PAPER Special Section on Blind Signal Processing: Independent Component Analysis and Signal Separation

# Robust Independent Component Analysis via Time-Delayed Cumulant Functions

Pando GEORGIEV[†,††], *Nonmember and* Andrzej CICHOCKI[†,†††], *Member*

## 1.

**SUMMARY** In this paper we consider blind source separation (BSS) problem of signals which are spatially uncorrelated of order four, but temporally correlated of order four (for instance speech or biomedical signals). For such type of signals we propose a new sufficient condition for separation using fourth order statistics, stating that the separation is possible, if the source signals have distinct normalized cumulant functions (depending on time delay). Using this condition we show that the BSS problem can be converted to a symmetric eigenvalue problem of a generalized cumulant matrix $\mathbf{Z}^{(4)}(\mathbf{b})$ depending on $L$-dimensional parameter $\mathbf{b}$, if this matrix has distinct eigenvalues. We prove that the set of parameters $\mathbf{b}$ which produce $\mathbf{Z}^{(4)}(\mathbf{b})$ with distinct eigenvalues form an open subset of $\mathbb{R}^L$, whose complement has a measure zero. We propose a new separating algorithm which uses Jacobi's method for joint diagonalization of cumulant matrices depending on time delay. We empasize the following two features of this algorithm: 1) The optimal number of matrices for joint diagonalization is 100 - 150 (established experimentally), which for large dimensional problems is much smaller than those of JADE; 2) It works well even if the signals from the above class are, additionally, white (of order two) with zero kurtosis (as shown by an example).
***key words:*** *Independent component analysis, blind source separation, eigenvalue decomposition, cumulant functions, join diagonalization.*

## 2. Introduction

The interest of blind signal processing, especially, the problem of Independent Component Analysis (ICA) and Blind Source Separation (BSS) has been increased recently, due to its potential applications in many areas, including brain signal processing and other biomedical signal processing, speech enhancement, wireless communication, geophysical data processing, data mining, etc.

The problem is formulated as follows: we can observe sensor signals $\mathbf{x}(k) = [x_1(k), \ldots, x_n(k)]^T$ which are modeled as

$$\mathbf{x}(k) = \mathbf{H}\mathbf{s}(k) + \mathbf{n}(k), \tag{1}$$

where $\mathbf{H}$ is $n \times n$ non-singular unknown mixing matrix, $\mathbf{s}(k) = [s_1(k), \ldots, s_n(k)]^T$ is a vector of unknown zero mean source signals and $\mathbf{n}(k)$ is a vector of additive zero mean noise. Our objective is to estimate the mixing matrix $\mathbf{H}$ and/or a separating matrix $\mathbf{W} = \mathbf{H}^{-1}$ and source signals assuming that they are uncorrelated or statistically independent.

In this paper we consider two general cases or models of source signals and additive noise. In the first model we assume that source signals are spatially uncorrelated and colored, and the noise is white with arbitrary distribution. In the second model we assume that source signals are statistically independent (or more generally, spatially uncorrelated of order 4) and colored in order 4, and the noise is white of order 4. Both models are applicable to speech and biomedical signals. More specifically, the second model is applicable to signals, which are spatially uncorrelated of order 4, but temporally uncorrelated of order 2 and correlated of order 4. Such type of signals arise in digital communications (see for instance [23]) and recently are considered in [20], where a blind deconvolution algorithm for such type of signals is developed.

We introduce a new sufficient condition for blind separation (see condition (**DNCF**$(P)$) below) requiring the sources to have different normalized cumulant functions of fourth order (depending on time delay) on a given set of delays $P$. This condition can be considered as a generalization of those ones described in [9] and [25] for second order statistics and used in [8]. It is interesting to mention that a related condition, sufficient for deconvolution problems, using only second order correlation functions and expressed by power spectral matrices is proposed in [19].

The use of second order statistics approach for blind separation of temporally correlated sources has been developed and analyzed by many researchers, including [1], [3, 4], [8-12], [18-19], [21], [24-25], etc.

We present two approaches to the BSS problem. The first one has an unified form for the second order statistic and high order statistics, and uses Eigen-Value Decomposition (EVD) of sum of covariance or cumulant matrices. Even for second order statistics this approach has some advantages that may not be found in others known results at the same time. It allows us to control

Manuscript received August 6, 2002.
Manuscript revised October 11, 2002.
[†]The authors are with the Lab. for Advanced Brain Signal Processing, Brain Science Institute, RIKEN, Wako-shi, Saitama 351-0198, Japan.
[††]On leave from the Sofia University "St. Kl. Ohridski", Bulgaria
[†††]On leave from the Warsaw University of Technology, Poland

successfulness of the separation by observing the distribution of the eigenvalues; it provides relative fast convergence (since several algorithm has been developed for the EVD with cubic convergence [15], [16]); it can solve large scale problems due to efficiency of available EVD algorithms; it extracts the components simultaneously; does not need the sources to be stationary; does not need that all but one signal to be Gaussian; and it is robust with respect to additive noise what often leads to smaller errors (cross-talking between estimated sources).

Such an approach (symmetric EVD) is suitable if the source signals are uncorrelated enough (for second order statistic approach) or are independent enough (for four order statistic approach). If not, more suitable are joint diagonalization of covariance matrices (SOBI algorithm) and joint diagonalization of cumulant matrices (JADE algorithm).

Here we propose a joint diagonalization of cumulant matrices of fourth order depending on time delays - this is our second approach to the BSS problem for colored signals of order four, allowing to extract signals even with zero kurtosis, which are white of order two. The computer experiments showed that the new algorithm is quite promising in high dimensional problems, when JADE algorithm fails due to high computation time. For small dimensions its performance is comparable with those of JADE and in some cases is better (for signals with kurtosis near to zero).

## 3. Covariance and Cumulant Matrices

We shall consider a concrete form of fourth order cumulants and note that generalizations to high order cumulants is straightforward.

