



**HAL**  
open science

# an iterative block procedure for the optimization of constrained contrast functions

Eric Moreau, Benoît Stoll

► **To cite this version:**

Eric Moreau, Benoît Stoll. an iterative block procedure for the optimization of constrained contrast functions. International Conference on Independent Component Analysis (ICA'99), Jan 1999, Aussois, France. hal-03792250

**HAL Id: hal-03792250**

**<https://hal.science/hal-03792250v1>**

Submitted on 29 Sep 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# AN ITERATIVE BLOCK PROCEDURE FOR THE OPTIMIZATION OF CONSTRAINED CONTRAST FUNCTIONS

*Eric MOREAU and Benoit STOLL*

MS-GESSY, ISITV, av. G. Pompidou, BP 56,  
F-83162 La Valette du Var Cedex, France  
e-mail: `moreau@isitiv.univ-tln.fr`

## ABSTRACT

We consider the source separation problem based on the maximization of contrast functions. A general iterative procedure for the optimization of any contrasts working on white vectors is proposed. It does not require any parametrization of the separating matrix. Comparisons with the ICA and JADE algorithms by computer simulations illustrate the good performances of this new algorithm and the potential interest of such an approach.

## 1. INTRODUCTION

We consider the source separation problem [5] which can be simply formulated as follows. Several inobservable linear spatial mixtures of some independent signals called “sources” are observed. The purpose is to recover the original sources assuming that the mixing system is unknown.

In the past ten years, numerous solutions to this problem have been proposed. In this paper we are concerned with block (off-line) algorithms based on high-order (higher than two) statistics. Related algorithms are JADE [1] and ICA [3]. The JADE algorithm is based on an algebraical diagonalization of a fourth-order cumulant matrix. The ICA algorithm is based on the optimization of a specific contrast function which is the sum of the squares of outputs auto-cumulants. Moreover this contrast requires white vectors, *i.e.* vectors whose covariance matrix is equal to the identity.

Since [3], numerous contrasts working on white vectors have been proposed, *e.g.* [6]-[9]. Our main interest in this paper is to propose a general simple iterative procedure for the optimization of any contrasts working on white vectors. Unlike ICA, the proposed algorithm does not require a parametrization of the separating matrix thanks to Givens (planar) rotations. Thus this leads to a “simplified” optimization procedure. The basic idea is to take into account the whitening constraint using a kind of “projected gradient” procedure.

## 2. PROBLEM FORMULATION

We consider the classical linear memoryless mixture model

$$\mathbf{x}(n) = \mathbf{G}\mathbf{a}(n) \quad (1)$$

where  $n \in \mathbb{Z}$  is the discrete time,  $\mathbf{a}(n)$  the  $(N,1)$  vector of  $N \in \mathbb{N}^* \setminus \{1\}$  inobservable *real* input signals  $a_i(n)$ ,  $i \in \{1, \dots, N\}$ , called sources,  $\mathbf{x}(n)$  the  $(N,1)$  vector of observed signals  $x_i(n)$ ,  $i \in \{1, \dots, N\}$ , and  $\mathbf{G}$  the  $(N,N)$  square mixing matrix, assumed *invertible*. For clarity, in this article we restrict our attention to the real case although the following derivations may be easily extended to the complex one.

The sources are assumed to satisfy the following three assumptions

**A1a.** “*Independence*” The sources  $a_i(n)$ ,  $i \in \{1, \dots, N\}$ , are zero-mean, unit power and statistically mutually independent;

**A1b.** “*Stationarity*”  $\mathbf{a}(n)$  is a random vector stationary up to order under consideration, *i.e.*  $\forall i \in \{1, \dots, N\}$ , the cumulant  $\text{Cum} \underbrace{[a_i(n), \dots, a_i(n)]}_{p \times}$  is an independent function of  $n$ , denoted by  $C_p[a_i]$ ;

**A1c.** “*At most one null cumulant*” Given the order  $p \in \mathbb{N}^*$ , the cumulants of sources are assumed to satisfy one of the two following conditions:

- c1.  $|C_p[a_1]| \geq \dots \geq |C_p[a_N]| > 0$ ;
- c2.  $|C_p[a_1]| \geq \dots \geq |C_p[a_{N-1}]| > |C_p[a_N]| = 0$ .

In particular this means that at most one of the cumulants  $C_p[a_i]$ ,  $i \in \{1, \dots, N\}$ , is null.

Notice that the unit power assumption and **A1c** can be satisfied without loss of generality because sources are inobservable and the mixture is unknown.

