**RESEARCH**       **Open Access**

# On the conditions for valid objective functions in blind separation of independent and dependent sources

Cesar F Caiafa[1,2]

## Abstract

It is well known that independent sources can be blindly detected and separated, one by one, from linear mixtures by identifying local extrema of certain *objective functions* (*contrasts*), like *negentropy*, *non-Gaussianity* (NG) measures, *kurtosis*, etc. It was also suggested by Donoho in 1981, and verified in practice by Caiafa et al., that some of these measures remain useful for particular cases with dependent sources, but not much work has been done in this respect and a rigorous theoretical ground still lacks. In this article, it is shown that, if a specific type of pairwise dependence among sources exists, called *linear conditional expectation (LCE)* law, then a family of objective functions are valid for their separation. Interestingly, this particular type of dependence arises in modeling material abundances in the spectral unmixing problem of remote sensed images. In this study, a theoretical novel approach is used to analyze *Shannon entropy* (SE), NG measure and absolute smoments of arbitrarily order $\beta$, i.e. *generic absolute moments* for the separation of sources allowing them to be dependent. We provide theoretical results that show the conditions under which sources are isolated by searching for a maximum or a minimum. Also, simple and efficient algorithms based on Parzen windows estimations of probability density functions and Newton–Raphson iterations are proposed for the separation of dependent or independent sources. A set of simulation results on synthetic data and an application to the blind spectral unmixing problem are provided in order to validate our theoretical results and compare these algorithms against FastICA and a very recently proposed algorithm for dependent sources, the bounded component analysis algorithm. It is shown that, for dependent sources verifying the LCE law, the *NG measure* provides the best separation results.

**Keywords:** Dependent component analysis (DCA), Independent component analysis (ICA), Blind source separation (BSS), Generic absolute (GA) moments, Entropy measures, Non-Gaussianity (NG)

## Introduction

In signal processing, a generic problem is how to separate signals that are linearly combined in the measurements. Blind source separation (BSS) consists on the task of isolating $n$ sources from a set of $m$ linear mixtures:

$$\mathbf{x}_t = \mathbf{A}\mathbf{s}_t, \tag{1}$$

where $\mathbf{x}_t \in \mathbb{R}^m$ is a column vector containing the *mixtures* (observed measurements), $\mathbf{s}_t \in \mathbb{R}^n$ is a column vector containing the unknown source signals (*sources*)

and $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the unknown *mixing matrix* containing the mixing coefficients. The parameter $t$ is an index that can be related to the position in time or space (pixel index) depending on the application. The model of Equation (1) is commonly referred as the *noiseless instantaneous mixing model*.

When sources are statistically independent and $m \geq n$ (overdetermined case), the problem is well posed in the sense that sources can be identified up to some unimportant indeterminacies [1]. This result allowed the development of a sort of independent component analysis (ICA) algorithms which were successfully and widely used in engineering problems [2-4]. Many criteria have been proposed in the context of ICA, for example, it is known that sources can be detected by identifying the local minima of

Correspondence: ccaiafa@iar.unlp.edu.ar
[1] Instituto Argentino de Radioastronomía (CCT La Plata, CONICET), C.C.5, 1894 Villa Elisa, Buenos Aires, Argentina
[2] Facultad de Ingeniería, Universidad de Buenos Aires, Paseo Colón 850, C1063ACV, Buenos Aires, Argentina

the SE, in the space of the mixing parameters, keeping the variance constant because of a classic result from information theory: the entropy of a sum of independent variables is larger than the entropy of individual variables [1,3,5-7]. More generally, Comon [1] has introduced the definition of *contrasts* for ICA, i.e. objective functions such that their global maxima corresponds to the separation of all sources. Besides *negentropy* (negative SE), other *contrasts* have been proposed for ICA as it is the case of higher order cumulants for which fourth order cumulant is a particular case [3,8-10], the convex perimeter for bounded sources [11], $L^2$-distance *non-Gaussianity* (NG) measure [12], least absolute end-point (LAE) [13], and others. For an up to date review of existing algorithms for ICA, the reader may refer to [2].

Unlike in the ICA case, the separation of dependent sources or dependent component analysis (DCA), has not been fully studied in the past and showed more difficulties. Hyvärinen & Hoyer [14] have proposed independent subspace analysis (ISA) as an extension of ICA where components in different subspaces are assumed independent whereas components in the same subspace have dependencies. When the sources and the mixing matrix are restricted to be non-negative, the problem can be seen as a non-negative matrix factorization (NMF) problem for which many algorithms have been developed [15]. However, NMF suffers from non-uniqueness of the solutions and the separation is not granted if not additional constraints are assumed, for example, by imposing sparsity of sources [16,17]. Bedini et al. [18] have developed algorithms for the separation of correlated sources found in astrophysical applications based on multiple-lag data covariance matrices, i.e. by exploiting the time structure of sources. In [12,19], an algorithm called MaxNG based on the maximization of a NG measure was proposed and tested on dependent sources extracted from remote sensed images. In [20,21], some DCA methods were tested on astrophysical sources. Cruces [11] proposed bounded component analysis (BCA) as an alternative method for BSS which relies on the bounded support of sources. In BCA, the separation is granted when the convex hull of the sources domain can be written as the cartesian product of the convex hulls of the individual source supports which is a very restrictive assumption. Recently, Eldogar [22] showed that a particular type of dependent sources generated according with a copula-*t* distribution are perfectly separated by BCA for a wide range of the correlation coefficient.

In all the previously mentioned DCA methods, when the independence assumption is relaxed, the success of the separation relies on alternative, and usually very restrictive conditions on sources. Moreover, it was suggested in [23], and verified in practice in [12,19,24], that some

measures used in ICA, such as *negentropy*, NG and *kurtosis*, remain useful for particular cases with dependent sources, but not much work has been done in this respect and a rigorous theoretical ground still lacks.

In this work, we propose a unified theoretical framework to study the capability of any objective function to detect each of the sources as a local maximum (or minimum) in the space of coefficients. From our analysis, it turns out that many new objective functions can be proposed owing this property if a particular type of dependence is verified among sources. In particular, we analyze *generic absolute (GA) moments*, SE and NG measure as valid objective functions. We introduce simple and efficient algorithms for the separation of sources using Parzen windows estimations of pdfs and Newton–Raphson (N–R) iteration. We analyze the performance of these measures and compare them against FastICA and BCA algorithm.

This article is organized as follows: in Section 2, the theoretical aspects are introduced followed by detailed analysis of the independent sources case (Section 3) and the dependent sources case (Section 4); in Section 5, using this theory, a set of particular cases are rigorously analyzed and illustrated by simulations; in Section 6, algorithms for source separation with $n = 2$ sources are derived by using Parzen windows estimation of pdfs and N–R iterative method; in Section 7, several simulation experiments are presented showing the performance of the proposed algorithms and comparing them against FastICA and BCA; finally, in Section 8, the main conclusions of this work and a discussion about our results is included.

**Notation and assumptions**

We use capital letters to denote random variables, for example, $S_1, S_2, \ldots, S_n$ are the random variables associated to the sources which have a *joint probability density function* (pdf) denoted by $f_{S_1 S_2 \ldots S_n}(s_1, s_2, \ldots, s_n)$. Obviously, when sources are independent, the joint pdf factorizes, i.e.

$$f_{S_1 S_2 \ldots S_n}(s_1, s_2, \ldots, s_n) = f_{S_1}(s_1) f_{S_2}(s_2) \ldots f_{S_n}(s_n), \quad (2)$$

where $f_{S_i}(s_i)$ is called the *marginal pdf* of variable $S_i$. In this work, we are also interested in the case of having dependent sources where such a factorization of the joint pdf does not exist.

We also define the *conditional pdf* of a random variable $S_1$ given that $S_2 = x$ as follows: $f_{S_1|S_2}(s_1, x) = f_{S_1,S_2}(S_1, x)/f_{S_2}(x)$. Accordingly, we define the *first and second order conditional expectations* as follows: $E[S_1|S_2 = x] = \int s_1 f_{S_1|S_2}(s_1, x) ds_1$ and $E[S_1^2|S_2 = x] = \int s_1^2 f_{S_1|S_2}(s_1, x) ds_1$. In the case of having only two sources, we can simplify the notation by using $E[S_1|x] \equiv E[S_1|S_2 = x]$ and $E[S_1^2|x] \equiv E[S_1^2|S_2 = x]$. Since conditioned expectations are functions of $x$, we use the

following simplified notation: $E'[S_1|x] \equiv \frac{d}{dx}E[S_1|x]$ and $E'[S_1^2|x] \equiv \frac{d}{dx}E[S_1^2|x]$.

In several parts of this article, when we apply the differentiation operator under the integral sign, i.e. $\frac{d}{d\tau}\int g(x,\tau)dx = \int \frac{d}{d\tau}g(x,\tau)dx$ we will assume that the function $g(x,\tau)$ is sufficiently nicely behaved in order to allow this operation. Basically, we assume that $g(x,\tau)$ and $\frac{d}{d\tau}g(x,\tau)$ are continuous for $x$ in the range of integration and there are upper bounds $|g(x,\tau)| \leq A(x)$ and $|\frac{d}{d\tau}g(x,\tau)| \leq B(x)$ independent of $\tau$ such that the integrals $\int A(x)dx$, $\int B(x)dx$ do exist.

### A motivation for this work: the blind spectral unmixing problem

Blind spectral unmixing is a specific application of BSS to the analysis of hyper-spectral remote sensed images. In this case, it is known that, at any fixed pixel, the linear mixing model of Equation (1) is valid. The vector of mixtures **x** represents the sensor measurements at different wavelengths, matrix **A** contains in its columns the spectral signatures of the endmembers (materials) existing in the covered area and vector **s** contains the endmember fractional abundances at the given pixel [25]. When the spectral signatures are unknown this is a blind problem and the objective is to estimate both, matrix **A** and the endmember fractional abundances **s**.

Most of the existing spectral unmixing algorithms exploit geometrical concepts by using the fact that, due to the linear mixing equation, the mixed pixels have to lie inside the convex hull of the endmembers. This convex hull forms a simplex in the spectral space, with the endmembers as spanning vertices. Some algorithms then try to look for the largest volume embedded simplex (e.g. simplex growing algorithm (SGA) [26], simplex-projection unmixing (SPU) [27]) or try to identify extreme points in the data cloud (e.g. vertex component analysis (VCA) [28]). Other recently proposed methods are based on the NMF with sparsity assumptions (e.g. S-measure constrained NMF algorithm (NMF-SMC) [29]).