Define a covariance matrix of the sensor (resp. source) signals by

$$\mathbf{R_x}(p) = E\{\mathbf{x}\mathbf{x}_p^T\}, \tag{2}$$

$$\text{(resp. } \mathbf{R_s}(p) = E\{\mathbf{s}\mathbf{s}_p^T\}), \tag{3}$$

where

$$\mathbf{x}_p = \mathbf{x}(t-p), \quad \mathbf{s}_p = \mathbf{s}(t-p), \tag{4}$$

and $E$ is the mathematical expectation.

Define a fourth order cumulant matrix $\mathbf{C}^{2,2}_{\mathbf{x},\mathbf{x}_p}$ of the sensor signals as follows:

$$\begin{aligned}\mathbf{C}^{2,2}_{\mathbf{x},\mathbf{x}_p} = \; & E\{\mathbf{x}\mathbf{x}^T\mathbf{x}_p^T\mathbf{x}_p\} - E\{\mathbf{x}\mathbf{x}^T\}\mathrm{tr}E\{\mathbf{x}_p\mathbf{x}_p^T\} \\ & -2E\{\mathbf{x}\mathbf{x}_p^T\}E\{\mathbf{x}_p\mathbf{x}^T\}\end{aligned} \tag{5}$$

and similarly, a fourth order cumulant matrix $\mathbf{C}^{2,2}_{\mathbf{s},\mathbf{s}_p}$ of the source signals $\mathbf{s}_i, i = 1, ..., n$. It is easy to see that the $(i, j)$-th element of $\mathbf{C}^{2,2}_{\mathbf{x},\mathbf{x}_p}$ is

$$C^{2,2}_{\mathbf{x},\mathbf{x}_p}(i,j) = \sum_{l=1}^{n} \mathbf{cum}\{x_i(t), x_j(t), x_l(t-p), x_l(t-p)\}$$

(see [5] for more general cumulant matrices).

For a given set $P = \{p_1, ..., p_L\}$ of time delays and vectors $\mathbf{b} = (b_1, ..., b_L)^T \in \mathbb{R}^L$ define the following matrices

$$\mathbf{X}^{(2)}(\mathbf{b}) = \sum_{i=1}^{L} b_i \mathbf{R_x}(p_i), \tag{6}$$

$$\mathbf{X}^{(4)}(\mathbf{b}) = \sum_{i=1}^{L} b_i \mathbf{C}^{2,2}_{\mathbf{x},\mathbf{x}_{p_i}} \tag{7}$$

and similarly for the source signals

$$\mathbf{S}^{(2)}(\mathbf{b}) = \sum_{i=1}^{L} b_i \mathbf{R}_s(p_i); \tag{8}$$

$$\mathbf{S}^{(4)}(\mathbf{b}) = \sum_{i=1}^{L} b_i \mathbf{C}^{2,2}_{\mathbf{s},\mathbf{s}_{p_i}}. \tag{9}$$

## 4. Robust Orthogonalization

In our method below we need the global mixing matrix to be orthogonal. The standard whitening procedure [18] is not acceptable, since it enhances the noise, especially when the number of sensors is equal to the number of sources and the problem is ill conditioned. We propose a preprocessing procedure, which is not sensitive to additive white noise. This orthogonalization procedure allows us to define a new orthogonal mixing matrix for the preprocesed data. The idea is to use time-delayed cumulant (resp. covariance) matrices which are not sensitive to additive white noise of order 4 (resp. of order 2) and construct a positive definite matrix from their linear combination (for sufficiently large number of samples). Such a problem for white noise of order 2 is solved in [3] by a finite-step global convergence algorithm [26].

Denote

$$\mathbf{cum}_{s_i}(p) = \mathbf{cum}\{s_i(t), s_i(t), s_i(t-p), s_i(t-p)\}$$

and assume that the vectors

$$\left\{\left(\mathbf{cum}_{s_i}(p_1), ..., \mathbf{cum}_{s_i}(p_L)\right)\right\}_{i=1}^{n}$$

(resp. vectors $\left(E\{s_i(t)s_i(t-p_1\}, ..., E\{s_i(t)s_i(t-p_L)\}\right)_{i=1}^{n}$) are linearly independent. These conditions are necessary in order to be realized the finite-step global convergence algorithm [26] in Step 1 of the algorithm below.

We summarize two robust orthogonalization algorithms: for cumulant matrices (which is new) and for covariance matrices (considered in [3]).

**Algorithm Outline: Robust Orthogonalization**

1. Find by the finite-step global convergence algorithm [26] a set of parameters $\{\alpha_i\}_{i=1}^L$ such that the matrix $\mathbf{C_x}(\boldsymbol{\alpha}) = \sum_{i=1}^L \alpha_i \mathbf{C}^{2,2}_{\mathbf{x},\mathbf{x}_{p_i}}$ (resp. $\mathbf{C_x}(\boldsymbol{\alpha}) = \sum_{i=1}^L \alpha_i \mathbf{R_x}(p_i)\big)$ is positive definite.
2. Perform an eigenvalue-decomposition of $\mathbf{C_x}(\boldsymbol{\alpha})$, $\mathbf{C_x}(\boldsymbol{\alpha}) = \mathbf{U_x}\boldsymbol{\Lambda}_\mathbf{x}\mathbf{U}_\mathbf{x}^T$, where the entries of diagonal matrix $\boldsymbol{\Lambda}_\mathbf{x}$ are the positive eigenvalues of $\mathbf{C_x}(\boldsymbol{\alpha})$ and compute the preprocessing matrix $\mathbf{Q} = \boldsymbol{\Lambda}_\mathbf{x}^{-\frac{1}{2}}\mathbf{U}_\mathbf{x}^T$.
3. Compute the pre-processed data $\mathbf{z}(k) = \mathbf{Qx}(k) = \mathbf{QHs}(k)$.

