数理解析研究所講究録 1246 巻 2002 年 19-30

On Some New Ideas and Algorithms for Independent Component Analysis

Pando Georgiev[†] and Andrzej Cichocki[‡]

Abstract

For every integer $p \geq 4$, even, we consider a specific optimization problem OP(p) arising from the blind source extraction problem (BSE) and prove that every local maximum of OP(p) is a solution of (BSE) in sense that it extracts one source signal from a linear mixture of unknown statistically independent signals. We construct an algorithm for solving OP(p) with rate of convergence $p-1$. We propose new sufficient conditions for separation of source signals, stating that the separation is possible, if the source signals have different autocorrelation or cumulant functions (depending on time delay). We show that the problem of blind source separation of signals can be converted to a symmetric eigenvalue problem of a generalized cumulant matrices if these matrices have distinct eigenvalues. We propose new algorithms, based on non-smooth analysis and optimization theory, which disperse the eigenvalues of these generalized cumulant matrices.

1 Introduction

The problem of independent component analysis is formulated as follows: we observe sensor signals (random variables) $\mathbf{x}(t) = [x_1(t), \ldots, x_m(t)]^T$ and want to represent them as linear mixture of random variables $\mathbf{s}(t) = [s_1(t), \ldots, s_n(t)]^T$, which are independent, as much as possible:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \quad t \geq 0, \tag{1}$$

where $\mathbf{A}$ is $n \times n$ non-singular matrix.

The problem of blind source extraction (BSE), which we shall consider, is formulated as follows: for given source signals (random variables) $\mathbf{x}(t) = [x_1(t), \ldots, x_m(t)]^T$ and knowing that they are obtained as a linear mixture (1), the task is to find $\mathbf{s}(t)$ and the matrix $\mathbf{A}$. In general this is impossible, but if $s_i, i = 1, ..., n$ are statistically independent and $\mathbf{A}$ is nonsingular, then this is possible up to permutation and scaling, i.e. we can obtain $\mathbf{ADP}$, where $\mathbf{D}$ and $\mathbf{P}$ are unknown diagonal and permutation matrices respectively. Therefore, we can obtain $d_i s_{p_i}(t)$, where p_i is a permutation of $\{1, ..., n\}$ (unknown) and d_i are scaling coefficients (unknown).

The literature about independent component analysis and BSE problem is huge (see for instance [16] and references therein).

Here we generalize the algorithm of Hyvarinen and Oja [17] and prove rigorously its conditions for convergence. Even as a mathematical problem this algorithm is interesting, since it provides an example of an algorithm with arbitrarily fast convergence (defined in the beginning).

Define the function $\varphi_p : \mathbb{R}^n \to \mathbb{R}$ by $\varphi_p(\mathbf{w}) = \mathbf{cum}_p(\mathbf{w}^T\mathbf{x})$ where $\mathbf{cum}_p$ means the self-cumulant of order p (see [22] for definition and properties of the cumulants). Consider the maximization problem

$$\text{OP(p)}: \quad \text{maximize } |\varphi_p(\mathbf{w})| \qquad \text{under constraint } \|\mathbf{w}\| = 1.$$

We shall see that this maximization problem has interesting properties, namely, it has exactly n solutions $\mathbf{w}_1, ..., \mathbf{w}_n$, which are orthonormal. We can recover the original source signals up to sign and permutation: $y_i(t) = \mathbf{w}_i^T\mathbf{x}(t) = \pm s_{r_i}(t), r_i \in \{1, ..., n\}$. We construct an algorithm for finding $\mathbf{w}_1, ..., \mathbf{w}_n$ one by one, which has rate of convergence $p-1$.

We note that the idea of maximizing of $\mathbf{cum}_4(\mathbf{w}^T\mathbf{x})$ in order to extract one source from a linear mixture is already considered in [9].

We need the following lemma, which is generalization of a lemma in [9] (considered there the case $p = 4$).

Lemma 1 *Consider the optimization problem: minimize (maximize)* $\sum_{i=1}^{n} k_i v_i^p$

subject to $\|\mathbf{v}\| = c > 0$, *where* $p > 2$ *is even, where* $\mathbf{v} = (v_1, ..., v_n)$.

Denote $I^+ = \{i \in \{1, ..., n\} : k_i > 0\}$, $I^- = \{i \in \{1, ..., n\} : k_i < 0\}$ *and* $e_i = (0, ...0, 1, 0, ..., 0)$, *(1 is the i−th place). Assume that* $I^+ \neq \emptyset$ *and* $I^- \neq \emptyset$.

Then the points of local minimum are exactly the vectors $m_i^{\pm} = \pm c e_i, i \in I^-$ *and the points of local maximum are exactly the vectors* $M_j^{\pm} = \pm c e_j, j \in I^+$.

Proof. Applying the Lagrange multipliers theorem for a point of a local optimum $\overline{v} = (\overline{v}_1, ..., \overline{v}_m)$, we write:

$$k_i p \overline{v}_i^{p-1} - 2\lambda \overline{v}_i = 0, i = 1, ..., m, \tag{2}$$

where λ is a Lagrange multiplyer.

Multiplying (1) by $\overline{v}_i$ and summing, we obtain:

$$p f_{opt.} = 2\lambda c^2,$$

where $f_{opt.}$ means the value of f at the local optimum. Hence

$$\lambda = \frac{p}{2c^2} f_{opt.}. \tag{3}$$

From (1) we obtain

$$\overline{v}_i (t_i p \overline{v}_i^{p-2} - \frac{p}{c^2} f_{opt.}) = 0$$

whence

$$\overline{v}_i \text{ is either } 0, \text{ or } \pm \left(\frac{f_{opt.}}{k_i c^2}\right)^{\frac{1}{p-2}}. \tag{4}$$

Case 1. Assume that $k_{i_0} < 0$ for some index i_0 and $\overline{v}$ is a local minimum. Then obviously $f_{loc.min.} < 0$. According to the second order optimality condition [1], a point x^0 is a local minimum if

$$h^T L''(x^0) h > 0 \quad \forall h \in K(x^0) = \{h : h^T x^0 = 0\}, h \neq 0,$$

where

$$L(x) = \sum_{i=1}^{n} k_i x_i^p - \lambda(\|x\|^2 - c^2)$$

is the Lagrange function.