As a BSS problem, sources in the blind spectral unmixing problem are clearly not independent since they are related to fractional abundances, in fact, they are constrained to sum up to one, i.e. $\sum_{i=1}^{n} S_i = 1$. Additionally, in order to avoid scale indeterminacy and remove constant values, we have to work with normalized sources $S_i$, i.e. $S_i \leftarrow (S_i - \mu_i)/\sigma_i^2$ where $\mu_i$ and $\sigma_i$ are the mean and standard deviation associated to source $i$. Under these conditions, the blind spectral unmixing task can be approached by solving a BSS problem with the additional constraint [19]:

$$\sum_{i=1}^{n} S_i = 0, \tag{3}$$

with $E[S_i] = 0, E[S_i^2] = 1 (i = 1, 2, \ldots n)$.

The dependence between any pair of normalized sources can be characterized, for example, by the conditional expectations $E[S_i|S_j]$ and $E[S_j|S_i]$ $(i \neq j)$. Clearly, when sources are independent, we have $E[S_i|S_j] = E[S_i] = 0$. On the other hand, when sources are dependent the conditional expectation $E[S_i|S_j]$ $(i \neq j)$ is a function of $S_j$. We can try, to determine these functions in order to satisfy Equation (3). By applying the conditional expectation to this equation with respect to $S_i$ $(i = 1, 2, \ldots, n)$ we obtain a system of $n$ equations with $n^2 - n$ unknown (the conditional expectations $E[S_i|S_j]$ $\forall i \neq j$). Thus, for $n \geq 2$ there is not a unique choice of the conditional expectations. Then, we have to obtain this information from the observation of real data. In [19], it was shown that hyper-spectral data can be well modeled as having linear conditional expectations (LCEs), i.e. $E[S_i|S_j] = aS_j + b$ (see [19]). The following theorem provides us the values of the constants $a$ and $b$.

**Theorem 1.** *([19], **Theorem 3**): Given a pair of dependent normalized sources $S_i$, $S_j$, if the conditional expectation $E[S_i|S_j]$ is linear in $S_j$, that is $E[S_i|S_j] = aS_j + b$, then $a = E[S_iS_j]$ and $b = 0$.*

In other words, in the blind spectral unmixing problem, the normalized sources can be modeled as verifying the following condition:

$$E[S_i|S_j] = \rho_{ij}S_j, \text{ with } \rho_{ij} = E[S_iS_j], \tag{4}$$

which is called the LCE law.

As we demonstrate in this article, when this particular type of dependence between sources is valid, their separation from linear mixtures can be obtained by maximizing (or minimizing) different types of objective functions.

### Relationship with previous works and new contributions

In [12], we have proposed the Parzen based NG measure and developed the MaxNG algorithm, a DCA method which showed to be useful to separate dependent sources extracted from images and also astrophysical dependent sources [20], outperforming classical ICA algorithms. However, in those articles the *NG measure* was not rigorously justified. Later, in [19], the *NG measure* was proposed as a solution for the blind spectral unmixing problem and a partial theoretical justification was given in terms of the LCE condition (guarantee of local extremum). In [24], we reported the results obtained by extending our method to other measures of NG such as Negentropy, and moment based measures, and applying them to synthetic and real datasets in the blind spectral unmixing problem context. However, a complete theoretical foundation was still lacking and several questions arose from those experimental results, for example: (1) why kurtosis based measure failed to separate one specific

type of dependent sources? or (2) given a particular measure, how to determine if it can be used for the separation of sources by local maximization or minimization?, and (3) on which conditions on sources the separation can be granted? In the present manuscript, we present a unified theoretical framework for the study of different measures for separation of independent and dependent sources verifying the LCE condition. The main contribution of this manuscript is to fill the gaps existing among previous works and give rigorous theoretical answers to the above open questions. In particular, it is shown that the kurtosis based measure (i.e. *GA moment* with $\beta = 4$) has zero second order derivative for the constrained dependent sources, which makes it useless for the separation as our empirical results showed in [24]. We also provide a precise condition (see Equation (33)) that establishes whether independent sources are separated by maximizing or minimizing the corresponding *GA moment* $\mu_\beta(\theta)$. On the other side, for dependent sources, it is necessary to know the second order conditional expectations, i.e. $E[S_i^2|S_j]$. Another contribution of the present article is a new algorithm based on a N–R search for local extrema which has quadratic convergence, i.e. being much faster than the algorithms proposed in [12,19,20] which used a steepest ascend method with fixed update step. Additionally, in the present manuscript we compare our algorithms against a recently proposed DCA algorithm, namely, the BCA [22], one state-of-the-art algorithm for DCA.

## Detection of sources by maximizing or minimizing objective functions

In this article, we focus only on sequential methods, also known as *deflation* methods, that extract normalized sources one after another by searching for local extrema of a predefined objective function. This simple idea was already used in the ICA context [2,3] and can be introduced as follows. When the matrix **A** in Equation (1) is full-column rank, the sources can be expressed as a linear combination of the mixtures by premultiplying Equation (1) with a pseudo inverse matrix such as the case of the Moore–Penrose pseudoinverse $\mathbf{A}^\dagger$, i.e.

$$\mathbf{s}_\tau = \mathbf{A}^\dagger \mathbf{x}_\tau. \tag{5}$$

Since matrix **A** is unknown, a reasonable strategy could be to search in the space of coefficients, for those points which correspond to each one of the sources. In other words, if we denote by $S_i$ the random variable associated to the source $s_{it}$ then we need to analyze the behavior of the *mixture random variable X* defined as:

$$X = \alpha_1 S_1 + \alpha_2 S_2 + \cdots + \alpha_n S_n. \tag{6}$$

We say that variable $S_i$ is separated from the mixture when all coefficients are zero except $\alpha_i$, i.e. $\alpha_i = 1$ and $\alpha_j = 0$ for every $j \neq i$. Here we arrive at the main

question we want to answer in this article: how can we discriminate between a single source compared to any linear combination of two or more sources?

We introduce some important objective functions that allow us to answer this question. Any valid candidate for an objective function should involve the pdf of the mixture $f_X(x)$ which depends on the mixture coefficients $\alpha_i$. In particular we consider the SE:

$$g_{SE} = -\int f_X(x) \log(f_X(x)) dx, \tag{7}$$

the NG measure defined as follows [12,19]:

$$g_{NG} = \int \left[ f_X(x) - \phi(x) \right]^2 dx, \tag{8}$$

where $\phi(x) = \frac{1}{\sqrt{(2\pi)}} \exp\left(-\frac{1}{2}x^2\right)$, and the *GA moment* of order $\beta$ which is defined as follows:

$$\mu_\beta = E[|X|^\beta] = \int |x|^\beta f_X(x) dx. \tag{9}$$

The SE is a well known measure already used in ICA, on the other side the *GA moment* was not used before for BSS except in the particular case when $\beta = 4$ which is closely related with the *kurtosis* [3,9]. In fact, *kurtosis* is defined as $\kappa = \mu_4/\mu_2^2$. In this sense, our analysis generalizes existing ICA methods providing insightful interpretations of the results. Note that, for $\beta = 2$ we obtain the second order moment of the variable which, in our case is fixed to $\mu_2 = 1$. But we can choose any order of moment $\beta$ provided that the integral in Equation (9) exists. In the following section we prove that these objective functions are valid for ICA.

## Independent sources case

Let us focus on Equation (6) and analyze wether it is possible to isolate a source $S_i$ from the rest by only looking at the statistical behavior of the mixture variable $X$, i.e. by studying how the pdf $f_X(x; \boldsymbol{\alpha})$ varies according to the mixing parameters in the vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$. For the ease of the presentation we start with the analysis of the two independent sources case. Since we have to generate all possible mixtures of two independent sources maintaining the variance constant, we use the following parameterization: $X = \alpha_1(\theta)S_1 + \alpha_2(\theta)S_2$ with $\alpha_1(\theta) = \cos(\theta)$ and $\alpha_2(\theta) = \sin(\theta)$ which means that the separation is obtained at $\theta = k\pi/2$ ($k = 0, \pm 1, \pm 2, \ldots$) (ignoring scaling ambiguity). We first introduce a new result characterizing the pdf of a mixture of independent random variables as a function of the mixing parameter $\theta$, i.e. $f_X(x; \theta)$.

**Lemma 1.** *(Two independent sources case): Given two zero-mean and unit-variance independent source variables $S_1$ and $S_2$, the pdf $f_X(x; \theta)$ of the mixture variable $X = \alpha_1(\theta)S_1 + \alpha_2(\theta)S_2$ with $\alpha_1(\theta) = \cos(\theta)$ and $\alpha(\theta) =$*

$\sin(\theta)$, *has zero-derivative with respect to $\theta$ for every $x$ at $\theta = k\pi/2$ ($k = 0, \pm 1, \pm 2, \ldots$), i.e.*

$$\left.\frac{\partial f_X(x;\theta)}{\partial \theta}\right|_{\theta=k\pi/2} = 0. \qquad (10)$$

*Proof.* Let us prove first the zero-derivative condition (maximum or minimum) at $\theta = \pi/2$ ($\boldsymbol{\alpha}_0 = (\alpha_1, \alpha_2) = (0,1)$) which corresponds to the separation of source $S_2$. In the neighborhood of $\boldsymbol{\alpha}_0$, i.e. $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0 + \boldsymbol{\delta}$, we can write the pdf of the mixture as the convolution:

$$f_X(x;\boldsymbol{\alpha}) = \frac{1}{\alpha_2} \int f_{S_1}(s_1) f_{S_2}\left(\frac{x - \alpha_1 s_1}{\alpha_2}\right) ds_1. \qquad (11)$$

By using the chain rule of derivatives we obtain

$$\frac{\partial f_X(x;\boldsymbol{\alpha})}{\partial \theta} = \frac{\partial f_X(x;\boldsymbol{\alpha})}{\partial \alpha_1}\alpha_1'(\theta) + \frac{\partial f_X(x;\boldsymbol{\alpha})}{\partial \alpha_2}\alpha_2'(\theta). \qquad (12)$$

We compute the partial derivatives in the last equation evaluated at $\alpha_1 = 0$ and $\alpha_2 = 1$ and, by inserting the derivative operator inside the integral, we obtain:

$$\left.\frac{\partial f_X(x;\boldsymbol{\alpha})}{\partial \alpha_1}\right|_{\alpha_1=0} = -f_{S_2}'(x)\int s_1 f_{S_1}(s_1)ds_1 = -f_{S_2}'(x)E[S_1] = 0, \qquad (13)$$

$$\left.\frac{\partial f_X(x;\boldsymbol{\alpha})}{\partial \alpha_2}\right|_{\alpha_2=1} = -f_{S_2}(x) - x f_{S_2}'(x) = -(x f_{S_2}(x))', \qquad (14)$$

Now, taking into account that $\alpha_1'(\pi/2) = 1$ and $\alpha_2'(\pi/2) = 0$, and using Equations (13) and (14) into Equation (12), we arrive at

$$\left.\frac{\partial f_X(x;\boldsymbol{\alpha})}{\partial \theta}\right|_{\theta=\pi/2} = 0 \times 1 - (x f_{S_2}(x))' \times 0 = 0, \qquad (15)$$

for every $x$. Using a similar procedure but working in a neighbouhood of $\boldsymbol{\alpha}_0 = (\alpha_1, \alpha_2) = (1,0)$, i.e. by considering $f_X(x;\boldsymbol{\alpha}) = \frac{1}{\alpha_1}\int f_{S_2}(s_2) f_{S_1}\left(\frac{x-\alpha_2 s_2}{\alpha_1}\right) ds_2$ instead of Equation (11), we can prove the zero-derivative condition at $\theta = 0$, which corresponds to the separation of source $S_1$. Finally, it is easy to see that the zero-derivative condition also holds for any integer multiple of $\pi/2$, because the resulting mixture becomes $\pm S_1$ or $\pm S_2$ and the same reasoning used before applies. Thus, the zero-derivative condition is valid for every $x$ at $\theta = k\pi/2$ ($k = 0, \pm 1, \pm 2, \ldots$) as claimed by this lemma. □

In Figure 1, a graphical interpretation of this lemma is shown in terms of the shape of the pdf for a mixture of two independent sub-Gaussian variables at $\theta_0 = \pi/2$ and $\theta = \theta_0 \pm \delta\theta$.

