BLIND SOURCE SEPARATION VIA SYMMETRIC EIGENVALUE DECOMPOSITION

Pando Georgiev[†] *and Andrzej Cichocki*[‡]

Brain Science Institute, RIKEN, Wako-shi, Saitama 351-01, Japan

[†]-On leave from the Sofia University "St. Kl. Ohridski", Bulgaria

E-mail: georgiev@bsp.brain.riken.go.jp.

[‡]-On leave from the Warsaw University of Technology, Poland

E-mail: cia@bsp.brain.riken.go.jp.

ABSTRACT

We propose a new sufficient condition for separation of colored source signals with temporal structure, stating that the separation is possible, if the source signals have different higher self-correlation functions of even order. We show that the problem of blind source separation of uncorrelated colored signals can be converted to a symmetric eigenvalue problem of a special covariance matrix $\mathbf{Z}(\mathbf{b}) = \sum_{i=1}^{L} b(p_i)\mathbf{R}_z(p_i)$ depending on L-dimensional parameter $\mathbf{b}$, if this matrix has distinct eigenvalues. We prove that the parameters $\mathbf{b}$ for which this is possible, form an open subset of $\mathbb{R}^L$, which complement has a Lebesgue measure zero. We use a robust orthogonalization of the mixing matrix, which is not sensitive to the white noise, and propose a new sufficient condition for that: the source signals to have linearly independent higher self-correlation functions of even order. We propose a new one-step algorithm, based on the non-smooth optimization theory, which disperses the eigenvalues of the matrix $\mathbf{Z}(\mathbf{b})$ providing sufficient distance between them.

1. INTRODUCTION

The interest of blind signal processing, especially, independent component analysis (ICA) 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. (see e.g. [1-6]).

The problem of blind source separation (BSS) is formulated as follows: we can observe sensor signals $\mathbf{x}(k) = [x_1(k), \ldots, x_m(k)]^T$ which are described as

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

where $\mathbf{H}$ is $m \times n$ full-rank unknown mixing matrix, $\mathbf{s}(k) = [s_1(k), \ldots, s_n(k)]^T$ $(n \leq m)$ is a vector of unknown zero mean colored (i.e. with temporal structure) source signals and $\mathbf{n}(k)$ is a vector of additive white noise. Our objective is to estimate the mixing matrix $\mathbf{H}$ and/or source signals simultaneously or sequentially one-by-one assuming that they are uncorrelated (not necessarily statistically independent) but arbitrarily distributed colored (not independent identically distributed), i.e. we assume that sources satisfy relation: $E\{s_i(k)\, s_i(k-p)\} \neq 0$ at least for some $p = 1, 2, \ldots$ and have different temporal structures [4], [5]. More generally, we introduce a new condition for decorrelation of high order and a new sufficient conditions for separation (see condition (**DAF**) below) requiring the sources to have different higher self-correlation functions of even order. This condition can be considered as a generalization of those one described in [5] (see also references therein for similar sufficient conditions for blind source separation). 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 [14].

In this paper we prove that BSS problem can be converted to symmetric eigenvector problem, for which the value of the kurtosis is unimportant. So, any algorithm for eigenvector problem can separate simultaneously uncorrelated colored sources with different temporal structures. It is worth to mention that simultaneous extraction of i.i.d. (independent and identically distributed) signals is also possible with global convergence (see for example [17]).

The use of second statistics approach for blind separation of temporally correlated sources has been developed and analyzed by many researchers, including Amari [2], Molgdey and Schuster [15], Pham and Garat [19], Belouchrani et al. [4], Belouchrani and Cichocki [3], Cichocki, Rutkowski, Barros, and Oh [6], Cichocki, and Thawonmas [7], Choi and Cichocki [8], Pearlmutter and Parra [18], Mueller et al.[16], etc. Moreover, it should be mentioned that recently several researchers have developed a number of efficient algorithms for sequential blind source extraction, especially works of Delfosse and Loubaton [10],

However, our approach has some advantages that may not be found in others at the same time. It is computationally simpler and more efficient than Joint Diagonalization approach since it employs powerful eigenvalue decomposition (EVD) [11]; it provides relative fast convergence (since several algorithm has been developed for the EVD with cubic convergence); it can solve large scale problem with hundreds or even thousands of sources due to efficiency of available EVD algorithms, it extracts the components simultaneously; does not assume non-zero kurtosis neither statistical independence of the sources; does not need the sources to be stationary; does not need that all but one signal should be Gaussian; and it is robust in respect to white additive noise what often leads to smaller errors (cross-talking between estimated sources).

