issue in blind source separation problem is separability, i.e. the existence of a matrix $\mathbf{B}$ such that the product $\mathbf{B}.\mathbf{A}$ separates the original signals. This depends purely on the structure of $\mathbf{A}$ (mixing matrix). In [24], it has been proved that if the rank of $\mathbf{A}$ equals d (i.e. full column rank) then there always exists such a Matrix $\mathbf{B}$. Also it has been shown that for separating all sources, there must exist at least d (the number of sources) sensors.

Another issue in BSS is blind identifiability (and indeterminacy) [15]. Complete identification of the mixture matrix is impossible because the exchange of a fixed scalar factor between a given source signal and the corresponding column of $\mathbf{A}$ doesn't affect the observations (scaling indeterminacy).

$$\mathbf{x}(t) = \mathbf{A}.\mathbf{s}(t) + \mathbf{n}(t) = \sum_{k=1}^{d} \frac{\mathbf{a}_k}{\alpha_k} \alpha_k s_k(t) + \mathbf{n}(t) \qquad (4)$$

where α_k is an arbitrary factor, and $\mathbf{a}_k$ denotes the k-th column of $\mathbf{A}$.

Indeterminacy is in the order of the separated signals (permutation indeterminacy). This is expressed by:

$$\hat{s}(t) = P.s(t) \qquad (5)$$

where $\mathbf{s}(t)$ and $\hat{\mathbf{s}}(t)$ are original and separated source signals and $\mathbf{P}$ is a permutation matrix. Generally, in the blind context, there are the following "waveform-preserving" relations between the original and the estimated mixing matrices and the source signals doublets ($\mathbf{A}$, $\mathbf{s}(t)$) and ($\hat{\mathbf{A}}$, $\hat{\mathbf{s}}(t)$) [15]:

$$\hat{\mathbf{A}} = \mathbf{A}.\Lambda^{-1}.\mathbf{P}^T \qquad (6)$$

$$\hat{\mathbf{s}}(t) = \mathbf{P}.\Lambda.\mathbf{s}(t) \qquad (7)$$

for some permutation matrix $\mathbf{P}$, and some nonsingular diagonal matrix Λ.

3. PRE-SEPARATION PROCEDURE

The main step in our approach for correlated sources is a pre-separation process. The observed data is decomposed into regular and predictable components, using Wold decomposition. In the predictable component, the combination of uncorrelated contributions of source signals is identified on whose basis $\mathbf{A}$ (and consequently the source signals) is estimated using second order statistics. In this section, Wold decomposition and its application on observation data are introduced. The description of identifying algorithm for mixing matrix will follow.

3.1. Wold Decomposition   An arbitrary process can be written as a sum:

$$s(t) = s_r(t) + s_p(t) \qquad (8)$$

where $s_r(t)$ and $s_p(t)$ are regular and predictable processes. This expansion is called Wold decomposition. In [25,26] it has been proved that the processes $s_p(t)$ and $s_r(t)$ are orthogonal:

$$E\{s_r(t+\tau)s_p^*(t)\} = 0 \qquad , \forall\, t, \qquad (9)$$

and have the same prediction filter as $s(t)$. Furthermore, $s_p(t)$ is comprised of complex exponentials:

$$s_p(t) = \mathbf{c}_0 + \sum_i \mathbf{c}_i.\exp(j\omega_i t) \qquad (10)$$

where $\mathbf{c}_i$'s are orthogonal zero-mean random variables. Hence, $s_p(t)$ has a line spectrum:

$$P_{s_p}(\omega) = \sum_i 2\,\pi\alpha_i\,\delta(\omega-\omega_i) \qquad (11)$$

$s_r(t)$ can be represented as the response of a minimum-phase innovation filter ($L(s)$) with a white process input. It has a smooth spectrum: $P_{s_r}(\omega) = |L(j\omega)|^2$ that satisfies the Paley-Wiener condition:

$$\int \frac{\left|\ln P_{s_r}(\omega)\right|}{1+\omega^2}\,d\omega \langle\infty.$$

3.2. Observation Decomposition   In this subsection a method is proposed for extracting and decomposing some information from the regular and predictable parts of the observation data. For simplicity, a special case of Model 2 with d = 2 and m = 2 is considered, that can be extended to general cases. So, we have the following model satisfying conditions expressed in II-A:

$$\mathbf{x}(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \mathbf{A}.\begin{bmatrix} s_1(t) \\ s_2(t) \end{bmatrix} + \begin{bmatrix} n_1(t) \\ n_2(t) \end{bmatrix} \qquad (12)$$

where $s_1(t)$ and $s_2(t)$ are the source signals and $A = \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \end{bmatrix}$ is the mixing matrix. Hence:

$$x_1(t) = \alpha_1 s_1(t) + \beta_1 s_2(t) + n_1(t) \qquad (13)$$

$$x_2(t) = \alpha_2 s_1(t) + \beta_2 s_2(t) + n_2(t)$$

Regular and predictable parts of source signal are indicated by $s_{ir}(t)$ and $s_{ip}(t)$ ($i = 1,2$):

$$s_i(t) = s_{ip}(t) + s_{ir}(t) \qquad (14)$$

where,

$$s_{1p}(t) = \sum_k \mathbf{a}_k.\exp(j\omega_{1k} t) \qquad (15)$$

$$s_{2p}(t) = \sum_l \mathbf{b}_l.\exp(j\omega_{2l} t) \qquad (16)$$

in which $\{\mathbf{a}_k\}$ and $\{\mathbf{b}_l\}$ are sets of orthogonal random variables, and $\{\omega_1\}$, $\{\omega_2\}$ are proper frequency sets. Also, since source signals are assumed jointly stationary, $\mathbf{a}_k$ and $\mathbf{b}_l$ corresponding to $\omega_{1k} \neq \omega_{2l}$ are orthogonal.

Using (13), (14) and the fact that regular and predictable parts in each signal are orthogonal, we obtain:

$$x_i(t) = x_{ip}(t) + x_{ir}(t) + n_i(t) \quad ; \; i = 1,2 \qquad (17)$$

where

$$x_{ip}(t) = \alpha_i s_{1p}(t) + \beta_i s_{2p}(t) \qquad ; \; i = 1,2 \qquad (18)$$

$$x_{ir}(t) = \alpha_i s_{1r}(t) + \beta_i s_{2r}(t) \qquad ; \; i = 1,2 \qquad (19)$$

From (15),(16),(18) following relation obtains:

$$x_{ip}(t) = \sum_q \mathbf{d}_{iq}.\exp(j\omega_q t) \quad ; \; i = 1,2 \qquad (20)$$

where $\{\mathbf{d}_{iq}\}$ are orthogonal random variables and

$$\{\omega_q\} = \{\omega_1\} \cup \{\omega_2\}.$$

From these equations, each observation signal has regular and predictable components, each corresponding to the combination of individual regular and predictable parts of source signals. A spectral method for separating these parts follows.