In the following, using this fundamental property of independent variables (Lemma 1), we can easily prove that $g_{SE}(\theta)$ and $\mu_\beta(\theta)$ have local extrema at the desirable separation points, i.e. $g_{SE}'(k\pi/2) = \mu_\beta'(k\pi/2) = 0$, for $k \in \mathbb{Z}$.

**Theorem 2.** *Local extrema of SE (independent sources case): Given two zero-mean and unit-norm source variables $S_i$ ($i = 1, 2$), the SE ($g_{SE}(\theta)$) of the mixture variable $X = \cos(\theta)S_1 + \sin(\theta)S_2$, has local extrema at $\theta = k\pi/2$ ($k \in \mathbb{Z}$).*

*Proof.* As in the proof of Lemma 1, here it suffices to prove that the derivative of the SE, with respect to the parameter $\theta$, vanishes at $\theta = \pi/2$. From the definition of SE in Equation (7), if we take its derivative with respect to $\theta$ we have:

$$g_{SE}'(\theta) = -\int \frac{df_X(x;\theta)}{d\theta}\left(\log(f_X(x;\theta)) + 1\right) dx. \qquad (16)$$

Now, using Lemma 1 we see that the derivative of the pdf is zero at $\theta = \pi/2$ for every $x$ $\left(\frac{df_X(x;\theta)}{d\theta} = 0 \forall x\right)$ and therefore we conclude that $g_{SE}'(\pi/2) = 0$. □

**Theorem 3.** *Local extrema of the GA moment measure (independent sources case):*
*Given two zero-mean and unit-norm source variables $S_i$ ($i = 1, 2$), the GA moment of order $\beta$, $\mu_\beta(\theta)$ of the mixture variable $X = \cos(\theta)S_1 + \sin(\theta)S_2$, has local extrema at $\theta = k\pi/2$ ($k \in \mathbb{Z}$).*

*Proof.* We need to prove that the derivative of Equation (9), with respect to the parameter $\theta$, is zero at $\theta = \pi/2$. If we take the derivative of this equation we obtain:

$$\mu_\beta'(\theta) = \int |x|^\beta \frac{df_X(x;\theta)}{d\theta} dx, \qquad (17)$$

where, by using the Lemma 1, we see that $\mu_\beta'(\pi/2) = 0$. □

It is interesting to note that, our Theorem 3 shows that local extrema are found at the desirable locations for any chosen value of parameter $\beta$.

In [19], it was shown that the NG measure defined in Equation (8) has also local extrema at the separation points. Moreover, it is clear that, by using the same line of reasoning we can prove the existence of local extrema
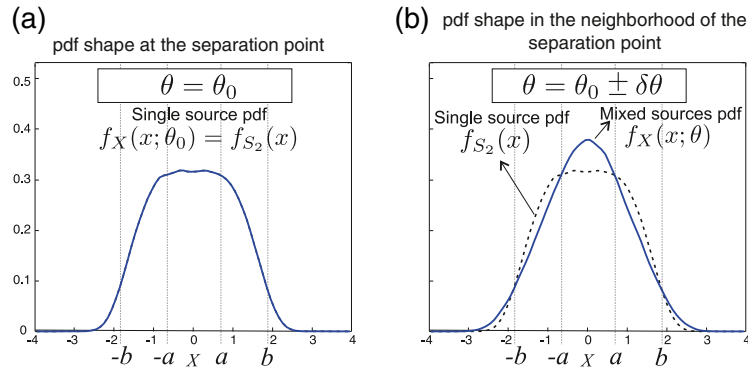
**Figure 1 pdf for the mixture variable $X = \alpha_1(\theta)S_1 + \alpha_2(\theta)S_2$ with sub-Gaussian sources at $\theta_0 = \pi/2$ (left) and at $\theta = \theta_0 \pm \delta\theta$ (right).** Note that, when $\theta = \pi/2$ the pdf corresponds to the source $S_2$, on the other side, when a perturbation on the mixing parameter $\theta$ is considered, i.e. $\theta = \theta_0 \pm \delta\theta$, a pdf with a shape closer to the Gaussian one is obtained. It is also noted that in the intervals $x \in \{(-\infty, -b) \cup (-a, a) \cup (b, +\infty)\}$ $(x \in \{(-b, -a) \cup (a, b)\})$ the pdf $f_X(x; \theta)$ attains its minimum (maximum) at $\theta = \pi/2$.

in many other objective functionals as, for example, for the case of Renyi entropy which was already proposed and studied for ICA [6,7].

It is important to note that, a local extremum at $(\alpha_1, \alpha_2) = (0, 1)$ is a necessary condition to separate source $S_2$ but it is not a sufficient condition. The existence of local extrema which do not correspond to a separation of sources, also known as mixing local extrema or spurious local extrema, was a topic of research in the ICA setting. Moreover, some theoretical results are available showing the existence of spurious local minima for the Entropy measure when sources has multimodal distributions [30,31]. Vrins and Verleysen [32] have shown that the kurtosis-based contrast functions are more robust than the information theoretic ones when the source distributions are multimodal.

### Relaxing independence: DCA

In the previous section, we have shown that Lemma 1 suffices to guarantee the validity of the SE, the *GA moment* and the NG measure as objective functions for ICA. We are interested now to look at the problem of separation of potentially dependent sources. Then, a natural question raises here: what kind of dependence should have the sources in order to guarantee the same behavior of the pdf as in the ICA scenario? The following result provides a necessary and sufficient condition.

**Lemma 2. *(n dependent sources case)*** *Given a set of zero-mean and unit-variance source variables $S_i$ ($i = 1, 2, \ldots, n$) the pdf $f_X(x; \boldsymbol{\alpha})$ of the mixture variable $X = \alpha_1 S_1 + \alpha_2 S_2 + \cdots + \alpha_n S_n$, constrained to the case of having unit-variance $E[X^2] = 1$, has local extrema for every x at the separation points ($\alpha_k = 1$ and $\alpha_i = 0$*

*for all $i \neq k$) iff the LCE law defined in Equation (4) holds.*

*Proof.* Here it is only necessary to prove the local extrema condition (maximum or minimum) for only one point so we arbitrarily choose the case $\alpha_n = 1$ and $\alpha_i = 0$ for all $i \neq n$. In this case, in the neighborhood of $\boldsymbol{\alpha}_0 = (0, 0, \ldots, 1)$, i.e. $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0 + \boldsymbol{\delta}$, we can write the pdf of the mixture as follows:

$$
\begin{aligned}
f_X(x; \boldsymbol{\alpha}) = \frac{1}{\alpha_n} \int \cdots \int & f_{S_1 \ldots S_n} \\
\times & \left( s_1, \ldots y, \frac{x - \alpha_1 s_1 - \cdots y - \alpha_{n-1} s_{n-1}}{\alpha_n} \right) \\
\times & \, ds_1 \ldots ds_{n-1}.
\end{aligned}
$$

(18)

Following the Lagrangean method, the condition for the existence of a local extrema point at $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ under the unit-variance constraint is as follows:

$$
\nabla L(\boldsymbol{\alpha}_0) = 0,
$$

(19)

with

$$
L(\boldsymbol{\alpha}) = f_X(x; \boldsymbol{\alpha}) + \lambda \left( \sum_{i=1}^{n} \alpha_i^2 + 2 \sum_{i<k} \rho_{ik} \alpha_i \alpha_k - 1 \right),
$$

(20)

where $\lambda$ is the Lagrange multiplier and $\rho_{ik} = E[S_i S_k]$ is the correlation coefficient between sources $i$ and $k$.

Now, we take the derivatives of the pdf in Equation (18) with respect to the coefficients $\alpha_i$ with $i = 1, 2, \ldots, n - 1$ which, evaluated at $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ give:

$$\frac{\partial f_X(x; \boldsymbol{\alpha})}{\partial \alpha_i}\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} = -\int \frac{\partial f_{S_i S_n}(s_i, x)}{\partial x} s_i ds_i$$

$$= -\left(E[S_i|S_n = x]f_{S_n(x)}\right)'. \quad (21)$$

Similarly, the derivative of the pdf in Equation (18) with respect to the coefficient $\alpha_n$ is:

$$\frac{\partial f_X(x; \boldsymbol{\alpha})}{\partial \alpha_n}\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} = -\left(x f_{S_n}(x)\right)'. \quad (22)$$

Using Equations (20), (21) and (22) in Equation (19) we obtain the following set of conditions:

$$\frac{\partial L}{\partial \alpha_i}\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} = -\left(E[S_i|S_n = x]f_{S_n(x)}\right)' + 2\lambda\rho_{in} = 0 \quad (23)$$

$$\text{for } i = 1, 2, \ldots, n - 1$$

$$\frac{\partial L}{\partial \alpha_n}\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} = -\left(x f_{S_n}(x)\right)' + 2\lambda = 0. \quad (24)$$

The last equation determines the Lagrange multiplier, i.e. $\lambda = (1/2)\left(x f_{S_n}(x)\right)'$ and, by inserting it into (23) we arrive to the desired condition:

$$E[S_i|S_n = s_n] = \rho_{in}s_n. \quad (25)$$

$\square$

It is important to note that the LCE condition is also valid for the particular case of independent sources, i.e. $E[S_1|S_2 = s_2] = E[S_1] = 0$ and $\rho = 0$. In Figure 2, some examples of sources are given indicating whether they follow or not the LCE law.