2. ROBUST ORTHOGONALIZATION

In our method below we need the global mixing matrix to be orthogonal. The standard whitening procedure is not acceptable, since it enhances the noise. We use a preprocessing procedure in, which is not sensitive to the white noise and which allows us to define a new semi-orthogonal (orthogonal if $m = n$) mixing matrix for the preprocesed data [3]. The idea is to use time-delayed correlation matrices that are not sensitive to additive white noise and construct a positive definite matrix from their linear combination (for sufficiently large number of samples), a problem solved in [3] by a finite-step global convergence algorithm [22].

Let us define a a time-delayed correlation matrix of the observation vector $\mathbf{x}(k)$ by

$$\mathbf{R}_x(p) = E\{\mathbf{x}(k)\mathbf{x}^T(k-p)\} \tag{2}$$

and a symmetric matrix $\tilde{\mathbf{R}}_x(p)$ by

$$\tilde{\mathbf{R}}_x(p) = \frac{1}{2}\left\{\mathbf{R}_x(p) + \mathbf{R}_x^T(p)\right\}. \tag{3}$$

Similarly we define analogous matrices $\mathbf{R}_s(p)$ and $\tilde{\mathbf{R}}_s(p)$ for the source signals $\mathbf{s}(k)$.

In a linear data model (1), the time-delayed correlation matrices of the observation vector $\mathbf{x}(k)$ for any $p \neq 0$ satisfy (due to the assumption of white noise) $\tilde{\mathbf{R}}_x(p) = \mathbf{A}\tilde{\mathbf{R}}_s(p)\mathbf{A}^T$.

Note that $\mathbf{R}_x(p)$ is symmetric, if $\mathbf{R}_s(p)$ is a diagonal matrix) but in order to avoid the effect of computational errors (which could destroy the symmetricity of $\mathbf{R}_x(p)$), we use (3).

The robust orthogonalization algorithm can be summarized as follows.

Algorithm Outline: Robust Orthogonalization

1. Find (by the method described in [3]), i.e. choose or estimate a set of parameters $\{\alpha_i\}_{i=1}^K$ such that the matrix $\mathbf{C}_x(\boldsymbol{\alpha}) = \sum_{i=1}^K \alpha_i \tilde{\mathbf{R}}_x(p_i)$ is positive definite.
2. Perform an eigenvalue-decomposition of $\mathbf{C}_x(\boldsymbol{\alpha})$, $\mathbf{C}_x(\boldsymbol{\alpha}) = \mathbf{U}_x\boldsymbol{\Lambda}_x\mathbf{U}_x^T$, where the entries of diagonal matrix $\boldsymbol{\Lambda}_x$ are the positive eigenvalues of $\mathbf{C}_x(\boldsymbol{\alpha})$ and compute the preprocessing matrix $Q = \boldsymbol{\Lambda}_x^{-\frac{1}{2}}\mathbf{U}_x^T$.
3. Compute the preprocesed data $\mathbf{z}(k) = \mathbf{Q}\mathbf{x}(k) = \mathbf{QHs}(k)$.

