GENERALIZED CONTRAST FUNCTIONS FOR BLIND SEPARATION OF OVERDETERMINED LINEAR MIXTURES WITH UNKNOWN NUMBER OF SOURCES

Mohamed SAHMOUDI

TSI, Telecom Paris, 37-39 rue Dareau, 75014 Paris
E-mail: sahmoudi@tsi.enst.fr

ABSTRACT

In this paper, we address the problem of blind separation of m independent sources from their n linear mixtures in the overdetermined systems ($n \geq m$) with unknown number of sources. After generalizing the definition of classical and nonsymmetrical contrast functions, we exhibit a wide class of generalized contrast functions using some super-additive functionals and concave functions. Two practical generalized contrasts based on the support Lebesgue measure (SLM) and the mutual information (MI) criteria are proposed and discussed. Finally, computer simulations illustrate the results and demonstrate all the interest we can find in considering a generalized contrast function.

1. INTRODUCTION

In *blind source separation* (BSS), the goal is to extract statistically independent but otherwise unknown source signals from their mixtures without knowing the mixing coefficients [2]. This kind of blind techniques have applications in several areas, such as data communications, speech processing, and various biomedical signal processing problems (MEG/EEG data) [7]. Formally the linear BSS model is of the form

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t), \;\; t = 1 \dots T \tag{1}$$

where $\mathbf{x} = (x_1, \cdots, x_n)^T$ is a random vector of observations, $\mathbf{s} = (s_1, \cdots, s_m)^T$ is a random vector of hidden sources with mutually independent components, and $\mathbf{A}$ is an $n \times m$ nonsingular mixing matrix. Define $\mathbf{B} = \mathbf{A}^{-1}$, which is usually called the demixing or separating matrix. The goal of BSS is to find a separating matrix $\mathbf{B}$ such that the output vector

$$\mathbf{y} = \mathbf{B}\mathbf{x} \tag{2}$$

contains accurate estimates of the m (less than n) original source signals. It is well known that $\mathbf{A}$ (thus $\mathbf{B}$) is identifiable up to ambiguity of order, sign and scaling [2]. This type of unmixing problems is also called independent component analysis (ICA) in signal processing engineering. Among the great number of approaches that have been proposed in the recent literature, we are primary concerned with contrast function approach, e.g. [2, 9, 12, 13]. In this field, contrast functions constitute separation criteria in the sense that their maximization solve the source separation problem. Thus, a typical BSS algorithm is primarily composed of a contrast function and an optimization procedure. Although many methods and algorithms have been proposed for the BSS problem, in their corresponding models it is usually assumed that the number of sources signals is known in advance. Typically it should be equal to the number of sensors and outputs. However, in practice, these assumptions do not often hold. In that case, we can extract the sources one by one using a deflation procedure [5]. Such algorithms work regardless of the source number, but the separation quality would increasingly degrade due to accumulated errors. Another way consists in achieving the blind separation in two stages. Firstly, the observations are pre-processed such that the sensor vector is transformed to a white vector and at the same time the dimensionality is reduced from n to m, and then an $m \times m$ orthogonal matrix is determined to obtain source separation. This scheme suffers from poor separation results for ill conditioned mixing matrix or weak sources. In [1] the authors confirm surprisingly by extensive experiments that one can use directly the natural gradient algorithm to learn an $n \times n$ nonsingular matrix, and show that among n outputs at convergence, there are m independent components, and each of the remaining $n - m$ component is a rescaled or rearranged copy of some independent component. Recently, a generalized natural gradient algorithm was proposed in [14] to perform the BSS in the unknown number of sources case.

The first main objective of this paper is to introduce a generalized contrast functions as a valuable extension of the existing classical [2] and nonsymmetrical [9] ones to perform overdetermined blind separation with m unknown. The second objective is the construction of a family of generalized contrast functions composing the super additivity property of some independence measure indexes and the convexity property of a real function to obtain a more regular criteria. In particular, we use the proposed approach to construct two practical generalized contrasts. The first one is based on the SLM of the outputs and the second is based on the MI criterion. Finally, we extend the Jacobi like algorithms to maximizing the generalized SLM contrast and we use the generalized natural algorithm proposed in [14] for the generalized MI contrast optimization. Notice that one of the main advantages of our approach will be the application to the overdetermined blind separation $(n > m)$ with an unknown number of sources.