Before proceeding with our additional results, we have to solve a technical problem because, as our sources are now dependent they are allowed to be correlated, thus the parameterization $X = \cos(\theta)S_1 + \sin(\theta)S_2$ does not longer preserve the variance of the mixture variable $X$. Let us consider the general linear mixture $X = \alpha_1 S_1 + \alpha_2 S_2$, if we are constrained to the unit-variance case $E[X^2] = 1$, then $\alpha_1^2 + \alpha_2^2 + 2\rho\alpha_1\alpha_2 = 1$, where we used $\rho = E[S_1 S_2]$ to denote the correlation coefficient between sources. Then, the following parameterization preserves the variance and uses only one parameter $\tau$:

$$\alpha_1(\tau) = \tau \text{ and } \alpha_2(\tau) = -\tau\rho + \sqrt{\tau^2(\rho^2 - 1) + 1}. \quad (26)$$

The following results can be considered as generalizations of Theorems 2 and 3 to the case of two dependent sources.

**Theorem 4.** *Local Extrema of SE (general case): Given two zero-mean and unit-norm source variables $S_i$ ($i = 1, 2$) following the LCE law with respect to $S_2$, the SE of the mixture variable $X = \alpha_1 S_1 + \alpha_2 S_2$ constrained to the unit-variance case $E[X^2] = 1$, has a local extremum at $(\alpha_1, \alpha_2) = (0, 1)$.*

*Proof.* The proof can be obtained identically to the proof of Theorem 2, taking into account the parameterization (26) and using the fact that the LCE condition implies the existence of local extrema of the pdf as stated by Lemma 2.
$\square$

**Theorem 5.** *Local extrema of the GA moment (general case): Given two zero-mean and unit-norm source variables $S_i$ ($i = 1, 2$) following the LCE law with respect to $S_2$, the GA moment of order $\beta$ of the mixture variable $X = \alpha_1 S_1 + \alpha_2 S_2$, constrained to the unit-variance case $E[X^2] = 1$, has a local extremum at $(\alpha_1, \alpha_2) = (0, 1)$.*

*Proof.* The proof can be obtained identically to the proof of Theorem 3, taking into account the parameterization (26) and using the fact that the LCE condition implies the existed of the local extrema of the pdf as stated by Lemma 2
$\square$

### Detailed analysis of SE and *GA moments*

In previous sections, we proved that some objective functions applied to a unit-variance mixture of sources verifying the LCE law, have local extrema when only one of the coefficients is non-zero, which means that we can separate those sources by searching for local extrema. Nevertheless, a more detailed analysis is required in order to determine if each local extremum corresponds to a maximum or a minimum.

Here, we compute the second order derivative of the objective function with respect to $\tau$ for the special cases of the SE and *GA moments* of order $\beta$. As we will show, the condition of a maximum or minimum depends on the second order conditional expectation of sources and on their marginal pdfs. First we need to compute the second order derivative of the pdf with respect to the parameter $\tau$ which is as follows (its derivation is included in Appendix 1):

$$\frac{\partial^2 f_X(x; \alpha_1(\tau), \alpha_2(\tau))}{\partial \tau^2}\bigg|_{\tau=0} = \left(f_{S_2}(x)E[S_1^2|x]\right)'' + (1 - 3\rho^2)$$

$$\times f_{S_2}(x) + x(1 - 5\rho^2)f'_{S_2}(x)$$

$$- \rho^2 x^2 f''_{S_2}(x). \quad (27)$$

We note that the second order derivative explicitly depends on the second order conditional expectation $E[S_1^2|S_2 = x]$ and the marginal pdf $f_{S_2}(x)$.

Using this result, we are able to obtain the second order derivatives of the objective function as follows:

(1) SE measure: to obtain the second order derivative of SE we take the derivative of Equation (16) with respect to the parameter $\tau$ arriving at:
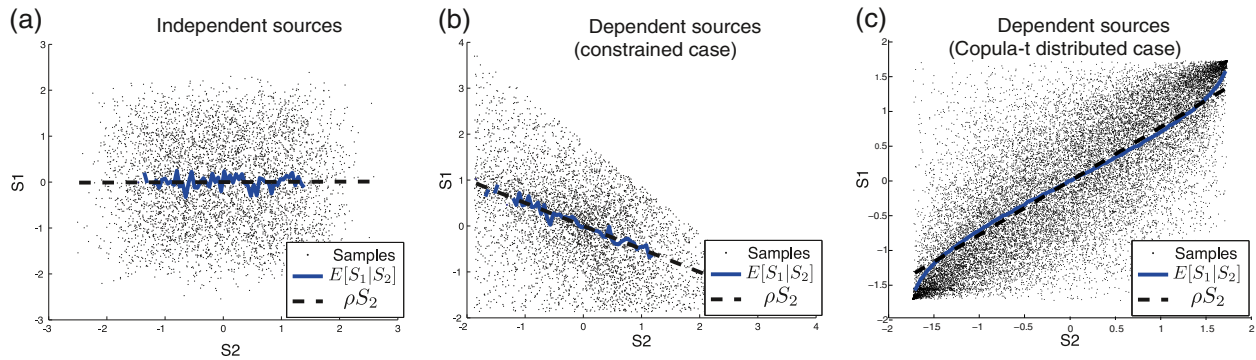
**Figure 2 Scatter plots for three examples of sources with estimations of the conditional expectation $E[S_1|S_2]$ and the linear regression given by $\rho S_2$: (a)** corresponds to a pair of independent sources (sub-Gaussian); **(b)** corresponds to a pair of dependent sources with negative correlation following the model of abundances in spectral unmixing problem (constrained case); and **(c)** shows a pair of dependent sources generated using a Copula-t distribution with correlation $\rho = 0.8$. In **(a)** and **(b)** the LCE condition holds while in **(c)** it is only approximately verified.

$$g''_{SE}(\tau) = -\int \frac{d^2 f_X(x;\tau)}{d\tau^2} \left( \log(f_{S_2}(x) + 1) \right) dx, \tag{28}$$

and, by using Equation (27) in the last equation and taking into account that the LCE law holds, i.e. $E[S_1|x] = \rho x$, we obtain (see its derivation in Appendix 2):

$$g''_{SE}(0) = \int \left( E\left[S_1^2|x\right] - \rho^2 x^2 \right) \frac{\left( f'_{S_2}(x) \right)^2}{f_{S_2}(x)} dx - \int E''\left[S_1^2|x\right] \\ \times f_{S_2}(x) dx + (3\rho^2 - 1). \tag{29}$$

(2) *GA moment*: To compute the second order derivative of the *GA moment* we need to take the derivative of Equation (17) with respect to the parameter $\tau$ reaching to:

$$\mu''_\beta(\tau) = \int |x|^\beta \frac{d^2 f_X(x;\tau)}{d\tau^2} dx. \tag{30}$$

Again, by using Equation (27) into the last equation and using the LCE law we obtain (see its derivation in Appendix 3):

$$\mu''_\beta(0) = \beta(\beta - 1) \int |x|^{\beta-2} f_{S_1}(x) E\left[S_1^2|x\right] dx - \beta\mu \\ \times \left(1 + \rho^2(\beta - 2)\right), \tag{31}$$

which is valid only when the integral $\int |x|^{\beta-2} f_{S_1}(x) E[S_1^2|x]\, dx$ exists.

## Some particular cases

In this section, we analyze selected examples to illustrate our theoretical results applied to different types of independent and dependent sources.

(1) *Independent sources*: Let us consider the simplest case of having two independent sources $S_1$ and $S_2$. We see that the LCE law (Equation (4)) holds since $\rho = 0$ and $E[S_1|S_2 = s_2] = E[S_1] = 0$ which means that SE, *GA moment* and *NG measure* have a local extrema at $\tau = 0$ using the parameterization of Equation (26). Additionally, we note that the second order conditional expectation is $E[S_1^2|S_2 = s_2] = E[S_1^2] = 1$ and then the second order derivative of SE using Equation (29) becomes:

$$g''_{SE}(0) = \int \frac{(f'_{S_2}(x))^2}{f_{S_2}(x)} dx - 1, \tag{32}$$

which is always greater than zero except for the Gaussian distribution for which is equal to zero $\left( \int \frac{(f'_{S_2}(x))^2}{f_{S_2}(x)} dx \text{ is the Fisher information} \right)$ (see for example [33], p. 23). This confirms the fact that, at the separation point, we have a local minimum of the SE. Now, using Equation (31) we evaluate the second order derivative of the *GA moment* which is

$$\mu''_\beta(0) = \beta \left[ (\beta - 1)\mu_{\beta-2} - \mu_\beta \right]. \tag{33}$$

Let us now analyze different cases corresponding to different values of $\beta$. For example, if we consider the fourth order moment case ($\beta = 4$), we obtain $\mu''4(0) = 4[3 - \mu_4]$ which means that, for sources with $\beta_4 > 3$ (super-Gaussian) the fourth order moment of the mixture has a minimum at $\tau = 0$. On the other hand, for sources with $\mu_4 < 3$ (sub-Gaussian), a maximum of the fourth order moment of the mixture is found. More interestingly, we can evaluate any arbitrarily order $\beta$ and Equation (33) will tell us if we need to search for a maximum or a minimum to attain the separation.

(2) *Uncorrelated but dependent sources*: We consider here two sources $S_1$ and $S_2$ generated as follows: $S_1 = N_1 N_2$ and $S_2 = N_2$, where $N_1$ and $N_2$ are independent non-Gaussian random variables with $E[N_1] = E[N_2] = 0$ and $E[N_1^2] = E[N_2^2] = 1$. We see that $S_1$ and $S_2$ are highly dependent but are uncorrelated because $\rho = E[S_1 S_2] = E[N_1 N_2^2] = E[N_1] E[N_2^2] = 0$. The first order conditional expectation is zero, i.e. $E[S_1|S_2 = s_2] = E[N_1] s_2 = 0$. We also compute the second order conditional expectation which is $E[S_1^2|S_2 = s_2] = E[N_1^2 N_2^2|N_2] = s_2^2 E[N_1^2] = s_2^2$. Then, by using Equation (29), the second order derivative of SE at $\tau = 0$ becomes:

$$g_{SE}''(0) = \int x^2 \frac{(f_{S_2}'(x))^2}{f_{S_2}(x)} dx - 3. \qquad (34)$$

It is interesting to note that SE could have a maximum at $\tau = 0$ if the integral in the last equation is smaller than three as in the case of our example in Figure 3d.

Regarding the *GA moment*, for these sources, Equation (31) becomes:

$$\mu_\beta''(0) = \beta(\beta - 2)\mu_\beta, \qquad (35)$$

and we conclude that we have a minimum at the separation point ($\mu''_\beta(0) > 0$) for every $\beta > 2$.