Remark 1 By defining a new mixing matrix as $\mathbf{A} = \mathbf{QHD}^{\frac{1}{2}}$, where $\mathbf{D} = \sum_{i=1}^K \alpha_i \tilde{\mathbf{R}}_s(p_i)$ is a diagonal (scaling) matrix with positive entries it is easy to show that $\mathbf{C}_z(\boldsymbol{\alpha}) = \mathbf{AA}^T = \mathbf{I}_m((m\times m)$ unit matrix), thus the matrix $\mathbf{A}$ is orthogonal, if $m = n$. This orthogonality condition is necessary for performing separation of signals using EVD. It should be noted that in contrast to the standard prewhithening procedure for our robust orthogonalization generally $E\{\mathbf{zz}^T\} \neq \mathbf{I}_m$, but it is possible to be arranged $E\{\mathbf{zz}^T\} = \mathbf{I}_m$, with a more complicated algorithm. Also, we have $\mathbf{z} = \mathbf{A}\tilde{\mathbf{s}} + \mathbf{Qn}$, where $\tilde{\mathbf{s}} = \mathbf{D}^{-\frac{1}{2}}\mathbf{s}$. But due to the scaling indeterminacy of the sources we may write in the sequel that $\mathbf{z} = \mathbf{As} + \tilde{\mathbf{n}}$ $(\tilde{\mathbf{n}} = \mathbf{Qn})$.

3. SUFFICIENT CONDITION FOR SIMULTANEOUS BLIND SOURCE SEPARATION OF COLORED SOURCES

We define higher autocorrelation functions of even order for the source signals by

$$r_i(\mathbf{p}) = E\{s_i(k)s_i(k-p_1)\cdots s_i(k-p_{2N})s_i(k-p_{2N+1})\},$$

where $\mathbf{p} = (p_1, ..., p_{2N+1}) \in \mathbb{R}^{2N+1}$ and introduce the following condition:

$$\forall i, j \neq i \;\; \exists \mathbf{p}(i,j) \in \mathbb{R}^{2N+1} : r_i(\mathbf{p}(i,j)) \neq r_j(\mathbf{p}(i,j)), \quad (\mathbf{DAF})$$

i.e. the sources have different higher autocorrelation functions of even order, at least for some discrete time delays $\mathbf{p}(i,j) = (p_1(i,j), ..., p_{2N+1}(i,j))$.

We shall say that these functions are *linearly independent*, if

$$\sum_{i=1}^n \mu_i r_i(\mathbf{p}) = 0 \;\; \forall \mathbf{p} \in \mathbb{R}^{2N+1} \text{ iff } \mu_i = 0 \;\; \forall i. \quad (\mathbf{LIAF})$$

Define a high order covariance matrix of sensor signals by

$$\mathbf{R}_z(\mathbf{p}) = E\{\mathbf{z}\mathbf{z}_{p_1}^T ... \mathbf{z}_{p_{2N}}\mathbf{z}_{p_{2N+1}}^T\},$$

and similarly, a high order covariance matrix of source signals by

$$\mathbf{R}_s(\mathbf{p}) = E\{\mathbf{s}\mathbf{s}_{p_1}^T ... \mathbf{s}_{p_{2N}}\mathbf{s}_{p_{2N+1}}^T\},$$

where $\mathbf{z}_p = \mathbf{z}(k-p), \mathbf{z} = \mathbf{z}(k), \mathbf{s}_p = \mathbf{s}(k-p), \mathbf{s} = \mathbf{s}(k)$.

We shall say that the source signals are *uncorrelated of order* $2N+1$, if $\mathbf{R}_s(\mathbf{p})$ are diagonal matrices for every $\mathbf{p} \in \mathbb{R}^{2N+1}$. Note that in this case the diagonal elements of $\mathbf{R}_s(\mathbf{p})$ are $r_i(\mathbf{p}), i = 1, ..., n$. If the source signals are statistically independent random variables, then this condition is satisfied, but the converse assertion is not always true. We say that *the sources are colored of order* $2N+1$, if for some vector $\mathbf{p}_0 \in \mathbb{R}^{2N+1}$ the matrix $\mathbf{R}_s(\mathbf{p}_0)$ is nonzero (diagonal) matrix.