The separation task is realized by computing a vector

$$\mathbf{y}(n) = \mathbf{H}\mathbf{x}(n) \quad (2)$$

in such a way that the different components of the  $(N, 1)$  vector  $\mathbf{y}(n)$  restore the different sources. Hence the separation problem is solved if the global matrix defined by

$$\mathbf{S} \stackrel{\text{def}}{=} \mathbf{H}\mathbf{G} \quad (3)$$

reads

$$\mathbf{S} = \mathbf{D}\mathbf{P} \quad (4)$$

where  $\mathbf{D}$  is an invertible diagonal matrix corresponding to arbitrary attenuations for restored sources and  $\mathbf{P}$  a permutation matrix corresponding to an arbitrary order of restitution. According to (2), (1) and (3) the output vector reads

$$\mathbf{y}(n) = \mathbf{S}\mathbf{a}(n). \quad (5)$$

### 3. ABOUT CONTRAST FUNCTIONS

In the following the explicit dependence of sources, observations and outputs vectors with the discrete time  $n$  will be omitted whenever no confusion is possible.

Before recalling the definition of a contrast, it is useful to define some notations. Let  $\mathcal{A}$  denotes the set of random vectors satisfying assumptions A1a, A1b and A1c. Let  $\mathcal{S}$  denotes the set of invertible matrices. The subset of  $\mathcal{S}$  of matrices  $\mathbf{S}$  of the form (4) is denoted by  $\mathcal{P}$  and the subset of  $\mathcal{P}$  of invertible diagonal matrices is denoted by  $\mathcal{D}$ . The set of random vectors  $\mathbf{y}(n)$  built from (5) where  $\mathbf{a}(n) \in \mathcal{A}$  and  $\mathbf{S} \in \mathcal{S}$  is denoted by  $\mathcal{Y}$ .

As initially defined [3] a contrast has to be a symmetrical and scale invariant function which must be maximized to get separation. Examples of such contrasts can be found e.g. in [3],[6]-[9].

In [9] it is shown that the symmetrical property of contrast is rather restrictive and not really of prime necessity. It is the reason why the following generalized definition of contrast was proposed.

**Definition.** “Non symmetrical contrast” A contrast on  $\mathcal{Y}$  is a multivariate mapping  $\mathcal{I}(\cdot)$  from the set  $\mathcal{Y}$  to  $\mathbb{R}$  which satisfies the following three requirements:

- R1.  $\forall \mathbf{y} \in \mathcal{Y}, \forall \mathbf{D} \in \mathcal{D}, \mathcal{I}(\mathbf{D}\mathbf{y}) = \mathcal{I}(\mathbf{y});$
- R2.  $\forall \mathbf{a} \in \mathcal{A}, \forall \mathbf{S} \in \mathcal{S}, \mathcal{I}(\mathbf{S}\mathbf{a}) \leq \mathcal{I}(\mathbf{a});$
- R3.  $\exists \mathcal{P}_d \subset \mathcal{P}, \mathcal{P}_d \neq \emptyset / \forall \mathbf{a} \in \mathcal{A}, \forall \mathbf{S} \in \mathcal{S}, \mathcal{I}(\mathbf{S}\mathbf{a}) = \mathcal{I}(\mathbf{a}) \Leftrightarrow \mathbf{S} \in \mathcal{P}_d.$

According to this definition, the maximization of a contrast is a sufficient condition for source separation. That is, if  $\mathcal{I}(\cdot)$  is such a contrast, then a sufficient condition of separation is  $\mathcal{I}(\mathbf{y}) = \mathcal{I}(\mathbf{a})$ .

Classical contrasts rely on the important notion of white vectors, *i.e.* vectors  $\mathbf{y}$  such that their covariance matrix  $\mathbf{R}_y \stackrel{\text{def}}{=} \mathbf{E}[\mathbf{y}\mathbf{y}^T]$  equals the identity. Because of

assumption A1a on sources then  $\mathbf{R}_a = \mathbf{I}$ . Hence  $\mathbf{R}_y = \mathbf{I}$  is equivalent to  $\mathbf{S}\mathbf{S}^T = \mathbf{I}$ , *i.e.*  $\mathbf{S}$  is an unitary matrix. Hence from now, we only consider such vectors. For this let  $\mathcal{U}$  denotes the set of unitary matrices and  $\mathcal{Y}_u$  the set of random vectors  $\mathbf{y}(n)$  built from (5) where  $\mathbf{a}(n) \in \mathcal{A}$  and  $\mathbf{S} \in \mathcal{U}$ .