The correlation functions of the observation data are given by:

$$r_{ij}^x(\tau) = E\{x_i(t+\tau)x_j^*(t)\} \qquad ; \qquad i,j = 1,2$$
$$= r_{ijr}^x(\tau) + r_{ijp}^x(\tau) + N_0.\delta(\tau) \qquad (21)$$

where N_0 is the noise variance and $r_{ijr}^x(\tau)$ and $r_{ijp}^x(\tau)$ are correlation functions of regular and predictable parts:

$$r_{ijr}^x(\tau) = E\{x_{ir}(t+\tau)x_{jr}^*(t)\} \qquad (22)$$

$$r_{ijp}^x(\tau) = E\{x_{ip}(t+\tau)x_{jp}^*(t)\}$$
$$= \sum_q E\{\mathbf{d}_{iq}\mathbf{d}_{iq}^*\}.\exp(j\omega_q t) \qquad (23)$$

Hence power spectral density (psd) and cross-spectral density (csd) functions of observations have the forms:

$$P_{ij}^x(\omega) = P_{ijr}^x(\omega) + P_{ijp}^x(\omega) + N_0 \qquad (24)$$

where,

$$P_{ijp}^x(\omega) = \sum_q 2\pi.E\{\mathbf{d}_{iq}\mathbf{d}_{jq}^*\}.\delta(\omega - \omega_q) \qquad (25)$$

As expected, the spectra of the predictable parts are pure impulsive. So, it is possible to detect and separate these components in the observation spectra. Although some little power of regular part and noise remain on separated predictable part spectra that are negligible. Consequently, correlation functions ($r_{ijp}^x(\tau)$) of the predictable

parts are obtained that will be used next.

3.3 Extracting Desired Information From Predictable Part
Rewriting predictable parts of source signals (15)-(16), considering $\{\Omega_n\}$ as the common frequency set, we obtain:

$$s_{1p}(t) = \sum_{k \neq n} \mathbf{a}_k.\exp(j\omega_{1k}t) + \sum_n \mathbf{a}_n.\exp(j\Omega_n t) \qquad (26)$$

$$s_{2p}(t) = \sum_{l \neq n} \mathbf{b}_l.\exp(j\omega_{2l}t) + \sum_n \mathbf{b}_n.\exp(j\Omega_n t) \qquad (27)$$

where:

(1) Random variables $\mathbf{a}_k$ and $\mathbf{b}_l$ corresponding to $\omega_{1k} \neq \omega_{2l}$ are orthogonal, i.e.:

$$E\{\mathbf{a}_k\mathbf{b}_l\} = 0 \qquad (28)$$

(2) Correlation of predictable signals $s_{1p(t)}$ and $s_{2p(t)}$ arises from correlation of random variables $\mathbf{a}_n$ and $\mathbf{b}_n$ corresponding to $\{\Omega_n\}$ (common frequency components of source signals).

(3) Removing common frequency components of source signals from $s_{1p(t)}$ and $s_{2p(t)}$ result in two residue signals, $\widetilde{s}_{1p(t)}$ and $\widetilde{s}_{2p(t)}$, that are uncorrelated:

$$\widetilde{s}_{1p}(t) = \sum_{k \neq n} \mathbf{a}_k.\exp(j\omega_{1k}t) \qquad (29)$$

$$\widetilde{s}_{2p}(t) = \sum_{l \neq n} \mathbf{b}_l.\exp(j\omega_{2l}t) \qquad (30)$$

$$E\{\widetilde{s}_{1p}(t+\tau).\widetilde{s}_{2p}^*(t)\} = 0 \qquad (31)$$

Hence,

$$x_{ip(t)} = [\alpha_i\widetilde{s}_{1p(t)} + \beta_i\widetilde{s}_{2p(t)}] +$$

$$\sum_n (\alpha_i\mathbf{a}_n + \beta_i\mathbf{b}_n).\exp(j\Omega_n t)$$

$$= \widetilde{x}_{ip(t)} + \sum_n (\alpha_i\mathbf{a}_n + \beta_i\mathbf{b}_n).\exp(j\Omega_n t)$$

$$\text{for } i = 1,2 \qquad (32)$$

and Relation 25 can be rewritten as:

$$P_{ijp}^{x}(\omega) = \sum_{q \neq n} 2\pi . E\{\mathbf{d}_{iq}\mathbf{d}_{jq}{}^{*}\} \delta(\omega - \omega_{q})$$
$$+ \sum_{n} 2\pi . E\{\mathbf{d}_{in}\mathbf{d}_{jn}{}^{*}\} \delta(\Omega - \Omega_{n}) \qquad (33)$$

Removing the terms corresponding to common frequency components:

$$\widetilde{P}_{ijp}^{x}(\omega) = \sum_{q \neq n} 2\pi . E\{\mathbf{d}_{iq}\mathbf{d}_{jq}{}^{*}\} \delta(\omega - \omega_{q}) \qquad (32)$$

from which the desired correlation functions are obtained:

$$\widetilde{r}_{ijp}^{x}(\tau) = F^{-1}\{\widetilde{P}_{ijp}^{x}(\omega)\} \qquad (34)$$

$$= E\{\widetilde{x}_{ip\,(t+\tau)}\widetilde{x}_{jp}{}^{*}{}_{(t)}\} \qquad \text{for } i = 1,2$$

$$= E\{(\alpha_{i}\widetilde{s}_{1p(t)} + \beta_{i}\,\widetilde{s}_{2p(t)}).(\alpha_{j}\widetilde{s}_{1p(t+\tau)} + \beta_{j}\,\widetilde{s}_{2p(t+\tau)})^{*}\}$$

At last, because $\widetilde{s}_{1p\,(t)}$ and $\widetilde{s}_{2p\,(t)}$, are uncorrelated, we get the following matrix form:

$$\widetilde{\mathbf{R}}_{p}^{x}(\tau) \overset{\Delta}{=} \begin{bmatrix} \widetilde{r}_{11p}^{x}(\tau) & \widetilde{r}_{12p}^{x}(\tau) \\ \widetilde{r}_{21p}^{x}(\tau) & \widetilde{r}_{22p}^{x}(\tau) \end{bmatrix}$$

$$= \begin{bmatrix} \alpha_{1} & \beta_{1} \\ \alpha_{2} & \beta_{2} \end{bmatrix} \begin{bmatrix} \widetilde{r}_{11p}^{s}(\tau) & 0 \\ 0 & \widetilde{r}_{22p}^{s}(\tau) \end{bmatrix} \begin{bmatrix} \alpha_{1} & \alpha_{2} \\ \beta_{1} & \beta_{2} \end{bmatrix}^{*}$$

$$= \mathbf{A}.\widetilde{\mathbf{R}}_{p}^{s}(\tau).\mathbf{A}^{H} \qquad (35)$$

where H denotes complex conjugate transpose, and

$$\widetilde{r}_{iip}^{s}(\tau) = E\{\widetilde{s}_{ip}(t+\tau)\widetilde{s}_{ip}{}^{*}(t)\} \quad ; \, i = 1,2 \qquad (36)$$

It is seen that in (35) the matrix, which is related to source signals, is diagonal (a desired condition). This representation is the basis of an algorithm for estimating mixing matrix $\mathbf{A}$.