For a given vector $\mathbf{b} \in \mathbb{R}^{L^{2N+1}}$ define

$$\mathbf{Z}(\mathbf{b}) = \sum_{i=1}^{2N+1}\sum_{p_i=1}^{L} b(\mathbf{p})\mathbf{R}_z(\mathbf{p}),$$

and similarly for the source signals

$$\mathbf{S}(\mathbf{b}) := \sum_{i=1}^{2N+1}\sum_{p_i=1}^{L} b(\mathbf{p})\mathbf{R}_s(\mathbf{p}).$$

Lemma 1 *If the mixing matrix* $\mathbf{A}$ *is orthogonal and matrix* $\mathbf{S}(\mathbf{b})$ *is a diagonal matrix, then the matrix* $\mathbf{Z}(\mathbf{b})$ *is symmetrical and can be decomposed as* $\mathbf{Z}(\mathbf{b}) = \mathbf{AS}(\mathbf{b})\mathbf{A}^T = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^T$. *Moreover, if the diagonal matrix* $\boldsymbol{\Lambda}$ *has distinct entries, then the mixing matrix can be estimated as* $\mathbf{A} = \mathbf{U}$ *up to multiplication with arbitrary permutation and diagonal nonsingular scaling matrices.*

Theorem 1 *Assume that the signals are colored and uncorrelated of order* $2N+1$*, condition (*$\mathbf{DAF}$*) is satisfied and the mixing matrix is orthogonal. Then*

(a) for any $L \geq \max_{1\leq i,j\leq n} p_{2N+1}(i,j)$*, there exists a vector* $\mathbf{b} \in \mathbb{R}^{L^{2N+1}}$ *such that the matrix* $\mathbf{Z}(\mathbf{b})$ *has distinct eigenvalues. Furthermore, the set* $B(L)$ *of all vectors* $\mathbf{b} \in \mathbb{R}^{L^{2N+1}}$ *with this property form an open subset of* $\mathbb{R}^{L^{2N+1}}$*, which complement has a Lebesgue measure zero.*

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

Proof. (a) Observe that the matrices $\mathbf{Z}(\mathbf{b})$ and $\mathbf{S}(\mathbf{b})$ have the same eigenvalues. It is easy to see that the complement of $B(L)$ is a finite union of subspaces of $\mathbb{R}^{L^{2N+1}}$. If we prove that $B(L)$ is nonempty, then every of these subspaces must be proper (i.e.

different from $\mathbb{R}^{L^{2N+1}}$), consequently, with a Lebesgue measure zero (with respect to $\mathbb{R}^{L^{2N+1}}$), therefore the complement of $B(L)$ must have a Lebesgue measure zero too.

Let $\{\sigma_i(\mathbf{b})\}_{i=1}^n$ be the diagonal elements of the matrix $\mathbf{S}(\mathbf{b})$, where $\mathbf{b} \in \mathbb{R}^{L^{2N+1}}$. Assume that two diagonal elements of the matrix $\mathbf{S}(\mathbf{b})$ 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 $b(\mathbf{p}(1,2))$ ($\mathbf{p}(1,2)$ is defined by the condition (**DAF**)). Then $\sigma_1(\mathbf{b}(1,2)) \neq \sigma_2(\mathbf{b}(1,2))$, because of the condition (**DAF**). If all diagonal elements of $\mathbf{S}(\mathbf{b}(1,2))$ 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 $b(\mathbf{p}(i,j))$ of the vector $\mathbf{b}(1,2)$ (keeping the other components the same) and obtain a new vector $\mathbf{b}(i,j)$ such that $\sigma_i(\mathbf{b}(i,j)) \neq \sigma_j(\mathbf{b}(i,j))$ (because of condition (**DAF**) and keeping $\sigma_1(\mathbf{b}(i,j)) \neq \sigma_2(\mathbf{b}(i,j))$. 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 $b(\mathbf{p}(k,r))$ keeping the pair-wise 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 $\mathbf{S}(\mathbf{b}^*)$ are distinct. This proves the non-emptiness of the set $B(L)$ and finishes the proof of (a).