Among the new contrasts (on  $\mathcal{Y}_u$ ) proposed in [9], three of them will be of particular interest in the following. The first one is

$$\mathcal{I}_1(\mathbf{y}) = \sum_{i=1}^N \gamma_i |\mathbf{C}_p[y_i]| \quad (6)$$

where the real numbers  $\gamma_i, i \in \{1, \dots, N\}$ , are assumed to satisfy

$$\gamma_1 \geq \dots \geq \gamma_N > 0. \quad (7)$$

The second contrast

$$\mathcal{I}_2(\mathbf{y}) = \varepsilon_p \sum_{i=1}^N \gamma_i \mathbf{C}_p[y_i] \quad (8)$$

requires an additional assumption on sources, that is

A1d. “Identical sign” When cumulants of sources satisfy condition c1 of A1c, we assume that  $\forall i \in \{1, \dots, N\}$ ,  $\text{sign}(\mathbf{C}_p[a_i]) = \varepsilon_p$  while when they satisfy condition c2 of A1c, we assume that  $\forall i \in \{1, \dots, N-1\}$ ,  $\text{sign}(\mathbf{C}_p[a_i]) = \varepsilon_p$  while by convention  $\text{sign}(\mathbf{C}_p[a_N]) = \varepsilon_p$ .

The third contrast which is perhaps the simplest one, is

$$\mathcal{I}_3(\mathbf{y}) = \varepsilon_4 \sum_{i=1}^N \gamma_i \mathbf{E}[y_i^4]. \quad (9)$$

It also requires assumption A1d.

Finally it has to be noticed that if source  $a_N$  is such that  $\mathbf{C}_p[a_N] = 0$  then in the three hereabove contrasts one can take  $\gamma_N = 0$  while the  $N-1$  other ones satisfy  $\gamma_1 \geq \dots \geq \gamma_{N-1} > 0$ .

### 4. OPTIMIZATION OF A CONSTRAINED CONTRAST

Contrasts on  $\mathcal{Y}_u$ , say  $\mathcal{I}(\cdot)$  for generality (e.g.  $\mathcal{I}_1(\cdot)$ ,  $\mathcal{I}_2(\cdot)$  and  $\mathcal{I}_3(\cdot)$  hereabove defined), assume that vectors  $\mathbf{y}$  are white. Two strategies can be considered for the derivation of a practical maximization algorithm.

The first (classical) one consists in a two stages separation:

- s1. Realize a whitening of the observations:  $\mathbf{x}_b = \mathbf{B}\mathbf{x}$  such that  $\mathbf{R}_{x_b} = \mathbf{I}$ .

**s2.** find a matrix  $\mathbf{H}$  maximizing  $\mathcal{I}(\mathbf{H}\mathbf{x}_b)$  under the constraint  $\mathbf{H}\mathbf{H}^T = \mathbf{I}$ .

The second strategy consists in a one stage separation:

**p1.** find a matrix  $\mathbf{H}$  maximizing  $\mathcal{I}(\mathbf{H}\mathbf{x})$  under the constraint  $\mathbf{R}_y = \mathbf{H}\mathbf{R}_x\mathbf{H}^T = \mathbf{I}$ .

Some algorithms have already been designed to realize the first strategy [3][7]. However, it seems that the second strategy have not yet been considered. Hence following the ideas of [7], we derive hereafter an algorithm corresponding to this second strategy.

A classical way in order to obtain a (local) maximum of a contrast is to use a gradient based algorithm leading to the following recursive scheme

$$\mathbf{H}' = \mathbf{H} + \lambda \dot{\mathbf{I}} \quad (10)$$

where  $\mathbf{H} = (H_{i,j})$  and  $\mathbf{H}'$  are respectively the separating matrices before and after the iteration,  $\lambda$  a positive constant and  $\dot{\mathbf{I}} = \partial\mathcal{I}/\partial\mathbf{H}$  the matrix whose  $(\ell, m)$  component is  $\partial\mathcal{I}/\partial H_{\ell,m}$ . This is referred to as a classical gradient approach.

However, and because parameters we are looking for are matrices, one can also consider a relative or natural gradient approach [2]. This leads to the following recursive scheme

$$\mathbf{H}' = \left( \mathbf{I} + \lambda \dot{\mathbf{I}} \mathbf{H}^T \right) \mathbf{H} . \quad (11)$$

Clearly if  $\mathbf{H}$  is an unitary matrix (as in the first strategy), (10) and (11) are identical. However if not (as in the second strategy) this leads to a different algorithm. In the following, we only consider the estimation in (11).