In our case, by (2) and (3) we obtain

$$\begin{aligned} h^T L''(\overline{v}) h &= \sum_{i=1}^{n} (p(p-1) k_i \overline{v}_i^{p-2} - 2\lambda) h_i^2 \\ &= \frac{p}{c^2} f_{loc.min.} \Big[(p-2) \sum_{i \in I} h_i^2 - \sum_{i \notin I} h_i^2\Big], \end{aligned} \tag{5}$$

where I is the set of those indeces i, for which $\overline{v}_i$ is different from 0.

We shall check this second order optimality condition for the points $m_{i_0}^{\pm}$. We have

$$K(m_{i_0}^{\pm}) = \{h : h_{i_0} = 0\},$$

therefore, for $h \in K(m_{i_0}^{\pm}), h \neq 0$ we have

$$h^T L''(m_{i_0}^{\pm})h > 0,$$

since $h_{i_0} = 0$ and $f_{loc.min.} < 0$.

By (4) it follows that any other vector with at least two nonzero elements is not a local minimum.

Case 2. Assume that $k_j > 0$ and $\overline{v}$ is a local maximum. We apply Case 1 to the function $-f$ and finish the proof. ■

Theorem 2 *Assume that the matrix* $\mathbf{A}$ *in (1) is orthogonal. Then*

(a) the maximization problem OP(p) has exactly n solutions $\mathbf{w}_1, ..., \mathbf{w}_n$*, which are orthonormal;*

(b) We can recover the original source signals up to sign and permutation: $y_i(t) = \mathbf{w}_i^T \mathbf{x}(t) = \pm s_{r_i}(t), r_i \in \{1, ..., n\}$.

Proof. Consider the maximization problem:
(DP(p)) maximize $|\psi_p(\mathbf{c})|$
under constraint $\|\mathbf{c}\| = 1$,
where $\psi_p(\mathbf{c}) = \mathbf{cum}_p \sum_{i=1}^n c_i s_i$.

It is easy to see that the problems (DP(p)) and OP(p) are equivalent in sense that $\mathbf{w}_0$ is a solution of OP(p) if and only if $\mathbf{c}_0 = \mathbf{A}\mathbf{w}_0$ is a solution of (DP(p)). By the properties of the cumulants [5], we have

$$\psi_p(\mathbf{c}) = \mathbf{cum}_p \left(\sum_{i=1}^n c_i s_i \right) = \sum_{i=1}^n c_i^p \mathbf{cum}_p(s_i),$$

Applying Lemma 1 we obtain that (DP(p)) has n solutions, which are exactly the vectors $\pm \mathbf{e}_i$ with ± 1 in $i-$th place and 0 in the other places. Now the conclusion of (a) is evident.

(b) The assertion is evident. ■

2 Algorithm with high order convergence speed

Denoting $y = \mathbf{w}^T \mathbf{x}$, $\mathbf{c} = \mathbf{A}^T \mathbf{w}$ we have $y = \mathbf{c}^T \mathbf{s} = \sum_{i=1}^n c_i s_i$.

Consider the following algorithm:

$$\mathbf{w}(l) = \frac{\varphi_p'(\mathbf{w}(l-1))}{\|\varphi_p'(\mathbf{w}(l-1))\|}, \quad l = 1, 2, ..., \tag{6}$$

where $\varphi_p(\mathbf{w}) = \mathbf{cum}_p(\mathbf{w}^T \mathbf{x})$.

Theorem 3 *Assume that* s_i *are statistically independent, zero mean signals. Let* $p \geq 4$*, even, be given,* $\mathbf{cum}_p(s_i) \neq 0, i = 1, ..., n$ *and let* $I(\mathbf{c}) := \arg \max_{1 \leq i \leq n} c_i |\mathbf{cum}_p(s_i)|^{\frac{1}{p-2}}$. *Let* W_0 *be the set of all elements* $\mathbf{w} \in \mathbb{R}^n$ *such that* $\|\mathbf{w}\| = 1$*, the set* $I(\mathbf{A}^T \mathbf{w})$ *contains only one element, say* $i(\mathbf{w})$*, and* $c_{i(\mathbf{w})} \neq 0$. *Then*

(a) The complement of W_0 *has measure zero.*

(b) If $\mathbf{w}(0) \in W_0$ *then* $\lim_{l \to \infty} y_l(t) = \pm s_{i_0}(t)$ *for every* $k = 1, 2, ...,$ *where* $y_l(t) = \mathbf{w}(l)^T \mathbf{x}(t)$ *and* $i_0 = i(\mathbf{w}(0))$.

(c) The rate of convergence in (b) is of order $p - 1$.

Proof.

(a) It is easy to see that the complement of W_0 is a finite union of proper subspaces of $\mathbb{R}^n$, therefore this union has measure zero.

(b), (c). By the properties of the cumulants, since $s_i, i = 1, ..., n$ are statistically independent (see [22]), we have:

$$\mathbf{cum}_p\left(\sum_{i=1}^{n} c_i s_i\right) = \sum_{i=1}^{n} c_i^p \mathbf{cum}_p(s_i).$$

Define $\psi_p(\mathbf{c}) = \sum_{i=1}^n c_i^p \mathbf{cum}_p(s_i)$, and $\mathbf{c}(l) = \mathbf{A}^T\mathbf{w}(l)$, where $\mathbf{w}(l)$ is given by (6).