2. GENERALIZED CONTRAST FUNCTIONS

Let us define some notations which will be useful in the following. $\mathcal{Y}_m$ the set of m-dimensional output vectors built from the BSS model (2). $\mathcal{S}_m$ the set of source vectors which are assumed stationary and independent. $\mathcal{C}_p$ denotes the set of $p \times p$ nonsingular matrix. $\mathcal{D}_p$ denotes the set of $p \times p$ invertible diagonal matrices. $\mathcal{P}_p$ denotes the set of $p \times p$ generalized permutation matrices. $\mathscr{S}_n$ denotes the set of n-dimensional vector, such that the m first components of which are the m original sources signals and the rest $n - m$ components consist of zero or a copies of some source signal. All matrices W such that for any $\mathbf{z} \in \mathscr{S}_n$, $W\mathbf{z}$ is a re-scaled version of a vector $\tilde{\mathbf{z}} \in \mathscr{S}_n$, form the set $\mathcal{W}_n$.

We first recall the initial definition of a contrast [2].

Definition 1 - ***Classical Contrast:*** *A contrast on $\mathcal{Y}_m$ is a multivariate mapping $\mathcal{I}$ from the site $\mathcal{Y}_m$ to $\mathbb{R}$, which satisfies the following three requirements:*
R1 - $\forall \mathbf{y} \in \mathcal{Y}_m, \forall \mathbf{C} \in \mathcal{P}_m, \mathcal{I}(\mathbf{Cy}) = \mathcal{I}(\mathbf{y})$;
R2 - $\forall \mathbf{s} \in \mathcal{S}_m, \forall \mathbf{C} \in \mathcal{C}_m, \mathcal{I}(\mathbf{Cs}) \leq \mathcal{I}(\mathbf{s})$;
R3 - $\forall \mathbf{s} \in \mathcal{S}_m, \forall \mathbf{C} \in \mathcal{C}_m, \mathcal{I}(\mathbf{Cs}) = \mathcal{I}(\mathbf{s}) \Longleftrightarrow \mathbf{C} \in \mathcal{P}_m$

Such contrast are symmetrical and scale invariant functions (R1) which have to be maximised (R2) to get separation (R3). According to this definition, numerous contrasts has been proposed, see e.g. [2]. In order to consider non symmetrical functions, Moreau have been proposed the following definition of a contrast [9]:

Definition 2 - ***Nonsymmetrical Contrast:*** *Let $\mathcal{P}_d$ be a non-empty set of $\mathcal{P}$. A contrast function on $(\mathcal{Y}_m, \mathcal{P}_d)$ is a multivariate mapping $\mathcal{I}$ from the set $\mathcal{Y}_m$ to $\mathbb{R}$, which satisfies the following three requirements:*
R1 - $\forall \mathbf{y} \in \mathcal{Y}_m, \forall \mathbf{D} \in \mathcal{D}_m, \mathcal{I}(\mathbf{Dy}) = \mathcal{I}(\mathbf{y})$;
R2 - $\forall \mathbf{s} \in \mathcal{S}_m, \forall \mathbf{C} \in \mathcal{C}_m, \mathcal{I}(\mathbf{Cs}) \leq \mathcal{I}(\mathbf{s})$;
R3 - $\forall \mathbf{s} \in \mathcal{S}_m, \forall \mathbf{C} \in \mathcal{C}_m, \exists \mathcal{P}_d \subset \mathcal{P}_m, \mathcal{P}_d \neq \varnothing \,/\, \mathcal{I}(\mathbf{Cs}) = \mathcal{I}(\mathbf{s}) \Longleftrightarrow \mathbf{C} \in \mathcal{P}_d$

From these definition, it is can be seen that a classical contrast have to be symmetrical function and invariant under a phase shift. This implies that they are invariant under all separation states. Hence the global maximization of such a contrast is a necessary and sufficient condition for source separation, whereas a nonsymmetrical contrast is merely invariant to non-zero scale transformation and its maximization no longer necessary and sufficient condition for source separation but a sufficient one. Now, in order to consider overdetermined mixture with unknown number of sources, we propose the main idea of this paper which lies in that all local maxima of the generalized contrast must correspond to source separation, whereas only the global ones of the classical and nonsymmetrical contrasts do so. In other words, we propose a new definition of the concept of contrast function such that the maximization lead to n local maxima where n outputs are composed of m independent components and $n - m$ redundant components, whereas no redundant component is allowed using the classical and nonsymmetrical contrast.

Hence, for this task, it becomes necessary to extend the notion of contrast functions. It is the reason why the following generalized definition is proposed.