(b) This follows from the well known facts of linear algebra [13]. ■

Remark 2 An advantage of using high order correlations is that it is possible two source signals to have the same autocorrelation functions of second order but to have different higher autocorrelation functions of even order.

Remark 3 It should be noted that ideal case under assumption that mixing matrix is orthogonal and $\mathbf{R}_s(p)$ are diagonal matrices, noise is white and uncorrelated with source signals the covariance matrix is symmetrical and standard symmetric eigenvalue decomposition (EVD) can be applied. If the matrix $\mathbf{Z}(b)$ is nonsymmetric (due to numerical errors) the following procedure can be applied. Construct symmetric matrix: $\tilde{\mathbf{Z}}(\mathbf{b}) = \frac{1}{2}[\mathbf{Z}(\mathbf{b}) + \mathbf{Z}^T(\mathbf{b})]$ and then apply the EVD.

Remark 4 If all time delays $p_1, ..., p_{2N+1}$ are different, then the high-order correlation matrices $E\{\mathbf{z}\mathbf{z}_{p_1}^T ... \mathbf{z}_{2N}\mathbf{z}_{2N+1}^T\}$ and consequently $\mathbf{Z}(\mathbf{b})$ are unbiased by the additive noise under condition that it is white (i.i.d.) and independent from the source signals.

Remark 5 A sufficient condition for the robust orthogonalization is condition (**LIAF**). In this case it is possible also to choose such set of parameters $\mathbf{b}$ that the matrix $\mathbf{Z}(\mathbf{b})$ or $\tilde{\mathbf{Z}}(\mathbf{b})$ to be positive definite.

4. ONE-STEP ALGORITHM, WHICH DISPERSES THE EIGENVALUES OF THE MATRIX $\mathbf{Z}(\mathbf{b})$

We present one step algorithm which ensures that all eigenvalues of the matrix $\mathbf{Z}(b)$ are different and dispersed as much as we want. For its derivation (which is omitted because of limited space) we use the notions and facts from the non-smooth analysis and the optimization theory, contained in [9] and [12].

Consider the function:

$$\varphi(\mathbf{b}) = \min_{1 \le i \le m-1} \{\lambda_i(\mathbf{b}) - \lambda_{i+1}(\mathbf{b})\}, \tag{4}$$

where $\lambda_i(\mathbf{b})$ are the eigenvalues (in decreasing order) of the matrix $\mathbf{Z}(\mathbf{b})$. This function is positively homogeneous, i.e. $\varphi(t\mathbf{b}) = t\varphi(\mathbf{b})$ for $t > 0$. So, it is enough to find an accent direction $\mathbf{d}$ of this function and then we can disperse the eigenvalues of the matrix $\mathbf{Z}(\mathbf{b})$ simply by multiplication. Below we propose an algorithm for finding an accent direction. We point out that this is not the steepest accent direction, although we can find this steepest accent direction with a more complicated algorithm.

For simplicity we shall consider here the case when $N = 0$, i.e. the autocorrelation functions are of second order. Also, from a practical point of view it is useful to work with a set of fixed and global time delays $p_1, ..., p_L$ for all signals. So, in the sequel we shall assume that $\mathbf{b} \in \mathbb{R}^L$, $\mathbf{Z}(\mathbf{b}) = \sum_{i=1}^L b_i \mathbf{R}_z(p_i)$ and the condition (**DAF**) is satisfied with respect to the set $P = \{p_1, ..., p_L\}$, i.e. $\forall i \neq j \;\; \exists k : r_i(p_k) \neq r_j(p_k)$. The algorithm is summarized as follows:

1. Start from arbitrary $\mathbf{b} \neq 0, \mathbf{b} \in \mathbb{R}^L$.

2. Perform an EVD of the matrix $\mathbf{Z}(\mathbf{b})$: $\mathbf{Z}(\mathbf{b}) = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, where $\mathbf{\Lambda}$ is a diagonal matrix whose diagonal elements are the eigenvalues of $\mathbf{Z}(\mathbf{b})$ and the columns of $\mathbf{U}$ are eigenvectors of $\mathbf{Z}(\mathbf{b})$. If $\varphi(\mathbf{b}) > 0$, then stop. Otherwise go to 3.

