RESEARCH

Separation of instantaneous mixtures of a particular set of dependent sources using classical ICA methods

Marc Castella[1*], Selwa Rafi[1], Pierre Comon[2] and Wojciech Pieczynski[1]

Abstract

This article deals with the problem of blind source separation in the case of a linear and instantaneous mixture. We first investigate the behavior of known independent component analysis (ICA) methods in the case where the independence assumption is violated: specific dependent sources are introduced and it is shown that, depending on the source vector, the separation may be successful or not. For sources which are a probability mixture of the previous dependent ones and of independent sources, we introduce an extended ICA model. More generally, depending on the value of a hidden latent process at the same time, the unknown components of the linear mixture are assumed either mutually independent or dependent. We propose for this model a separation method which combines: (i) a classical ICA separation performed using the set of samples whose components are conditionally independent, and (ii) a method for estimation of the latent process. The latter task is performed by iterative conditional estimation (ICE). It is an estimation technique in the case of incomplete data, which is particularly appealing because it requires only weak conditions.

Keywords: Blind source separation, Dependent sources, Independent Component Analysis (ICA), Higher order statistics, Iterative Conditional Estimation (ICE)

1 Introduction

For the last decades, blind source separation (BSS) has been an active research problem: this popularity comes from the wide panel of potential applications such as audio processing, telecommunications, biology, etc. In the case of a linear multi-input/multi-output (MIMO) instantaneous system, BSS corresponds to independent component analysis (ICA), which is now a well recognized concept [1]. Contrary to other frameworks where techniques take advantage of a strong information on the diversity, for instance through the knowledge of the array manifold in antenna array processing, the core assumption in ICA is much milder and reduces to the statistical mutual independence between the inputs. However, the latter assumption is not mandatory in BSS. For instance, in the case of static mixtures, sources can be separated

if they are only decorrelated, provided that their nonstationarity or their color can be exploited. Other properties such as the fact that sources belong to a finite alphabet can alternatively be utilized [2,3] and do not require statistical independence. We consider in this article the case of dependent sources without assuming nonstationarity nor color.

To the best of authors' knowledge, only few references have tackled the issue of dependent source separation [4-15], although the interest in dependent sources has been witnessed by studies in various applied domains such as cosmology [6,13,14], biology/medicine [7,8,16], feature extraction [17]. Among the interesting proposed extensions of ICA to dependent components, we should mention tree-dependent models [11] and models with dependence in variance profiles [12]. Contrary to the mentioned articles, our approach is based on the selection of an appropriately chosen sub-sample of the available data, which then feeds the entry of a classical ICA method. Among ICA or BSS methods, one can distinguish two approaches: some methods recover the sources one by

*Correspondence: marc.castella@it-sudparis.eu
[1]Institut Mines-Télécom/Télécom Sudparis, CNRS UMR 5157 SAMOVAR, 9 rue Charles Fourier, 91011 Évry Cedex, France
Full list of author information is available at the end of the article

Springer

one, which is what we refer to as multi-input/single-output (MISO) approaches. These approaches are often used in conjunction with a so-called deflation procedure [18,19]. In contrast, other approaches, which will be referred to as MIMO recover all the sources simultaneously.

Inspired from [20-23], we investigate in a first part of the article the behavior of the kurtosis contrast function: this criterion is well-known in MISO BSS approaches and we study some of its properties in some specific cases of dependent sources.

In a second part of the article, we investigate a particular model which combines an ICA model with a probabilistic model on the sources, making them either dependent or independent at different time instants. Our method exploits the "independent part" of the source components. Although it is possible to refine our model by introducing a temporal dependence, it assumes neither nonstationarity nor color of the sources. We would like to outline the difference between our study and [17]: the latter assumes a conditional independence of the sources, whereas, depending on a hidden process, we assume either conditional independence or dependence. The proposed separation method which is introduced relies on iterative conditional estimation (ICE), which has been introduced recently [24].

The considered model and notations are specified in Section 2. In Section 3, specific dependent sources are introduced and the behavior of the kurtosis contrast function is investigated. Then, Section 4 introduces a genuine model of dependent sources, for which separation is possible. The principles of our method and a discussion on ICE are provided in Section 5. The algorithm is precisely described in Section 6, where a parallel is also made with the accept-reject random generation method. Some simulations are provided in Section 7 and Section 8 concludes authors' study.

2 Mixture model

2.1 Linear mixture

We consider a set of T samples of vector *observations*. At each time instant $t \in \{1, \ldots, T\}$ the observed vector is denoted by $\mathbf{x}(t) \triangleq (x_1(t), \ldots, x_N(t))^{\mathrm{T}}$. We assume that these observations result from a linear mixture of N unknown and unobserved *source* signals. More precisely and in other words, there exists a matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ and a vector valued process $\mathbf{s}(t) \triangleq (s_1(t), \ldots, s_N(t))^{\mathrm{T}}$ such that:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t), \qquad \forall t \in \{1, \ldots, T\}. \tag{1}$$

Let $\mathbf{X} \triangleq (\mathbf{x}(1), \ldots, \mathbf{x}(T))$ be the $N \times T$ matrix with all samples of the observations and $\mathbf{S} \triangleq (\mathbf{s}(1), \ldots, \mathbf{s}(T))$ be the $N \times T$ matrix with all sources samples. The matrix $\mathbf{A}$ is unknown and the objective consists in recovering $\mathbf{S}$ from $\mathbf{X}$ only: this is the so-called *blind source separation*

problem. We will assume here that $\mathbf{A}$ is a square left-invertible matrix and the problem thus reduces to the estimation of $\mathbf{A}$ or its inverse. A solution has been developed for long and is known as ICA [1]. It generally requires two assumptions: the source components should be non Gaussian—except possibly one of them—and they should be statistically mutually independent. With these assumptions, it is known that one can estimate a matrix $\mathbf{B} \in \mathbb{R}^{N \times N}$ such that $\mathbf{y}(t) = \mathbf{B}\mathbf{x}(t)$ restores the sources up to some ambiguities, namely ordering and scaling factors. In this article, with no loss of generality, we assume that the sources are zero-mean and have unit power. Finally, note that if $\mathbf{A}$ is a tall matrix (i.e., there are more observations than sources), a dimensionality reduction technique such as the principal component analysis (PCA) can be used to obtain a mixture with as many observations as sources.

2.2 Notations

In the following, $\mathbf{B}$ denotes the estimated inverse of $\mathbf{A}$ and is referred to as the separating matrix. Defining $\mathbf{G} \triangleq \mathbf{B}\mathbf{A}$ the combined mixing-separating matrix, the BSS problem is solved if $\mathbf{G}$ is a so-called trivial matrix, i.e., the product of a diagonal matrix with a permutation: these are well known ambiguities of BSS.

In Section 3, we will study separation criteria as functions of $\mathbf{G}$. Source separation sometimes proceeds iteratively, extracting one source at a time (e.g., deflation approach). In this case (and particularly in Section 3), we will write $y(t) = \mathbf{b}\mathbf{x}(t) = \mathbf{g}\mathbf{s}(t)$ where $\mathbf{b}$ and $\mathbf{g} = \mathbf{b}\mathbf{A}$, respectively, correspond to a row of $\mathbf{B}$ and $\mathbf{G}$ and $y(t)$ denotes the only output of the separating algorithm. In this case, the separation criteria are considered as functions of $\mathbf{g}$.