Since $\varphi_p(\mathbf{w}) = \psi_p(\mathbf{A}^T\mathbf{w})$, using the chain rule for differentiating the composite function $\psi_p(\mathbf{A}^T\mathbf{w})$ with respect to $\mathbf{w}$, we obtain:

$$\varphi_p'(\mathbf{w}) = \mathbf{A}\psi_p'(\mathbf{c}) = p\mathbf{A}\left(c_1^{p-1}\mathbf{cum}_p(s_1), ..., c_n^{p-1}\mathbf{cum}_p(s_n)\right)^T, \tag{7}$$

therefore

$$\begin{aligned} \mathbf{Ac}(l) &= \mathbf{w}(l) = \frac{\varphi_p'(\mathbf{w}(l-1))}{\|\varphi_p'(\mathbf{w}(l-1))\|} \\ &= \frac{\mathbf{A}\psi_p'(\mathbf{c}(l-1))}{\|\mathbf{A}\psi_p'(\mathbf{c}(l-1))\|}. \\ &= \frac{\mathbf{A}\psi'(\mathbf{c}(l-1))}{\|\psi'(\mathbf{c}(l-1))\|}, \end{aligned} \tag{8}$$

since $\mathbf{A}$ is orthogonal.

Multiplying (8) by $\mathbf{A}^T$ we obtain

$$\mathbf{c}(l) = \frac{\psi'(\mathbf{c}(l-1))}{\|\psi'(\mathbf{c}(l-1))\|}, \tag{9}$$

or

$$c_i(l) = \frac{\sigma_i p |\mathbf{cum}_p(s_i)| c_i(l-1)^{p-1}}{\|\psi'(\mathbf{c}(l-1))\|} \quad \forall i = 1, ..., n, \tag{10}$$

where $\sigma_i = sign\ \mathbf{cum}_p(s_i)$.

From (10) we obtain

$$c_i(l)|\mathbf{cum}_p(s_i)|^{\frac{1}{p-2}} = \frac{\sigma_i\left[c_i(l-1)|\mathbf{cum}_p(s_i)|^{\frac{1}{p-2}}\right]^{p-1}}{\|\left(c_1(l-1)^{p-1}\mathbf{cum}_p(s_1), ..., c_n(l-1)^{p-1}\mathbf{cum}_p(s_n)\right)^T\|}. \tag{11}$$

From (11) it follows by induction that if the initial conditions satisfy $c_i(0)\mathbf{cum}_p(s_i) \neq 0$ for some $i \in \{1, ..., n\}$, then the denominator in (11) is not zero for every $l \geq 0$, so $\mathbf{w}(l)$ in (8) is well defined. Hence, if $\mathbf{w}(0) \in W_0$, we obtain:

$$\frac{c_i(l)}{c_{i_0}(l)}\left(\frac{|\mathbf{cum}_p(s_i)|}{|\mathbf{cum}_p(s_{i_0})|}\right)^{\frac{1}{p-2}} = \frac{\sigma_i}{\sigma_{i_0}}\left[\frac{c_i(l-1)}{c_{i_0}(l-1)}\left(\frac{|\mathbf{cum}_p(s_i)|}{|\mathbf{cum}_p(s_{i_0})|}\right)^{\frac{1}{p-2}}\right]^{p-1} \tag{12}$$

Now by (12) one proves easily by induction that for every $l \geq 0$ and every $i = 1, ..., n$

$$\frac{c_i(l)}{c_{i_0}(l)} \left(\frac{|\mathbf{cum}_p(s_i)|}{|\mathbf{cum}_p(s_{i_0})|}\right)^{\frac{1}{p-2}} = \left(\frac{\sigma_i}{\sigma_{i_0}}\right)^l \left[\frac{c_i(0)}{c_{i_0}(0)} \left(\frac{|\mathbf{cum}_p(s_i)|}{|\mathbf{cum}_p(s_{i_0})|}\right)^{\frac{1}{p-2}}\right]^{(p-1)^l} . \tag{13}$$

From (13) it follows that $c_i(l) \to 0$ when $l \to \infty$ for $i \neq i_0$ and $|c_{i_0}(l)| \to 1$ (since $\|\mathbf{c}(l)\| = 1$), as the speed of the convergence is $p - 1$. Thus we proved:

$$\lim_{l\to\infty} y_l(t) = \lim_{l\to\infty} \mathbf{w}^T(l)\mathbf{x}(t) = \lim_{l\to\infty} \mathbf{c}^T(l)\mathbf{s}(t) = \pm s_{i_0}(t) \quad \forall k = 1, 2, ...$$

Similar algorithm can be used for cumulants involving time delays:

$$\mathbf{cum}_p\{s_i, s_i, s_i(. - p), s_i(. - p)\}.$$

This will allows to extract signals having non-zero cumulants (depending on time delays).