(3) *A simplified model for material abundances in spectral unmixing (dependent sources)*: A simple model to generate a special type of sources which are dependent, correlated and constrained to have their sum constant is as follows [19]. First, we generate $P > 2$ independent, nonnegative random variables $N_1, N_2, \ldots, N_P$; then, we define the following random variables: $U_i = N_i / \sum_{p=1}^{P} N_p$, for $i = 1, 2, \ldots, P$. We note that these signals meet the constraint $\sum_{i=1}^{P} U_i = 1$ as in the spectral unmixing application. Now, we define our sources by normalizing two of these constrained sources, i.e.: $S_i = (U_i - \bar{U}_i)/\sigma_{U_i}$, $i = 1, 2$. It is not hard to prove that these sources meet the LCE law since $E[S_1 S_2] = \rho = -1/(P - 1)$ and $E[S_1|S_2 = s_2] = \rho s_2 = -1/(P - 1)s_2$. Additionally, It is not difficult to prove that, for this particular type of sources we have constant *GA moment* of order $\beta = 4$ which makes it not suitable as an objective function for this case. This behavior was already observed in [24] but not theoretical explanation was available until now. In Section 7.3, we generate data and test ICA/DCA algorithms using a more realistic model for material abundances in hyperspectral images by computing directly the material percentages per pixel in a real image.

In order to illustrate these theoretical results, in Figure 3, plots for SE, the NG measure, and *GA moments* with several values of $\beta$, are shown for the following types of datasets using a sample size of $T = 10^6$ (except for SE and NG for which we used $T = 10^4$): (a) Independent sub-Gaussian sources, generated by applying the function $\sinh(u)^{-1}$ to zero-mean Gaussian independent signals; (b) Independent super-Gaussian sources, generated by applying the function $\sinh(u)$ to zero-mean Gaussian independent signals; (c) Independent bimodal sources, where each of the independent sources were generated by mixing two Gaussians with $(\mu_1, \sigma_1) = (0.5, 0.2)$ and $(\mu_2, \sigma_2) = (-0.5, 0.2)$, respectively; (d) Dependent uncorrelated sources, generated by using $s_{1t} = n_{1t} n_{2t}$ and $s_{2t} = n_{2t}$ where $n_{1t}, n_{2t}$ were generated as independent zero-mean uniform distributions; and (e) Dependent constrained sources, generated by using $s_i(t) = \frac{n_{it}}{\sum_{p=1}^{4} n_{p_t}}$ with $i = 1, 2$, where signals $n_{p_t}$ ($p = 1, 2, \ldots y, 4$) were generated using independent uniform distributions in $[0, 1]$.

We see that for the cases (a), (b), (c) and (e), the separation of each source is attained at the minima of the SE and the maxima of the *NG measure*. Interestingly, sources in case (d) (dependent and correlated) shows that one of the sources is detected at one maximum of the SE and one minimum of the NG measure. It is important to note that the SE have also spurious local minima for the case of bimodal distributions (case (c)). This behavior in information theoretic measures was already analyzed in [30-32] for the independent sources case. On the other hand, in our results, we see that the NG measure and *GA moments* are more robust having no spurious local extrema. We also note that, for Sub-Gaussian independent sources (a), the *GA moment* measure have local minima at source locations, on the other side, for super-Gaussian sources, they are located at local maxima. Nevertheless, it is important to note that for large order ($\beta = 4$ and $\beta = 7$) one local maxima is less evident because moments of a large order are affected by outliers (see scatter plot in Figure 3b). In the case (e), we observe *GA moments* provide a local maximum for $\beta = 3$ and local minima for $\beta = 7, 10$, and, for $\beta = 4$ the second order derivative is in theory zero and for that reason the local extrema are not clear.

## Parzen windows based algorithms for source separation

Parzen windows method is a non-parametric technique used to estimate a pdf based on a set of samples [34]. Using Parzen windows we can obtain the following estimators for SE [6] and the NG measure [12]:

$$\hat{g}_{SE}(\theta) = -\sum_{t_1=1}^{T} \log \left[ \frac{1}{Th} \sum_{t_2=1}^{T} \phi \left( \frac{y_{t_1}(\theta) - y_{t_2}(\theta)}{h} \right) \right], \qquad (36)$$
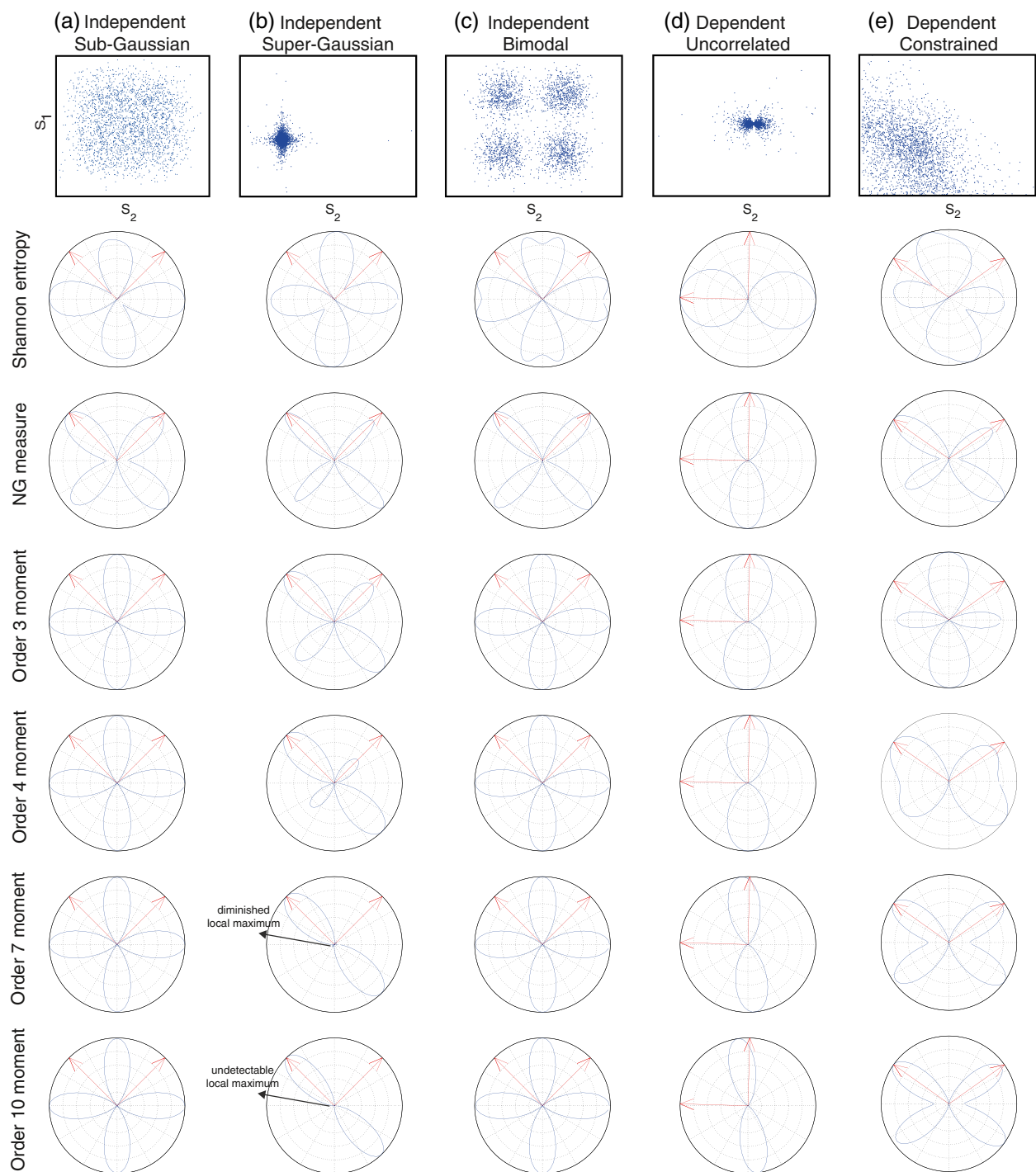
**Figure 3 Computation of SE, NG measure and GA moments for different types of independent and dependent sources $S_1$ and $S_2$.** After a de-correlation step (whitening) the measures are computed using the polar parameterization $y(\theta) = cos(\theta)X_1 + sin(\theta)X_2$ where $X_1$ and $X_2$ are the whitened variables. The corresponding scatter plots are shown in the 1st row. The position of theoretical positions (in polar coordinates) are shown as red arrows. The measures were normalized in order to cover the range [0, 1]. We used signals with a total number of samples $T = 10^6$ but we used only a subset of 10,000 samples to compute SE and *NG measure* to avoid the extremely high computational demand. For the generation of sub-Gaussian and super-Gaussian sources we used the transformation $sinh^{-1}(x)$ and $sinh(x)$ applied to a Gaussian variable $x$, respectively.

$$\hat{g}_{NG}(\theta) = -\frac{2}{T\sqrt{h^2+1}} \sum_{t_1=1}^{T} \phi\left(\frac{y_{t_1}(\theta)}{\sqrt{h^2+1}}\right) + \frac{1}{T^2 h\sqrt{2}} \sum_{t_1=1}^{T}$$

$$\sum_{t_2=1}^{T} \phi\left(\frac{y_{t_1}(\theta) - y_{t_2}(\theta)}{\sqrt{2}h}\right) + \frac{1}{2\sqrt{\pi}},$$

$$(37)$$

where $T$ is the number of samples,

$$y_t(\theta) = \cos(\theta)x_{1t} + \sin(\theta)x_{2t}, \qquad (38)$$

is the projected variable[a] sampled at time $t$ ($x_1$ and $x_2$ are assumed uncorrelated, i.e. whitened), $\phi(\cdot)$ is the kernel function (typically a Gaussian kernel) and $h$ is a parameter which determines the size of the windows (we adopt $h = 1.06 \times T^{-\frac{1}{5}}$ as determined by the minimum mean integrated square error (MISE) [34]). From Equations (36) and (37) we see that their computational complexity is quadratic in terms of the number of available samples ($\mathcal{O}(T^2)$).

On the other hand, for the estimation of *GA moments* we can use the ergodic average formula:

$$\hat{\mu}_\beta(\theta) = \frac{1}{T} \sum_{t=1}^{T} |y_t(\theta)|^\beta. \qquad (39)$$

Clearly, a big advantage of *GA moments* over the other measures is its lower computational cost since it is linear in the number of samples, i.e. $\mathcal{O}(T)$.