Finally, Cum $\{.\}$ denotes the cumulant of a set of random variables (see e.g., [1]) and Var$\{.\}$ denotes the variance of a random variable. For a random vector $\mathbf{s} = (s_1, \ldots, s_N)^{\mathrm{T}}$ and for any multi-index $\mathbf{i} = (i_1, \ldots, i_N)$, we introduce the notation:

$$\kappa_{\mathbf{i}}^{(\mathbf{s})} \triangleq \mathrm{Cum} \left\{ \underbrace{s_1, \ldots, s_1}_{i_1}, \ldots, \underbrace{s_N, \ldots, s_N}_{i_N} \right\}$$

We denote by $\mathcal{N}(\mu, \sigma^2)$ the Gaussian law with mean μ and variance σ^2 and by $\mathcal{L}(\lambda)$ the Laplace (i.e., double-exponential) distribution with zero-mean and scale parameter λ. The symbol $\sim$ denotes the law followed by a random variable or equality between probability distributions; the conditional distribution of a random variable (or vector) $\mathbf{r}$ knowing $\mathbf{X}$ under a parameter value θ is denoted by $\mathbb{P}(\mathbf{r} \mid \mathbf{X}; \theta)$. Finally, we denote by $\delta(.)$ the Kronecker function which equals 1 if the argument in parenthesis holds true and 0 otherwise.

3 A class of dependent sources

In this section, we introduce particular dependent sources based on products of independent signals. These models were shown to be useful when dealing with underdetermined [20] and nonlinear [23] mixtures. We assume that all processes are stationary. At each time instant t, the vectors $\mathbf{s}(t)$ and $\mathbf{x}(t)$ are realizations of random vectors. Since no confusion is possible, in this section only, we drop the time index t and these vectors are denoted, respectively, by $\mathbf{s}$ and $\mathbf{x}$.

3.1 Three dependent sources

3.1.1 Specific sources and properties

Binary phase shift keying (BPSK) signals have specificity that will allow us to obtain source vectors with interesting properties. By definition, BPSK sources take values $s = +1$ or $s = -1$ with equal probability $1/2$. We define the following source vector:

A1. Let ϵ be a BPSK random variable and a a real-valued non Gaussian random variable with non-zero fourth-order cumulant $\kappa_4^{(a)} \neq 0$. We assume also that a is independent of ϵ and $\mathbb{E}\{a\} = \mathbb{E}\{a^3\} = 0$, $\mathbb{E}\{a^2\} = 1$. Then, we define the source vector $\mathbf{s} \triangleq (s_1, s_2, s_3)^{\mathsf{T}}$ as follows:

$$s_1 = a \qquad s_2 = a\epsilon \qquad s_3 = \epsilon$$

Interestingly, the following lemma holds true:

Lemma 1. *The sources s_1, s_2, s_3 defined by A1 are zero-mean, unit variance, mutually dependent, decorrelated and their fourth-order cross-cumulants values are such that:*

$$\kappa_{0,0,4}^{(\mathbf{s})} = -2$$
$$\kappa_{4,0,0}^{(\mathbf{s})} = \kappa_{0,4,0}^{(\mathbf{s})} = \kappa_4^{(a)} \neq 0$$
$$\kappa_{2,2,0}^{(\mathbf{s})} = Var\{a^2\} = \kappa_{4,0,0}^{(\mathbf{s})} + 2$$
$$\kappa_{i_1,i_2,i_3}^{(\mathbf{s})} = 0 \text{ for any other } i_1 + i_2 + i_3 = 4.$$

Proof. From A1, $s_1 = a$ and $s_3 = \epsilon$ have zero-mean by definition and so does $s_2 = a\epsilon$ by independence of a and ϵ. Hence $\mathbb{E}\{s_1\} = \mathbb{E}\{s_2\} = \mathbb{E}\{s_3\} = 0$ and for such centered random variables, it is known that cumulants can be expressed in terms of moments as follows:

$$\text{Cum}\left\{s_i, s_j\right\} = \mathbb{E}\{s_i s_j\} \tag{2}$$

$$\begin{aligned}
\text{Cum}\left\{s_i, s_j, s_k, s_l\right\} = {} & \mathbb{E}\{s_i s_j s_k s_l\} - \mathbb{E}\{s_i s_j\}\mathbb{E}\{s_k s_l\} \\
& - \mathbb{E}\{s_i s_k\}\mathbb{E}\{s_j s_l\} - \mathbb{E}\{s_i s_l\}\mathbb{E}\{s_j s_k\}
\end{aligned} \tag{3}$$

It is then possible to check all cases of Equations (2) and (3), using again A1. We obtain that (2) vanishes for $i \neq j$

and the decorrelation of the sources follows. The values of the fourth-order cumulants in the lemma are obtained similarly.

On the other hand, the third order cross-cumulant reads:

$$\text{Cum}\left\{s_1, s_2, s_3\right\} = \mathbb{E}\{s_1 s_2 s_3\} = \mathbb{E}\{a^2 \epsilon^2\} = \mathbb{E}\{a^2\} = 1 > 0$$

and this proves that s_1, s_2, s_3 are mutually dependent. □

Depending on $s_1 = a$, more can be proved about the source vector defined by A1. For example, if the probability distribution of a is symmetric, then s_2 and s_3 are independent. On the contrary s_1 and s_2 are generally not independent. An even more specific case is obtained when $s_1 = a$ is itself BPSK. Using in Lemma 1 the fact that, in the latter specific case $s_2^2 = a^2 = 1$, and calculating also the pairwise probability functions, we obtain the following result:

Lemma 2. *Consider the source vector defined by A1. If in addition $s_1 = a$ is BPSK, the source vector $\mathbf{s}$ satisfies:*

- *each component s_i ($i = 1, 2$ or 3) is BPSK,*
- *(s_1, s_2, s_3) are mutually dependent,*
- *(s_1, s_2, s_3) are pairwise independent and hence decorrelated,*
- *all fourth order cross cumulants of $\mathbf{s}$ vanish, that is:*

$$\kappa_{4,0,0}^{(\mathbf{s})} = \kappa_{0,4,0}^{(\mathbf{s})} = \kappa_{0,0,4}^{(\mathbf{s})} = -2 \tag{4}$$
$$\kappa_{i_1,i_2,i_3}^{(\mathbf{s})} = 0 \text{ for any other } i_1 + i_2 + i_3 = 4.$$

3.1.2 Properties of the kurtosis contrast function with sources satisfying A1

Consider a source vector $\mathbf{s}$ which satisfies Assumption A1. If in addition $s_1 = a$ is BPSK, the arguments in [22] prove that a mixture of such a source vector can be separated by many classical ICA algorithms such as CoM2 [25], JADE [26], and FastICA [27]: this follows straightforwardly from Lemma 2 and the fact that the corresponding algorithms rely only on the vanishing of the cross-cumulants of the sources, that is on Equation (4). We now extend this result to more general distributions of a and state the following proposition[a]:

Proposition 1. *Let $y = \mathbf{gs}$ where the vector of sources is defined by A1. If $\kappa_4^{(a)} = \kappa_{4,0,0}^{(\mathbf{s})} \neq 2$ the function*

$$\mathbf{g} \mapsto |\kappa_4^{(y)}|^\alpha, \quad \alpha \geq 1 \tag{5}$$

defines a MISO contrast function, that is, its maximization over the set of unit norm vectors ($\|\mathbf{g}\|^2 = 1$) leads to a vector $\mathbf{g}$ with only one non-zero component. More precisely, depending on the fourth order cumulant $\kappa_4^{(a)} = \kappa_{4,0,0}^{(\mathbf{s})}$ of s_1:

- If $\kappa_4^{(a)} > 2$, the maximization of (5) leads to extraction of either $s_1 = a$ or $s_3 = \epsilon$,
- if $-2 < \kappa_4^{(a)} < 2$, the maximization of (5) leads to extraction of $s_3 = \epsilon$,
- if $\kappa_4^{(a)} = -2$, the maximization of (5) leads to extraction of either one of the sources s_1, s_2 or s_3,
- if $\kappa_4^{(a)} = 2$, (5) does not define any contrast function and its maximization leads either to extraction of s_3 or to a mixture of s_1 and s_2 only.