3. Let $\lambda_i(\mathbf{b}), i \in I \subset \{1, ..., m\}$ be the set of non-distinct eigenvalues of $\mathbf{Z}(\mathbf{b})$, i.e. every $\lambda_i(\mathbf{b})$ has multiplicity $m_i \ge 2$. Calculate the L-dimensional vectors

$$\mathbf{w}_{i,k} = [\mathbf{u}_{i,k}^T \mathbf{R}_z(p_1)\mathbf{u}_{i,k}, ..., \mathbf{u}_{i,k}^T \mathbf{R}_z(p_L)\mathbf{u}_{i,k}]^T,$$

where $\mathbf{u}_{i,k}, k = 1, ..., m_i$ are the eigenvectors among the columns of $\mathbf{U}$ corresponding to the eigenvalue $\lambda_i, i \in I$;

4. Choose a vector (denoted by $\mathbf{d}$) with maximal norm among the vectors $\{\mathbf{w}_{i,k} - \mathbf{w}_{i,r} : k, r = 1, ..., m_i, i \in I\}$ and compute the new vector as $\mathbf{b}_* = \mathbf{b} + \theta\mathbf{d}$, where $0 < \theta \le 1$.

Then $\varphi(\mathbf{b}_*) > 0$, i.e. the eigenvalues of $\mathbf{Z}(\mathbf{b}_*)$ are different. If they are not dispersed enough, we take the matrix $\mathbf{Z}(t\mathbf{b}_*)$ for an appropriate $t > 1$.

In the following theorem we show how to disperse two equal eigenvalues. This gives an idea why direction $\mathbf{d}$ in the above algorithm has such a form.

Theorem 2 *Assume that the eigenvalues* $\lambda_i(\mathbf{b}), i = 1, ..., m$ *of the matrix* $\mathbf{Z}(\mathbf{b}), \mathbf{b} \in \mathbb{R}^L$ *are ordered in decreasing order,* $\lambda_k(\mathbf{b})$ *has multiplicity 2, i.e.* $\lambda_k(\mathbf{b}) = \lambda_{k+1}(\mathbf{b})$ *for some k, and* $\mathbf{u}_k$ *and* $\mathbf{u}_{k+1}$ *are two unit linearly independent eigenvectors of* $\mathbf{Z}(\mathbf{b})$ *corresponding to* $\lambda_k(\mathbf{b})$. *Then, for any* $\theta \neq 0$ *we have* $\lambda_k(\mathbf{b} + \theta\mathbf{d}) \neq \lambda_{k+1}(\mathbf{b} + \theta\mathbf{d})$, *where the components of* $\mathbf{d}$ *are* $d_i = \mathbf{u}_k^T \mathbf{R}_z(p_i)\mathbf{u}_k - \mathbf{u}_{k+1}^T \mathbf{R}_z(p_i)\mathbf{u}_{k+1}$.

Proof. Since $\mathbf{A}$ is orthogonal, the eigenvalues of the matrices $\mathbf{Z}(\mathbf{b} + \theta\mathbf{d})$ and $\mathbf{S}(\mathbf{b} + \theta\mathbf{d}) = \sum_{i=1}^L (b_i + \theta d_i)\mathbf{R}_s(p_i)$ coincide. We have

$$\mathbf{A}^T\mathbf{u}_k = \alpha_1\mathbf{e}_k + \alpha_2\mathbf{e}_{k+1}, \quad \mathbf{A}^T\mathbf{u}_{k+1} = \beta_1\mathbf{e}_k + \beta_2\mathbf{e}_{k+1},$$