Let us now consider specifically contrasts  $\mathcal{I}_1(\cdot)$ ,  $\mathcal{I}_2(\cdot)$  and  $\mathcal{I}_3(\cdot)$ . Differentiating  $\mathcal{I}_1(\cdot)$  and  $\mathcal{I}_2(\cdot)$  with respect to  $H_{\ell,m}$  leads to

$$\frac{\partial\mathcal{I}_k}{\partial H_{\ell,m}} = \sum_{i=1}^N \gamma_i \beta_{k,i} \frac{\partial\mathcal{C}_p[y_i]}{\partial H_{\ell,m}}, \quad k = 1, 2 \quad (12)$$

where  $\beta_{1,i} = \text{sgn}(\mathcal{C}_p[y_i])$  is assumed constant and  $\beta_{2,i} = \varepsilon_p$ . According to (2), we have

$$\frac{\partial\mathcal{C}_p[y_i]}{\partial H_{\ell,m}} = p \delta[i, \ell] \text{Cum}[\underbrace{y_i, \dots, y_i}_{(p-1)\times}, x_m] \quad (13)$$

where  $\delta[i, \ell] = 1$  if  $\ell = i$  and 0 otherwise. Using (13) in (12), we finally have

$$\frac{\partial\mathcal{I}_k}{\partial H_{\ell,m}} = p \gamma_\ell \beta_{k,\ell} \text{Cum}[\underbrace{y_\ell, \dots, y_\ell}_{(p-1)\times}, x_m]. \quad (14)$$

Now differentiating  $\mathcal{I}_3(\cdot)$  with respect to  $H_{\ell,m}$  one has

$$\frac{\partial\mathcal{I}_3}{\partial H_{\ell,m}} = 4\varepsilon_2 \gamma_\ell \mathbb{E}[y_\ell^3 x_m^*]. \quad (15)$$

Recalling that  $\mathbf{H}$  has to respect certain constraints throughout the iterations ( $\mathbf{H}\mathbf{R}_x\mathbf{H}^T = \mathbf{I}$  for the second strategy considered), it is easily seen that the iterative procedures in (10) (or (11) in other respects) do not keep these constraints, *i.e.*  $\mathbf{H}'\mathbf{R}_x\mathbf{H}'^T \neq \mathbf{I}$ .

That is why a second step of ‘‘projection’’ is needed in the algorithm. It corresponds to a normalization of  $\mathbf{H}'$  in order to maintain the constraint throughout the iterations. This normalization will be denoted by  $\mathcal{N}(\mathbf{H}')$  in the following.

Because  $\mathbf{H}'\mathbf{R}_x\mathbf{H}'^T \neq \mathbf{I}$ , the reconstructed signals  $\mathbf{y}' = \mathbf{H}'\mathbf{x}$  built from  $\mathbf{H}'$  does not have a covariance matrix equal to the identity. Thus one would like to find a vector  $\mathbf{y}$  which is the closest to  $\mathbf{y}'$  while having an identity covariance matrix. For this task let us define an error vector as

$$\mathbf{e} = \mathbf{y}' - \mathbf{y} = (\mathbf{H}' - \mathbf{H}) \mathbf{x} . \quad (16)$$

Our objective is now to determine a matrix  $\mathbf{H}$  satisfying  $\mathbf{H}\mathbf{R}_x\mathbf{H}^T = \mathbf{I}$  and such that the energy of  $\mathbf{e}$  is minimum. We have the following result

**Proposition 1.** *A matrix  $\mathbf{H}$  satisfying  $\mathbf{H}\mathbf{R}_x\mathbf{H}^T = \mathbf{I}$  which minimizes the error energy  $E_e = \mathbb{E}[\mathbf{e}^T \mathbf{e}]$  where  $\mathbf{e} = (\mathbf{H}' - \mathbf{H}) \mathbf{x}$  for a given matrix  $\mathbf{H}'$ , is given by  $\mathbf{H} = \mathbf{U}_t \mathbf{V}_t^T \mathbf{M}^{-1}$  where  $\mathbf{M}$  is a square root of  $\mathbf{R}_x$  and  $\mathbf{U}_t$  and  $\mathbf{V}_t$  are the unitary matrices derived from the SVD of  $\mathbf{H}'\mathbf{M}$  as  $\mathbf{H}'\mathbf{M} = \mathbf{U}_t \mathbf{D}_t \mathbf{V}_t^T$ . •*