3 Cumulant matrices

We will consider the case of cumulants of order four for simplicity. Define (like in [6], but here using time delays) a 4-th order cumulant matrix $\mathbf{C}^{2,2}_{\mathbf{x},\mathbf{x}_p}(\mathbf{B})$ of the sensor signals as follows:

$$\begin{aligned} \mathbf{C}^{2,2}_{\mathbf{x},\mathbf{x}_p}(\mathbf{B}) &= E\{\mathbf{x}\mathbf{x}^T\mathbf{x}_p^T\mathbf{B}\mathbf{x}_p\} - E\{\mathbf{x}\mathbf{x}^T\}\mathrm{tr}(\mathbf{B}E\{\mathbf{x}_p\mathbf{x}_p^T\}) \\ &\quad -E\{\mathbf{x}\mathbf{x}_p^T\}\mathbf{B}E\{\mathbf{x}_p\mathbf{x}^T\} - E\{\mathbf{x}\mathbf{x}_p^T\}\mathbf{B}^T E\{\mathbf{x}_p\mathbf{x}^T\} \end{aligned} \tag{14}$$

where $\mathbf{B} \in \mathbb{R}^{n^2}$ is a matrix, $\mathbf{x}_p = \mathbf{x}(t - p), \mathbf{x} = \mathbf{x}(t), \mathbf{s}_p = \mathbf{s}(t - p), \mathbf{s} = \mathbf{s}(t)$ and E is the mathematical expectation (with respect to t). Similarly, define a fourth order cumulant matrix $\mathbf{C}^{2,2}_{\mathbf{s},\mathbf{s}_p}(\mathbf{B})$ of the source signals $\mathbf{s}_i, i = 1, ..., n$.

Assume that the additive noise $\mathbf{n}$ has independent Gaussian components (with zero means), which are independent also with $s_i, i = 1, ..., n$. It is easy to see that the (i, j)-th element of $\mathbf{C}^{2,2}_{\mathbf{x},\mathbf{x}_p}(\mathbf{B})$ is

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

where $\mathbf{cum}\{x_i(t), x_j(t), x_k(t - p), x_l(t - p)\}$ denotes the fourth order cumulant.

In the sequel we suppose that $\mathbf{s}_i,$, $i = 1, ..., n$ are statistically independent. Then we have:

$$\mathbf{C}^{2,2}_{\mathbf{x},\mathbf{x}_p}(\mathbf{B}) = \mathbf{H}\Delta(\mathbf{B})\mathbf{H}^T, \tag{15}$$

$$\Delta(\mathbf{B}) = diag\{\mathbf{cum}_{s_1}(p)\mathbf{h}_{*1}^T\mathbf{B}\mathbf{h}_{*1}, ..., \mathbf{cum}_{s_n}(p)\mathbf{h}_{*n}^T\mathbf{B}\mathbf{h}_{*n}\}, \tag{16}$$

where $\mathbf{cum}_{s_i}(p) = \mathbf{cum}\{s_i(k), s_i(k), s_i(k - p), s_i(k - p)\}$ and $\mathbf{h}_{*i}$ denotes the i-th column of $\mathbf{H}$. Therefore, if the mixing matrix $\mathbf{H}$ is orthogonal, we can separate the sources by eigenvalue decomposition of $\mathbf{C}^{2,2}_{\mathbf{x},\mathbf{x}_p}(\mathbf{B})$ (if its eigenvalues are distinct), which estimates $\mathbf{H}$ up to multiplication with permutation and diagonal sign matrices. Below we show how to disperse its eigenvalues, if

4 Subdifferential algorithms for dispersing the eigenvalues

Cardoso [6] mentioned (without proof), that the set $\mathcal{B}$ of matrices $\mathbf{B}$ for which the eigenvalues of $\mathbf{C}^{2,2}_{\mathbf{z},\mathbf{z}_p}(\mathbf{B})$ are distinct, has full measure, that is, its complement (in $\mathbb{R}^{n^2}$) has measure zero. This fact is an easy exercise, based on the observations that 1): $\mathcal{B}$ is non-empty, due to the non-singularity of $\mathbf{H}$, and 2): the complement of $\mathcal{B}$ in $\mathbb{R}^{n^2}$ is a finite union of proper subspaces, therefore, with measure zero. This suggests random choice of matrix $\mathbf{B}$ in order to achieve separation. But, as it was mentioned by Cardoso in [6], we need more in practice, because the algorithms use only sample estimates of the cumulant matrices and a small error in the sample estimate could produce a large deviation of the eigenvectors, if the eigenvalues are not well enough separated.

We present algorithm which ensures enough separation of the eigenvalues of the matrix $\mathbf{C}^{2,2}_{\mathbf{z},\mathbf{z}_p}(\mathbf{B})$. 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 [8] and [11].

Consider the function:

$$\varphi(\mathbf{B}) = \min_{1\le i\le m-1}\{\lambda_i(\mathbf{B}) - \lambda_{i+1}(\mathbf{B})\}, \tag{17}$$

where $\lambda_i(\mathbf{B})$ are the eigenvalues (in decreasing order) of the matrix $\mathbf{C}^{2,2}_{\mathbf{z},\mathbf{z}_p}(\mathbf{B})$. Every eigenvalue can be expressed by Fisher's minimax theorem [18].

This function is positively homogeneous, i.e. $\varphi(t\mathbf{B}) = t\varphi(\mathbf{B})$ for $t > 0$. It is enough to find an accent direction $\mathbf{d}$ of this function to achieve separation of the eigenvalues. 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.

The following lemma gives accent directions of a nonsmooth function.

Lemma 4 *Assume that $f : \mathbb{R}^L \to \mathbb{R}$ is a locally Lipschitz function, regular in sense of Clarke [8]. 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 [8]), we have:

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

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

$$\partial\varphi(\mathbf{B}) = \overline{co}\{\mathbf{R}(\mathbf{u},\mathbf{v}) : \mathbf{v}\in\mathbf{V}_i, \mathbf{u}\in\mathbf{V}_i, i\in I_0\}, \tag{18}$$

where $\mathbf{R}(\mathbf{u},\mathbf{v})$ is a matrix with (k,l)-th element $\mathbf{v}^T\mathbf{Q}_{k,l}\mathbf{v} - \mathbf{u}^T\mathbf{Q}_{k,l}\mathbf{u}$, I_0 is the set where the minimum in (17) is attained, $\mathbf{V}_i$ is the set of all unit eigenvectors of $\mathbf{C}^{2,2}_{\mathbf{z},\mathbf{z}_p}(\mathbf{B})$ corresponding to the eigenvalue $\lambda_i(\mathbf{B})$, $\mathbf{Q}_{k,l}$ is the matrix with (r,s)-th element $\mathbf{cum}(z_r(t), z_s(t), z_k(t-p), z_l(t-p))$ and $\overline{co}$ denotes the closed convex hull. The derivation uses subdifferential calculus for functions of sup-type, applied in our case for the function in Fisher minimax formula.