4. BLIND SOURCE SEPARATION ALGORITHM

In this section, an algorithm for estimating $\mathbf{A}$ (and recovering source signals) is proposed which is based on the model embedded in Equation 32 and restated in Equation 37 using second order statistics.

$$\widetilde{x}_{ip}(t) = \alpha_{i}\,\widetilde{s}_{1p}(t) + \beta_{i}\,\widetilde{s}_{2p}(t) \qquad ; \, i = 1,2 \qquad (37)$$

Steps of the algorithm are following:

4.1 Orthogonalization Although $\widetilde{s}_{1p\,(t)}$ and $\widetilde{s}_{2p\,(t)}$ are uncorrelated and based on the discussion in II.B, we can assume that:

$$\widetilde{\mathbf{R}}_{p}^{s}(0) = \mathbf{I} \qquad (38)$$

According to Equation 37, $\widetilde{x}_{1p\,(t)}$ and $\widetilde{x}_{2p\,(t)}$ are correlated. Hence, we apply Karhunen-Loeve orthogonalization transformation on $\widetilde{x}_{1p\,(t)}$ and $\widetilde{x}_{2p\,(t)}$. The orthogonalizer matrix is obtained from eigendecomposition of matrix $\widetilde{\mathbf{R}}_{p}^{x}(\tau)$ at $\tau = 0$. If eigenvalues of $\widetilde{\mathbf{R}}_{p}^{x}(0)$ are denoted by $\lambda_{1} \& \lambda_{2}$ and $\mathbf{v}_{1} \& \mathbf{v}_{2}$ are the corresponding eigenvectors, the orthogonalization matrix $\mathbf{T}$, defined by:

$$\mathbf{T} = [\frac{1}{\lambda_{1}}\mathbf{v_{1}}, \frac{1}{\lambda_{2}}\mathbf{v_{2}}]^{H} \qquad (39)$$

satisfies:

$$\mathbf{T}.\widetilde{\mathbf{R}}_{p}^{x}(0).\mathbf{T}^{H} = \mathbf{I} \qquad (40)$$

Also from (35),(38),(40), it is seen that :

$$\mathbf{T}.\mathbf{A}.\widetilde{\mathbf{R}}_{p}^{s}(0).\mathbf{A}^{H}.\mathbf{T}^{H} = \mathbf{T}.\mathbf{A}.\mathbf{A}^{H}.\mathbf{T}^{H} = \mathbf{I} \qquad (41)$$

This equation shows that matrix $\mathbf{U} = \mathbf{T}.\mathbf{A}$, is a unitary matrix. As a consequence mixing matrix $\mathbf{A}$ can be factored as:

$$\mathbf{A} = \mathbf{T}^{-1}.\mathbf{U} \qquad (42)$$

It is important to note here that this orthogonalization procedure changes the problem from the determination of mixing matrix $\mathbf{A}$ to that of a unitary matrix $\mathbf{U}$.

4.2. Estimation of $\mathbf{U}$, $\mathbf{A}$ and $\mathbf{s}(t)$

By applying orthogonalization matrix $\mathbf{T}$ to equation (35) for some $\tau \neq 0$,

$$\overline{\mathbf{R}}_p^x(\tau) \stackrel{\Delta}{=} \mathbf{T}.\widetilde{\mathbf{R}}_p^x(\tau).\mathbf{T}^H \qquad ; \ \forall \tau \neq 0$$
$$= \mathbf{T}.\mathbf{A}.\widetilde{\mathbf{R}}_p^s(\tau).\mathbf{A}^H.\mathbf{T}^H \qquad (43)$$

Hence,

$$\overline{\mathbf{R}}_p^x(\tau) = \mathbf{U}.\widetilde{\mathbf{R}}_p^s(\tau).\mathbf{U}^H \qquad (44)$$

where matrix $\overline{\mathbf{R}}_p^x(\tau)$ is called orthogonal correlation matrix.

Since $\mathbf{U}$ is unitary and $\widetilde{\mathbf{R}}_p^s(\tau)$ is diagonal, equation (44) states that orthogonal correlation matrix $\overline{\mathbf{R}}_p^x(\tau)$ is diagonalized by the unitary transformation $\mathbf{U}$ (unitary diagonalization). In other words unitary matrix $\mathbf{U}$ can be specified by unitary diagonalizing of orthogonal correlation matrix $\overline{\mathbf{R}}_p^x(\tau)$ for some lag $\tau \neq 0$. This property is a result of spectral theorem for hermitian matrices [27]. It states that a hermitian (and/or normal) matrix $\mathbf{M}$ is unitarily diagonalizable ,i.e. there exists a unitary matrix $\mathbf{U}$ and a diagonal matrix $\mathbf{D}$ such that $\mathbf{M} = \mathbf{U}.\mathbf{D}.\mathbf{U}^H$.

Since $\overline{\mathbf{R}}_p^x(\tau)$ is hermitian (and normal) and equation (44) is satisfied , the existence of unitary matrix $\mathbf{V}$ such that for any time lag τ and a diagonal matrix $\mathbf{D}$,

$$\mathbf{V}^H.\overline{\mathbf{R}}_p^x(\tau).\mathbf{V} = \mathbf{D} \qquad (45)$$

is guaranteed. The essential aim is finding a unique unitary matrix $\mathbf{U}$ that diagonalizes $\overline{\mathbf{R}}_p^x(\tau)$ for all time lags τ.

A method for attaining this aim, as used in most BSS approaches that exploit statistical properties, is joint diagonalization(JD) method (see the Appendix) which operates as simultaneous diagonalization of the set $\{\overline{\mathbf{R}}_p^x(\tau_i) \, | \, i = 1,2,...,K\}$ of K orthogonal correlation matrices and is described in the following theorem[14]:

THEOREM. Let $\tau_1, \tau_2, ..., \tau_K$ be K nonzero time lags, and let $\mathbf{V}$ be a unitary matrix such that :

$$\forall 1 \leq k \leq K \quad \mathbf{V}^H.\overline{\mathbf{R}}_p^x(\tau).\mathbf{V} = \mathrm{diag}.[d_1(k), ..., d_n(k)] \qquad (46)$$

$$\forall 1 \leq i \neq j \leq n \quad \exists k, \ni \ d_i(k) \neq d_j(k) \qquad (47)$$

then:

- $\mathbf{V}$ is essentially equal to $\mathbf{U}$ (desired unique unitary matrix)
- A permutation can be operated on diagonal elements of $\mathrm{diag}.[d_1(k), ..., d_n(k)]$.(In our case n = 2)

This theorem is a consequence of "essential uniqueness of joint diagonalization" theorem (see Appendix) which states that a unique unitary matrix $\mathbf{U}$ can be determined if for at least a $\tau \neq 0$, eigenvalues of $\overline{\mathbf{R}}_p^x(\tau)$ are distinct, a condition that is surely satisfied for sources with different spectra.

After determination of a unique unitary matrix $\mathbf{U}$, $\mathbf{A}$ can be computed from $\mathbf{A} = \mathbf{T}^{-1}.\mathbf{U}$, and consequently the source signals are estimated as

$$\mathbf{s}(t) = \mathbf{A}^{-1}.\mathbf{x}(t) \qquad (48)$$

It is important to note that for computing $\mathbf{s}(t)$, we use observation data $\mathbf{x}(t)$ (not $\widetilde{\mathbf{x}}(t)$), so there isn't any information loss.

5. SIMULATION RESULTS

In this section, the performance of the proposed method is investigated via computer simulation results.

Source signals $\left(s_{i\,(t)\,;\,i\,=\,1,2}\right)$ are composed of regular and predictable parts. Regular component in each source consists of a zero-mean normal process (independent from the other) and a uniform process (common between two sources). Predictable parts consist of random amplitudes sinusoidal functions with some common frequency and correlated amplitude components. These signals are mixed by an arbitrary 2×2 mixing matrix $\mathbf{A}$, and corrupted by AWGN, to obtain the observation signals $\left(x_{i\,(t)\,;\,i\,=\,1,2}\right)$. Then algorithm is applied on observed data and the estimation of $\mathbf{A}$, $\hat{\mathbf{A}}$, is obtained. This procedure is repeated for G=500 independent trials.

To evaluate the approach, a performance index is introduced. Since, it is ideally expected that the product $\mathbf{Q} \overset{\Delta}{=} \hat{\mathbf{A}}^{-1}.\mathbf{A}$ (called unmixing matrix) equals identity matrix $\mathbf{I}_{2 \times 2}$ (complete separation), the following performance index (PI) is applied,

$$ PI = 10.\log_{10} \left[\frac{1}{G} \sum_{g=1}^{G} \left\| \hat{\mathbf{A}}^{-1}.\mathbf{A} - \mathbf{I} \right\|_F^2 \right] \qquad (49) $$

where $\left\| . \right\|_F$ is the Frobenius norm. Actually, PI shows the distance between unmixing matrix $\mathbf{Q}$ (obtained in experiment) and the ideal identity matrix $\mathbf{I}_{2 \times 2}$.