Definition 3 - ***Generalized Contrast:*** *We define a generalized contrast function on $(\mathcal{Y}_n, \bar{\mathcal{W}}_n)$ as a multivariate mapping $\mathcal{I}$ from the set $\mathcal{Y}_n$ to $\mathbb{R}$, which satisfies the following three requirements:*
R1 - $\forall \mathbf{y} \in \mathcal{Y}_n, \forall \mathbf{D} \in \mathcal{D}_n, \mathcal{I}(\mathbf{Dy}) = \mathcal{I}(\mathbf{y})$;
R2 - $\forall \mathbf{y} \in \mathcal{Y}_n, \exists \epsilon > 0, \mathcal{I}(\mathbf{y} + \epsilon \mathbf{y}) \leq \mathcal{I}(\mathbf{y}) \Rightarrow \exists\, \mathbf{z} \in \mathscr{S}_n$ *and* $\mathbf{D} \in \mathcal{D}_n$ *such that* $\mathbf{y} = \mathbf{Dz}$;
R3 - $\forall \mathbf{z} \in \mathscr{S}_n, \forall M \in \mathcal{C}_n, \exists\, \bar{\mathcal{W}}_n \subset \mathcal{W}_n, \mathcal{I}(M\mathbf{z}) = \mathcal{I}(\mathbf{z}) \Leftrightarrow M \in \bar{\mathcal{W}}_n$

Obviously, when a generalized contrast function is maximized, the outputs provide all the source signals up to ordering and non-zero scaling. It is also be noticed that contrasts in the classical and nonsymmetrical sense are contrasts in the generalized sense as special instances. A generalized contrast function $\mathcal{I}(\mathbf{y}, \mathbf{B})$ of the n output components reaches its local maxima if and only if $\mathbf{y} = \mathbf{Pz}$, where $\mathbf{P} \in \mathcal{P}_n$ (an $n \times n$ generalized permutation matrix) and $\mathbf{z} \in \mathscr{S}_n$. Hence the maximization of the generalized contrast function converge when the output $\mathbf{y}$ contains m independent components that are the separated sources and $n - m$ redundant components.

Theorem 1 *Maximizing a generalized contrast function, all the source signals are recovered at least once and the composite BSS system matrix $\mathbf{C} = \mathbf{BA}$ takes the following form*

$$\mathbf{C} = \mathbf{P} \begin{pmatrix} \mathbf{I}_m \\ -- \\ \mathbf{\Delta}_{n-m} \end{pmatrix} \quad (3)$$

where $\mathbf{P}$ is a generalized permutation matrix and each column vector δ_i of $\mathbf{\Delta}_{n-m}$ is either a null vector or a column of the identity matrix. In other words, we can straightforward that a generalized contrast of the BSS problem meaning

$$\mathcal{C}(\mathbf{y}; \mathbf{B}) = 0 \quad \Leftrightarrow \quad \mathbf{C} = \mathbf{P} \left(\mathbf{I}_m \mid \mathbf{\Delta}_{n-m} \right)^T$$

Such empirical findings were first reported by Cichocki et al. in [1] and from optimization point of view in [14] to extend the gradient algorithm to the unknown source number case. Here we present the theoretical justification behind it from the viewpoint of contrast function.

Remark 1 *Because of lack of place, the proofs are omitted and reported in a full-length journal paper version.*

3. GENERALIZED CONTRAST CONSTRUCTION

Definition 4 *A functional $\mathcal{F}$ of the distribution of a random variable X, denoted by $\mathcal{F}(X)$, is said to be scale equivariant if $\mathcal{F}(aX) = |a|\mathcal{F}(X)$ for any real number a.*

Note that if $\mathcal{F}$ is scale equivariant, then $|\mathcal{F}|$ is also scale equivariant. Hence, we can without loss of generality assume in this work $\mathcal{F} \geq 0$.

3.1. Super-Additivity and Concavity based Contrasts

Definition 5 *A functional $\mathcal{G}$ of the distribution of a random variable X, denoted by $\mathcal{G}(X)$, is said to be σ-super-additive if*

$$\mathcal{G}^{\sigma}(X+Y) \geq \mathcal{G}^{\sigma}(X) + \mathcal{G}^{\sigma}(Y) \tag{4}$$

for any two independent random variables X and Y. For $\sigma = 1$, this equation coincides with the classical super-additivity.