**Remark 1:** In practice instead of $\mathbf{R_x}(p_i)$ and $\mathbf{C}^{2,2}_{\mathbf{x},\mathbf{x}_{p_i}}$, it is better to use

$$\tilde{\mathbf{R}}_\mathbf{x}(p_i)) = \frac{1}{2}\Big(\mathbf{R_x}(p_i) + \mathbf{R_x}(p_i)^T\Big)$$

and

$$\tilde{\mathbf{C}}^{2,2}_{\mathbf{x},\mathbf{x}_{p_i}} = \frac{1}{2}\Big(\mathbf{C}^{2,2}_{\mathbf{x},\mathbf{x}_{p_i}} + \Big(\mathbf{C}^{2,2}_{\mathbf{x},\mathbf{x}_{p_i}}\Big)^T\Big)$$

respectively (due to computational errors, which could destroy symmetricity of $\mathbf{R_x}(p_i)$ and $\mathbf{C}^{2,2}_{\mathbf{x},\mathbf{x}_{p_i}}$).

By defining a new mixing matrix as $\mathbf{A} = \mathbf{QHD}^{\frac{1}{2}}$, where $\mathbf{D} = \sum_{i=1}^L \alpha_i \mathbf{C}^{2,2}_{\mathbf{s},\mathbf{s}_{p_i}}$ (resp. $\mathbf{D} = \sum_{i=1}^L \alpha_i \mathbf{R_s}(p_i))$ is a diagonal (scaling) matrix with positive entries, we see that $\mathbf{C_z}(\boldsymbol{\alpha}) = \mathbf{AA}^T = \mathbf{I}_n$ (unit matrix), so, the matrix $\mathbf{A}$ is orthogonal. This orthogonality condition is necessary for performing separation of signals using either symmetric EVD, or joint diagonalization. It should be noted that in contrast to the standard prewhithening procedure for our robust orthogonalization generally $E\{\mathbf{zz}^T\} \neq \mathbf{I}_n$, but we have $\sum_{i=1}^L \alpha_i \mathbf{C}^{2,2}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}_{p_i}} = \mathbf{I}_n$, where $\tilde{\mathbf{s}} = \mathbf{D}^{-\frac{1}{2}}\mathbf{s}$ and $\sum_{i=1}^L \alpha_i \mathbf{C}^{2,2}_{\mathbf{z},\mathbf{z}_{p_i}} = \mathbf{I}_n$ (resp. $\sum_{i=1}^L \alpha_i \mathbf{R}_{\tilde{\mathbf{s}}}(p_i) = \mathbf{I}_n$ and $\sum_{i=1}^L \alpha_i \mathbf{R_z}(p_i) = \mathbf{I}_n$).

So, our model is

$$\mathbf{z} = \mathbf{A}\tilde{\mathbf{s}} + \mathbf{Qn}, \quad \text{where} \quad \tilde{\mathbf{s}} = \mathbf{D}^{-\frac{1}{2}}\mathbf{s}. \tag{10}$$

Note that using the standard pre-whitening procedure [18], we have the model (10) with $\mathbf{D} = E\{\mathbf{ss}^T\}$.

Further we shall use the following notation

$$\mathbf{Z}^{(2)}(\mathbf{b}) = \sum_{i=1}^L b_i \mathbf{R_z}(p_i), \quad \mathbf{Z}^{(4)}(\mathbf{b}) = \sum_{i=1}^L b_i \mathbf{C}^{2,2}_{\mathbf{z},\mathbf{z}_{p_i}}$$

and similarly for the source signals

$$\tilde{\mathbf{S}}^{(2)}(\mathbf{b}) = \sum_{i=1}^L b_i \mathbf{R}_{\tilde{\mathbf{s}}}(p_i); \quad \tilde{\mathbf{S}}^{(4)}(\mathbf{b}) = \sum_{i=1}^L b_i \mathbf{C}^{2,2}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}_{p_i}}.$$

## 5. Sufficient Conditions for Simultaneous Blind Source Separation

Let $P = \{p_1, ..., p_L\}$ be a set of positive integers with $L$ elements. We introduce the following conditions:

$$\forall i, j \neq i \quad \exists l_{i,j} \in \{1, ..., L\}:$$
$$E\{\tilde{s}_i(t)\tilde{s}_i(t - p_{l_{i,j}})\} \neq E\{\tilde{s}_j(t)\tilde{s}_j(t - p_{l_{i,j}})\} \qquad (\mathbf{DNAF}(P))$$

i.e. the sources have different normalized autocorrelation functions on $P$;

$$\forall i, j \neq i \ \exists l_{i,j} \in \{1, ..., L\}: \ \mathbf{cum}_{\tilde{s}_i}(p_{l_{i,j}}) \neq \mathbf{cum}_{\tilde{s}_j}(p_{l_{i,j}}), \qquad (\mathbf{DNCF}(P))$$

where

$$\mathbf{cum}_{\tilde{s}_i}(p) = \mathbf{cum}\{\tilde{s}_i(t), \tilde{s}_i(t), \tilde{s}_i(t-p), \tilde{s}_i(t-p)\}$$

i.e. the sources have different normalized cumulant functions of fourth order on the set $P$.

We recall that the source signals are *uncorrelated*, if $\mathbf{R}_s(p)$ are diagonal matrices for every $p \geq 0$. If the source signals are statistically independent, then this condition is satisfied, but the converse assertion is not always true. Note that the diagonal elements of $\mathbf{R}_s(p)$ are $E\{s_i(t)s_i(t-p)\}$. We say that the source signals are *colored*, if for some $p_0 \geq 1$ the matrix $\mathbf{R}_s(p_0)$ has a nonzero diagonal element. We shall say that the source signals are *uncorrelated of order* 4, if $\mathbf{C}^{2,2}_{\mathbf{s},\mathbf{s}_p}$ are diagonal matrices for every $p \geq 0$ with diagonal elements $\mathbf{cum}_{s_i}(p)$. If the source signals are statistically independent, then this condition is satisfied, but the converse assertion is not always true. We shall say that the sources are *colored of order* 4, if for some $p_0 \geq 1$, and some index $i_0$, $\mathbf{cum}_{s_{i_0}}(p_0)$ is nonzero. So, if $s_i, i = 1, ..., n$ are uncorrelated of order 4 and colored of order 4, then for some $p_0 \geq 1$, the matrix $\mathbf{C}^{2,2}_{\mathbf{s},\mathbf{s}_{p_0}}$ is a nonzero diagonal matrix.