We performed and compared two experiments: without pre-separation process (experiment #1) and with pre-separation process (experiment #2). The experiments were executed under noise free and SNR=3,5,8,10 (dB) conditions for various number of correlation matrices used in JD algorithm, and with different correlation coefficient of original source signals. Results are illustrated in Figures 1-8.

Figures 1-3, show performance index (PI)(in dB) versus the number of jointly diagonalized correlation matrices for experiments #1 and #2. In these figures correlation coefficient is 0.5, and each figure has been plotted for a fixed SNR. In Figures 4 and 5, corresponding to experiment #1 and #2, the performance indexes (in dB) versus several SNR (in dB) have been plotted for some constant number of jointly diagonalized correlation matrices with correlation coefficient 0.5. In Figures

6-8, SNR and the number of jointly diagonalized correlation matrices have been kept constant, and correlation coefficient of source signals has been varied from 0.1 to 0.9. These figures illustrate PI versus correlation coefficient for each experiments #1 and #2.

Almost in all figures, better performance of
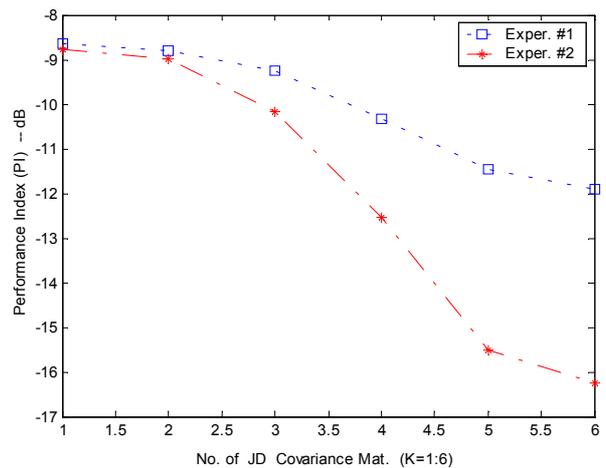
Figure 1. Performance versus number of JD covariance matrices (K = 1:6): [Noise Free] and [Correlation Coefficient = 0.5].
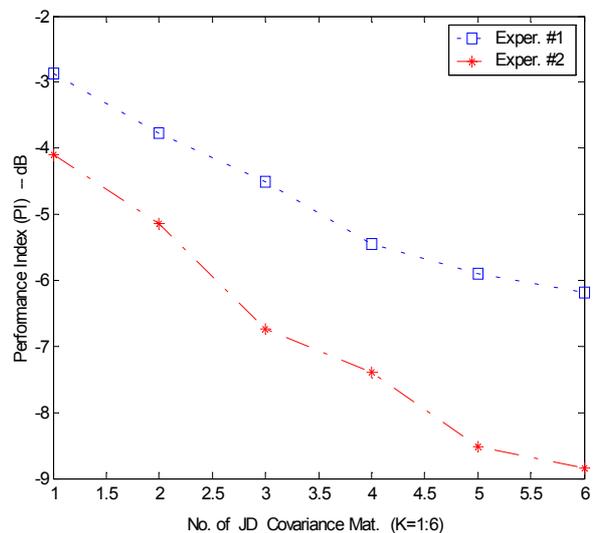
Figure 2. Performance versus number of JD Covariance Matrices (K = 1:6): [SNR = 3 dB] and [Correlation Coefficient = 0.5].
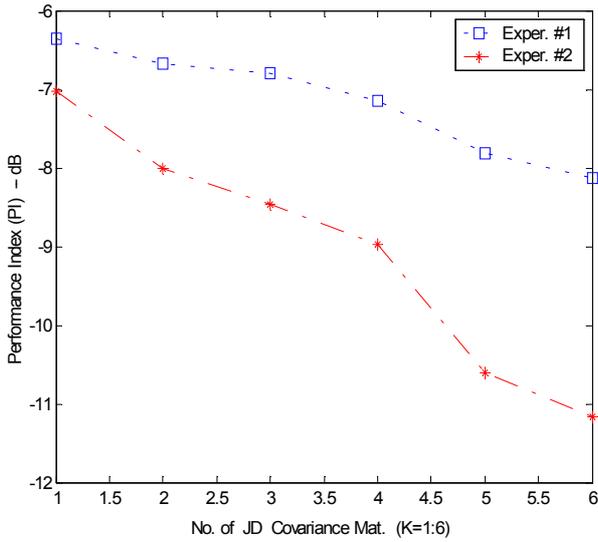
Figure 3. Performance versus number of JD covariance matrices (K = 1:6): [SNR = 8 dB] and [Correlation Coefficient = 0.5].
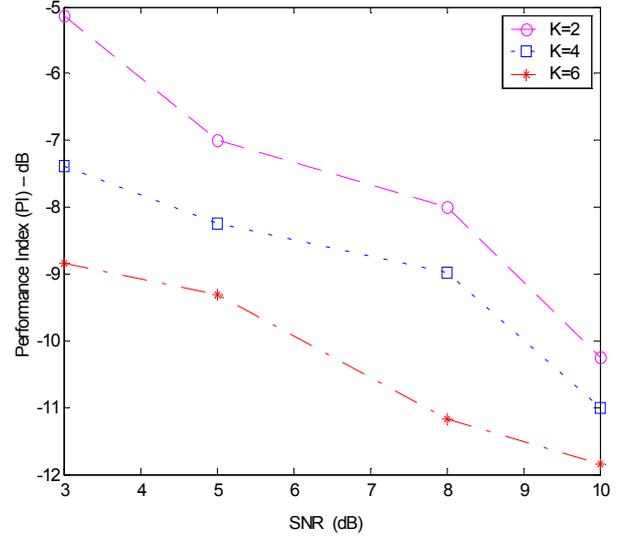
Figure 5. Performance Versus SNR for Experiment #2: [K(No. of JD Covariance Mat.) = 2,4,6] and [Correlation Coefficient = 0.5].
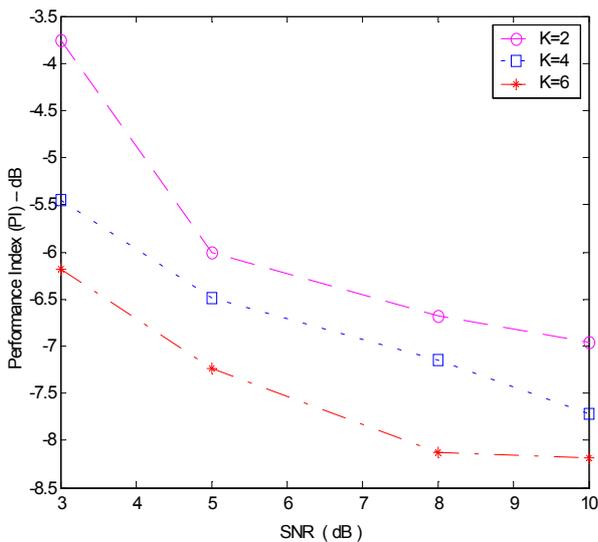
Figure 4. Performance versus SNR for Experiment #1: [K(No. of JD Covariance Mat.) = 2,4,6] and [Correlation Coefficient = 0.5].
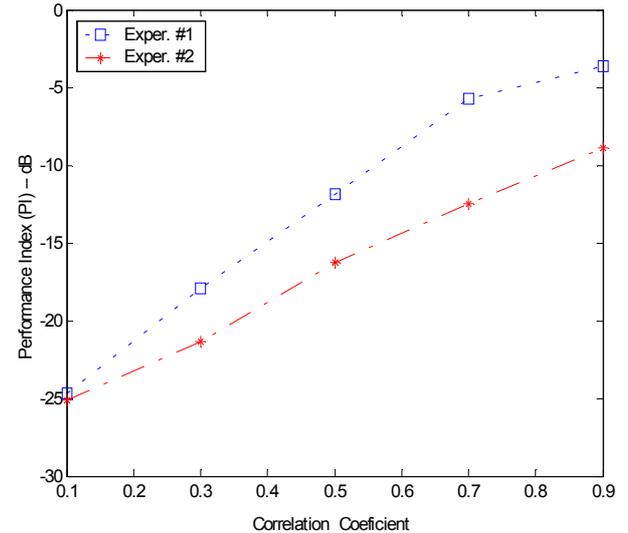
Figure 6. Performance versus correlation coefficient: [Noise Free] and [K (No. Of JD Covariance Mat.) = 6].