where $\mathbf{e}_k = (0, ...0, 1, 0, ..., 0)$, 1 is in the $k-$th place and $\alpha_1^2 + \alpha_2^2 = \beta_1^2 + \beta_2^2 = 1$. Denote by $\sigma_k(\mathbf{b} + \theta\mathbf{d})$ and $r_k(p_i)$ the k-th diagonal element of the matrices $\mathbf{S}(\mathbf{b} + \theta\mathbf{d})$ and $\mathbf{R}_s(p_i)$ respectively. Observe that $d_i = (\alpha_1^2 - \beta_1^2)r_k(p_i) + (\alpha_2^2 - \beta_2^2)r_{k+1}(p_i)$. Then we have:

$$\begin{aligned} &\sigma_k(\mathbf{b} + \theta\mathbf{d}) - \sigma_{k+1}(\mathbf{b} + \theta\mathbf{d}) \\ &= \textstyle\sum_{i=1}^L (b_i + \theta d_i)(r_k(p_i) - r_{k+1}(p_i)) \\ &= \theta(\alpha_1^2 - \beta_1^2) \textstyle\sum_{i=1}^L (r_k(p_i) - r_{k+1}(p_i))^2 \neq 0 \end{aligned}$$

Note that $\alpha_1^2 \neq \beta_1^2$ since $\mathbf{u}_k$ and $\mathbf{u}_{k+1}$ are linearly independent and the last sum is nonzero, due to condition (**DAF**), which is satisfied, as we assumed, with respect to the set $\{p_1, ..., p_L\}$. ■

The following lemma gives accent directions of a nonsmooth function.

Lemma 2. *Assume that $f : \mathbb{R}^L \to \mathbb{R}$ is a locally Lipschitz function, regular in sense of Clarke [9]. Let $\partial f(\mathbf{b})$ mean the Clarke subdifferential of f at $\mathbf{b}$ and $\mathbf{d} \in \partial f(\mathbf{b})$ be any nonzero element of $\partial f(\mathbf{b})$. Then $\sup_{t>0} f(\mathbf{b} + t\mathbf{d}) > f(\mathbf{b})$, i.e. $\mathbf{d}$ is direction in which the function can increase strictly.*

Proof. By the properties of the regular locally Lipschitz functions (see [9]), we have:

$$\lim_{t\to 0_+} \frac{f(\mathbf{b}+t\mathbf{d})-f(\mathbf{b})}{t} = f'(\mathbf{b};\mathbf{d})$$
$$= \max_{\mathbf{c}\in\partial f(\mathbf{b})} \mathbf{d}^T\mathbf{c} \geq \|\mathbf{d}\|^2 > 0. \quad \blacksquare$$

The derivation of the subdifferential $\partial\varphi(\mathbf{b})$ is complicated and is given by the formula:

$$\partial\varphi(\mathbf{b}) = \overline{co}\{(\mathbf{v}^T\mathbf{R}_z(p_1)\mathbf{v}, ..., \mathbf{v}^T\mathbf{R}_z(p_L)\mathbf{v}) - (\mathbf{u}^T\mathbf{R}_z(p_1)\mathbf{u}, ..., \mathbf{u}^T\mathbf{R}_z(p_L)\mathbf{u}) : \mathbf{v} \in \mathbf{V}_i, \mathbf{u} \in \mathbf{V}_i, i \in I_0\}, \quad (5)$$

where I_0 is the set where the minimum in (4) is attained, $\mathbf{V}_i$ is the set of all unit eigenvectors corresponding to the eigenvalue λ_i and $\overline{co}$ denotes the closed convex hull. The above algorithm is based on formula (5) and Lemma 2.

5. CONCLUSIONS

We have formulated a new sufficient condition for blind separation of signals and a new sufficient condition for robust orthogonalization of the mixing matrix, based on higher order autocorrelation functions. Moreover, we have presented a new algorithm for BSS of colored sources based on symmetrical eigenvalue decomposition, using a new procedure for dispersion of eigenvalues, derived by non-smooth analysis and optimization. The proposed algorithm is robust with respect to white additive noise. Furthermore, the algorithm is suitable for large scale problem due to efficiency of several recently developed procedures for the eigenvalue decomposition. Most of the assertions in this paper are proved rigorous mathematically.