*Proof.* We have from (16)

$$\begin{aligned} E_e &= \mathbb{E}[\mathbf{x}^T (\mathbf{H}' - \mathbf{H})^T (\mathbf{H}' - \mathbf{H}) \mathbf{x}] \\ &= \mathbb{E}[\text{Tr}\{(\mathbf{H}' - \mathbf{H}) \mathbf{x} \mathbf{x}^T (\mathbf{H}' - \mathbf{H})^T\}] \\ &= \text{Tr}\{(\mathbf{H}' - \mathbf{H}) \mathbf{R}_x (\mathbf{H}' - \mathbf{H})^T\} \\ &= \text{Tr}\{\mathbf{H}'\mathbf{R}_x\mathbf{H}'^T + \mathbf{H}\mathbf{R}_x\mathbf{H}^T - \mathbf{H}'\mathbf{R}_x\mathbf{H}^T - \mathbf{H}\mathbf{R}_x\mathbf{H}'^T\} \\ &= C - E'_e \end{aligned} \quad (17)$$

where  $C = \text{Tr}\{\mathbf{H}'\mathbf{R}_x\mathbf{H}'^T\} + \text{Tr}\{\mathbf{H}\mathbf{R}_x\mathbf{H}^T\}$  (which is also equal to  $\text{Tr}\{\mathbf{H}'\mathbf{R}_x\mathbf{H}'^T\} + \text{Tr}\{\mathbf{I}\}$ ) is a constant term since it does not depend on  $\mathbf{H}$  and where

$$\begin{aligned} E'_e &= \text{Tr}\{\mathbf{H}'\mathbf{R}_x\mathbf{H}^T + \mathbf{H}\mathbf{R}_x\mathbf{H}'^T\} \\ &= \text{Tr}\{\mathbf{H}'\mathbf{R}_x\mathbf{H}^T + (\mathbf{H}'\mathbf{R}_x\mathbf{H}^T)^T\} \\ &= 2 \text{Tr}\{\mathbf{H}'\mathbf{R}_x\mathbf{H}^T\} . \end{aligned} \quad (18)$$

Because  $C$  is constant, the minimization of  $E_e$  is equivalent to the maximization of  $E'_e$  in (18). Let us now reformulate the problem. Let  $\mathbf{M}$  be a square root of  $\mathbf{R}_x$  then we have to find a maximum of

$$E'_e = 2 \operatorname{Tr}\{\mathbf{H}'\mathbf{M}\mathbf{M}^T\mathbf{H}^T\} = 2 \operatorname{Tr}\{\mathbf{B}\mathbf{A}^T\}$$

where  $\mathbf{B} = \mathbf{H}'\mathbf{M}$  and  $\mathbf{A} = \mathbf{H}\mathbf{M}$  has to satisfy  $\mathbf{A}\mathbf{A}^T = \mathbf{I}$ , i.e.  $\mathbf{A}$  is unitary.

This latter problem is a classical algebraical one [4]. Let us briefly recall the solution according to our notation. If  $\mathbf{B} = \mathbf{U}_t\mathbf{D}_t\mathbf{V}_t^T$  is a singular value decomposition of  $\mathbf{B}$ , then

$$\operatorname{Tr}\{\mathbf{B}\mathbf{A}^T\} = \operatorname{Tr}\{\mathbf{U}_t\mathbf{D}_t\mathbf{V}_t^T\mathbf{A}^T\} = \operatorname{Tr}\{\mathbf{D}_t\mathbf{V}_t^T\mathbf{A}^T\mathbf{U}_t\}.$$

Thus  $\operatorname{Tr}\{\mathbf{B}\mathbf{A}^T\} = \sum_{i=1}^N \sigma_i(\mathbf{B})T_{i,i}$  where  $\sigma_i(\mathbf{B})$  are the singular values of  $\mathbf{B}$  and  $\mathbf{T} = (T_{i,j}) = \mathbf{V}_t^T\mathbf{A}^T\mathbf{U}_t$  is a unitary matrix. This sum is clearly maximized when  $T_{i,i} = 1$  for all  $i$ , that is, when  $\mathbf{A} = \mathbf{U}_t\mathbf{V}_t^T$ . Finally, because  $\mathbf{H} = \mathbf{A}\mathbf{M}^{-1}$ , the matrix  $\mathbf{H}$  maximizing  $E'_e$  is thus  $\mathbf{H} = \mathbf{U}_t\mathbf{V}_t^T\mathbf{M}^{-1}$ . •

Hence according to the hereabove Proposition, if we consider the following decomposition of the covariance matrix  $\mathbf{R}_x = \mathbf{M}\mathbf{M}^T$ , then the constraint can be written as  $\mathbf{U}_d\mathbf{U}_d^T = \mathbf{I}$  where  $\mathbf{U}_d = \mathbf{H}\mathbf{M}$ . Further, let us consider the SVD of  $\mathbf{H}'\mathbf{M}$  as  $\mathbf{H}'\mathbf{M} = \mathbf{U}_t\mathbf{D}_t\mathbf{V}_t^T$  where matrices  $\mathbf{U}_t$  and  $\mathbf{V}_t$  are unitary and  $\mathbf{D}_t$  is diagonal, then we can impose

$$\mathcal{N}(\mathbf{H}') = \mathbf{U}_t\mathbf{V}_t^T\mathbf{M}^{-1} \quad (19)$$

which corresponds to the best (in the mean square error sense) choice of a matrix leading an output vector  $\mathbf{y}$  with the required constraint.