proposed algorithm is evident. In Figures 1-3, it is obvious that the performances of two experiments become better as the number of the jointly diagonalized correlation matrices is increased, but this has a limit (as it is seen from small different between PI for K=5 and K = 6).

Figures 4 and 5 show improvement in performance by increasing SNR. From Figures 6-8, it is revealed that the performance index of two experiments are close for small correlation coefficients, and as the correlation coefficient is increased, the PI of two
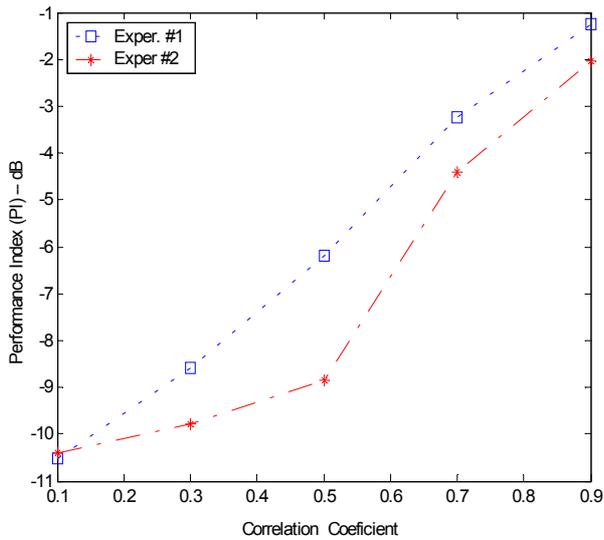
Figure 7. Performance versus correlation coefficient: [SNR = 3 dB] and [K(No. of JD Covariance Mat.) = 6].
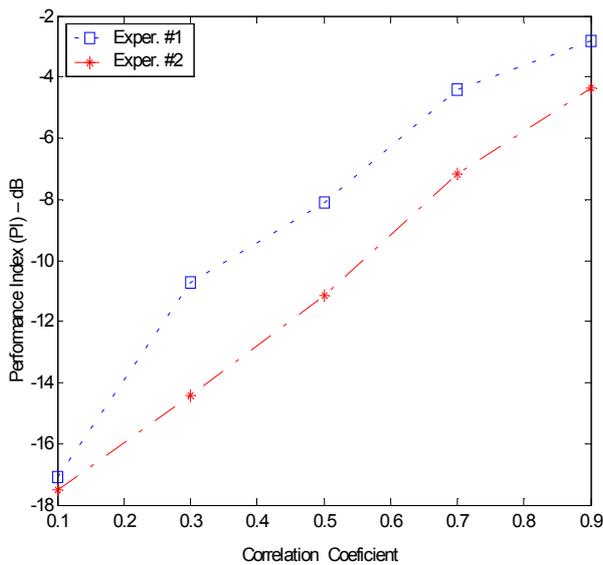
Figure 8. Performance versus correlation coefficient: [SNR = 8 dB] and [K (No. Of JD Covariance Mat.) = 6.

experiments are decreased, but the performance in experiment #2 is better than that in experiment #1, particularly for intermediate correlation coefficients.

For visible demonstration of proposed

algorithm another experiments were performed by considering two simple signals plotted in Figure 9. The combination signals $(x_i(t)\ ;\ i = 1,2)$ in Figures 10 and 13 corresponding to free noise and SNR = 5dB cases, have obtained by applying an arbitrary matrix **A** on the original signals. Figures 11 and 14 show separated signals, which have been extracted using proposed algorithm. Differences of the original signals and the separated signals have been plotted in Figures 12 and 15.

Because there are two indeterminacies in the solution of problem, noticed in section II-B, it must be mentioned that in these experiments the separated signals have been modified knowing mixture matrix **A** for better comparing with original signals and computing their differences. So it seems that this criterion is not very suitable for general and more complicated cases although it depicts better the performance of the algorithm for simple cases. However, the effectiveness of the proposed algorithm is evident in these experiments, too.

6. CONCLUSION

In this paper, an approach for solving BSS problem in the cases where source signals are correlated is introduced without additional assumptions on signal or mixing matrix structures.

An important step of this BSS algorithm is a pre-separation procedure where based on Wold decomposition principle, the information of predictable part of source signals (i.e. uncorrelated

parts of predictable signals) is derived. The diagonal structure of the correlation matrix of this part is essential for next step of algorithm where by using the second-order based method and JD technique; separation process is completed by

estimating $\hat{\mathbf{A}}$ and recovering $\hat{\mathbf{s}}(t) = \begin{bmatrix} \hat{s}_1(t) \\ \hat{s}_2(t) \end{bmatrix}$.