The following algorithm is based on formula (18) and Lemma (4).

1. Start from arbitrary $\mathbf{B}\in\mathbb{R}^{n^2}$.

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

3. Let $\lambda_i(\mathbf{B}), i \in I \subset \{1, ..., m_i\}$ be the set of non-distinct eigenvalues of $\mathbf{C}^{2,2}_{\mathbf{z},\mathbf{z}_p}(\mathbf{B})$; i.e. every $\lambda_i(\mathbf{B}), i \in I$ has multiplicity $m_i \geq 2$. Calculate the $n \times n$-matrices $\mathbf{W}_{i,j}$ with (k,l)-th element

$$W_{i,j}(k,l) = \mathbf{u}^T_{i,j}\mathbf{Q}_{k,l}\mathbf{u}_{i,j}$$

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

4. Choose a matrix (denoted by $\mathbf{D}$) with maximal norm among the matrices $\{\mathbf{W}_{i,j} - \mathbf{W}_{i,r} : j, r = 1, ..., m_i, i \in I\}$ and compute the new matrix as $\mathbf{B}_* = \mathbf{B} + \theta\mathbf{D}$, where $0 < \theta \leq 1$.

Then $\varphi(\mathbf{B}_*) > 0$, i.e. the eigenvalues of $\mathbf{C}^{2,2}_{\mathbf{z},\mathbf{z}_p}(\mathbf{B}_*)$ are distinct.

In the following theorem we show how to disperse two equal eigenvalues. This is an independent proof for the validity of the above algorithm for this case and gives an idea why the matrix $\mathbf{D}$ has such a form.

Theorem 5 *Assume that the eigenvalues $\lambda_i(\mathbf{B}), i = 1, ..., m$ of the matrix $\mathbf{C}^{2,2}_{\mathbf{z},\mathbf{z}_p}(\mathbf{B}), \mathbf{B} \in \mathbb{R}^{n^2}$ are ordered in decreasing order, $\lambda_j(\mathbf{B})$ has multiplicity 2, i.e. $\lambda_j(\mathbf{B}) = \lambda_{j+1}(\mathbf{B})$ for some j, $\mathbf{u}_j$, $\mathbf{u}_{j+1}$ are two unit linearly independent eigenvectors of $\mathbf{C}^{2,2}_{\mathbf{z},\mathbf{z}_p}(\mathbf{B})$ corresponding to $\lambda_j(\mathbf{B})$ and $\mathbf{cum}_{s_i}(p).\mathbf{cum}_{s_{i+1}}(p) \neq 0$. Then, for any $\theta \neq 0$ we have $\lambda_j(\mathbf{B} + \theta\mathbf{D}) \neq \lambda_{j+1}(\mathbf{B} + \theta\mathbf{D})$, where the components of $\mathbf{D}$ are*

$$D_{k,l} = \mathbf{u}^T_j\mathbf{Q}_{k,l}\mathbf{u}_j - \mathbf{u}^T_{j+1}\mathbf{Q}_{k,l}\mathbf{u}_{j+1} \tag{19}$$

and $Q_{k,l}$ are defined after formula (18).

Proof. We have

$$\mathbf{C}^{2,2}_{\mathbf{z},\mathbf{z}_p}(\mathbf{B}) = \mathbf{A}\Delta(\mathbf{B})\mathbf{A}^T, \tag{20}$$

$$\Delta(\mathbf{B}) = diag\left(\mathbf{cum}_{s_1}(p)\mathbf{a}^T_{*1}\mathbf{B}\mathbf{a}_{*1}, ..., \mathbf{cum}_{s_n}(p)\mathbf{a}^T_{*n}\mathbf{B}\mathbf{a}_{*n}\right). \tag{21}$$

Since $\mathbf{A}$ is orthogonal, the eigenvalues of the matrices $\mathbf{C}^{2,2}_{\mathbf{z},\mathbf{z}_p}(\mathbf{B} + \theta\mathbf{D})$ and $\Delta(\mathbf{B} + \theta\mathbf{D})$ coincide. Since $\lambda_j(\mathbf{B})$ has multiplicity 2, we have

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

for some $\alpha_1, \alpha_2, \beta_1, \beta_2$, where $\mathbf{e}_j = (0, ...0, 1, 0, ..., 0)$, 1 is in the $j-$th place and $\alpha_1^2 + \alpha_2^2 = \beta_1^2 + \beta_2^2 = 1$. By (19) we have

$$\begin{aligned} D_{k,l} &= \mathbf{u}^T_j\mathbf{Q}_{k,l}\mathbf{u}_j - \mathbf{u}^T_{j+1}\mathbf{Q}_{k,l}\mathbf{u}_{j+1} = \\ &\sum_{r,s=1}^{n} \mathbf{cum}\{z_r(t), z_s(t), z_k(t-p), z_l(t-p)\}(u_{j,r}u_{j,s} - u_{j+1,r}u_{j+1,s}). \end{aligned}$$