Assume that the additive noise $\mathbf{n}$ has independent components with zero means, which are independent also with $s_i, i = 1, ..., n$. Recall that a signal $s$ is *white* (resp. *white of order 4*) if

$$E\{s(t)s(t-p)\} = 0, \quad \forall p \geq 1$$

$$(\text{resp.}\mathbf{cum}\{s(t-p_1), s(t-p_2), s(t-p_3), s(t-p_4)\} = 0$$

for every $p_i \geq 1, i = 1, ..., 4$) (see [22]).

The proof of the following lemma is straightforward and is omitted.

**Lemma 1:** Assume that the model (10) is satisfied, the noise $\mathbf{n}$ is white with zero mean, and $\tilde{\mathbf{S}}^{(2)}(\mathbf{b})$ is a diagonal matrix (resp. the noise $\mathbf{n}$ is white of order 4 and zero mean, and $\tilde{\mathbf{S}}^{(4)}(\mathbf{b})$ is a diagonal matrix).

Then the matrix $\mathbf{Z}^{(2)}(\mathbf{b})$ (resp. $\mathbf{Z}^{(4)}(\mathbf{b})$) is symmetrical and can be decomposed as $\mathbf{Z}^{(2)}(\mathbf{b}) = \mathbf{A}\tilde{\mathbf{S}}^{(2)}(\mathbf{b})\mathbf{A}^T = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$ (resp. $\mathbf{Z}^{(4)}(\mathbf{b}) = \mathbf{A}\tilde{\mathbf{S}}^{(4)}(\mathbf{b})\mathbf{A}^T = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$), where $\mathbf{U}$ is an orthogonal matrix and $\boldsymbol{\Lambda}$ is a diagonal matrix. If the diagonal elements of $\boldsymbol{\Lambda}$ are distinct, then the mixing matrix can be estimated as $\hat{\mathbf{A}} = \mathbf{U}$ up to multiplication with arbitrary permutation and diagonal nonsingular scaling matrices.

**Theorem 1:** Assume that the model (10) is satisfied, and (i): the source signals are colored and uncorrelated, the noise $\mathbf{n}$ is white with zero mean and condition ($\mathbf{DNAF}(P)$) is satisfied (resp. (ii): the source signals are colored of order 4 and uncorrelated of order 4, condition ($\mathbf{DNCF}(P)$) is satisfied and the noise $\mathbf{n}$ is white of order 4 with zero mean). Then

(a) there exists a vector $\mathbf{b} \in \mathbb{R}^L$ such that the matrix $\mathbf{Z}^{(2)}(\mathbf{b})$ (resp. $\mathbf{Z}^{(4)}(\mathbf{b})$) has distinct eigenvalues. Furthermore, the set $B(L)$ of all vectors $\mathbf{b} \in \mathbb{R}^L$ with this property form an open subset of $\mathbb{R}^L$, whose complement has a measure zero.

(b) If $\mathbf{U}$ is given from an $EVD$ of the matrix $\mathbf{Z}^{(2)}(\mathbf{b})$ (resp. the matrix $\mathbf{Z}^{(4)}(\mathbf{b})$) for some $\mathbf{b} \in B(L)$, i.e. $\mathbf{Z}^{(2)}(\mathbf{b}) = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, (resp. $\mathbf{Z}^{(4)}(\mathbf{b}) = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$), then the estimating mixing matrix is $\hat{\mathbf{A}} = \mathbf{U}$ and the separating matrix is $\mathbf{W} = \hat{\mathbf{A}}^T = \mathbf{U}^T$ (up to multiplication with arbitrary permutation and diagonal nonsingular scaling matrices).

**Proof.** We shall prove the theorem under condition (i) (the proof is similar under condition (ii)).

(a) Since $\tilde{s}_i, i = 1, ..., n$ are uncorrelated, $\tilde{\mathbf{S}}^{(2)}(\mathbf{b})$ is a diagonal matrix and by Lemma 1, $\mathbf{Z}^{(2)}(\mathbf{b}) = \mathbf{A}\tilde{\mathbf{S}}^{(2)}(\mathbf{b})\mathbf{A}^T$. Observe that the matrices $\mathbf{Z}^{(2)}(\mathbf{b})$ and $\tilde{\mathbf{S}}^{(2)}(\mathbf{b})$ have the same eigenvalues. It is easy to see that the complement of $B(L)$ is a finite union of subspaces of $\mathbb{R}^L$. If we prove that $B(L)$ is nonempty, then every of these subspaces must be proper (i.e. different from $\mathbb{R}^L$), consequently, with a measure zero (with respect to $\mathbb{R}^L$), therefore the complement of $B(L)$ must have a measure zero too.