Remark 1. Since a is zero-mean, unit-variance, the fourth-order cumulant satisfies $\kappa_4^{(a)} \geq -2$. This follows from $\mathrm{Var}\{a^2\} = \mathbb{E}\{a^4\} - \mathbb{E}\{a^2\}^2 \geq 0$ and Equation (3). All cases are hence given in the above proposition.

Remark 2. The above result characterizes the *global* maximum of the criterion. However, one should remember that most optimization algorithms search for a *local* maximum only and may therefore fail to reach the global maximum: however, in simulations (Section 3.3), we did not observe convergence to any spurious local maximum. The same remark holds for Propositions 2, 3, and 4.

Proof. First note that for all $\alpha \geq 1$, the criterion in (5) reaches its maxima for the same values of $\mathbf{g}$. We hence consider $\alpha = 1$ in the proof. Using $y = \mathbf{gs}$, the multilinearity of the cumulants and Lemma 1 which holds for sources satisfying A1, we obtain:

$$\kappa_4^{(y)} = \kappa_4^{(a)}(g_1^4 + g_2^4) - 2g_3^4 + \left(\kappa_4^{(a)} + 2\right) g_1^2 g_2^2$$

Since the maxima of the criterion (5) under the constraint $\|\mathbf{g}\|^2 = g_1^2 + g_2^2 + g_3^2 = 1$ are either minima or maxima of $\kappa_4^{(y)}$ under the same constraint, we introduce the following Lagrangian:

$$\mathcal{L} = \kappa_4^{(a)}(g_1^4 + g_2^4) - 2g_3^4 + \left(\kappa_4^{(a)} + 2\right) g_1^2 g_2^2 - \lambda(g_1^2 + g_2^2 + g_3^2 - 1) \tag{6}$$

The sought extrema necessarily satisfy:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{g}} = 0 \quad \text{and:} \quad \frac{\partial \mathcal{L}}{\partial \lambda} = 0 \tag{7}$$

The corresponding system is polynomial. Using well-known algebraic techniques now implemented in many computer algebra systems [28], one can get a system of triangular equations equivalent to (7), whose solutions can be given explicitly. For $\kappa_4^{(a)} \neq 2$, there are 26 solutions to (7), some of them being complex-valued depending on $\kappa_4^{(a)}$. In Table 1, we give the real-valued solutions to (7) and the corresponding value of $\kappa_4^{(y)}$. We also indicate for which values of $\kappa_4^{(a)}$ the solutions hold.

For all values of $\kappa_4^{(a)} \neq 2$ in Proposition 1 one can then consider all potential maxima in Table 1 and check which one maximizes $|\kappa_4^{(y)}|$.

For $\kappa_4^{(a)} = 2$, then $\kappa_4^{(y)} = 2(g_1^2 + g_2^2)^2 - 2g_3^4$ and, using again the Lagrangian, it can be checked by hand that the extrema of $\kappa_4^{(y)}$ subject to $\|\mathbf{g}\|^2 = 1$ satisfy either $g_1 = g_2 = 0, g_3^2 = 1$ or $g_1^2 + g_2^2 = 1, g_3 = 0$. □

Amazingly, the above result does not hold any longer if one considers a mixture of only the first two components of the sources given by Assumption A1.

Proposition 2. *Let $y = \mathbf{gs}$ where the vector of sources is given by the first two components $\mathbf{s} = (s_1, s_2)^{\mathrm{T}}$ of the sources defined by A1. The function in Equation (5) satisfies:*

- If $\kappa_4^{(a)} < -\frac{2}{7}$ or $\kappa_4^{(a)} > 2$, it is a contrast function and its maximization leads to extraction of either s_1 or s_2 (that is: $\mathbf{g} = (\pm 1, 0)$ or $\mathbf{g} = (0, \pm 1)$),
- if $-\frac{2}{7} \leq \kappa_4^{(a)} \leq 2$, it is not a contrast function and its maximization leads to a non-separating solution of the type $\mathbf{g} = (\pm\frac{1}{\sqrt{2}}, \pm\frac{1}{\sqrt{2}})$.

Table 1 List of all 26 solutions to Equation (7) for $\kappa_4^{(a)} \neq 2$ (second column), corresponding value of $\kappa_4^{(y)}$ (third column) and condition for the solution g to be real-valued (last column)

Separating	$\mathbf{g}^{\mathrm{T}} = (g_1, g_2, g_3)$	$\kappa_4^{(y)}$	Number of solutions	Solution in $\mathbb{R}$
yes	$(\pm 1, 0, 0)$	$\kappa_4^{(a)}$	2	
yes	$(0, \pm 1, 0)$	$\kappa_4^{(a)}$	2	
no	$\left(\pm\frac{1}{\sqrt{2}}, \pm\frac{1}{\sqrt{2}}, 0\right)$	$\frac{3\kappa_4^{(a)}+2}{4}$	4	
yes	$(0, 0, \pm 1)$	-2	2	
no	$\left(0, \pm\sqrt{\frac{-2}{\kappa_4^{(a)}-2}}, \pm\sqrt{\frac{\kappa_4^{(a)}}{\kappa_4^{(a)}-2}}\right)$	$\frac{2\kappa_4^{(a)}}{2-\kappa_4^{(a)}}$	4	$\kappa_4^{(a)} \leq 0$
no	$\left(\pm\sqrt{\frac{-2}{\kappa_4^{(a)}-2}}, 0, \pm\sqrt{\frac{\kappa_4^{(a)}}{\kappa_4^{(a)}-2}}\right)$	$\frac{2\kappa_4^{(a)}}{2-\kappa_4^{(a)}}$	4	$\kappa_4^{(a)} \leq 0$
no	$\left(\pm\sqrt{\frac{-4}{3\kappa_4^{(a)}-6}}, \pm\sqrt{\frac{-4}{3\kappa_4^{(a)}-6}}, \pm\sqrt{\frac{3\kappa_4^{(a)}+2}{3\kappa_4^{(a)}-6}}\right)$	$\frac{2(3\kappa_4^{(a)}+2)}{3(2-\kappa_4^{(a)})}$	8	$\kappa_4^{(a)} \leq -\frac{2}{3}$

Proof. The proof is similar to the proof of Proposition 1. Indeed, we have $\kappa_4^{(y)} = \kappa_4^{(a)}(g_1^4 + g_2^4) + \left(\kappa_4^{(a)} + 2\right) g_1^2 g_2^2$ and the Lagrangian reads $\mathcal{L} = \kappa_4^{(a)}(g_1^4 + g_2^4) + \left(\kappa_4^{(a)} + 2\right) g_1^2 g_2^2 - \lambda(g_1^2 + g_2^2 - 1)$. Using a computer algebra system, one can then check that (7) is satisfied at 8 points. These points correspond to values of (g_1, g_2) and $\kappa_4^{(y)}$ which are given in the first three rows of Table 1 (precisely those rows for which $g_3 = 0$). Then, (5) is a contrast function if and only if its maximization yields a separating solution, that is if and only if $|\kappa_4^{(a)}| > |\frac{3\kappa_4^{(a)} + 2}{4}|$. The proposition then follows easily. □

3.2 Pairwise and mutual independence
3.2.1 Pairwise independent sources
We now investigate the particular case of pairwise independent sources and introduce the following source vector:

A2. $\mathbf{s} = (s_1, s_2, s_3, s_4)^T$ where s_1, s_2 and s_3 are independent BPSK and $s_4 = s_1 s_2 s_3$.