Simulation results show effectiveness of the algorithm.
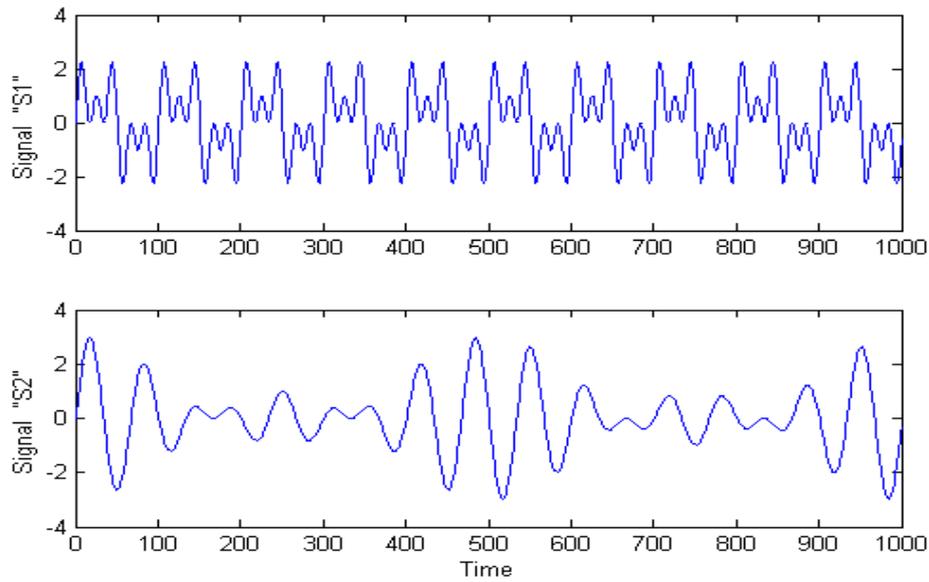
Figure 9. Original signals $S_1(t)$ and $S_2(t)$.
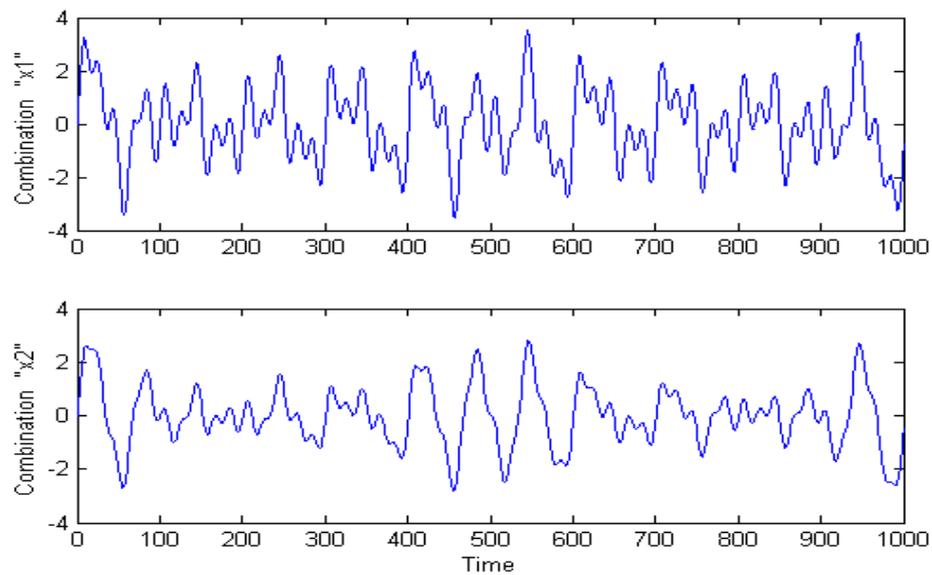
Figure 10. Two combinations of original signals $X_1(t)$ and $X_2(t)$ [Noise Free].

7. APPENDIX

Joint Diagonalization   Joint diagonalization of several matrices defines a kind of "average eigenstructure" shared by several matrices. It has been studied in [28] from the point of view of
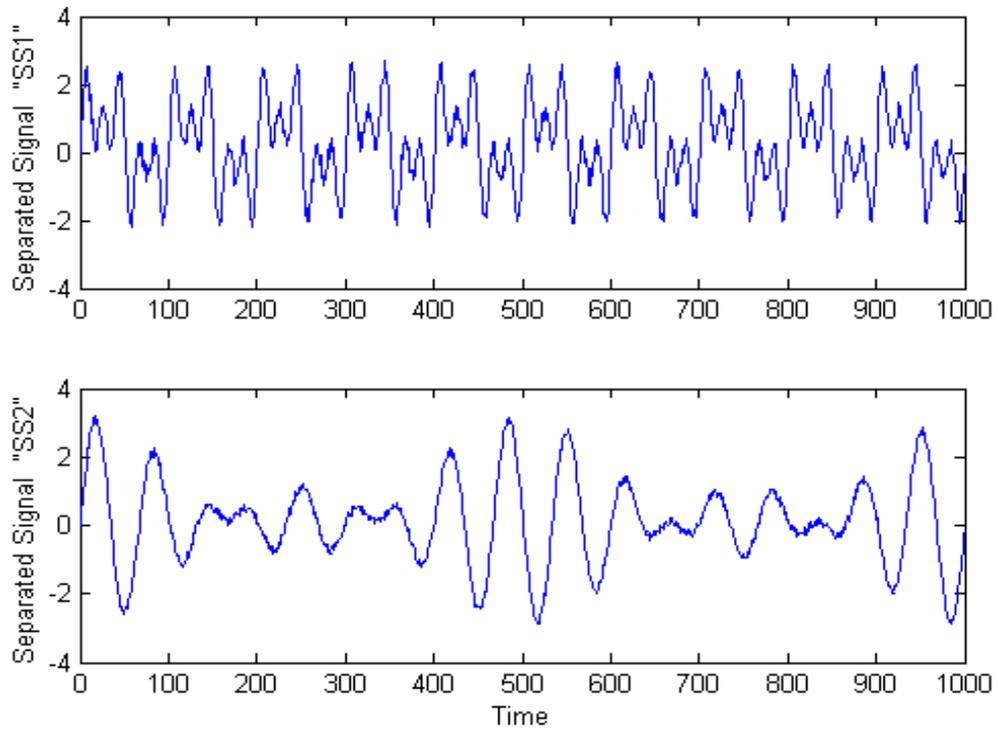
Figure11. Separated signals SS_1 (t) and SS2 (t).
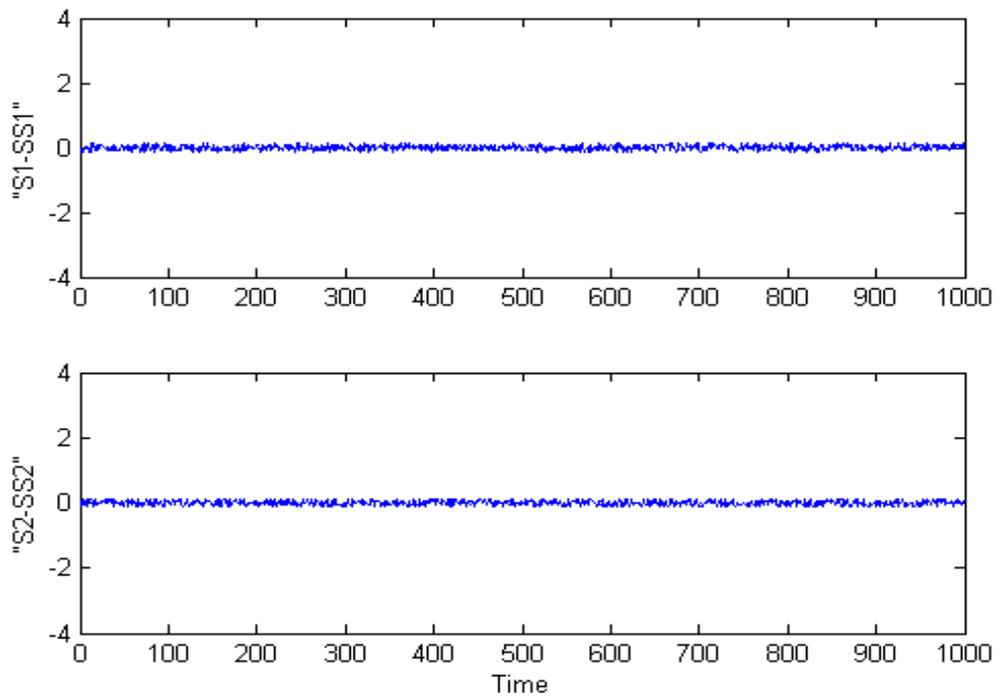
Figure 12. Difference of original signal with separated signals [Noise Free].
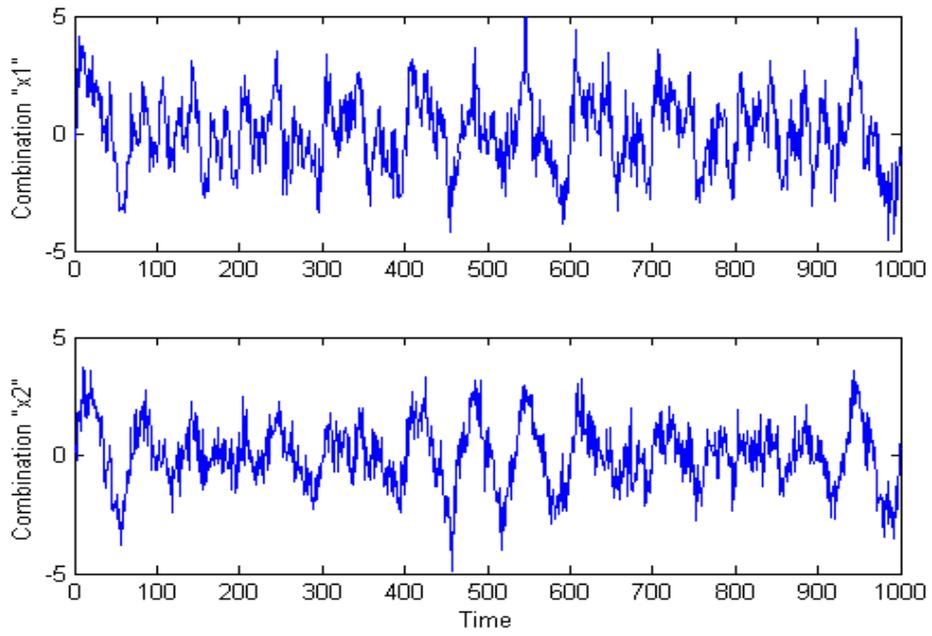
Figure 13. Two combination of original signals X(t) and X(t) [SNR 5 dB].
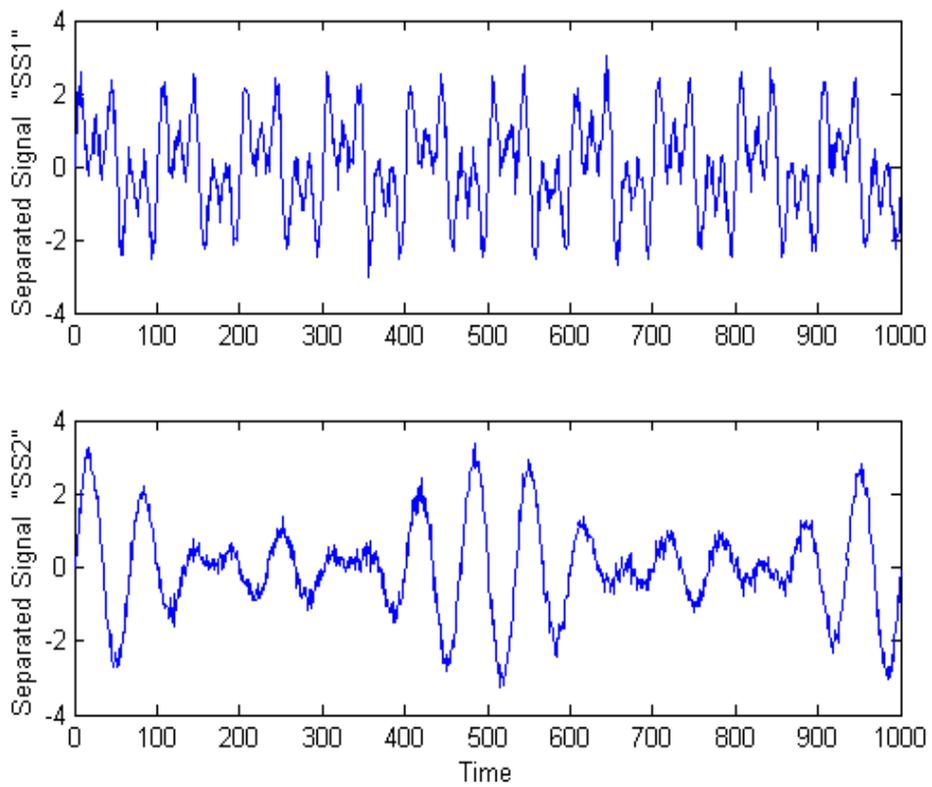
Figure 14. Separated signals: SS$_1$(t) and SS$_2$(t) [SNR 5dB].