6. REFERENCES

[1] Amari S. and Cichocki A. "Adaptive blind signal processing - neural network approaches". *Proceedings IEEE* (invited paper), 86(10): 2016-2048, 1998.

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

[3] Belouchrani A. and Cichocki A. "Robust whitening procedure in blind separation context". *Electronics Letters,* 36(24): 2050-2051, 2000.

[4] Belouchrani A., Meraim K.A., Cardoso J.-F. and Moulines E. "A blind source separation technique using second order statistics." *IEEE Trans. on Signal Processing,* 45: 434-444, 1997.

[5] Chang C., Ding Z, Yau S.F. and Chan F.H.Y. "A matrix-pencil approach to blind separation of colored nonstationary signals". *IEEE Trans. Signal Proc.,* 48(3): 900-907, 2000.

[6] Cichocki A., Rutkowski T., Barros A.K. and Oh S.-H. " Blind extraction of temporally correlated but statistically dependent acoustic signals". *Proc. IEEE Workshop on Neural Networks for Signal Processing*, NNSP '2000, Sydney, Australia, pages 455-464, Dec. 2000.

[7] Cichocki A., and Thawonmas R. " On-line algorithm for blind signal extraction of arbitrarily distributed, but temporally correlated sources using second order statistics". *Neural Processing Letters* 12: 91-98, 2000.

[8] Choi S. and Cichocki A. "Blind separation of nonstationary and temporally correlated sources from noisy mixtures". *IEEE Workshop on Neural Networks for Signal Processing, NNSP'2000*, Sydney, Australia, pages 405-414, Dec. 2000.

[9] Clarke F.S. *Optimization and Nonsmooth Analysis*. J. Wiley and Sons, 1983.

[10] Delfosse N. and Loubaton P. "Adaptive blind separation of independent sources: a deflation approach". *Signal Processing*, 45: 59 - 83, 1995.

[11] Edelman A., Arias T.A. and Smith A.T. "The geometry of algorithms with orthogonality constraints". *SIAM J. Matrix Anal. Appl.*, 20(2): 303-353, 1998.

[12] Georgiev P.Gr. "Submonotone Mappings in Banach Spaces and Applications". *Set-Valued Analysis*, 5: 1-35, 1997.

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

[14] Hua Y. and Tugnait J.K. "Blind Identifiability of FIR-MIMO Systems with Colored Input Using Second Order Statistics". *IEEE Signal Processing Letters*, 7(12): 348-350, 2000.

[15] Molgedey L. and Schuster H.G. "Separation of a mixture of independent signals using time-delayed correlations. *Physical. Review Letters*, 72(23): 3634-3637, 1994.

[16] Mueller K.-R., Philips P., and Ziehe A. "JADETD: Combining higher-order statistics and temporal information for blind source separation (with noise)". *Proc. IEEE International Conference on Independent Component Analysis and Signal Separation*, Aussois, France, pages 312-316, Jan.1999.

[17] Paradias C.B. "Globally Convergent Blind Source Separation Based on Multiuser Kurtosis Maximization Criterion". *IEEE Trans. Signal Processing*, 48(12): 3508-3519, 2000.

[18] Pearlmutter B.A. and Parra L.C. "Maximum Likelihood blind source separation, A context-sensitive generalization of ICA". *Proc. NIPS 96*, MIT Press, 9, pages 613-619, 1997.

[19] Pham D.T. and Garat P. "Blind separation of mixtures of independent sources through a quasi-maximum likelihood approach". *IEEE Trans. on Signal Processing*, 45(7): 1712-1725, 1997.

[20] D.T. Pham and J.F. Cardoso, "Blind separation of instantaneous mixtures of nonstationary sources", in Proc. ICA-2000, pp. 187-192, 2000.

[21] D.T. Pham, "Joint approximate diagonalization of positive definite Hermitian matrices", Tech. Rep. LMC/IMAG, University of Grenoble, France 2000.

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