Let $\{\sigma_i(\mathbf{b})\}_{i=1}^n$ be the diagonal elements of the matrix $\tilde{\mathbf{S}}^{(2)}(\mathbf{b})$, where $\mathbf{b} \in \mathbb{R}^L$. Assume that two diagonal elements of the matrix $\tilde{\mathbf{S}}^{(2)}(\mathbf{b})$ are equal, for example $\sigma_1(\mathbf{b}) = \sigma_2(\mathbf{b})$. Let $\mathbf{b}(1,2)$ be a vector, which is different from $\mathbf{b}$ only in the component $b_{l_{1,2}}$ ($l_{1,2}$ is defined by the condition ($\mathbf{DNAF}(P)$)). Then $\sigma_1(\mathbf{b}(1,2)) \neq \sigma_2(\mathbf{b}(1,2))$, because of the condition ($\mathbf{DNAF}(P)$). If all diagonal elements of $\tilde{\mathbf{S}}^{(2)}(\mathbf{b}(1,2))$ are different, we finish the proof. If not, suppose that $\sigma_i(\mathbf{b}(1,2)) = \sigma_j(\mathbf{b}(1,2))$ for some indexes $i$ and $j$. We can change a little the component $b_{l_{i,j}}$ of the vector $\mathbf{b}(1,2)$ (keeping the other components the same) such that for the new vector $\mathbf{b}(i,j)$ to be satisfied $\sigma_i(\mathbf{b}(i,j)) \neq \sigma_j(\mathbf{b}(i,j))$ (because of condition ($\mathbf{DNAF}(P)$) and $\sigma_1(\mathbf{b}(i,j)) \neq \sigma_2(\mathbf{b}(i,j))$. Continuing in such a way, for any couple $(k,r), k \neq r$ for which $\sigma_k(\mathbf{b}(k',r')) = \sigma_r(\mathbf{b}(k',r'))$ (where $\mathbf{b}(k',r')$ is the vector considered in the previous step), we make small change of $b_{l_{k,r}}$ keeping the pairwise difference of the diagonal elements considered in the previous steps and obtain vector $\mathbf{b}(k,r)$ for which $\sigma_k(\mathbf{b}(k,r)) \neq \sigma_r(\mathbf{b}(k,r))$. So, after finite number of steps we obtain a vector $\mathbf{b}^*$ for which the diagonal elements of $\tilde{\mathbf{S}}^{(2)}(\mathbf{b}^*)$ are distinct. This proves the nonemptiness of the set $B(L)$ and finishes the proof of (a).

(b) This follows from the well known facts of linear algebra [16]. ■

**Corrolary 1:** Under assumption (ii) of Theorem 1, an estimation of the mixing matrix is possible from the EVD of the cumulant matrix $\mathbf{C}^{2,2}_{\mathbf{z},\mathbf{z}_p}$, if the sources have different cumulants of fourth order for a fixed time delay $p$, i.e., if

$$\begin{aligned}&\mathbf{cum}(\tilde{s}_i(t), \tilde{s}_i(t), \tilde{s}_i(t-p), \tilde{s}_i(t-p)) \neq \\ &\mathbf{cum}(\tilde{s}_j(t), \tilde{s}_j(t), \tilde{s}_j(t-p), \tilde{s}_j(t-p))\end{aligned} \tag{11}$$

for every $i \neq j$. When $p = 0$, (11) means that the source signals have different kurtosis; in this case the conclusion is also true, if the noise is Gaussian.

**Remark 2:** In case of usual pre-whitening, the matrix $\mathbf{D}$ in (10) is equal to $E\{\mathbf{s}\mathbf{s}^T\}$ and, when all source signals are colored of order 2, we recover the identifiability conditions presented in [25], Theorem 2.

**Remark 3:** Theorem 1 gives another proof of the mathematical foundation of the SOBI algorithm [4].

**Remark 4:** If the matrix $\mathbf{Z}^{(2)}(b)$ (resp. $\mathbf{Z}^{(4)}(b)$) is non-symmetric (due to some numerical errors and finite number of samples) the following procedure can be applied. Construct symmetric matrix: $\mathbf{M}^{(2)}(\mathbf{b}) = \frac{1}{2}[\mathbf{Z}^{(2)}(\mathbf{b}) + \mathbf{Z}^{(2)}(\mathbf{b})^T]$ (resp. $\mathbf{M}^{(4)}(\mathbf{b}) = \frac{1}{2}[\mathbf{Z}^{(4)}(\mathbf{b}) + \mathbf{Z}^{(4)}(\mathbf{b})^T]$) and work with it instead of $\mathbf{Z}^{(2)}(b)$ (resp. $\mathbf{Z}^{(4)}(b)$).

**Remark 5:** In practical situations, when the sources are supposed to be very different (i.e. to have different autocorrelation (resp. cumulant) functions for almost all delays $p$), the set $P$ can be chosen to consist of only one element and to take trials for different $p$, until obtaining distinct eigenvalues of the matrices $\mathbf{Z}^{(2)}(\mathbf{b})$ (resp. $\mathbf{Z}^{(4)}(\mathbf{b})$).

## 6. Joint diagonalization of cumulant matrices depending on time delays - simulation results

We use the Jacobi algorithm for joint diagonalization of several matrices (see [6], [7], [5]) but here we use time-delayed cumulant matrices $\mathbf{C}^{2,2}_{\mathbf{x},\mathbf{x}_p}$ defined by (5) (after standard prewhitening).

Our algorithm, called JADETD[†] was tested on several benchmarks and below we present three of them. We use the package "ICALAB for Signal Processing" developed at the laboratory for Advance Brain Signal Processing, BSI, RIKEN[††]. As a measure of efficiency

[†]matlab code available upon request
[††]free downloaded at http://www.bsp.brain.riken.go.jp

of an algorithm we use the following performance index [10], which measures how close is the global matrix $\mathbf{G} = \mathbf{WH}$ ($\mathbf{W}$ separating matrix, $\mathbf{H}$ mixing matrix) to a (scaled) permutation matrix:

$$PI = \frac{1}{n-1} \sum_{i=1}^{n} \left\{ \sum_{k=1}^{n} \frac{g_{ik}^2}{\max_j g_{ij}^2} - 1 \right\}$$

We noticed by several experiments that the optimal number of cumulant matrices for joint diagonalization giving best result measured by this performance index is between 100 and 150.