numerical analysis.

The "off" of a $n \times n$ matrix $\mathbf{M}$ is defined as:

$$\text{off}(\mathbf{M}) \overset{\Delta}{=} \sum_{1 \le i \ne j \le n} \left| M_{ij} \right|^2 \qquad (50)$$

where M_{ij} denotes the (i,j)-th entry of matrix $\mathbf{M}$. The unitary diagonalization of a single matrix $\mathbf{M}$ by a unitary matrix $\mathbf{V}$ is equivalent to zeroing $\text{off}(\mathbf{V}^H \mathbf{M} \mathbf{V})$. In addition, if a matrix $\mathbf{M}$ is in the form $\mathbf{M} = \mathbf{U}\mathbf{D}\mathbf{U}^H$, where $\mathbf{U}$ is unitary and $\mathbf{D}$ is diagonal with distinct diagonal elements, then it may be unitarily diagonalized only by matrices that are essentially equal to $\mathbf{U}$, that is, if

$\text{off}\ (\mathbf{V}^H \mathbf{M} \mathbf{V}) = 0$.

Similarly, joint diagonalization of a set $M = \{\mathbf{M}_k\}_{k=1}^K$ of K, $n \times n$ matrices is achieved by any unitary matrix $\mathbf{V}$ which minimize the "joint diagonality" criterion, that is defined as:

$$\Gamma(\mathbf{M}, \mathbf{V}) \overset{\Delta}{=} \sum_{k=1}^K \text{off}(\mathbf{V}^H \mathbf{M}_k \mathbf{V}) \qquad (51)$$

the unitary matrix $\mathbf{V}$ that minimizes (51) over the set of all unitary matrices is called "joint diagonalizer" of the set M. At this issue, the uniqueness of a joint diagonalizer is more interested for BSS problems, which following theorem exists:

THEOREM. *Essential Uniqueness of Joint*
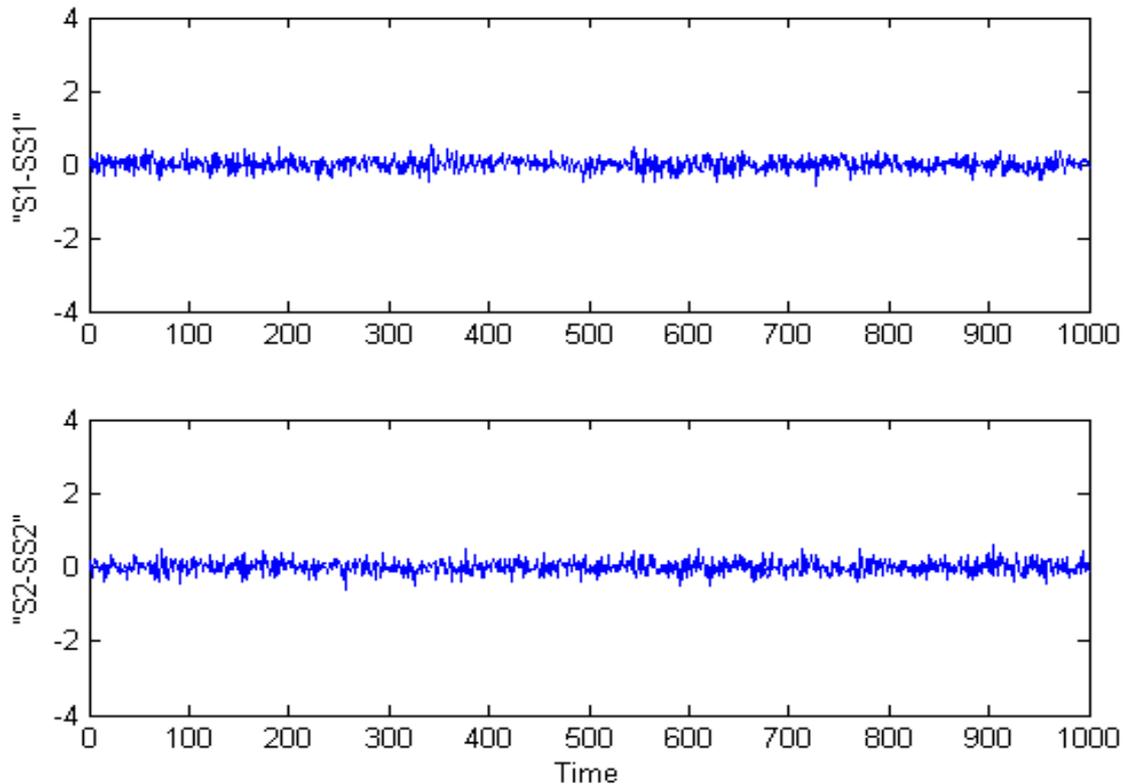
Figure 15. Difference of original signals and separated signals [SNR. 5dB].