Let us now summarize the One Stage Iterative Block Separation algorithm (OSIBS):

#### OSIBS algorithm

1. Estimate the covariance matrix of the observations:  $\hat{\mathbf{R}}_x$ .
2. Compute the SVD of  $\hat{\mathbf{R}}_x$ :  $\hat{\mathbf{R}}_x = \mathbf{U}_r\mathbf{D}_r\mathbf{U}_r^T$ .
3. Compute a square root of  $\hat{\mathbf{R}}_x$ : for example  $\mathbf{M} = \mathbf{U}_r\mathbf{D}_r^{1/2}\mathbf{U}_r^T$ .
4. Depending on *a priori* information about sources, choose a contrast:  $\mathcal{I}_1(\cdot)$ ,  $\mathcal{I}_2(\cdot)$  or  $\mathcal{I}_3(\cdot)$ , say  $\mathcal{I}(\cdot)$ .
5. Choose an initial value for  $\mathbf{H}$  (e.g.  $\mathbf{H} = \mathbf{I}$ ) and choose a real constant  $\lambda$ . Initialize the outputs:  $\mathbf{y} = \mathbf{H}\mathbf{x}$  and then iterate (a)-(d) until convergence:

- (a) Estimate the derivative of a contrast  $\hat{\mathcal{I}}$ .
- (b) Compute  $\mathbf{H}' = (\mathbf{I} + \lambda \hat{\mathcal{I}} \mathbf{H}^T) \mathbf{H}$ .
- (c) Compute  $\mathbf{H} = \mathcal{N}(\mathbf{H}')$  as defined in (19).
- (d) Compute the estimated source signals  $\mathbf{y} = \mathbf{H}\mathbf{x}$ .

## 5. COMPUTER SIMULATIONS

In order to illustrate the performances of the OSIBS algorithm, some computer simulations are now presented. Contrast  $\mathcal{I}_3(\cdot)$  is used. We consider the case of two sources ( $N = 2$ ) and the two following mixing matrices

$$\mathbf{G}_1 = \begin{pmatrix} 1 & 0.5 \\ 0.4 & 1 \end{pmatrix}, \quad \mathbf{G}_2 = \begin{pmatrix} 1 & 0.95 \\ 0.9 & 1 \end{pmatrix}. \quad (20)$$

A noisy model is considered where the additive noise on each channel is assumed to be a zero-mean, i.i.d. Gaussian random signal with equal power.

The performances of the algorithms are associated to an index/measure defined on the global matrix  $\mathbf{S} = (S_{i,j})$  according to

$$\operatorname{ind}(\mathbf{S}) = \frac{1}{2} \left[ \sum_i \left( \sum_j \frac{|S_{i,j}|^2}{\max_\ell |S_{i,\ell}|^2} - 1 \right) + \sum_j \left( \sum_i \frac{|S_{i,j}|^2}{\max_\ell |S_{\ell,j}|^2} - 1 \right) \right]. \quad (21)$$

Indeed, this non-negative index is zero if  $\mathbf{S}$  satisfies (4) and a small value indicates the proximity to the desired solutions.

For the input sources, two cases are considered:

- i*) the two sources are 2-PAM signals (taking their values in  $\{-1, +1\}$  with equal probability)
- ii*) the two sources are 4-PAM signals (taking their values in  $\{-3/\sqrt{5}, -1/\sqrt{5}, 1/\sqrt{5}, 3/\sqrt{5}\}$  with equal probability).

Concerning the algorithms, we compare 1) the JADE algorithm (Version 1.5 Dec. 1997. for the real case) represented by a dashdot line with plus points; 2) the ICA algorithm represented by a dotted line with x-mark points and 3) the OSIBS algorithm represented by a solid line with circle points.

#### Case 1: Index versus data number

First we plot the mean and the standard deviation (STD) of the estimated index over 500 independent runs as a function of the data number. In Fig.1 for case *i*) of sources and Fig.2 for case *ii*) (resp. Fig.3

and Fig.4 for the same respective case of sources) the mixture  $\mathbf{G}_1$  (resp.  $\mathbf{G}_2$ ) is used and the noise power is kept constant to 0dB (resp.  $-20$ dB). The figures show that performances of the OSIBS algorithm are a bit better in the presented cases.