The first benchmark consists of five signals (see ICALAB for Signal Processing in http://www.bsp.brain.riken.go.jp, benchmark Gnband.mat), spatially uncorrelated of order 4, white of order 2, colored of order 4, two of them with kurtosis near to zero (displayed in Fig.1), mixed with a random matrix. This matrix is not given, since for another random matrices produced in ICALAB for Signal Processing, the results are approximately the same (the reader could repeat this experiment). Our algorithm gives smallest performance index on this benchmark compared with other algorithms.

Performance Index

JADETD (Joint Approximate Diagonalization of 100 cumulant matrices $\mathbf{C}^{2,2}_{\mathbf{x},\mathbf{x}_p}, p = 1, ..., 100$):
0.0287627730

JADETD (120 cumulant matrices):
0.0271277802

JADETD (148 cumulant matrices):
0.0305472923

Fixed Point Algorithm [17]: 0.2811041266

JADE [6]: 0.2571942434

Natural Gradient (online) [2]: 0.3052515540

Natural Gradient (flexible ICA) [13]: 0.2354635225

AMUSE [24]: 0.5096235026

SOBI [4]: 0.3611529007

The second benchmark consists of five signals (see ICALAB for Signal Processing in http://www.bsp.brain.riken.go.jp, benchmark nband.mat), which are spatially uncorrelated of order 4, colored of order 4, with kurtosis very near to zero, mixed with a random matrix. We compare high order statistics algorithms on this benchmark.

Performance Index

JADETD: 0.0052936352

JADE [6]: 0.3318943056

Fixed Point Algorithm [17]: extracted only 1 component;

Natural Gradient (online) [2]: 0.3453449005

Natural Gradient (flexible ICA) [13]: 0.2499995186

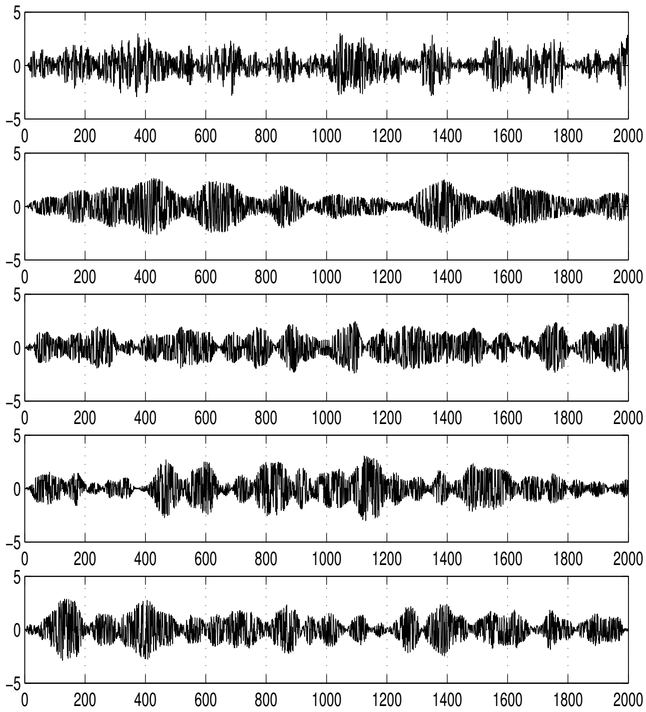

**Fig. 1** Independent signals which are white of order 2, but colored of order 4, two of them with kurtosis near to zero