This case has been considered in [20], where it has been shown that

$$\forall i \in \{1, \ldots, 4\}, \quad \text{Cum } \{s_i, s_i, s_i, s_i\} = -2, \quad (8)$$
$$\text{Cum } \{s_1, s_2, s_3, s_4\} = 1$$

and all other cross-cumulants vanish. The latter cumulant value shows that the sources are mutually dependent; although it can be shown that they are pairwise independent. It should be clear that pairwise independence is not equivalent to mutual independence but in an ICA context, it is relevant to recall the following proposition, which is a direct consequence of Darmois' theorem ([25], p. 294):

Property 1. *Let $\mathbf{s}$ be a random vector with mutually independent components, and $\mathbf{x} = \mathbf{Gs}$. Then the mutual independence of the entries of $\mathbf{x}$ is equivalent to their pairwise independence.*

Based on this proposition, the ICA algorithm in [25] searches for an output vector with pairwise independent components. Let us stress that this holds only if the source vector has *mutually* independent components: pairwise independence is indeed not sufficient to ensure identifiability as we will see in following section.

3.2.2 Pairwise independence is not sufficient
We first have the following preliminary result:

Lemma 3. *Let $y = \mathbf{gs}$ where the vector of sources is defined by A2. Assume that the vector (s_1, s_2, s_3) takes all 2^3*

possible values. If the signal y has values in $\{-1, +1\}$, then $\mathbf{g} = (g_1, g_2, g_3, g_4)$ is either one of the solutions below:

$$\begin{cases} \exists i \in \{1, \ldots, 4\} & g_i = \pm 1, \quad \text{and:} \quad \forall j \neq i, g_j = 0 \\ \exists i \in \{1, \ldots, 4\} & g_i = \pm 1/2, \quad \text{and:} \quad \forall j \neq i, g_j = -g_i \end{cases} \quad (9)$$

Proof. If $y = \mathbf{gs}$, using the fact that $s_i^2 = 1$ for $i = 1, \ldots, 4$, we have with the particular sources given by A2:

$$\begin{aligned} y^2 &= g_1^2 + g_2^2 + g_3^2 + g_4^2 \\ &\quad + 2\left[(g_1 g_2 + g_3 g_4) s_1 s_2 + (g_1 g_3 + g_2 g_4) s_1 s_3 \right. \\ &\quad \left. + (g_2 g_3 + g_1 g_4) s_2 s_3 \right] \end{aligned}$$

Since (s_1, s_2, s_3) take all possible values in $\{-1, 1\}^3$, we deduce from $y^2 = 1$ that the following equations necessarily hold:

$$\begin{cases} g_1^2 + g_2^2 + g_3^2 + g_4^2 = 1 \\ g_1 g_2 + g_3 g_4 = g_1 g_3 + g_2 g_4 = g_2 g_3 + g_1 g_4 = 0 \end{cases} \quad (10)$$

First observe that values given in (9) indeed satisfy (10). Yet, if a polynomial system of N equations of degree d in N variables admits a finite number of solutions[b], then there can be at most d^N distinct solutions. Hence we have found them all in (9), since (9) provides us with 16 solutions for (g_1, g_2, g_3, g_4). □

Using the above result, we are now able to specify the output of classical ICA algorithms when applied to a mixture of sources which satisfy A2.

Constant modulus and contrasts based on fourth order cumulants The constant modulus (CM) criterion is one of the most known criteria for BSS. In the real valued case, it simplifies to:

$$J_{\text{CM}}(\mathbf{g}) \triangleq \mathbb{E}\left\{ (y^2 - 1)^2 \right\} \qquad \text{with: } y = \mathbf{gs} \quad (11)$$

Proposition 3. *For the sources given by A2, the minimization of the constant modulus criterion with respect to $\mathbf{g}$ leads to either one of the solutions given by Equation (9).*

Proof. We know that the minimum value of the constant modulus criterion is zero and that this value can be reached (for $\mathbf{g}$ having one entry being ± 1 and other entries zero). Moreover, the vanishing of the constant modulus criterion implies that $y^2 - 1 = 0$ almost surely and one can then apply Lemma 3. □

A connection can now be established with the fourth-order auto-cumulant if we impose the following constraint:

$$\mathbb{E}\{y^2\} = 1 \quad \text{(or equivalently } \|\mathbf{g}\|^2 = 1 \text{ since } y = \mathbf{gs})$$

$$(12)$$

Because of the scaling ambiguity of BSS, the above normalization can be freely imposed. Under (12), we have $\kappa_4^{(y)} = \mathbb{E}\left\{\left(y^2 - 1\right)^2\right\} - 2$ and minimizing $J_{CM}(\mathbf{g})$ thus amounts to maximizing $-\kappa_4^{(y)}$. Unfortunately, since $\kappa_4^{(y)}$ may be positive or negative, no simple relation between $|\kappa_4^{(y)}|$ and $J_{CM}(\mathbf{g})$ can be deduced from the above equation. Recall that usual results on contrasts do not apply here since source s_4 depends on the others ([1], pp. 83–85). However, we can state:

Proposition 4. *Let $y = \mathbf{g}s$ where the vector of sources is defined by A2. Then, under the constraint (12) ($\|\mathbf{g}\| = 1$), we have:*

(i) The maximization of $\mathbf{g} \mapsto -\kappa_4^{(y)}$ leads to either one of the solutions given by Equation (9).

(ii) $|\kappa_4^{(y)}| \leq 2$ and the equality $|\kappa_4^{(y)}| = 2$ holds true if and only if $\mathbf{g}$ is one of the solutions given in Equation (9).

Proof. Part (i) follows from the arguments given above. In addition, using multilinearity of the cumulants and (8), we have for $y = \mathbf{g}s$:

$$\kappa_4^{(y)} = -2\left(g_1^4 + g_2^4 + g_3^4 + g_4^4\right) + 24\left(g_1 g_2 g_3 g_4\right) \quad (13)$$

The result then follows straightforwardly from the study of the polynomial function in Equation (13). Indeed, optimizing (13) leads to the following Lagrangian:

$$\mathcal{L} = -2\sum_{i=1}^4 g_i^4 + 24\prod_{i=1}^4 g_i - \lambda\left(\sum_{i=1}^4 g_i^2 - 1\right) \quad (14)$$

After solving the polynomial system which cancels the Jacobian of the above expression, one can check that all solutions are such that $|\kappa_4^{(y)}| \leq 2$. Part (ii) of the proposition easily follows. □

3.3 Simulations

3.3.1 Context
We illustrate with a few simulations Propositions 1 and 2. The random variable a in Assumption A1 has been generated as the following mixture of Gaussians:

$$a \sim \frac{1}{2}\mathcal{N}(\mu, \sigma^2) + \frac{1}{2}\mathcal{N}(-\mu, \sigma^2) \quad (15)$$

μ can take any value in $]0, 1[$ and we have set $\sigma^2 = 1 - \mu^2$. The latter choice ensures that $\mathbb{E}\{a^2\} = 1$, whereas we obviously have $\mathbb{E}\{a\} = \mathbb{E}\{a^3\} = 0$. Finally, we have $\kappa_4^{(a)} = -2\mu^4$ and we choose the particular values $\mu = 0.5$ and $\mu = 0.9$, which corresponds respectively to $\kappa_4^{(a)} = -0.125$ and $\kappa_4^{(a)} \approx -1.3122$.

We generated mixtures of the sources given by Assumption A1: we mixed the three sources s_1, s_2, s_3 or the two sources s_1, s_2 only with a matrix $\mathbf{A}$ randomly generated in $\mathbb{R}^{3\times3}$ or $\mathbb{R}^{2\times2}$, respectively.