### Case 2: Index versus noise power

Now the data number is held constant to  $N_d = 500$  and we plot the mean and STD of the estimated index over 500 independent runs as a function of the noise power. In Fig.5 for case *i*) of sources and Fig.6 for case *ii*) (resp. Fig.7 and Fig.8 for the same respective case of sources) the mixture  $\mathbf{G}_1$  (resp.  $\mathbf{G}_2$ ) is used. The figures show that the three algorithms have relatively similar performances with a little advantage in favour of OSIBS.

## 6. CONCLUSION

We have proposed an iterative procedure for the maximization of any kind of contrasts involving white vectors. It does not require a parametrization of the separating matrix which thus leads to a “simplest” algorithm. A comparison based on computer simulations illustrates the potential interest of such an approach at least in using the contrast  $\mathcal{I}_3(\cdot)$ .

## 7. REFERENCES

- [1] J.F. Cardoso and A. Souloumiac, “Blind Beamforming for non Gaussian Signals”, *IEE Proceedings-F*, Vol. 40, pp 362-370, 1993.
- [2] J.F. Cardoso and B.H. Laheld, “Equivariant adaptive source separation”, *IEEE Trans. Signal Processing*, Vol. 44, No. 12, pp 3017-3030, Dec. 1996.
- [3] P. Comon, “Independent component analysis, a new concept?”, *Signal Processing*, Vol. 36, pp 287-314, 1994.
- [4] R.A. Horn and C.R. Johnson, *Matrix analysis*, Cambridge University Press, 1985.
- [5] C. Jutten and J. Herault, “Blind separation of sources, Part I: An adaptative algorithm based on neuromimetic architecture”, *Signal Processing*, Vol. 24, pp 1-10, 1991.
- [6] E. Moreau and O. Macchi, “High order contrasts for self-adaptive source separation”, *International Journal of Adaptive Control and Signal Processing*, Vol. 10, No. 1, pp 19-46, Jan. 1996.

- [7] E. Moreau, “Criteria for complex sources separation”, in Proc. *EUSIPCO'96, European Signal Processing Conference*, Trieste, Italy, Vol. II, pp 931-934, September 1996.
- [8] E. Moreau and J.-C. Pesquet, “Generalized contrasts for multichannel blind deconvolution of linear systems”, *IEEE Signal Processing Letters*, Vol. 4, No. 6, pp 182-183, June 1997.
- [9] E. Moreau and N. Thirion-Moreau, “Non symmetrical contrasts for source separation”, *IEEE Trans. Signal Processing*, submitted, December 1997.

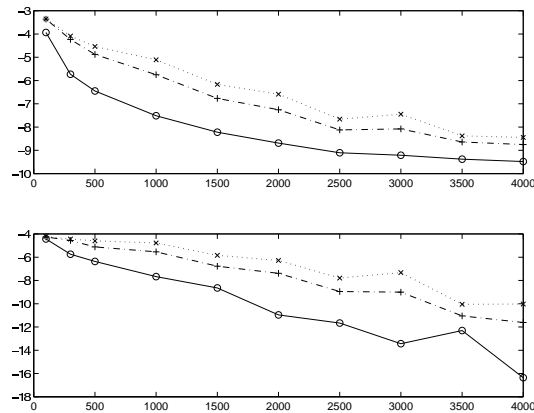


Figure 1: For case *i*) of sources and matrix  $\mathbf{G}_1$ , comparison of the experimental index w.r.t. data number.

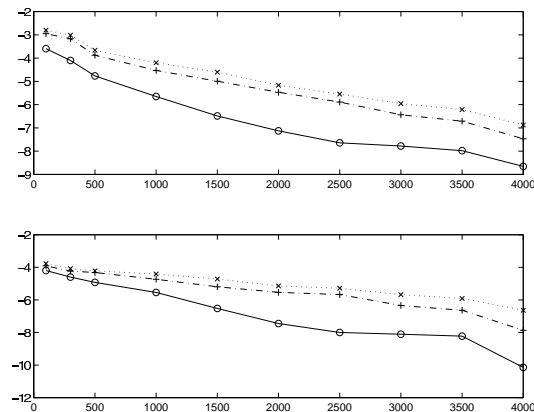


Figure 2: For case *ii*) of sources and matrix  $\mathbf{G}_1$ , comparison of the experimental index w.r.t. data number.

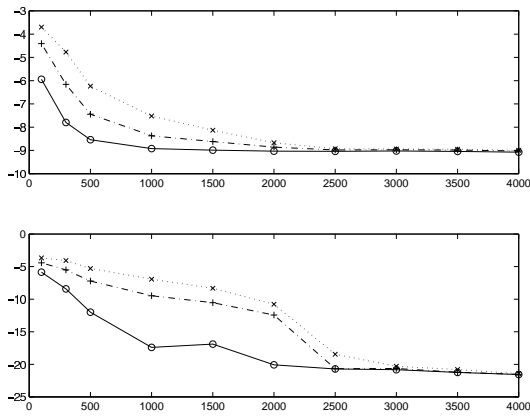


Figure 3: For case *i)* of sources and matrix  $G_2$ , comparison of the experimental index w.r.t. data number.