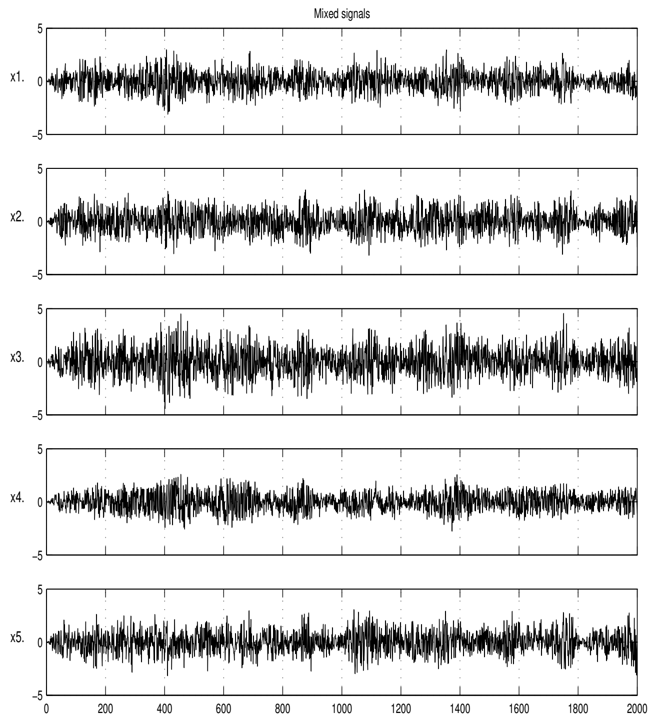


**Fig. 2** Mixed Signals

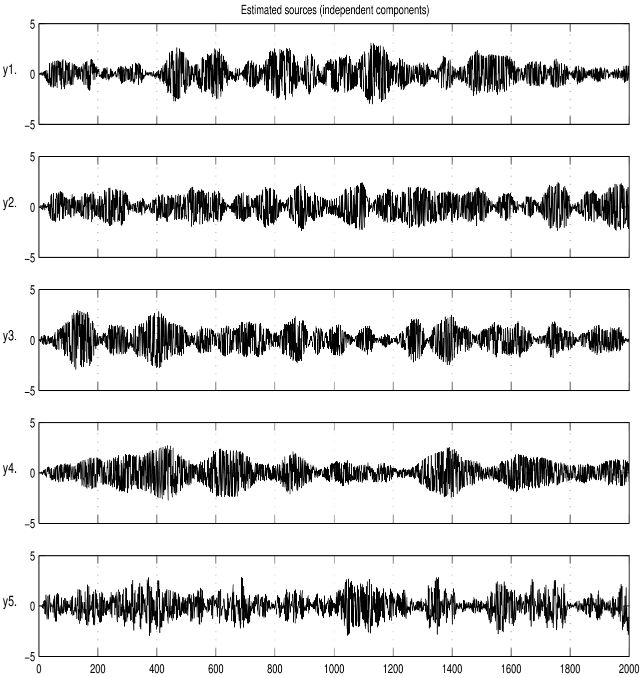


**Fig. 3** Estimated Signals with JADETD

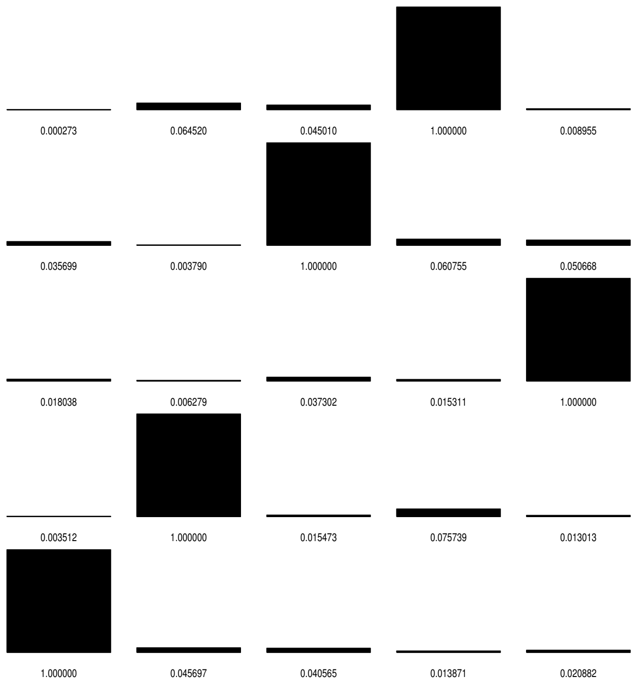


**Fig. 4** Normalized global matrix $\mathbf{G} = \mathbf{WH}$: each squared element is divided by the maximal squared element of the row

We note that second order statistics algorithms AMUSE and SOBI give better performance index (0.0006042555 and 0.0005480406 respectively) since the signals are colored of order two and uncorrelated.

The third benchmark† consists of 50 speech signals. They were mixed with a random matrix and the algorithm JADETD was run. The running time was 284.19 sec. (processor Pentium III, 750 Mhz), the performance index was 0.07340976173257005. JADE for the same benchmark run more than 3 hours and since there was no result (133 sweeps) it was stopped.

As we mention before, in lot of the experiments which we produced, the optimal number of matrices for joint diagonalization in JADETD was between 100 and 150. This number, for high dimensional problems, is much less than the number of the matrices used in the original JADE algorithm.This observation suggests that for high dimensional problems the proposed JADETD is a suitable replacement of the JADE algorithm.

## 7. Conclusion

We develop an unified approach by second and high order statistics to BSS problem, converting it to a symmetric eigenvalue problem. This approach is very suitable for high dimensional problems when the source signals are independent enough, since the symmetric EVD algorithms are very fast. Moreover, we can control the successfulness of the separation by monitoring the eigenvalues of the cumulant matrices. If the source signals are not independent enough, we propose joint diagonalization algorithm for cumulant matrices depending on time delays. This separating algorithm could be considered as a replacement of JADE for high dimensional problems, since it gives similar or better performance as JADE (for small dimensional problems), while for high dimensional problems (i.e. for $n > 30$) JADE is not suitable, due to the big computational time. Other advantages of our approach are the robustness to additive noise and the possibility to separate signals with zero kurtosis, which are white of order two but colored of order four.

### References

[1] S. Amari, "ICA of temporally correlated signals - learning algorithm". Proc. ICA '99: International workshop on blind signal separation, and independent component analysis, Aussois, France, pp. 13-18, Jan. 1999.

[2] S. Amari, A. Cichocki, and H.H. Yang, "A new learning algorithm for blind signal separation", in Advances in Neural Information Processing Systems, NIPS 1995, Vol. 8, pp. 757-763, MIT Press: Cambridge, MA, 1996.

[3] A. Belouchrani and A. Cichocki, "Robust whitening procedure in blind separation context". Electronics Letters, Vol. 36 No.24, pp. 2050-2051, 2000.

†available upon request

[4] A. Belouchrani, K. A. Meraim, J.-P. Cardoso, and E. Moulines, “A blind source separation technique using second order statistics.” IEEE Trans. on Signal Processing, Vol. 45, no. 2, pp. 434-44, Feb. 1997

[5] J.-F. Cardoso, “High-order contrasts for independent component analysis” Neural Computation, vol. 11, no 1, pp. 157–192, Jan. 1999.

[6] J.-F. Cardoso, A, Souloumiac, “Blind beamforming for non Gaussian signals” IEEEE Proceedings-F, 140 (1993), pp. 362-370.

[7] J.-F. Cardoso, A, Souloumiac, “Jacobi angels for simultanious diagonalization”, SIAM J. Matrix Analysis and Applications, Vol. 17, No.1, pp. 161-164, 1996.

[8] C. Chang, Z. Ding, S. F. Yau, and F. H. Y. Chan, “A matrix-pencil approach to blind separation of colored nonstationary signals”. IEEE Trans. Signal Proc., Vol. 48, No. 3, pp. 900-907, Mar. 2000.

[9] C. Chang, S.F. Yau, P. Kwok, F.H.Y. Chan, F.K. Lam, “Uncorrelated and component analysis for blind source separation”, Circuits Systems and Signal Processsing, Vol. 18, No.3, pp. 225-239, 1999.

[10] A. Cichocki and S. Amari, Adaptive Blind Signal and Image Processing, Wiley, April 2002.