As usual, in order to simplify the search of the maximum (or minimum), we first apply a whitening filter, i.e. $\mathbf{x}_t \leftarrow \mathbf{T}\mathbf{x}_t$ after which we obtain $E[\mathbf{x}\mathbf{x}^T] = \mathbf{I}$. The filter matrix is given by $\mathbf{T} = \Lambda^{-\frac{1}{2}}\mathbf{U}^T$ with $\Lambda$ and $\mathbf{U}$ being the diagonal matrix of singular values and the matrix of singular vectors of the covariance matrix $\mathbf{C}_{\mathbf{xx}} = E[\mathbf{x}\mathbf{x}^T]$, respectively [3,12].

The search for a local extrema $\theta^*$ can be done by iteratively evaluating the objective function and/or its derivatives at a current estimate $\theta^{(k)}$ and by generating a sequence $\theta^{(1)}, \theta^{(2)}, \ldots y, \theta^{(k)}$ that converges to $\theta^*$. Note that the derivatives of the measures can be easily computed from Equations (36), (37) and (39). The simplest way to generate this sequence could be to use a steepest ascend/descend method, i.e. $\theta^{(k+1)} = \theta^{(k)} \pm \epsilon g'(\theta^{(k)})$. In this case the step size $\epsilon$ must be chosen in order to guarantee the convergence in few steps which is not a simple task. To avoid this problem, we consider here a simple and efficient algorithm based in the N–R iteration which, in the one dimensional case, is equivalent to the steepest ascend/descend method with an adaptive step size defined by $\epsilon_k = \frac{1}{|g''(\theta^{(k)})|}$, i.e.:

$$\theta^{(k+1)} = \theta^{(k)} \pm \frac{g'(\theta^{(k)})}{|g''(\theta^{(k)})|}, \qquad (40)$$

where $g(\theta)$ could be any of $\hat{g}_{SE}(\theta)$, $\hat{g}_{NG}(\theta)$ or $\hat{\mu}_\beta(\theta)$, and the sign "+" or "−" must be chosen for the case of a maximum or minimum, respectively; $g'(\theta)$ and $g''(\theta)$ are the first and second order derivatives, respectively, whose formulae can be derived from Equations (36), (37) or (39), providing similar computation complexity. A great advantage of the N–R algorithm is that it is proven to converge quickly in general (quadratic convergence). A potential drawback of the N–R method is that a close to zero second order derivative can make the method diverge. Anyway, our simulations showed always very fast convergence suggesting that the zero second order derivative condition is not likely to occur in general.[b]

### Algorithm 1: DCA algorithm (two-sources case)
**Require:** mixtures $\mathbf{x}_t$ ($t = 1, 2, \ldots, T$) (centered), tolerance *tol*, max. # of Iterations $K_{max}$, attempts $N_{att}$.
**Ensure:** estimated sources $\hat{s}_{1t}$ and $\hat{s}_{2t}$.

```
1:  C_xx = 1/T Σ_{t=1}^T x_t x_t^T; Covariance matrix.
2:  UΛV^T = C_xx; Singular Value
    Decomposition SVD.
3:  x_t = Λ^{-1/2} U^T x_t, (t = 1, 2, ...y, T); Whitening.
4:  Search for first extremum
5:  θ^(0) = 2πu; Initialization: u is a random
    number uniformly distributed in [0,1].
6:  δ_θ = +∞, k = 0;
7:  while δ_θ > tol and k < K_max do
8:      θ^(k+1) = θ^(k) ± g'(θ^(k))/|g''(θ^(k))|; N--R iteration^c.
9:      δ_θ = |θ^(k+1) − θ^(k)|;
10:     k = k + 1;
11: endwhile
12: θ_1 = θ^(k−1); First local extremum found
13: Search for second extremum
14: θ^(0) = θ_1 + π/2; Initialization
15: Repeat STEPs 6-11;
16: n = 1;^d
17: while |θ^(k−1) − θ_1| < tol and n < N_att do
18:     θ^(0) = 2πu; Initialization: u is a
        random number uniformly
        distributed in [0,1].
19:     Repeat STEPs 6-11;
20:     n = n + 1;
21: endwhile
22: θ_2 = θ^(k−1); Second local extremum found
23: return ŝ_it = cos(θ_i)x_1t + sin(θ_i)x_2t (i = 1, 2);
```

In Algorithm 1, the algorithm for the case $n = m = 2$ (two mixtures and two sources) is shown. In this case, after the first local extremum is found, the algorithm searches for the second local extrema starting from an initial guess $\theta^{(0)} = \theta_1 + \pi/2$ which, in the case of having independent sources, would correspond exactly to the location of the second source (orthogonal case). It is noted that, in the general dependent sources case, it is possible that this procedure results in finding the same local extremum again. In order to avoid this situation, the algorithm re-start

the local extrema search by using different random initial guesses until the proper local extremum is found. The maximum number of attempts $N_{att}$ is a parameter which was set to $N_{att} = 20$ in our simulations.

It is important to highlight that, if we generalize Algorithm 1 to the case of arbitrary number of sources and $m = n > 2$, we may apply a deflation step by eliminating every local extrema after they are detected preventing from multiple detections. However, this deflation step is not trivial in the dependent case since the sources are not orthogonal and the classical deflation technique used in ICA is not longer valid. For the particular case of the *NG measure*, in [12] a special deflation step was developed by transforming the data in order to make it Gaussian at the location of any detected source.

We highlight that computing the derivatives of the SE based on Parzen windows produces numerically unstable results because $\frac{d}{d\theta} \log(f(x,\theta)) = \frac{1}{f(x,\theta)} \frac{df(x,\theta)}{d\theta}$, thus, the errors in the estimation of the pdf are amplified in the derivative. On the other hand, the estimation of the derivatives for *GA moments* and *NG measure* do not suffer this problem and showed to be numerically stable in our simulations.

## Source separation experiments
### Separation performance evaluation on different datasets
In this section, we show the results of applying our N–R algorithm based on *GA moments* (order $\beta = 0.5, 1, 1.5, 2.5, \ldots, 10$) and *NG measure* (MaxNG) compared with FastICA[e] (with $g(x) = x^3$ and $g(x) = \tanh(x)$ nonlinearities) and the BCA algorithm recently proposed in [22]. FastICA is a classic, very fast algorithm developed for ICA, on the other side, BCA algorithm is a powerful geometric method for ICA/DCA based on the idea that the mixture of bounded sources increases the volume of the support of random variables. BCA obtains the separation by minimizing the volume of the support of estimated sources by assuming that the support of the sources is equal to the cartesian product of the individual supports [11]. The last condition is valid for independent sources and can be seen as a strong condition for dependent sources, for instance, sources found in the blind spectral unmixing do not meet this condition as Figure 2 illustrates.

In Figure 4, we present the performance results in terms of the obtained signal to interference ratio (SIR) which is defined as $\text{SIR}_i = -10 \log_{10} \left( \frac{1}{T} \sum_{t=1}^{T} (\hat{s}_{it} - s_{it})^2 \right)$. We used the following datasets: (a) Independent Sub-Gaussian sources, generated by applying the function $\sinh(u)^{-1}$ to zero-mean Gaussian independent signals; (b) Independent Super-Gaussian sources, generated by applying the function $\sinh(u)$ to zero-mean Gaussian independent signals; (c) Independent bimodal sources,

where each of the independent sources were generated by mixing two Gaussians with $(\mu_1, \sigma_1) = (0.5, 0.2)$ and $(\mu_2, \sigma_2) = (-0.5, 0.2)$, respectively; (d) Independent and uniformly distributed zero-mean sources; (e) Dependent constrained sources, generated by using $s_{it} = \frac{n_{it}}{\sum_{p=1}^{4} n_{p_t}}$ with $i = 1, 2$, where signals $n_{p_t}$ ($p = 1, 2, \ldots, 4$) were generated as independent uniform distributions in $[0, 1]$. (f) Dependent sources with Copula-t distributions, where $s_{1t}$ and $s_{2t}$ were generated from a Copula-t with 4 degrees of freedom and with linear correlation $\rho = 0.8$ which makes them highly dependent.[f] We observe that, for the case of Sub-Gaussian independent sources (a), *GA moments* with $\beta = 3, 4, \ldots, 10$ give a similar performance as FastICA and MaxNG. For the case (b) (Super-Gaussian independent sources), the performance of *GA moments* is slighter less than FastICA and MaxNG. For bimodal independent sources (c) and uniformly distributed independent sources (d), the performance of *GA moment* is similar to FastICA and MaxNG for values $\beta = 1.0, 1.5, 2.5, \ldots, 6.5$. For constrained dependent sources (e), the best performance is obtained for $\beta = 6.0, 6.5, \ldots, 10$ and MaxNG with a SIR of approximately 40 dB. It is noted that the LCE condition holds exactly, thus the separation is almost perfect by using *NG measure*. On the other hand, in case (f) sources modelled with Copula-t distribution with correlation $\rho = 0.8$ where the LCE condition holds only approximately as the Figure 2c illustrated, for this reason, the quality of separation by using the *NG measure* is degraded (SIR of approximately 20 dB) and BCA outperforms all the other methods because sources fulfil the BCA conditions. It is important to mention that dataset (e) does not fulfil the assumptions for FastICA (independence) neither for BCA (support of sources is not equal to the cartesian product of individual supports). It is also interesting to note that for $\beta = 4$, the performance drops because the second order derivative is zero (not a maximum neither a minimum). It is clear that, thse lower performance of BCA for cases (a), (b), (e) and (d) can be attributed to the fact that these sources do not fulfil the conditions for BCA, i.e. or they have not bounded support or the support of sources can not be written as the cartesian product of individual supports.