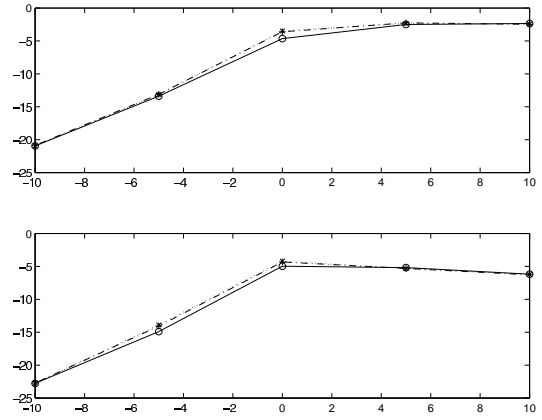


Figure 6: For case *ii)* of sources and matrix  $G_1$ , comparison of the experimental index w.r.t. noise power.

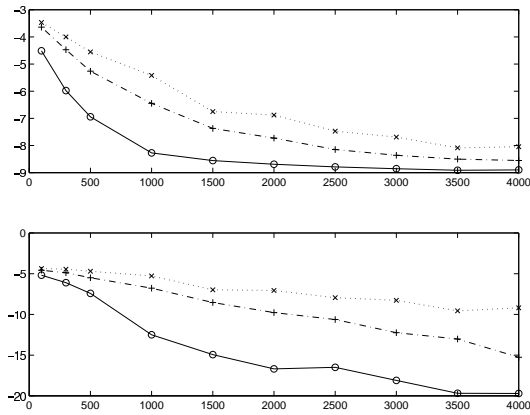


Figure 4: For case *ii)* of sources and matrix  $G_2$ , comparison of the experimental index w.r.t. data number.

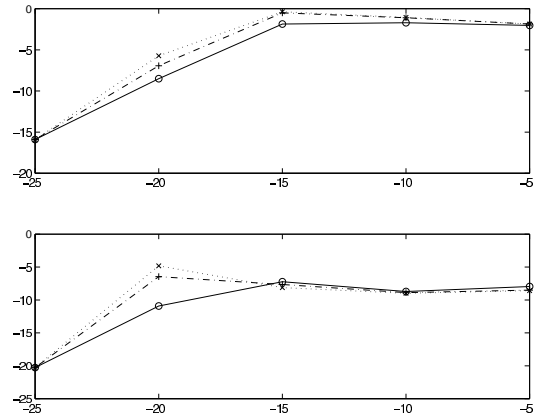


Figure 7: For case *i)* of sources and matrix  $G_2$ , comparison of the experimental index w.r.t. noise power.

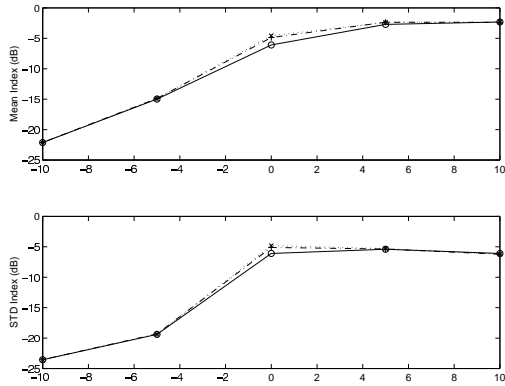


Figure 5: For case *i)* of sources and matrix  $G_1$ , comparison of the experimental index w.r.t. noise power.

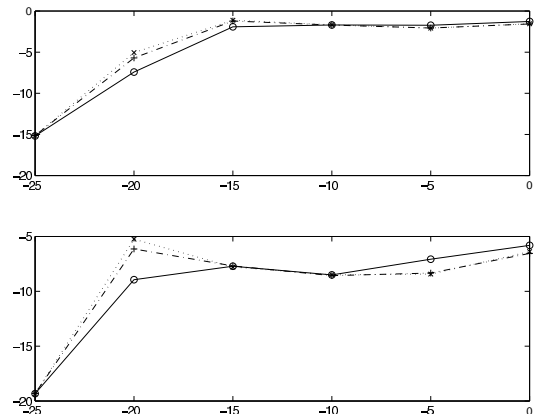


Figure 8: For case *ii)* of sources and matrix  $G_2$ , comparison of the experimental index w.r.t. noise power.