3.3.2 Algorithm
We used the algorithm CoM2 described in [25] and ([1], Chap. 5). It relies on the following MIMO extension of criterion (5):

$$\mathbf{G} \mapsto \sum_{i=1}^N \left|\kappa_4^{(y_i)}\right|^\alpha, \qquad \alpha \geq 1. \quad (16)$$

The algorithm in [25] first operates a prewhitening and the maximization of the above criterion is performed over the set of orthogonal matrices. From the results of Propositions 1 and 2 we expect the separation results which are given in Table 2, depending on $\kappa_4^{(a)}$ and the number of mixed sources. In Table 2, G is said separating when Above, $\mathbf{G}$ is said separating when $\mathbf{G} = \mathbf{PD}$ where $\mathbf{P}$ is a permutation matrix and $\mathbf{D} = \text{diag}(\pm1, \ldots, \pm1)$ and $\mathbf{G}$ extracts s_3 when $\mathbf{G} = \mathbf{P}\begin{pmatrix} * & * & 0 \\ * & * & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and the elements denoted $*$ have absolute value $1/\sqrt{2}$.

3.3.3 Results
We provide some results illustrating Propositions 1 and 2 through the behavior of the algorithm CoM2. Let us define the following performance criterion:

$$\tau_k = \min_i \left(1 - \frac{|G_{ik}|^2}{\sum_{j=1}^N |G_{ij}|^2}\right)$$

We have $0 \leq \tau_k \leq 1$ and τ_k is close to zero whenever source k is well separated.

We performed 100 Monte-Carlo runs both in the case where the three sources A1 and in the case where only s_1, s_2 are mixed. The average results are provided in Tables 3 and 4. We see in Table 3 that τ_3 is small in all cases, indicating that the source s_3 is indeed separated by the algorithm. On the contrary, one can see in Tables 3 and 4 that τ_1, τ_2 are small only for $\mu = 0.9$: this corresponds to a value of $\kappa_4^{(a)}$ for which the obtained $\mathbf{G}$ is theoretically separating. On the contrary, for $\mu = 0.5$, corresponding to a value of $\kappa_4^{(a)}$ for which the obtained $\mathbf{G}$ does not theoretically separate s_1 and s_2, the values of τ_1 and τ_2 are close to 0.5.

Table 2 Separation result provided by CoM2 for dependent sources satisfying A1

	$\kappa_4^{(a)} \in [\frac{-2}{7}, 2]$	$\kappa_4^{(a)} \notin [\frac{-2}{7}, 2]$
s_1 and s_2 only	**G** mixing, $G_{ij} = \pm\frac{1}{\sqrt{2}}$	**G** separating
s_1, s_2 and s_3	**G** extracts s_3	**G** separating

Table 3 Separation results of randomly driven mixtures of the three sources given by Assumption A1, average values over 100 Monte-Carlo runs

Number of samples	$T = 1000$		$T = 10000$	
parameter of the sources	$\mu = 0.5$	$\mu = 0.9$	$\mu = 0.5$	$\mu = 0.9$
τ_1	0.4995	0.0109	0.5000	0.0051
τ_2	0.5001	0.0161	0.5000	0.0051
τ_3	0.0117	0.0053	2.80e-4	2.76e-5

4 Extended ICA model

We have seen that, depending on the value of $\kappa_4^{(a)}$, the classical optimization criteria in Equations (5) and (16) are not contrasts any longer for the first two sources given by Assumption A1. We now introduce a new statistical model of dependent sources, which consists in a probability mixture of sources. One component of the probability mixture satisfies the requirement of ICA, whereas the other component of the probability mixture is dependent. As an interesting example of dependent sources, we will consider the first two sources defined by A1, where a is the mixture of Gaussians proposed in Section 3.3 with $\mu = 0.5$: this choice is justified by the previous results, which state that such sources cannot be separated by classical algorithms. We show that for our model, the separation is possible based on ICA and on the subset of samples where ICA assumptions are satisfied.

4.1 Latent variables

We first extend ICA methods and relax the independence assumption. The basic idea consists in introducing a hidden process $r(t)$ such that, depending on the particular value of $r(t)$ at instant t, the independence assumption is relaxed at time t. In this article, we will assume that $r(t)$ can take two values only in the set $\{0, 1\}$. Let $\mathbf{r} \triangleq (r(1), \ldots, r(T))$. We assume more precisely:

> **A3.** Conditionally on $\mathbf{r}$, the components $\mathbf{s}(1), \ldots, \mathbf{s}(T)$ of $\mathbf{S}$ at different times are independent and for all $t \in \{1, \ldots, T\}$ we have $\mathbb{P}(\mathbf{s}(t) \mid \mathbf{r}) = \mathbb{P}(\mathbf{s}(t) \mid r(t))$.
> **A4.** Conditionally on $r(t)$, when $r(t) = 0$, the components of $\mathbf{s}(t)$ are mutually independent and non Gaussian, except possibly one of them;

Table 4 Separation results of randomly driven mixtures of mixtures of the first two sources (s_1, s_2) given by Assumption A1, average values over 100 Monte-Carlo runs

Number of samples	$T = 1000$		$T = 10000$	
parameter of the sources	$\mu = 0.5$	$\mu = 0.9$	$\mu = 0.5$	$\mu = 0.9$
τ_1	0.4997	4.36e-4	0.5000	4.72e-5
τ_2	0.5003	4.38e-4	0.5000	4.73e-5

> **A5.** Conditionally on $r(t)$, when $r(t) = 1$, the components of $\mathbf{s}(t)$ are dependent.

One can see that, conditionally on $r(t) = 0$, the source components $\mathbf{s}(t)$ at time t satisfy the usual assumptions required by ICA. In a BSS context, if $\mathbf{r}$ were known, one could easily apply ICA techniques by discarding the time instants where the sources do not satisfy the ICA assumptions. To be more precise, let us define:

$$\mathcal{I}_0 \triangleq \{t \in \{1, \ldots, T\} \mid r(t) = 0\} \tag{17}$$

$$\text{and: } \mathbf{X}_0 \triangleq (\mathbf{x}(t))_{t \in \mathcal{I}_0} \tag{18}$$

The set $\mathcal{I}_0$ is the set of time instants where the components of $\mathbf{s}(t)$ are independent and non Gaussian, except possibly one of them. Then the subset $\mathbf{X}_0$ of the whole set $\mathbf{X}$ of the observations satisfies the assumptions usually required by ICA techniques. The core idea of our method consists in performing alternatively and iteratively an estimation of $\mathbf{B}$ (corresponding to $\mathbf{A}^{-1}$) and of the hidden data $\mathbf{r}$.

4.2 Typical sources separated by the proposed method

The sources that we consider satisfy Assumptions A3., A4., and A5. given previously in Section 4.1. We now detail our particular choices for A4., A5., which have been used in simulations and in the sequel to illustrate the assumptions. These particular choices are denoted hereunder by A4.' and A5.' or A5.".

We have considered two particular examples with $N = 2$ sources. In both cases, Assumption A4. particularizes to:

> **A4.'** When $r(t) = 0$, the components of $\mathbf{s}(t)$ are mutually independent, uniformly distributed on $[-\sqrt{3}, \sqrt{3}]$ (that is: zero-mean and unit variance).

4.2.1 Example 1

As a first example, we particularize Assumption A5. as follows:

> **A5.'** When $r(t) = 1$, then $s_1(t) = a(t)$ and $s_2(t) = \epsilon(t)a(t)$, where $\epsilon(t)$ is an independent BPSK random variable and $a(t)$ is an independent zero-mean, unit-variance random variable.