Lemma 1 *Let σ and α are two positive real numbers. Then, the σ-super-additivity implies the α-super-additivity if and only if $\sigma \leq \alpha$. In particular, it implies 2-super-additivity if and only if $\sigma \leq 2$.*

In a similar way, as the results that we have been presented in [12], we can prove the following its generalization.

Theorem 2 *Let $\mathcal{G}$ is a 2-super-additive and scale equivariant functional. Then for all increasing and concave real function g, the following objective function*

$$\mathcal{C}_g(y_1, \cdots, y_n; \mathbf{B}) = -\sum_{i=1}^{n} g[\mathcal{G}(y_i)] + g[|\det(\mathbf{B})|] \tag{5}$$

is a generalized contrast for blind separation of overdetermined linear mixtures with unknown source number.

Thus, $\mathcal{C}_g(\mathbf{y}; \mathbf{B})$ constitutes a family of objective functions and one needs only to find a 2-super-additive scale equivariant functional $\mathcal{G}$ to achieve the BSS. One can consider functions such as $g(u) = \lambda \ln(u)$, where $\lambda \geq 0$.

Remark 2 (Key remark.) *The main advantage of using an appropriate increasing concave function g is that the contrast function becomes more regular and easier to optimize.*

Remark 3 *When the source number is unknown, both the classical and the nonsymmetrical contrasts do not work, and it is necessary to use the generalized contrast. Since the number n of observations is always at hand, the separating matrix $\mathbf{B}$ can be imposed to be an $n \times n$ nonsingular matrix that we decompose as $\mathbf{B} = (\mathbf{B}_1 \mid \mathbf{B}_2)^T$ where $\mathbf{B}_1$ is composed of the first m rows of $\mathbf{B}$ and $\mathbf{B}_2$ is the submatrix composed of the remaining $n-m$ rows. Inspired from the mutual information contrast structure [14], it can be conjectured that the technical assumptions in theorem 2 can be replaced equivalently by the necessary and sufficient following conditions*

$$\begin{cases} \mathcal{C}_g(y_1, \cdots, y_n; \mathbf{B}) = \mathcal{C}_g(y_1, \cdots, y_m; \mathbf{B}_1) + \sum_{k=1}^{n-m} g[\mathcal{G}(y_{m+k})] \\ g[\mathcal{G}(y_{m+k})] \geq \max_{j=1,\ldots,m} g[\mathcal{G}([\mathbf{B}_2\mathbf{A}]_{m+k,j} s_j)] \end{cases}$$

3.2. Support Lebesgue Measure (SLM) Contrast

Here we present a criterion for the extraction of the sources whose density has the minimum SLM. Note that in this section, we will assume implicitly that the signal sources has measurable densities with finite SLM.

Definition 6 - ***Density Support Measure:*** *Let us denote the density of the output $\mathbf{y}$ as $f(\mathbf{y})$. We define the measure of the density support set of a random variable $\mathbf{y}$ as*

$$\mathcal{L}(\mathbf{y}) = \int_{\{y; f(y)>0\}} d\mu = \mu\,\{y; f(y) > 0\} \tag{6}$$

where $\mu(.)$ denotes the Lebesgue measure.

Unfortunately, this definition does not allow us to show the super-additivity property of the SLM criteria. Nevertheless, another probabilistic definition may be proposed.

Definition 7
1- *For any measure μ on $\mathbb{R}^d$, its support $\mathcal{S}_\mu$ is defined to be the set of $x \in \mathbb{R}^d$ such that $\mu(G) > 0$ for any open set G containing x. The support set $\mathcal{S}_\mu$ is a closed set.*
2- *For any random variable X on $\mathbb{R}^d$, the support of its distribution P_X is called the support of X and denoted by $\mathcal{S}_X = \mathcal{S}(X)$. It is the smallest closed set F satisfying $P(X \in F) = 1$.*

The following lemma is basic in using supports of random variables for BSS in the next.