[11] A. Cichocki, and R. Thawonmas “On-line algorithm for blind signal extraction of arbitrarily distributed, but temporally correlated sources using second order statistics”, Neural Processing Letters Vol. 12, pp. 91-98, Aug. 2000.

[12] S. Choi and A. Cichocki “Blind separation of nonstationary and temporally correlated sources from noisy mixtures”, IEEE Workshop on Neural Networks for Signal Processing, NNSP’2000, Sydney, Australia, pp. 405-414, Dec. 2000.

[13] S. Choi, A. Cichocki, and S. Amari, “Flexible independent component analysis”, Journal of VLSI Signal Processing, Vol. 26, No. 1/2, pp. 25-38, 2000.

[14] S. Cruces, L. Castedo and A.Cichocki, “Novel blind source separation algorithms using cumlant”, 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing, Istanbul, Turkey, pp. 3152-3155, Jun. 2000.

[15] A. Edelman, T. A. Arias, and A. T. Smith, “The geometry of algorithms with orthogonality constraints”, SIAM J. Matrix Anal. Appl., Vo. 20, No.2, pp. 303-353, 1998.

[16] G. H. Golub and C. F. Van Loan Matrix Computation, J. Hopkins Univ. Press, 1989.

[17] A. Hyvarinen and E. Oja, “A fast fixed point algorithm for independent component analysis”, Nural Computation, Vol. 9, No. 7, pp.1483-1492, 1997.

[18] A. Hyvarinen, J. Karhunen, and E. Oja, Independent Component Analysis, John Wiley & Sons, 2001.

[19] Y. Hua and J. K. Tugnait, “Blind Identifiability of FIR-MIMO Systems with Colored Input Using Second Order Statistics”, IEEE Signal Processing Letters, Vol. 7, No. 12, pp. 348-350, Dec. 2000.

[20] M. Kawamoto, Y. Inouye, A. Mansour, and R.-W. Liu, “Blind deconvolution algorithm for MINO-FIR systems driven by fourth-order colored signals”, 3-rd Int. Conf. on Independent Component Analysis and Signal Separation, San-Diego, USA, pp. 692-697, Dec. 2001.

[21] L. Molgedey, and H. G. Schuster, “Separation of a mixture of independent signals using time-delayed correlations”, Physical Review Letters, Vol. 72, No. 23, pp. 3634-3637, 1994.

[22] L. Nikias and A. P. Petropulu, “Higher-order spectra analysis: a nonlinear signal processing framework”, PTR Prentice-Hall, 1993.

[23] J. G. Proakis, “Diginal Communications”, NY: McGraw-Hill, New York, 1995.

[24] L. Tong, V.C. Soon, R. Liu, and Y. Huang, “AMUSE: a new blind identification algorithm,” Proc. ISCAS, New Orleans, LA, 1990.

[25] L. Tong, R. Liu, V.C. Soon, and Y.F. Huang, “Indeterminacy and Identifiability of Blind Identification”, IEEE Transactions on Circuits and Systems, Vol. 38, No.5, pp. 499-509, May 1991.

[26] L. Tong, Y. Inouye, and R. Liu “ A finite-step global algorithm for the parameter estimation of multichannel MA processes”. IEEE Trans. Signal Proc., 40(10), pp. 2547-2558, 1992.

**Pando G. Georgiev** received his MS, PhD and “Doctor of Mathematical Sciences” degrees in Mathematics (Operations Research) from Sofia University “St. Kl. Ohridski”, Bulgaria, in 1982, 1987 and 2001 respectively. He has been with the Department of Probability, Operations Research and Statistics at Faculty of Mathematics and Informatics, Sofia University “St. Kl. Ohridski”, Bulgaria, as Assistant Professor (1989-1994) and since 1994, as Associate Professor. He was a Visiting Professor at the University of Rome II, Italy (CNR grants), the International Center for Theoretical Physics, Trieste, Italy (ICTP grant, six months), the University of Pau, France (NATO grant, three months), Hirosaki University, Japan (JSPS grant, nine months). Since October 2000 he has been working as a research scientist at the Laboratory of Advanced Brain Signal Processing, Brain Science Institute, RIKEN, Wako-shi. His interests include Optimization and Operations Research, Signal and Image Processing, Independent Component Analysis and Blind Signal Separation, Neural Networks, Probability and Statistics, Numerical Methods and Algorithms, Nonlinear and Nonsmooth Analysis. He is a member of AMS, IEEE, UBM.

**Andrzej Cichocki** received the M.Sc.(with honors), Ph.D., and Habilitate Doctorate (Dr.Sc.) degrees, all in electrical engineering, from Warsaw University of Technology (Poland) in 1972, 1975, and 1982, respectively. Since 1972, he has been with the Institute of Theory of Electrical Engineering and Electrical Measurements at the Warsaw University of Technology, where he became a full Professor in 1991. He is the coauthor of three books: Adaptive Blind Signal and Image Processing - Learning Algorithms and Applications, Wiley, 2002, MOS Switched-Capacitor and Continuous-Time Integrated Circuits and Systems (Springer-Verlag, 1989) and Neural Networks for Optimization and Signal Processing (Teubner-Wiley,1993/94) and more than 150 research papers. He spent at University Erlangen-Nuernberg (Germany) a few years as Alexander Humboldt Research Fellow and Guest Professor.

Since 1995 he has been working in the Brain Science Institute RIKEN (Japan), as a team leader for laboratory for Open Information Systems and currently as a head of laboratory for Advanced Brain Signal Processing. His current research interests

include neural networks, signal and image processing, especially analysis and processing of multi-sensory biomedical data.
He is currently associate editor of IEEE Transaction on Neural Networks and Fluctuation and Noise Letters. He was 3 years the member of the Technical Committee for Neural Network for Signal Processing and recently the member of core group who established a new IEEE Circuits and Systems Technical Committee for Blind Signal Processing.
More information about his research activities can be found at http://www.bsp.brain.riken.go.jp/