In simulations, we choose $a(t)$ as a mixture of Gaussians whose distribution is given in Equation (15). A typical realization of a distribution satisfying A3., A4.' and A5.' is illustrated by the simulated values shown in Figure 1a. Conditionally on $r(t) = 1$, the vector $(s_1(t), s_2(t))^{\mathrm{T}}$ corresponds to the first two sources in Assumption A1, which means that for $r(t) = 1$, $(s_1(t), s_2(t))^{\mathrm{T}}$ lie on one of the two bisectors of the horizontal and vertical axes. According to the discussion in Sections 3.1 and 3.3, for $\kappa_4^{(a)} \in [-\frac{2}{7}, 2]$, algorithms based on the kurtosis contrast
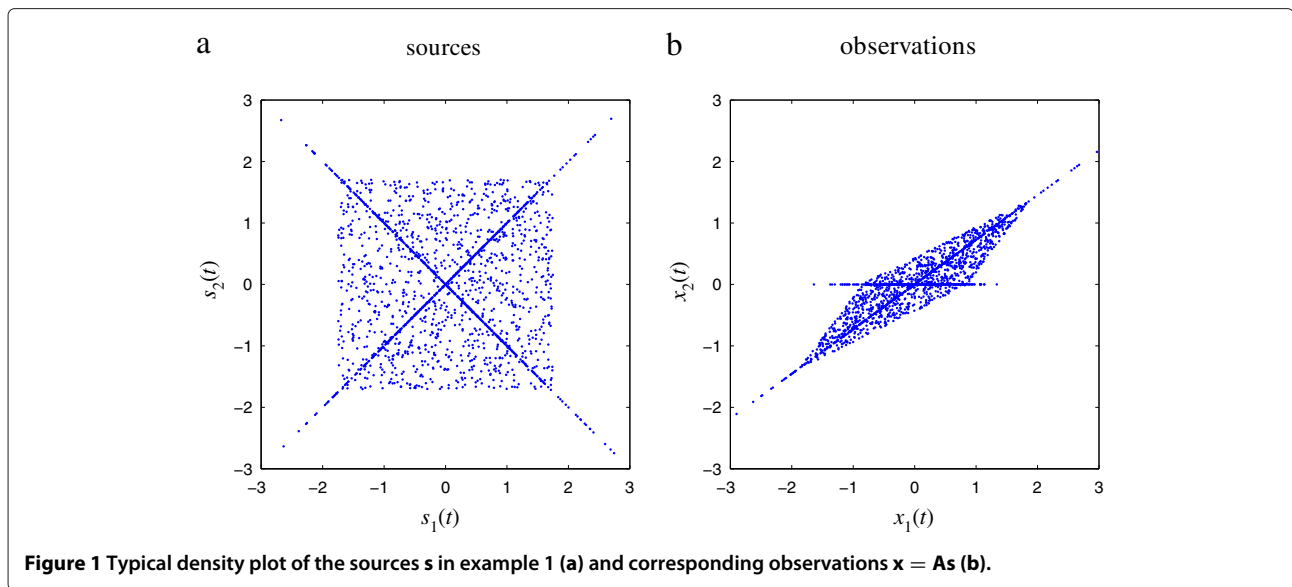
Figure 1 Typical density plot of the sources s in example 1 (a) and corresponding observations x = As (b).

functions (such as CoM2) should not separate any linear mixture of $(s_1(t), s_2(t))$T based on the samples where $r(t) = 1$.

On the contrary, considering the samples $\mathbf{X}_0$ only amounts to removing the set of dependent points that lie on the two bisectors of Figure 2a. In such a case, the remaining samples in $\mathbf{X}_0$ satisfy the usual requirement for ICA and any ICA algorithm should succeed in separating a linear mixture of the sources.

4.2.2 Example 2
The previous example is an extreme case where, conditionally on $r(t) = 1$, either $s_1(t) = s_2(t)$ or $s_1(t) = -s_2(t)$. Thus, the set of samples where $s_1(t)$ and $s_2(t)$ are dependent lies on the two bisectors of the horizontal and vertical axes. We considered in this example the case where, conditionally on $r(t) = 1$, the components of $(s_1(t), s_2(t))^{\mathrm{T}}$ are dependent but have a continuous joint density. We particularize Assumption A5. as follows:

A5." Let $u_1 \sim \mathcal{N}(0, \sigma_\lambda)$ and: $u_2 \sim \mathcal{L}(\lambda)$ be independent random variables where λ is a positive parameter and $\sigma_\lambda^2 = 2(1 - \frac{1}{\lambda^2})$. When $r(t) = 1$, $\mathbb{P}(\mathbf{s}(t) \mid r(t) = 1)$ is such that:

$$\begin{pmatrix} s_1(t) \\ s_2(t) \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} u_1 + u_2 \\ u_1 - u_2 \end{pmatrix}$$

To say it differently, the two components of

$$\mathbf{u} = \frac{1}{\sqrt{2}} \begin{pmatrix} s_1(t) + s_2(t) \\ s_1(t) - s_2(t) \end{pmatrix} \tag{19}$$
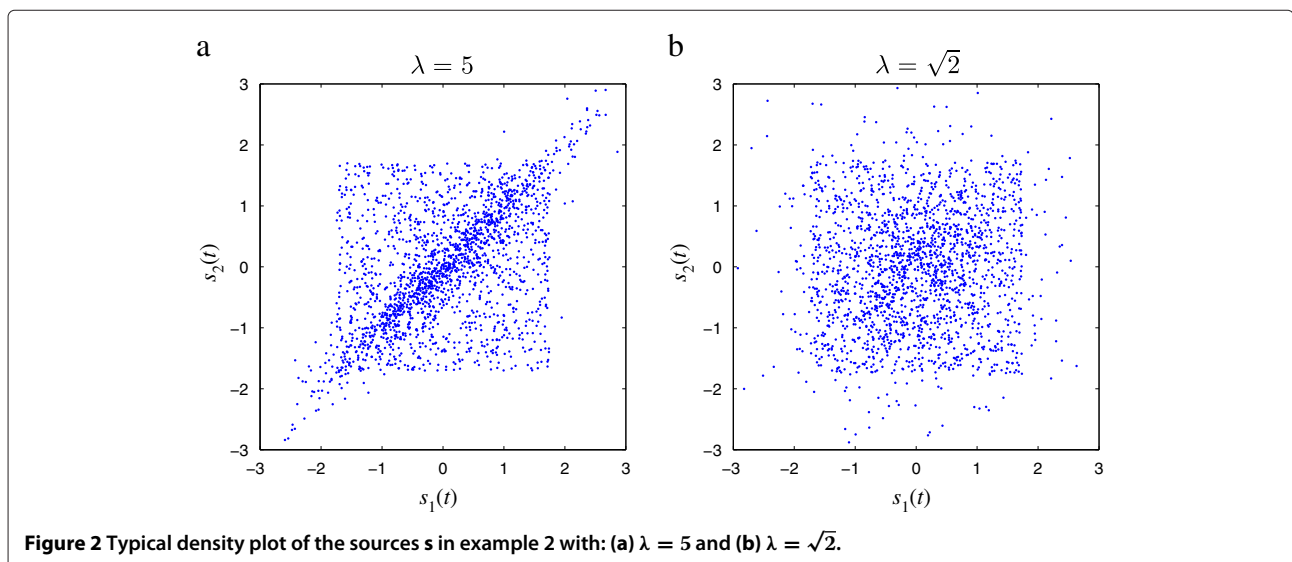
Figure 2 Typical density plot of the sources s in example 2 with: (a) $\lambda = 5$ and (b) $\lambda = \sqrt{2}$.

are independent, the first one being Gaussian and the second one being Laplace distributed. One can verify that the choice of σ_λ^2 ensures that, conditionally on $r(t) = 1$, the sources are unit-variance. Such a distribution density is illustrated by simulated values in Figure 2a with $\lambda = 5$ and in Figure 2b with $\lambda = \sqrt{2}$. Considering $\mathbf{X}_0$ only amounts to removing the cloud set of dependent points on the distributions. Visually, this seems much more difficult in Figure 2b than in Figure 2a, and much more difficult in this example than in Example 1 in Figure 1a. We will discuss further in Section 7 the influence of a good knowledge of $\mathbf{r}$: surprisingly, it is not necessarily crucial in our method to have a good knowledge of $\mathbf{r}$.