Lemma 2 *If X and Y are independent random variables on $\mathbb{R}^d$, then $\mathcal{S}_{X+Y}$ is the closure of $\{x+y : x \in \mathcal{S}_X, y \in \mathcal{S}_Y\}$, that is, $\mathcal{S}_{X+Y} = \overline{\mathcal{S}_X + \mathcal{S}_Y}$*

***Proof* 1** *If $x \in \mathcal{S}_X$ and $y \in \mathcal{S}_Y$, then $x+y \in \mathcal{S}_{X+Y}$, since, for any $\epsilon > 0$,*

$$P\{|X+Y-x-y| < \epsilon\} \geq P\{|X-x| < \frac{\epsilon}{2}\} P\{|Y-y| < \frac{\epsilon}{2}\} > 0$$

Hence $\mathcal{S}_{X+Y} \supset \overline{\mathcal{S}_X + \mathcal{S}_Y}$.
Conversely, if K_1 and K_2 are both compact, then K_1+K_2 is compact. Consequently, $\mathcal{S}_X+\mathcal{S}_Y$ is the union of a countable number of compact sets, hence it is a Borel set. We have

$$P\{X+Y \in \mathcal{S}_X + \mathcal{S}_Y\} \geq P\{X \in \mathcal{S}_X\}P\{Y \in \mathcal{S}_Y\} = 1.$$

Hence $\overline{\mathcal{S}_X + \mathcal{S}_Y}$ is a closed set with P_{X+Y}-measure 1. Therefore it contains $\mathcal{S}_{X+Y}$.

Proposition 1 *Let X and Y be two independent random variables with finite support Lebesgue measure (SLM) of densities $\mathcal{L}(X) \stackrel{\text{def}}{=} \mu(\mathcal{S}_X)$ and $\mathcal{L}(Y)$, where $\mu(.)$ denote the Lebesgue measure. Then the SLM is (super- and sub-) additive*

$$\mathcal{L}(X+Y) = \mathcal{L}(X) + \mathcal{L}(Y) \tag{7}$$

This result follows directly from the above lemma 2.

Theorem 3 *For all increasing and concave real function g, the following support Lebesgue measure criterion (SLM)*

$$\begin{aligned} \mathcal{S}(\mathbf{y};\mathbf{B}) &= -\sum_{i=1}^{n} g[\mathcal{L}(y_i)] + g[|\det(\mathbf{B})|] \\ &= -\sum_{i=1}^{n} g\left(\mu\{y_i; f(y_i) > 0\}\right) + g(|\det(\mathbf{B})|). \end{aligned}$$

is a generalized contrast for blind separation of overdetermined linear mixtures with unknown source number.

***Proof* 2** *From the above proposition 1, we have equivalently shown that the density support Lebesgue measure defined in (6) is super-additive, then from lemma 1 it is also 2-super-additive. We can see also that the density support measure is equivariant, then from theorem (2) the SLM is a generalized contrast function.*

For maximizing a generalized contrast function $\mathcal{C}(\mathbf{y};\mathbf{B})$ with respect to $\mathbf{B}$, gradient based methods may be applied if it is always possible to compute the required data statistics. This strongly depends on the functional form and a complicate functional form may lead to unimplementable gradient BSS algorithm. Indeed, to apply such algorithm, we must know the differential of the used criterion (SLM in this section), that is, its variation resulting from a small deviation in its argument. Unfortunately, it is not easy to use this approach because the differential of the SLM cost function is a more difficult problem in our knowledge. For convenience we will use algebraic Jacobi like algorithms (e.g. see [11]). Recall that this method consist of estimating $\mathbf{B}$ as a product of Givens rotations according to

$$\mathbf{B} = \prod_{\#sweeps} \prod_{1 \leq p < q \leq n} \mathbf{\Omega}_{pq}(\theta)$$

Where $\mathbf{\Omega}_{pq}(\theta)$ is the elementary Givens rotation defined as orthogonal matrix where all diagonal elements are 1 except for the two elements $c = \cos(\theta)$ in rows (and columns) p and q. Likewise, all off-diagonal elements of $\mathbf{\Omega}_{pq}(\theta)$ are 0 except for the two elements $s = \sin(\theta)$ and $-s$ at positions (p,q) and (q,p), respectively. The maximization of $\mathcal{S}(\mathbf{y};\mathbf{B}) = \mathcal{S}(\mathbf{y};\mathbf{\Omega}_{pq}(\theta)) = \mathcal{S}(\mathbf{y};\theta)$ is done numerically by searching θ using a fine grid into $[0, 2\pi]$.
A well-known problem of multimodal source separation is the existence of spurious maxima for usual BSS contrasts. This is e.g. the case when using the opposite of the output marginal entropy [13]. In order to avoid the existence of spurious maxima, it is interesting to derive contrasts that are convex on each quadrant $\mathcal{Q}_p$. In the following, the convexity of the generalized contrast function $\mathcal{S}(\mathbf{y};\theta)$ for bounded sources is proven in all quadrants.