Hence we obtain:

$$\mathbf{D} = \mathbf{A}diag\left\{\mathbf{cum}_{s_1}(p)\mathbf{a}^T_{*1}\mathbf{V}\mathbf{a}_{*1}, ..., \mathbf{cum}_{s_n}(p)\mathbf{a}^T_{*n}\mathbf{V}\mathbf{a}_{*n}\right\}\mathbf{A}^T,$$

where $\mathbf{V}$ is a matrix with elements $V_{r,s} = u_{j,r}u_{j,s} - u_{j+1,r}u_{j+1,s}$ and $\mathbf{a}_{*i}$ is the i-th column of $\mathbf{A}$. By (20), (21) we have

$$\begin{aligned} \mathbf{C}^{2,2}_{\mathbf{z},\mathbf{z}_p}(\mathbf{B} + \theta\mathbf{D}) &= \mathbf{A}diag\{\mathbf{cum}_{s_1}(p)\mathbf{a}^T_{*1}(\mathbf{B} + \theta\mathbf{D})\mathbf{a}_{*1}, ..., \mathbf{cum}_{s_n}(p)\mathbf{a}^T_{*n}(\mathbf{B} + \theta\mathbf{D})\mathbf{a}_{*n}\}\mathbf{A}^T \\ &= \mathbf{A}(\mathbf{C}^{2,2}_{\mathbf{z},\mathbf{z}_p}(\mathbf{B}) + \theta diag\{\mathbf{cum}_{s_1}(p)\mathbf{a}^T_{*1}\mathbf{D}\mathbf{a}_{*1}, ..., \mathbf{cum}_{s_n}\mathbf{a}^T_{*n}\mathbf{D}\mathbf{a}_{*n}\})\mathbf{A}^T. \end{aligned}$$

The theorem will be proved, if

$$\mathbf{cum}_{s_j}(p)\mathbf{a}_{*j}^T\mathbf{D}\mathbf{a}_{*j} \neq \mathbf{cum}_{s_{j+1}}(p)\mathbf{a}_{*j+1}^T\mathbf{D}\mathbf{a}_{*j+1}.$$

We calculate:

$$\begin{aligned}
\mathbf{cum}_{s_j}(p)\mathbf{a}_{*j}^T\mathbf{D}\mathbf{a}_{*j} &= \mathbf{cum}_{s_j}(p)\mathbf{a}_{*j}^T\mathbf{A}diag\{\mathbf{cum}_{s_1}(p)\mathbf{a}_{*1}^T\mathbf{V}\mathbf{a}_{*1},...,\mathbf{cum}_{s_n}(p)\mathbf{a}_{*n}^T\mathbf{V}\mathbf{a}_{*n}\}\mathbf{A}^T\mathbf{a}_{*j} \\
&= \mathbf{cum}_{s_j}(p)\mathbf{e}_j^T diag\{\mathbf{cum}_{s_1}(p)\mathbf{a}_{*1}^T\mathbf{V}\mathbf{a}_{*1},...,\mathbf{cum}_{s_n}(p)\mathbf{cum}_{s_n}(p)\mathbf{a}_{*n}^T\mathbf{V}\mathbf{a}_{*n}\}\mathbf{e}_j \\
&= \mathbf{cum}_{s_j}(p)^2\mathbf{a}_{*j}^T\mathbf{V}\mathbf{a}_{*j}\mathbf{cum}_{s_j}(p)^2\sum_{k,l=1}^{n} a_{k,j}u_{j,k}a_{l,j}u_{j,l} - a_{k,j}u_{j+1,k}a_{l,j}u_{j+1,l} \\
&= \mathbf{cum}_{s_j}(p)^2(\alpha_1^2 - \beta_1^2).
\end{aligned}$$

Analogously we obtain:

$$\mathbf{cum}_{s_{j+1}}(p)\mathbf{a}_{*j+1}^T\mathbf{D}\mathbf{a}_{*j+1} = \mathbf{cum}_{s_{j+1}}(p)^2(\alpha_2^2 - \beta_2^2).$$

Since $\alpha_1^2 - \beta_1^2 = -(\alpha_2^2 - \beta_2^2)$, we obtain

$$\mathbf{cum}_{s_j}(p)\mathbf{a}_{*j}^T\mathbf{D}\mathbf{a}_{*j} - \mathbf{cum}_{s_{j+1}}(p)\mathbf{a}_{*j+1}^T\mathbf{D}\mathbf{a}_{*j+1} = (\alpha_1^2 - \beta_1^2)(\mathbf{cum}_{s_j}(p)^2 + \mathbf{cum}_{s_{j+1}}(p)^2).$$

Note that $\alpha_1^2 \neq \beta_1^2$, since $\mathbf{u}_j$ and $\mathbf{u}_{j+1}$ are linearly independent. So, the above expression is nonzero, which finishes the proof of the theorem. ∎

5 Blind Source Extraction using time delays

In this section we prove that BSS problem can be converted to a symmetric eigenvector problem. So, any algorithm for eigenvector problem can be used to estimate the mixing matrix and therefore, to separate simultaneously sources with different temporal structures, or different cumulants.

The use of second statistics approach for blind separation of temporally correlated sources has been developed and analyzed by many researchers, including [3]-[5], [19]-[23], etc.

Our approach has unified form for the second order statistic and high order statistics using matrix cumulants with time delays and it allows to extract colored or signals with different cumulant functions. Even for second order statistics our approach has some advantages that may not be found in others known results at the same time. It allows us to control successfulness of the separation by observing the eigenvalues; it provides relative fast convergence (since several algorithm has been developed for the EVD with cubic convergence [10]); 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).