### Robustness to the sample size *T*
We have theoretically proved that several objective functions are valid to separate sources verifying the LCE condition. Nevertheless, in practice, the *GA moments* and the *NG measure* are estimated from available samples which implies that the measures are sensible to the size of the dataset $T$. In Figure 5, the robustness of the measures is shown by evaluating the mean SIR of the separation versus the sample size $T$. In the small dataset size case, the errors on the estimation of the moments and their derivatives
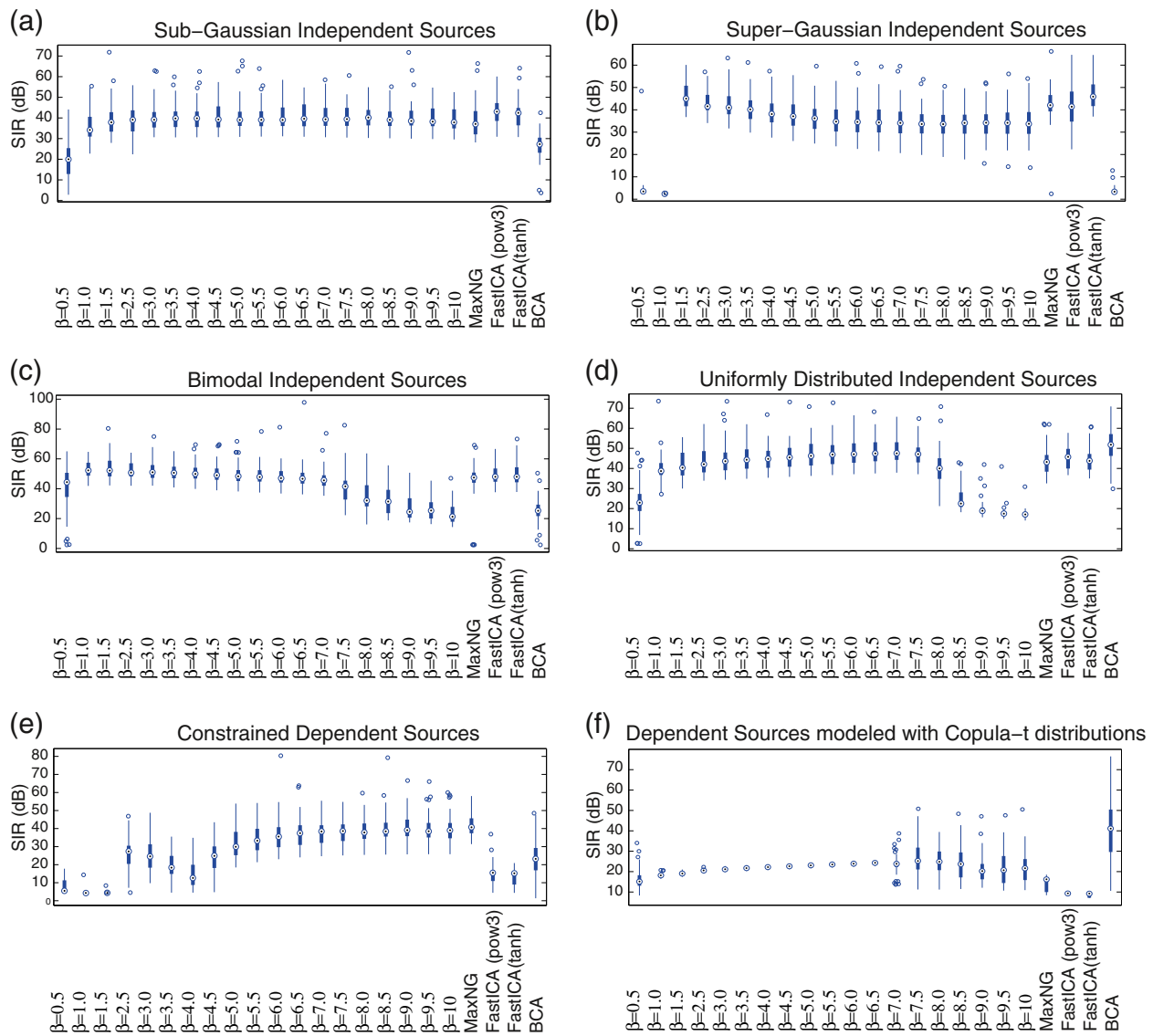
**Figure 4 Results of applying *GA moment* (order $\beta$ = 0.5,1,1.5,2.5, . . . ,10) and *NG measure* (MaxNG) compared with FastICA (with $g(x) = x^3$ and $g(x) = \tanh(x)$ nonlinearities) and the BCA algorithm proposed in [22] for the case of $n = 2$ sources and $m = 2$ mixtures.** The mixing matrix $\mathbf{A} \in \mathbb{R}^{2 \times 2}$ was generated using independent Gaussian random numbers. We performed a total number of 50 Monte Carlo simulations and we use a total number of $T = 8,000$ samples in each case. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually with circles.

can be significant, on the other side, the *NG measure* showed to be significantly more robust.

**Blind spectral unmixing example**

In order to have a realistic set of sources for testing our method in the context of the blind spectral unmixing problem, we used a set of material abundances generated as follows. Based on a real ground-truth (see Figure 6 (left)) of a selected area of Rome city, we assign a source to each one of the classes. For the estimation of each source (abundance) we divide the map in $8 \times 8$ pixel subareas

and we calculated the material abundances as the percentages of the classes within each subarea. As a result we obtained nine sources with a total of $T = 2814$ ($67 \times 42$) samples each (in Figure 7a scatter plots for some examples of pair of sources are shown). In Figure 7b, the performance results are shown for MaxNG, FastICA and BCA algorithms applied to different combinations of two sources and using randomly selected mixing matrices over a total of 50 simulations. We note that the results with *GA moments* are not included because their performance was poor (similar to FastICA). We think this is because the
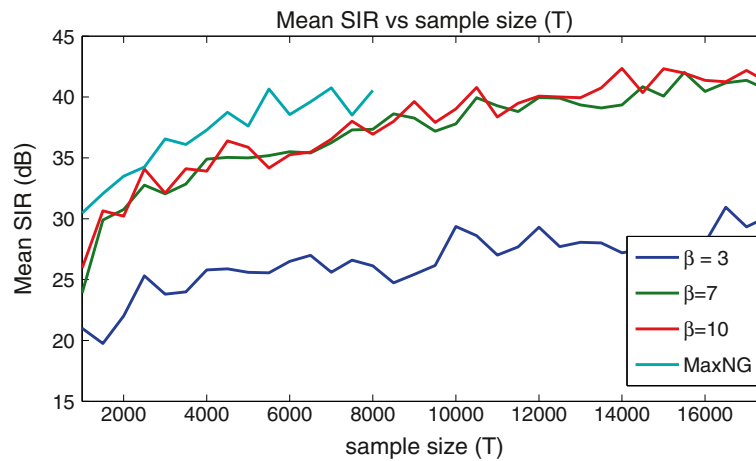
**Figure 5 Robustness to sample size *T*: Mean SIR values versus *T* was computed for the case of constrained sources.** The *NG measure* showed to be the most robust measure.

sample size is too small ($T = 2814$) and the distributions are very irregular. On the other hand MaxNG showed the best performance. BCA and FastICA has lower than MaxNG because sources does not fulfill the conditions required by the algorithms i.e. they are not independent and their support can not be written as the cartesian product of individual supports.

## Conclusions and discussion

This article contributes to shed light on the theoretical aspects of the separation of independent and dependent sources based on the maximization (or minimization) of objective functions by filling the gaps existing among previous works and giving rigorous theoretical answers to important questions. Furthermore, this new theoretical framework opens the possibility to analyze new objective functions for BSS problems. We have shown that, under the LCE assumption, several objective functions such as *GA moments*, *NG measure* and SE are valid for the separation of dependent sources. However, among these measures, we showed that *GA moments* are less robust to the sample size *T* than the *NG measure* but has much lower computational complexity. We have also shown that simple and efficient algorithms can be developed based on these measures by using Parzen windows technique combined with a N–R iterative search of local extrema. Nevertheless, it was noted that estimations of derivatives of the SE, based on Parzen windows, becomes numerically unstable.

Another disadvantage of the *GA moments* is that additional information about the sources is needed in order to determine if the separation is obtained at a maximum or a minimum. When sources are independent, we can determine the sign of the second order derivative by just evaluating Equation (33) which can be done quickly

and easily from data. On the other side, for dependent sources, it is necessary to know the second order conditional expectations, i.e. $E[S_i^2|S_j]$. Additionally, it is needed to chose the proper order $\beta$ which could be not simple and it is out of scope of this article. On the other hand, the *NG measure* does not require any extra parameter, it is very robust to the sample size *T* and usually the separation is obtained at local maxima (except in pathological cases as shown in our example in Figure 3d).

As a main conclusion, we have found that the separation of dependent sources is possible but additional constraints, or assumptions, on the type of dependence among sources must be taken into account. For example, if we know that the support of sources can be written as the cartesian product of the individual supports, then an elegant and very efficient method is to apply the BCA algorithm, or if sources have LCE, as in the case of abundances in the blind spectral unmixing application, then the methods presented in this article are the most appropriate.

## Appendix 1

Applying the differentiator operator under the integral sign in Equation (18) for the case of $n = 2$ sources, we to obtain the partial derivatives of the pdf evaluated at $(\alpha_1, \alpha_2) = (0, 1)$ as follows:

$$\frac{\partial f_X(x)}{\partial \alpha_1}\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} = -\left(f_{S_2}(x)E[S_1|x]\right)', \tag{41}$$

$$\frac{\partial^2 f_X(x)}{\partial \alpha_1^2}\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} = \left(f_{S_2}(x)E[S_1^2|x]\right)'', \tag{42}$$

$$\frac{\partial f_X(x)}{\partial \alpha_2}\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} = -\left(xf_{S_2}(x)\right)', \tag{43}$$
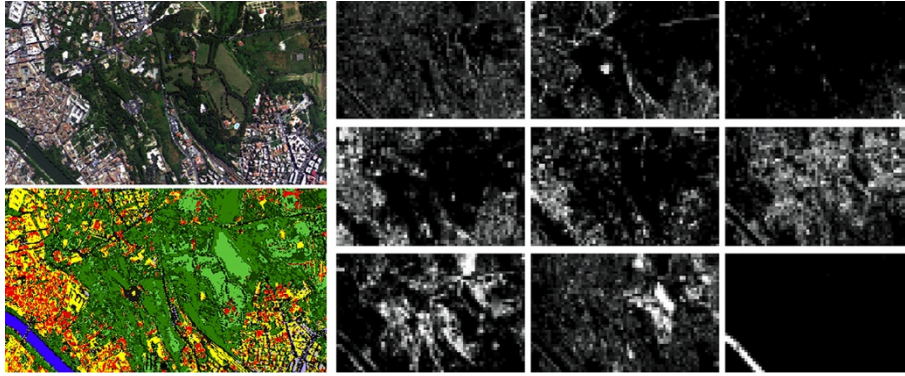
**Figure 6 Real radiometrically corrected hyper spectral image of a Rome city area as provided by the Airborne Laboratory for Environmental Research at IIA-CNR in Rome, Italy [19].** This 540 × 337 pixels image was obtained with an airborne imaging spectrometer containing 102 channels. The RGB version (left-upper), the classification map or ground-truth considering pure pixels (left-bottom) and the nine material abundances (67 × 42 pixels) computed by using a 8 × 8 window (right) are shown.

$$\frac{\partial^2 f_X(x)}{\partial \alpha_2^2}\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} = 2f_{S_2}(x) + 4xf'_{S_2}(x) + x^2 f_{S_2}(x), \qquad (44)$$

$$\frac{\partial^2 f_X(x)}{\partial \alpha_1 \alpha_2}\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0} = 2\left(f_{S_2}(x)E[\,S_1|x]\,\right)' + x\left(f_{S_2}(x)E[\,S_1|x]\,\right)'', \qquad (45)$$

Using the chain rule of derivatives we have that

$$\frac{d^2 f_X(x;\tau)}{d\tau^2} = \frac{\partial^2 f}{\partial \alpha_1^2}\left(\alpha'_1(\tau)\right)^2 + 2\frac{\partial^2 f}{\partial \alpha_1 \partial \alpha_2}\alpha'_1(\tau)\alpha'_2(\tau) + \frac{\partial^2 f}{\partial \alpha_2^2}$$
$$\times \left(\alpha'_2(\tau)\right)^2 + \frac{\partial f}{\partial \alpha_1}\alpha''_1 + \frac{\partial f}{\partial \alpha_2}\alpha''_2. \qquad (46)$$

And, using the fact that

$$\alpha'_1(0) = 1, \alpha''_1(0) = 0, \alpha'_2(0) = -\rho, \alpha''_2(0) = \rho^2 - 1; \qquad (47)$$

we obtain the desired result of Equation (27).