Diagonalization: Let $M = \{\mathbf{M}_k\}_{k=1}^{K}$ be a set of K matrices where, for $1 \le k \le K$, matrix $\mathbf{M}_k$ is in the form $\mathbf{M}_k = \mathbf{U}\mathbf{D}_k\mathbf{U}^H$ with $\mathbf{U}$ a unitary matrix, and $\mathbf{D}_k = \text{Diag}.[d_1(k),...d_n(k)]$. Any "joint diagonalizer" of M is essentially equal to $\mathbf{U}$ if and only if

$$\forall 1 \le i \ne j \le n \;, \exists k \ni 1 \le k \le K \quad d_i(k) \ne d_j(k) \tag{52}$$

8. REFERENCES

1. Krim, H., and Viberg, M., "Two Decade of Array Signal Processing Research, The Parametric Approach", *IEEE Signal Processing Magazine*, Vol. 13, (July 1996), 67-94.
2. Gonen, E., and Mendel, J. M., "Applications of Cummulants to Array Processing. III. Blind Beam Forming for Coherent Signals", *IEEE Tran. Signal Processing*, Vol. 45, No. 9, (September 1997), 2252-2264.
3. Cichocki, A., Rutkowski, T., Barros, A. K., and Sang-Hoon Oh, "A Blind Extraction of Temporally Correlated but Statistically Dependent Acoustic Signals", *Proceedings of the 2000 IEEE Signal Processing Society Workshop*, Vol. 1, (2000), 445.
4. Demoment, G., " Image Reconstruction and Restoration: Overview of Common Estimation Structures and Problems", *IEEE Transactions of Acoustics, Speech and Signal Processing*, Vol. 37, No. 12, (1989), 2024-2036.
5. Van Der Veen, A.-J., "Algebraic Methods for Deterministic Blind Beam Forming", *Proceedings of the IEEE*, Vol. 86, No. 10, (October 1998), 1987-2008.
6. Anand, K., Mathew, G., and Reddy, V. U., "Blind Separation of Multiple Co-Channel BPSK Signals Arriving at an Antenna Array", *IEEE Signal Processing Letters*, Vol. 2, (September 1995), 176-178.
7. Berrah, C. M., "Parametric Yield Estimation for a MOSFET Integrated Circuit", *IEEE International Symposium on Circuits and Systems*, (1990), Vol. 3, 2260-2263.
8. Parra, L., Spence, C., and De Vries, B., "Convolutive Blind Source Separation Based on Multiple Decorrelation", *Proceedings of the 1998 IEEE Signal Processing Society Workshop*, (1998), 22-32.
9. Sabala, I., Cichocki, A. and Amari, S. I., "Relationships Between Instantaneous Blind Source Separation and Multichannel Blind Deconvolution", *Neural Networks Proceedings, IEEE World Congress on Computational Intelligence*, Volume 1, (1998), 39-44.
10. Comon, P., "Independent Component Analysis, a New Concept", *Signal Processing*, Vol. 36, No. 3, (April 1994), 287-314.
11. Cardoso, J. F., "Blind Signal Separation: Statistical Principles", *Proceedings of the IEEE*, Vol. 86, No. 10, (October 1998), 2009-2025.
12. Pham, D. T., "Blind Separation of Instantaneous Mixture of Sources via an Independent Component Analysis", *IEEE Transactions on Signal Processing*, Vol. 44, No. 11, (November 1996), 2768-2779.
13. Belouchrani, A., Abed-Meraim, K., Cardoso, J. F. and Moulines, E., "A Blind Source Separation Technique Using Second-Order Statistics", *IEEE Transactions on Signal Processing*, Volume 45, Number 2, (February 1997), 434-444.
14. Tong, L., Liu, R. W., Soon, V. C. and Huang, Y. F., "Indeterminacy and Identifiability of Blind Identification", *IEEE Transactions of Circuits and Systems*, Vol. 38, No. 5, (May 1991), 499-509.
15. Agee, B. G., Schell, S. V. and Gardner, W. A., "Spectral Self-Coherence Restoral: a New Approach to Blind Adaptive Signal Extraction Using Antenna Arrays", *Proceedings of the IEEE*, Vol. 78, No. 4, (April 1990), 753-767.
16. Ferreol, A. and Chevalier, P., "On the Fourth-Order Cumulants Estimation for the HO Blind Separation of Cyclostationary Sources", *ICASSP-98*, Vol. 4, (1998), 2313-2316.
17. Van Der Veen, A. J. and Paulraj, A., "An Analytical Constant Modulus Algorithm", *IEEE Transactions Signal Processing*, Vol. 44, No. 5, (May 1996), 1136-1155.
18. Treichler, J. R. and Agee, B. G., "A New Approach to Multipath Correction of Constant Modulus Signals", *IEEE Transactions ASSP*, Vol. 31, (April 1983), 459-471.
19. Torlak, M., Hansen, L. K. and Xu, G., "A Fast Bind Sequence Estimation in Digital Communications", *Conference on Signals, Systems and Computers*, Vol. 1, (1997), 833-837.
20. Schmidt, R. O., "Multiple Emitter Location and Signal Parameter Estimation", *IEEE Transactions of Antennas Propagation*, Vol. 34, (March 1991), 276-280.
21. Roy, R. and Kailath, T., "ESPRIT-Estimation of Signal Parameters Via Rotational Invariance Techniques", *IEEE Transactions on Acoustics, Speech and Signal Processing*, Vol. 37, No. 7, (July 1989), 984-995.
22. Shan, T. J. and Wax, M. and Kailath, T., "On Spatial Smoothing for Direction of Arrival Estimation of Coherent Signals", *IEEE Transactions of ASSP*, Vol. 33, (April 1985), 806-811.
23. Cao, X. R. and Liu, R. W., "General Approach to Blind Source Separation", *IEEE Transactions on Signal Processing*, Vol. 44, No. 3, (March 1996), 562-571.
24. Papoulis, A., "Predictable Processes and Wold's Decomposition: A Review", *IEEE Transactions of ASSP*, Vol. 33, No. 4, (August 1985).

25. Papoulis, A., "Probability, Random Variables, and Statistic Process", Third Ed., McGraw-Hill Book Com., NY, (1991).

26. Golub, G. H. and Loan, C. F. V., "Matrix Computations", Baltimore MD: Johns Hopkins University Press, MD, (1989).

27. Bunse-Gerstner, A. Byers, R. and Mehrmann, V., "Numerical Methods for Simultaneous Diagonalization", *SIAM J. Matrix Anal. Applications*, Vol. 14, (1993), 927-949.