In our method below we need the global mixing matrix to be orthogonal. We use either standard orthogonalization procedure, or robust (to additive white noise) one [4], when it is possible. We perform such an orthogonalization by a linear transformation $\mathbf{z} = \mathbf{Q}\mathbf{x}$ such that the matrix $\mathbf{A} = \mathbf{Q}\mathbf{H}$ is orthogonal, i.e. $\mathbf{A}^T\mathbf{A} = \mathbf{I}$, so our model is $\mathbf{z}(k) = \mathbf{A}\mathbf{s}(k) + \tilde{\mathbf{n}}(k)$ $(\tilde{\mathbf{n}} = \mathbf{Q}\mathbf{n})$.

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

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

$$\forall i, j \neq i \quad \exists l_{i,j} \in \{1, ..., L\} :$$
$$E\{s_i(t)s_i(t - p_{l_{i,j}})\} \neq E\{s_j(t)s_j(t - p_{l_{i,j}})\} \qquad (\mathbf{DAF}(P))$$

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

$$\forall i, j \neq i \exists l_{i,j} \in \{1, ..., L\} : \ \mathbf{cum}_{s_i}(p_{l_{i,j}}) \neq \mathbf{cum}_{s_j}(p_{l_{i,j}}), \qquad (\mathbf{DCF}(P))$$

where

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

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

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

$$\mathbf{R_z}(p) = E\{\mathbf{z}\mathbf{z}_p^T\}, \text{ (resp. } \mathbf{R_s}(p) = E\{\mathbf{s}\mathbf{s}_p^T\}), \tag{22}$$

where

$$\mathbf{z}_p = \mathbf{z}(k-p), \mathbf{z} = \mathbf{z}(k), \mathbf{s}_p = \mathbf{s}(k-p), \mathbf{s} = \mathbf{s}(k). \tag{23}$$

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

$$\mathbf{C}^{2,2}_{\mathbf{z},\mathbf{z}_p} = \quad E\{\mathbf{z}\mathbf{z}^T\mathbf{z}_p^T\mathbf{z}_p\} - E\{\mathbf{z}\mathbf{z}^T\}\mathrm{tr}E\{\mathbf{z}_p\mathbf{z}_p^T\} - 2E\{\mathbf{z}\mathbf{z}_p^T\}E\{\mathbf{z}_p\mathbf{z}^T\} \tag{24}$$

where E is the mathematical expectation (with respect to k in (23)) 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 $\mathbf{C}^{2,2}_{\mathbf{z},\mathbf{z}_p}$ is

$$C^{2,2}_{\mathbf{z},\mathbf{z}_p}(i,j) = \sum_{l=1}^{n} \mathbf{cum}\{z_i(k), z_j(k), z_l(k-p), z_l(k-p)\}.$$

It is clear that $\mathbf{C}^{2,2}_{\mathbf{z},\mathbf{z}_p} = \mathbf{C}^{2,2}_{\mathbf{z},\mathbf{z}_p}(\mathbf{I}_n)$, according to (14), where $\mathbf{I}_n$ is the unit $n \times n$ matrix.
For a given vector $\mathbf{b} \in \mathbb{R}^L$ define the following matrices for the chosen set P of time delays:

$$\mathbf{Z}(\mathbf{b}) \quad = \quad \sum_{i=1}^{L} b_i \mathbf{R}_z(p_i); \quad \tilde{\mathbf{Z}}(\mathbf{b}) = \sum_{i=1}^{L} b_i \mathbf{C}^{2,2}_{\mathbf{z},\mathbf{z}_{p_i}} \tag{25}$$

and similarly for the source signals

$$\mathbf{S}(\mathbf{b}) \quad = \quad \sum_{i=1}^{L} b_i \mathbf{R}_s(p_i); \quad \tilde{\mathbf{S}}(\mathbf{b}) = \sum_{i=1}^{L} b_i \mathbf{C}^{2,2}_{\mathbf{s},\mathbf{s}_{p_i}}. \tag{26}$$

We recall that the source signals are *uncorrelated*, if $\mathbf{R}_s(p)$ are diagonal matrices for every $p \geq 1$. 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 $E\{\mathbf{s}_i(k)\mathbf{s}_i(k-p)\}$. 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 $p \geq 1$ 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 $p_0 \geq 1$, $\mathbf{cum}_{s_i}(p_0)$ 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

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

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

for every $p_i \geq 1, i = 1, ..., 4$).

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

Lemma 6 *Assume that the mixing matrix* $\mathbf{A}$ *is orthogonal, the noise* $\mathbf{n}$ *is white and* $\mathbf{S}(\mathbf{b})$ *is a diagonal matrix (resp. the noise* $\mathbf{n}$ *is white of order 4 and* $\tilde{\mathbf{S}}(\mathbf{b})$ *is a diagonal matrix). Then the matrix* $\mathbf{Z}(\mathbf{b})$ *(resp.* $\tilde{\mathbf{Z}}(\mathbf{b})$*) is symmetrical and can be decomposed as* $\mathbf{Z}(\mathbf{b}) = \mathbf{A}\mathbf{S}(\mathbf{b})\mathbf{A}^T = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$ *(resp.* $\tilde{\mathbf{Z}}(\mathbf{b}) = \mathbf{A}\tilde{\mathbf{S}}(\mathbf{b})\mathbf{A}^T = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$*), where* $\mathbf{U}$ *is an orthogonal matrix and* $\mathbf{\Lambda}$ *is a diagonal matrix. If the diagonal elements of* $\mathbf{\Lambda}$ *are distinct, then the mixing matrix can be estimated as* $\hat{\mathbf{A}} = \mathbf{U}$ *up to multiplication with arbitrary permutation and diagonal sign matrices.*