Proposition 2 *For $g(t) = \ln(t)$, the generalized contrast function $\mathcal{S}(\mathbf{y};\theta)$ is convex for θ in each quadrant $\mathcal{Q}_p$; $p \in \{1,2,3,4\}$.*

***Proof* 3** *We remark that the criterion $\mathcal{S}(\mathbf{y};\theta)$ is a result of many operations which preserve convexity, like pointwise addition of functions and the composition with another convex function. We should note that $-\ln(.)$, $\ln|\det(.)|$ and $\mathcal{L}(.)$ (see [13]) are convex functions in all quadrants.*

Then, the so called SLM separating algorithm is composed of the following steps:

◇ Optimization

- `Initialization`: $\mathbf{B}_0 = \mathbf{I}$;
- `Whitening`: find a $n \times n$ matrix $\mathbf{W}$ so that the sample covariance matrix of the whitened mixtures $\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t)$ equals the identity matrix [11]. This can be done by PCA methods [1];
- `Iteration`: compute an estimate $\hat{\mathbf{B}}$ ($n \times n$ matrix) that maximize the SLM cost function using the Jacobi rotations parametrization;

◇ Postprocessing

- `Source number estimation`: apply the result presented in [14] to estimate the source number m as the rank of a data matrix $\mathbf{X} = [\mathbf{x}(t+1), \cdots, \mathbf{x}(t+K)]^T$ where K is a large samples number, for example, $K = n$ or $2n$;
- `Source extraction`: since the desired signals appear at the first m channels, to extract the original sources we use the global system matrix $\mathbf{C}_1 = \hat{\mathbf{B}}_1\hat{\mathbf{A}}$, where $\hat{\mathbf{B}}_1$ is the submatrix composed of the first m rows of $\hat{\mathbf{B}}$ and $\hat{\mathbf{A}}$ is the estimate of the mixing matrix $\mathbf{A}$ computed as a pseudo inverse of $\hat{\mathbf{B}}$.

In our knowledge there is no Mathematical results concerning the convergence of such Jacobi algorithm. However, according to the experiences of several authors [2] and the author of the present paper [11], convergence is achieved in practice. Using the order statistics theory, Pham showed in [8], that the opposite of the sum of the log-marginal support widths of the output distributions is a global contrast for source separation. In [4], the authors introduce the minimum support criterion as a zero-order limiting case of the Renyi's entropy contrast. Recently, the authors of [13] derive a convex contrast, based on the support width of a single output distribution to separate bounded sources. Complementarily to these existing utilizations of the support measure in the BSS problem, we generalize these kind of criteria using some increasing concave image of the support Lebesgue measure. The aims of this generalization is to enforce the cost function to be more regular and suitable for nice optimization to handle the BSS problem of overdetermined systems with unknown source number. In addition this freedom of the g concave function choice can be exploited to handle some other questions in the BSS problem, like as the robustness against a non-gaussian noise and the multimodality of the sources distributions. The proposed contrast allows the separation even when the sources are non identically distributed.

3.3. Mutual Information Contrast

Let $\mathcal{G}(Y) = e^{H(Y)}$ define the square root of the entropy power (where h denote the differential entropy [3] and let $g(t) = \ln(t)$ the concave logarithm function. Then $\mathcal{G}$ is scale equivariant and it is shown in [3] that this information measure is 2-super-additive and more over the inequality (4) can be an equality (for $\sigma = 2$) only if both X and Y are Gaussian. Then we have the following result.

Theorem 4 *The following mutual information based criterion*

$$\begin{aligned}\mathcal{I}(y_1, \cdots, y_n; \mathbf{B}) &= -\sum_{i=1}^{n} \ln[\mathcal{G}(y_i)] + \ln|\det(\mathbf{B})| \\ &= -\sum_{i=1}^{n} H(y_i) + \ln|\det(\mathbf{B})| \quad (8)\end{aligned}$$

is a generalized contrast for blind separation of overdetermined linear mixtures with unknown source number.

Thus using theorem 2, we get a proof that the mutual information contrast can be extended to be a generalized contrast for the blind separation of the overdetermined mixtures.