5 Separation method for the extended ICA model

5.1 Complete and incomplete data

Let us denote by $\theta = (\mathbf{B}, \eta)$ the set of parameters to be estimated from the data: in this notation, we stress that θ consists of the separating matrix $\mathbf{B}$ and of the parameter vector η which characterize the distribution of $\mathbf{r}$. Let us call $(\mathbf{r}, \mathbf{X})$ the set of *complete* data, whereas $\mathbf{X}$ alone is the set of *incomplete* data: since $\mathbf{r}$ is a hidden process, the model described in Section 4 corresponds to the situation where only incomplete data is available for estimation of the searched parameters θ. Note that the adjective *blind* is used to emphasize that $\mathbf{S}$ is unavailable, whereas *incomplete* emphasizes that $\mathbf{r}$ is unavailable.

5.2 Iterative conditional estimation

Iterative conditional estimation (ICE) is an iterative estimation method that applies in the context of incomplete data and that has been proposed in the problem of image segmentation [24,29,30].

Another well-known iterative estimation technique is the expectation-maximization (EM) algorithm, which is based on the maximum likelihood estimator. Contrary to EM, the underlying estimator in ICE can be of any kind. This makes ICE more widely applicable in cases where the likelihood computation or maximization becomes intractable [29,31], for example in the non gaussian case. In the case where the maximum likelihood is chosen as the underlying estimator, ICE show similarities with EM (see e.g., [32] for an experimental comparative study). It has been proven in [33] that in the case of distributions belonging to the exponential family, EM is one particular case of ICE. Finally, the interest of ICE and its convergence in the case of data that are probability mixtures has been showed in [34].

We now shortly describe ICE. The prerequisites in order to apply ICE are the following:

- there exist an estimator from complete data $\hat{\theta}(\mathbf{r}, \mathbf{X})$,
- one is able either to calculate $\mathbb{E}\{\hat{\theta}(\mathbf{r}, \mathbf{X}) \mid \mathbf{X}; \theta\}$ or to draw random variables according to $\mathbb{P}(\mathbf{r} \mid \mathbf{X}; \theta)$.

Starting from an initial guess of the parameters, the method consists in finding iteratively a sequence of estimates of θ, where each estimate is based on the previous one. More precisely, if $\hat{\theta}^{[0]}$ is the first guess, the sequence of ICE estimates is defined by:

$$\hat{\theta}^{[q]} = \mathbb{E}\{\hat{\theta}(\mathbf{r}, \mathbf{X}) \mid \mathbf{X}; \hat{\theta}^{[q-1]}\}, \tag{20}$$

where $\mathbb{E}\{. \mid \mathbf{X}; \hat{\theta}^{[q-1]}\}$ denotes the expectation conditionally on $\mathbf{X}$ and with parameter values $\hat{\theta}^{[q-1]}$. If the above conditional expectation cannot be computed, it can be replaced by a sample mean, that is (20) can be replaced by:

$$\hat{\theta}^{[p]} = \frac{1}{K} \sum_{k=1}^{K} \hat{\theta}(\mathbf{r}^{(k)}, \mathbf{X}), \tag{21}$$

where $K \in \mathbb{N}^*$ is fixed and each $\mathbf{r}^{(k)}$ is drawn according to the a posteriori law $\mathbb{P}(\mathbf{r} \mid \mathbf{X}; \hat{\theta}^{[q-1]})$. Note that if θ is vector-valued, (20) can be used for those components for which it can be computed, and (21) can be used otherwise.

Remark that the two conditions requested in order to apply ICE are very weak, which is the reason for our interest in ICE. In fact concerning the first one, there would be no hope to perform incomplete data estimation if no complete data estimator exists, whereas the second requirement consists only in being able to simulate random values according to the a posteriori law.

5.3 Applicability of ICE and assumed distributions

In this section, we give details about how the two conditions given in 5.2 for applicability of ICE are fulfilled.

First, as explained in Section 4.1, knowing $\mathbf{r}$ provides an easy way of estimating a separating matrix by considering as in (18) the subset $\mathbf{X}_0$ of the samples. A complete data estimator $\hat{\theta}(\mathbf{r}, \mathbf{X})$ hence exists.

To use ICE, one should additionally know the law $\mathbb{P}(\mathbf{r} \mid \mathbf{X}; \eta)$. Since $\mathbf{X} = \mathbf{AS}$, we have

$$\mathbb{P}(\mathbf{r} \mid \mathbf{X}; \eta) = \frac{\mathbb{P}(\mathbf{r}, \mathbf{X}; \eta)}{\mathbb{P}(\mathbf{X}; \eta)} = \frac{\mathbb{P}(\mathbf{r}, \mathbf{S}; \eta)}{\mathbb{P}(\mathbf{S}; \eta)} \tag{22}$$

and $\mathbb{P}(\mathbf{r} \mid, \mathbf{X}; \eta)$ is identical to the law $\mathbb{P}(\mathbf{r} \mid \mathbf{S}; \eta)$, where $\mathbf{S} = \mathbf{A}^{-1}\mathbf{X} = \mathbf{BX}$. An expression for the latter law is available if a model is assumed for $\mathbb{P}(\mathbf{S}, \mathbf{r}; \eta)$. Importantly, the chosen model used in the ICE estimation algorithm can be different from the model followed by the simulated data that are processed by the algorithm. This is crucial, because actual distributions are generally unknown in practical BSS problems. In particular, we here choose a model which follows Assumptions A3., A4., and A5., but which is different from the particular Assumptions A4.', A5.', and A5." in Section 4.2. More precisely, translating

Assumption A3. only, the joint distribution of $(\mathbf{r}, \mathbf{S})$ under the parameter value η reads:

$$\mathbb{P}(\mathbf{r}, \mathbf{S}; \eta) = \mathbb{P}(\mathbf{r}; \eta) \prod_{t=1}^{T} \mathbb{P}(\mathbf{s}(t) \mid r(t); \eta) \qquad (23)$$

The above equation holds both for the assumed distribution and for the distribution of the simulated data. On the contrary, the expressions of $\mathbb{P}(\mathbf{r}; \eta)$ and $\mathbb{P}(\mathbf{s}(t) \mid r(t); \eta)$ that are assumed in ICE differ from the ones of the simulated data and they are given in the next paragraphs.

5.3.1 Assumed $\mathbb{P}(\mathbf{s}(t) \mid r(t))$ in ICE
First, note that in the following, similarly to the real data distribution, $\mathbb{P}(\mathbf{s}(t) \mid r(t); \eta) = \mathbb{P}(\mathbf{s}(t) \mid r(t))$ does not depend on the parameters η to be estimated.

Experimentally, we observed that, for robustness reasons, the distribution $\mathbb{P}(\mathbf{s}(t) \mid r(t))$ assumed in ICE should not have a bounded support or be too specific. For this reason, we assumed the following distributions:

- $\mathbb{P}(\mathbf{s}(t) \mid r(t) = 0) \sim \mathcal{N}(0, 1)$,
- $\mathbb{P}(\mathbf{s}(t) \mid r(t) = 1)$ is such that both components of $\mathbf{u}$ defined in (19) have the distribution $\frac{1}{2}\mathcal{L}(\lambda) + \frac{1}{2}\mathcal{N}\left(0, \sigma_\lambda^2\right)$.

A typical realization of the distribution of $(s_1(t), s_2(t))^{\mathrm{T}}$ assumed in ICE is given in Figure 3. This distribution is of course different from the ones of the simulated data in Figures 1 and 2. However, simulations will confirm that it is a reasonable choice.

Conditionally on $r(t) = 0$, $(s_1(t), s_2(t))^{\mathrm{T}}$ is assumed an independent, zero-mean, unit-variance Gaussian vector: this is an uninformative distribution having a density with non bounded support. Conditionally on $r(t) = 1$,

each component of $\mathbf{u}$ in the assumed distribution is a mixture of a Laplace and a Gaussian law. Experimentally, and from the observation of Figures 1 and 3, the latter choice seems a relevant approximation of the data model from Example 1 in Section 4.2.1. The parameter λ should make a compromise between a good fit to the data (λ large) and robustness of the method (λ small). Also, the assumed conditional distribution $\mathbb{P}(\mathbf{s}(t) \mid r(t) = 1)$ is invariant with respect to any permutation and sign change. Such a symmetry is necessary in our method because ICA algorithms leave permutation ambiguities. For this reason, the same distribution has been considered to model the data source signals generated in Example 2 (Section 4.2.2).