Theorem 7 *Assume that the mixing matrix* $\mathbf{A}$ *is orthogonal and (i): the source signals are colored and uncorrelated, the noise* $\mathbf{n}$ *is white and condition* $(\mathbf{DAF}(P))$ *is satisfied (resp. (ii): the source signals are colored of order 4 and uncorrelated of order 4, condition* $(\mathbf{DCF}(P))$ *is satisfied and the noise* $\mathbf{n}$ *is white of order 4). Then*

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

(b) If $\mathbf{U}$ *is given from an EVD of the matrix* $\mathbf{Z}(\mathbf{b})$ *(resp. the matrix* $\tilde{\mathbf{Z}}(\mathbf{b})$*) for some* $\mathbf{b} \in B(L)$*, i.e.* $\mathbf{Z}(\mathbf{b}) = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$*, (resp.* $\tilde{\mathbf{Z}}(\mathbf{b}) = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$*), 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 sign matrices).*

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

(a) Since $s_i, i = 1, ..., n$ are uncorrelated, $\mathbf{S}(\mathbf{b})$ is a diagonal matrix and by Lemma 1, $\mathbf{Z}(\mathbf{b}) = \mathbf{A}\mathbf{S}(\mathbf{b})\mathbf{A}^T$. 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 $\mathbb{R}^L$. 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 $\mathbf{S}(\mathbf{b})$, where $\mathbf{b} \in \mathbb{R}^L$. 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 $b_{l_{1,2}}$ ($l_{1,2}$ is defined by the condition $(\mathbf{DAF}(P))$). Then $\sigma_1(\mathbf{b}(1,2)) \neq \sigma_2(\mathbf{b}(1,2))$, because of the condition $(\mathbf{DAF}(P))$. 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 $b_{l_{i,j}}$ 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 $(\mathbf{DAF}(P))$ 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 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 [14]. ∎

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

6 Subdifferential dispersing algorithm for sources with distinct cumulant functions

We assume in section that the mixing matrix is orthogonal (after robust ortogonalization [4]) and the source signals have distinct cumulant functions, i.e. $\mathbf{cum}_{s_i}(p)$ and $\mathbf{cum}_{s_j}(p)$ as functions of p are different for any $i, j \neq i$). (If the mixing matrix is not orthogonal, the normalized cumulant functions should be different.) This condition could be considered as large generalization of the conditions given in [23], [7].

We have proved that the set of vectors $\mathbf{b} \in \mathbb{R}^L$ for which the matrix $\tilde{\mathbf{Z}}(\mathbf{b})$ (resp. $\mathbf{Z}(\mathbf{b})$) has distinct eigenvalues and an EVD of it leads to separation of the mixing signals, has a full measure (i.e. its complement has measure zero). This suggests a random choice of $\mathbf{b}$ to achieve separation. Nevertheless, here we propose a subdifferential dispersing algorithm for the eigenvalues of $\tilde{\mathbf{Z}}(\mathbf{b})$. When these cumulant functions are very different (i.e. if we can find apriori a small set of indexes $P = \{p_1, ..., p_L\}$ such that $\mathbf{cum}_{s_i}(p_{i,j}) \neq \mathbf{cum}_{s_i}(p_{i,j})$ for some $p_{i,j} \in P$) the following algorithm has an advantage for large scale problems (because the parameter here is vector, not matrix). Consider the function: $\psi(\mathbf{b}) = \min_{1 \leq i \leq m-1}\{\lambda_i(\mathbf{b}) - \lambda_{i+1}(\mathbf{b})\}$, where $\lambda_i(\mathbf{b})$ are the eigenvalues (in decreasing order) of the matrix $\tilde{\mathbf{Z}}(\mathbf{b})$. This function has properties similar to those of (17) and the algorithm for its maximization is the same as those of φ with the only difference that $\mathbf{B}$ in the above algorithm is replaced with the vector $\mathbf{b}$, and $\mathbf{W}_{i,j}$ are replaced with the vectors $\mathbf{w}_{i,k} = [\mathbf{u}_{i,k}^T \mathbf{C}_{\mathbf{z},\mathbf{z}_{p_1}}^{2,2}(\mathbf{I}_n)\mathbf{u}_{i,k}, ..., \mathbf{u}_{i,k}^T \mathbf{C}_{\mathbf{z},\mathbf{z}_{p_L}}^{2,2}(\mathbf{I}_n)\mathbf{u}_{i,k}]^T$.

The subdifferential $\partial\psi(\mathbf{b})$ is given by the formula:

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

where I_0 is the set where the minimum in the definition of ψ 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.

References

[1] V.M. Alekseev, V.M. Tihomirov and S.V. Fomin, *Optimal Control*, Nauka, 1979 (in Russian).

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

[3] 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.

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

[5] 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.

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

[7] 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.

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

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

[10] 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.

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

[12] P. Gr. Georgiev, A. Cichocki "Robust Blind Source Separation via Symmetric Eigenvalue Decomposition" (submitted).

[13] P. Gr. Georgiev, A. Cichocki, " Blind Source Separation via Symmetric Eigenvalue Decomposition", in Proc. ISSPA (Kuala Lumpur, Malaysia), pp. 17-20, 2001.

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

[15] 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.

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

[17] A. Hyvarinen and E. Oja *A fast fixed-point algorithm for independent component analysis*, Neural Comp., 9: 1483-1492, 1997.

[18] J. R. Magnus, H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, 1999.

[19] 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.

[20] K.-R. Mueller, P. Philips and A. Ziehe "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, pp. 312-316, Jan. 1999.

[21] D. T. Pham and P. Garat "Blind separation of mixtures of independent sources through a quasi-maximum likelihood approach". *IEEE Trans. on Signal Processing*, Vol. 45, No. 7, pp. 1712-1725, 1997.

[22] A. N. Shiryaev, *Probability*, Graduate Texts in Mathematics, Springer, 1996.

[23] 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