To maximize a generalized contrast, whitening based approaches has some disadvantages, too. The most notable of these is that for ill-conditioned mixing matrices and weak sources the separation results may be poor. Therefore, some other algorithms have been developed that learn the separating matrix $\mathbf{B}$ directly [1]. Of particular note is the natural gradient algorithm which is based on the above MI criterion (8) and generalized in [14] as

$$\mathbf{B}(t+1) = \mathbf{B}(t) + \eta(t)\{\mathbf{R} - \varphi[\mathbf{y}(t)]\mathbf{y}(t)^T\}\mathbf{B}(t) \quad (9)$$

in which

$$\mathbf{R} = \mathbf{E}\{\varphi[\mathbf{y}(t)]\mathbf{y}(t)^T\}|_{\mathbf{y}(t)=\mathbf{Pz}(t)} \quad (10)$$

and $\eta(t)$ is a positive learning rate, where $\mathbf{P} \in \mathcal{P}_n$, $\mathbf{z} \in \mathscr{S}_n$ and $\varphi[.]$ is an appropriate class of nonlinear functions that replace the unknown source score functions. Finally, it should be noticed that the separating matrix must be of the form $n \times n$ to perform overdetermined BSS whether the source number m is known or not. We note also that the constraint $\mathbf{y}(t) = \mathbf{Pz}(t)$ used in (10) ensuring that the stationary condition be hold for the desired output $\mathbf{y}(t) = \mathbf{Pz}(t)$ and then the gradient algorithm (9) convergence .

4. SIMULATION RESULTS

We report some of our simulations results to illustrate the performance of our procedures.

First experiment. *Separation of a noisy overdetermined mixture with unknown source number using the SLM contrast.*
For a comparison purpose, we consider here the same second example presented in [4] with five sources in the overdetermined case ($n = 8$). In this experiment we mixed 150 sample of five ($m = 5$) binary sources through a random mixing matrix in presence of additive white Gaussian noise with a maximum SNR of 10 dB. By using the generalized Jacobi algorithm as presented above, we minimized the support Lebesgue measure of the output. Thus, the algorithm converged to a global system matrix

$$\mathbf{C} = \begin{pmatrix} -0.99 & 0.01 & 0.00 & 0.03 & 0.01 & 0.01 & 0.99 & 0.00 \\ 0.06 & 0.97 & -0.01 & 0.01 & 0.89 & 0.00 & -0.06 & 0.00 \\ 0.02 & -0.07 & 0.00 & 0.00 & -0.05 & 0.02 & -0.02 & 0.00 \\ 0.02 & -0.07 & 0.09 & 0.86 & -0.01 & 0.00 & -0.02 & 0.01 \\ 0.02 & -0.07 & 0.98 & 0.01 & -0.00 & 0.00 & -0.01 & 0.00 \end{pmatrix}^T$$

and all the five sources are extracted. One can observe that as expected, the first five components corresponds to the five original sources, whereas the 6-th and the 8-th corresponds to zero vectors and the 7-th component consist of a copy of the first source.

Second experiment. *Separation of a noisy mixture of Gaussian and non-Gaussian sources with known source number using* SOBI, JADE, SLM *and the* GENERALIZED NATURAL GRADIENT *algorithm (9).*

In this example, we consider four zero-mean unit-variance signals as sources. The first three sources are i.i.d: s_1 has a Gaussian distribution, s_2 a Uniform distribution and s_3 an Alpha-Stable distribution (with $\alpha = 1.5$ and dispersion $\gamma = 1$). The fourth signal s_4 is Gaussian filtered with an AR1 low-pass filter: $s_4 = 0.7s_4(t-1) + \epsilon(t)$, where $\epsilon(t)$ is a Gaussian signal. All signals are 5000 points long. The mixing matrix is randomly generated at each iteration.

To measure the quality of separation we will use Amari's error criterion as a performance index (PI) defined as

$$PI = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{m}\frac{|C_{i,j}|}{\max_k |C_{i,k}|}-1\right)+\frac{1}{m}\sum_{j=1}^{m}\left(\sum_{i=1}^{n}\frac{|C_{i,j}|}{\max_k |C_{k,j}|}-1\right)$$

where $\mathbf{C} = (C_{i,j})_{1\le i\le n\ ;\ 1\le j\le m} = \mathbf{BA}$ is the global BSS system. After convergence we can compare the performance index PI of each method as shown in Figure 1. This

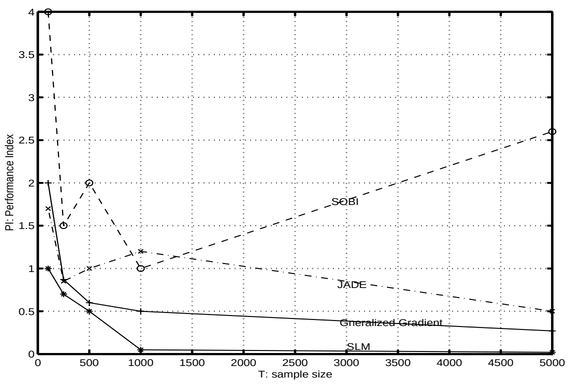

Fig. 1. The performance index versus the sample size.