## Appendix 2
The second order derivative of the SE at $\tau = 0$ is:

$$g''_{SE}(0) = \int \frac{d^2 f_X(x;0)}{d\tau^2}\left(\log(f_X(x;\tau)) + 1\right)dx. \qquad (48)$$

In the following, in order to simplify the notation we replace $f(x) \equiv f_{S2}(x)$ and $g_{SE} \equiv -\int f(x)\log(f(x))dx$.

Now, by using Equation (27) into (48) and, taking into account the following results:

$$\int \left(f(x)E[\,S_1^2|x]\,\right)' \frac{f'(x)}{f(x)}dx = -\int E[\,S_1^2|x]\frac{(f'(x))^2}{f(x)}dx$$
$$+ \int E''[\,S_1^2|x]f(x)dx,$$

$$\int xf'(x)\left(\log(f(x)) + 1\right)dx = g_{SE},$$

$$\int x^2 f''(x)\left(\log(f(x)) + 1\right)dx = -2g_{SE} - \int x^2 \frac{(f'(x))^2}{f(x)}dx,$$



(a) Examples of pairs of sources

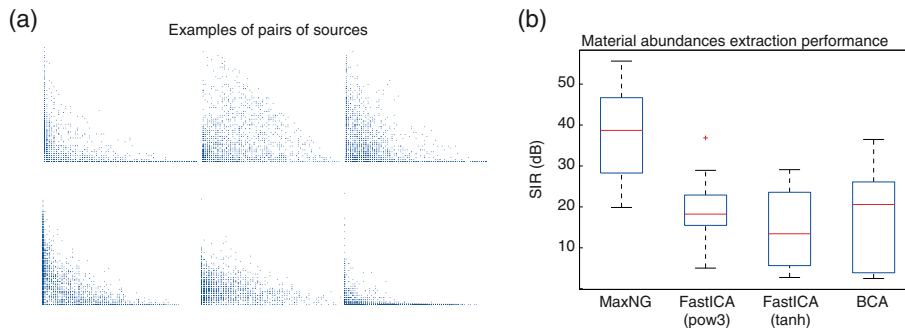(b) Material abundances extraction performance

**Figure 7 Results for blind spectral unmixing based on DCA.** Material abundances computed based on a real world map (Rome city image in Figure 6) were artificially and randomly mixed and separated by MaxNG, FastICA and BCA. In **(a)** selected examples of normalized sources pairs are shown. In **(b)** the performance results are shown in terms of the obtained SIR.

we finally arrive at the desire result of Equation (29).

## Appendix 3

By using Equation (27) into (30) and, taking into account the following results:

$$\int \left(f(x)E\left[S_1^2|x\right]\right)'' |x|^\beta dx = \beta(\beta-1)\int f(x)E\left[S_1^2|x\right]|x|^{\beta-2}dx,$$

$$\int x|x|^\beta f'(x)dx = -(\beta+1)\mu_\beta,$$

$$\int x^2|x|^\beta f''(x)dx = (\beta+2)(\beta+1)\mu_\beta,$$

we finally arrive at the desire result of Equation (31).

## Endnotes

[a]For ease of the presentation, we consider here only the case of two sources which correspond to have only one parameter $\theta$. For the case of $n > 2$ a hyper-spheric coordinate system can be used as shown in [12].

[b]In order to solve the problem of possible zero second order derivatives, more sophisticated methods well known in the literature can be implemented as, for example, by using the Conjugated Gradient method.

[c]$g'(.)$ and $g''(.)$ are the first and second order derivatives of a selected measure and can be computed by taking derivatives on Equations (36), (37) or (39) for the case of SE, NG or GA moment, respectively. Sign '+' and sign '-' correspond to maximum or minimum search, respectively.

[d]If the same local extremum is found then a new search starts (up to $N_{att}$ attempts).

[e]FastICA package was downloaded from the author's webpage http://research.ics.tkk.fi/ica/fastica/.

[f]We used the Matlab command s=copularnd('t',0.8,4,T).

### References

1. P Comon, Independent component analysis, a new concept. Signal Process. **36**(3), 287–314 (1994)
2. P Comon, C Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. (Academic Press, Oxford Burlington, 2010)
3. A Hyvärinen, J Karhunen, E Oja, *Independent Component Analysis*. (Wiley, New York, 2001)
4. A Cichocki, SI Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. (Wiley, Chichester, 2002)
5. S Cruces, A Cichocki, S Amari, in *6th International Work-Conference on Artificial and Natural Neural Networks: Bio-inspired Applications of Connectionism-Part II*. The minimum entropy and cumulants based contrast functions for blind source extraction, (Granada, 2001)), pp. 786–793
6. D Erdogmus, I Hild, E Kenneth, J Principe, Blind source separation using Renyi's [alpha]-marginal entropies. Neurocomputing. **49**(1–4), 25–38 (2002)
7. D Pham, F Vrins, M Verleysen, On the risk of using Renyi's entropy for blind source separation. IEEE Trans. Signal Process. **56**(10), 4611–4620 (2008)
8. J Cardoso, High-order contrasts for independent component analysis. Neural Comput. **11**, 157–192 (1999)
9. A Hyvärinen, Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural Netw. **10**(3), 626–634 (2002)
10. V Zarzoso, R Phlypo, P Comon, A contrast for independent component analysis with priors on the source kurtosis signs. IEEE Signal Process. Lett. **15**, 501–504 (2008)
11. S Cruces, Bounded component analysis of linear mixtures: a criterion of minimum convex perimeter. IEEE Trans. Signal Process. **58**(4), 2141–2154 (2010)
12. C Caiafa, A Proto. Separation of statistically dependent sources using an L2-distance non-Gaussianity measure Signal Process. **86**(11), 3404–3420 (2006)
13. J Lee, F Vrins, M Verleysen, Blind source separation based on endpoint estimation with application to the MLSP 2006 data competition. Neurocomputing. **72**(1–3), 47–56 (2008)
14. A Hyvärinen, P Hoyer, Emergence of phase- and shift-invariant features by decomposition of natural images into independent feature subspaces. Neural Comput. **12**(7), 1705–1720 (2000)
15. DD Lee, HS Seung, Learning the parts of objects by nonnegative matrix factorization. Nature. **401**, 788–791 (1999)
16. F Theis, K Stadlthanner, T Tanaka, First results on uniqueness of sparse non-negative matrix factorization. IEEE Trans. Image Process. **15**, 81–88 (2005)
17. C Caiafa, A Cichocki, Estimation of sparse nonnegative sources from noisy overcomplete mixtures using MAP. Neural Comput. **21**(12), 3487–3518 (2009)
18. L Bedini, D Herranz, E Salerno, C Baccigalupi, E Kuruoğlu, A Tonazzini, Separation of correlated astrophysical sources using multiple-lag data covariance matrices. EURASIP J. Appl. Signal Process. **2005**, 2400–2412 (2005)
19. C Caiafa, E Salerno, A Proto, L Fiumi, Blind spectral unmixing by local maximization of non-Gaussianity. Signal Process. **88**, 50–68 (2008)
20. C Caiafa, E Kuruoglu, in *Bayesian Inference and Maximum Entropy Methods in Science and Engineering(AIP Conference Proceedings)*, vol. 872. A minimax entropy method for blind separation of dependent components in astrophysical images, (Paris, 2006)), pp. 81–88
21. E Kuruoglu, in *Latent Variable Analysis and Signal Separation*. Dependent component analysis for cosmology: a case study (St. Malo , 2010)), pp. 538–545
22. A Erdogan, in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2012, ICASSP 2012*. A family of bounded component analysis algorithms, (Kyoto, 2012)), pp. 1–4
23. DL Donoho, On minimum entropy deconvolution. Appl. Time Ser. Anal. **2**, 564–608 (1981)
24. C Caiafa, E Salerno, A Proto, in *Knowledge-Based Intelligent Information and Engineering Systems*. Blind source separation applied to spectral unmixing: comparing different measures of nongaussianity, (Vietri Sul Mare, 2010)), pp. 1–8
25. N Keshava, JF Mustard, Spectral unmixing. IEEE Signal Process. Mag. **19**, 44–57 (2002)
26. CI Chang, CC Wu, W Liu, YC Ouyang, A new growing method for simplex-based endmember extraction algorithm. IEEE Trans. Geosci. Remote Sens. **44**(10), 2804–2819 (2006)
27. R Heylen, D Burazerovic, P Scheunders, Fully constrained least squares spectral unmixing by simplex projection. IEEE Trans. Geosci. Remote Sens. **49**(11), 4112–4122 (2011)
28. JMP Nascimento, JMB Dias, Vertex component analysis: a fast algorithm to unmix hyperspectral data. IEEE Trans. Geosci. Remote Sens. **43**(4), 898–910 (2005)

29.  Z Yang, G Zhou, S Xie, S Ding, JM Yang, J Zhang, Blind spectral unmixing based on sparse nonnegative matrix factorization. IEEE Trans. Image Process. **20**(4), 1112–1125 (2011)
30.  F Vrins, D Pham, M Verleysen, Mixing and non-mixing local minima of the entropy contrast for blind source separation. IEEE Trans. Inf. Theory. **53**(3), 1030–1042 (2007)
31.  DT Pham, F Vrins, Local minima of information-theoretic criteria in blind source separation. IEEE Signal Process. Lett. **12**(11), 788–791 (2005)
32.  F Vrins, M Verleysen, Information theoretic versus cumulant-based contrasts for multimodal source separation. IEEE Signal Process. Lett. 10. **12**(3), 190–193 (2005)
33.  O Johnson, *Information Theory and Central Limit Theorem*. ((Imperial College Press, River Edge London, 2004)
34.  BW Silverman, *Density Estimation for Statistics and Data Analysis*. (Chapman &amp Hall/Crc, Boca Raton, 1985)