Figure show that SOBI and JADE fail to separate this mixture, whereas the generalized gradient continues to produce good BSS solutions and that the SLM algorithm present the best performance. Note that the higher-order methods are able to separate mixture if at most one source is Gaussian, then JADE cannot separate s_1 and s_4. Similarly, second-order methods can only perform separation if the sources have different autocorrelation functions. Then, SOBI cannot separate s_1, s_2 and s_3 and the only source that can be recovered is s_4.

5. CONCLUSION

In this paper, we introduced a generalized definition of contrast function, which includes the existing classical and nonsymmetrical contrast as two special cases, to handle the problem of source separation of overdetermined linear mixtures with unknown source number. With the adjustment of the separating matrix form to be of dimension $n \times n$, we justified the fact that a natural gradient or Jacobi like algorithms can be extended to maximizing the considered generalized contrast function. These contrasts are shown to have no spurious local maxima, i.e. all the local maxima are relevant from the source separation point of view; they all correspond to non-mixing BSS solutions. However some question still open for this generalized case. Indeed, the effect of sources multimodality will be theoretically investigated in our futur works. Also the local stability of the generalized natural gradient algorithm will be analyzed and similarly convergence of the generalized Jacobi algorithm must be discussed.

6. REFERENCES

[1] A. Cichocki, J. Karhunen, W. Kasprzak and R. Vigario, ”Neural networks for blind separation with unknown number of sources”, *Neurocomputing*, Vol. 24, pp. 55-93, 1999.

[2] P. Comon, ”Independent Component Analysis, a new Concept ?”, *Signal Proc.*, Vol. 36, pp. 287-314, 1994.

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, New York, Wiley, 1991.

[4] S. Cruces and I. Duran, ”The minimum support criterion for blind signal extraction”, LNCS-3195, Springer-Verlag, pp. 57-64, ICA'2004, Granada, Spain, 2004.

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

[6] P.J. Huber, ”Projection Pursuit”, *The Annals of Statistics*, vol. 13, No. 2, pp. 435-475, 1985.

[7] E. Oja, J. Karhunen, A. Hyvarinen, R. Vigario and J. Hurri, ”Neural independent component analysis - approaches and applications”, in: S. Amari, N. Kasabov (Eds), Brain-Like Computing and Intelligent Information Systems, Springer, Singapor, pp. 167-188, 1997.

[8] D.-T. Pham, ”Blind separation of instantenaous mixtrures of sources based on order statistics”, *IEEE Trans. on Signal Processing*, vol. 48, no. 2, pp. 363.375, 2000.

[9] E. Moreau and N. Thirion-Moreau, ”Nonsymmetrical Contrasts for Sources Separation”, *IEEE Trans. on Signal Proc.*, vol. 47, No. 8, Aug. 1999.

[10] M. Sahmoudi, K. Abed-Meraim and M. Benidir ”Blind Separation of Heavy-Tailed Signals Using Normalized Statistics”, *in Proc. of ICA'2004*, Granada, Spain, Sept. 2004.

[11] M. Sahmoudi, K. Abed-Meraim and M. Benidir, “Blind Separation of Impulsive Alpha-Stable Sources Using Minimum Dispersion Criterion”, *in IEEE Signal Processing Letters*, Vol. 12, No. 4, April 2005.

[12] M. Sahmoudi and K. Abed-Meraim, ”Investigations on Contrast Functions for Blind Source Separation Based on Non-Gaussianity and Sparsity Measures”, *accepted for publication in the Proc. of the IEEE International Conference ISSPA'2005*, Sydney, Australia, 31th Aug. - 3rd Sept. 2005.

[13] F. Vrins, M. Verleysen and Ch. Jutten, ”SWM : A Class of Convex Contrasts for Source Separation”, *in Proc of the IEEE Int. Conference ICASSP'2005*, USA, May 2005.

[14] J. M. Ye, X. L. Zhu and X. D. Zhang, ”Adaptive blind separation with unknown number of sources”, *Neural Computation*, Vol. 16, No. 8, pp. 1641-1660, 2004.