5.3.2 Assumed $\mathbb{P}(\mathbf{r}; \eta)$ in ICE
We propose two different models for the latent process $\mathbf{r}$.
I.i.d. latent process
The simplest case is when $\mathbf{r}$ is an i.i.d. Bernoulli process, that is η is a scalar parameter in $[0, 1]$ and:

$$\mathbb{P}(\mathbf{r}; \eta) = \prod_{t=1}^{T} \mathbb{P}(r(t); \eta) \qquad (24)$$
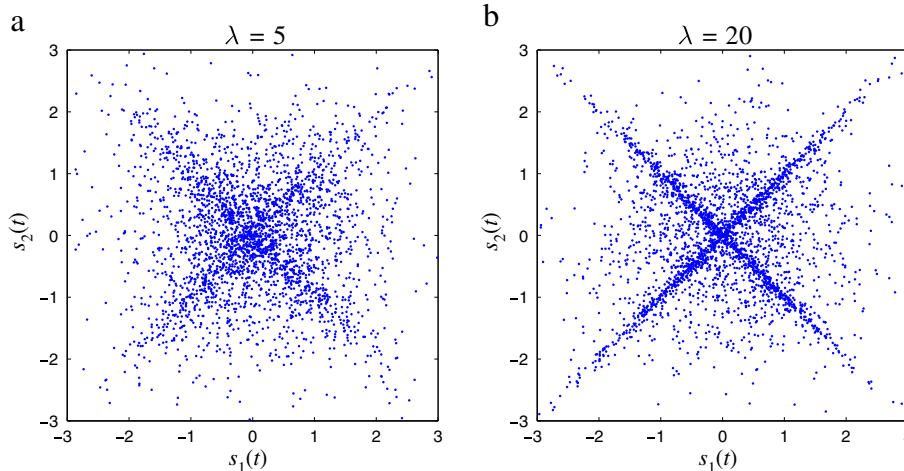
with for all $t \in \{1, \ldots, T\}$:

$$\mathbb{P}(r(t) = 0) = \eta \ \text{ and } \ \mathbb{P}(r(t) = 1) = 1 - \eta \qquad (25)$$

Markov latent process
We can alternatively consider that $\mathbf{r}$ is a stationary Markov Chain, that is:

$$\mathbb{P}(\mathbf{r}) = \mathbb{P}(r(1)) \prod_{t=2}^{T} \mathbb{P}(r(t) \mid r(t-1)),$$

where $\mathbb{P}(r(t) \mid r(t-1))$ is given by a transition matrix independent of t. In this case, the parameters η consist

Figure 3 Typical density plot of sources assumed in the ICE algorithm with: (a)$\lambda = 5$ **and (b)**$\lambda = 20$**.**

of the different transition probabilities and the initial probabilities. The main advantage of considering a Markov model is that the posterior transitions $\mathbb{P}(r(t) \mid r(t-1), \mathbf{X}; \eta)$ can be calculated by an efficient forward-backward algorithm [35]. A sampling of the hidden process according to the posterior law $\mathbb{P}(\mathbf{r} \mid \mathbf{X}; \eta)$ is hence possible, making the ICE method applicable [29].

6 Combined ICA/ICE separation algorithm

In this section, we first detail our separation algorithm which combines ICA and ICE. We then interpret ICE in terms of random variable generation.

6.1 Algorithm

We denote by ICA one among all possible ICA algorithms [1] such as JADE [26], CoM2 [25], FastICA [27], etc. Given a set of observation samples $\mathbf{X}$, the separating matrix estimated by the ICA algorithm is denoted by $\mathbf{B} = \mathtt{ICA}(\mathbf{X})$. With these notations, a complete data estimator of the separating matrix is provided by $\mathbf{B} = \mathtt{ICA}(\mathbf{X}_0)$, where $\mathbf{X}_0$ has been defined in Equation (18). Our algorithm consists in an ICE estimation of the parameters $\theta = (\mathbf{B}, \eta)$. The parameters η which characterize $\mathbf{r}$ are estimated according to (20), whereas the separating matrix $\mathbf{B}$ is estimated according to (21) with $K = 1$. Here is a summary of the algorithm:

Initialize the parameters $\hat{\theta}^{[0]} = \left(\hat{\mathbf{B}}^{[0]}, \hat{\eta}^{[0]} \right)$. For $q = 1, 2, \ldots, q_{\max}$, repeat:

1. calculate $\mathbb{P}\left(\mathbf{r} \mid \mathbf{X}; \hat{\theta}^{[q]} \right)$ and draw $\hat{\mathbf{r}}^{[q]}$ according to this distribution,
2. update the separating matrix:

   - set: $\hat{\mathcal{I}}_0^{[q]} = \left\{ t \mid \hat{r}^{[q]}(t) = 0 \right\}$ and $\hat{\mathbf{X}}_0^{[q]} = (x(t))_{t \in \hat{\mathcal{I}}_0^{[q]}}$
   - $\hat{\mathbf{B}}^{[q+1]} = \mathtt{ICA}\left(\hat{\mathbf{X}}_0^{[q]} \right)$.

3. update the parameters η of the process $\mathbf{r}$.

Steps 1 and 3 of the above algorithm are further detailed in the following sections. Note that, if the parameters η are available as an additional information, they need not be estimated and step 3 of the algorithm is not necessary.

6.2 Details in the case of an i.i.d. latent process

We here detail the three steps of our algorithm in the case where $\mathbf{r}$ is i.i.d. As we will see, in this case, ICE is akin to generating a set of random samples that satisfy the usual assumptions of ICA with an accept-reject method.

6.2.1 Step 1

In the case where $\mathbf{r}$ is i.i.d., we have from (22), (23), and (24):

$$
\begin{aligned}
\mathbb{P}(\mathbf{r} \mid \mathbf{X}; \eta) &= \prod_{t=1}^{T} \frac{\mathbb{P}(r(t); \eta)\mathbb{P}(\mathbf{s}(t) \mid r(t))}{\mathbb{P}(\mathbf{s}(t))} \\
&= \prod_{t=1}^{T} \mathbb{P}(r(t) \mid \mathbf{s}(t); \eta),
\end{aligned}
$$

where $\mathbb{P}(r(t); \eta)$ has been given in Equation (25) and $\mathbb{P}(\mathbf{s}(t) \mid r(t))$ has been described in Section 5.3.1. Writing $\alpha_t = \mathbb{P}(r(t) = 0 \mid \mathbf{X}; \eta)$ (and hence $1 - \alpha_t = \mathbb{P}(r(t) = 1 \mid \mathbf{X}; \eta)$), it follows from the above equation that in step 1, all samples of $\hat{\mathbf{r}}^{[q]}$ are independent and such that

$$
\begin{cases}
\hat{r}^{[q]}(t) = 0 & \text{with probability } \alpha_t \\
\hat{r}^{[q]}(t) = 1 & \text{with probability } 1 - \alpha_t.
\end{cases}
$$

6.2.2 Step 2: accept-reject random variable generation

One can see that, when selecting $\hat{\mathbf{X}}_0^{[q]}$ in the second step of our algorithm, some samples are kept, others are thrown away. A close parallel can be drawn with random variable generation by the acceptance-rejection method.

For clarity, let us define $\mathbb{P}_0(\mathbf{s}(t)) = \mathbb{P}(\mathbf{s}(t) \mid r(t) = 0)$ and $\mathbb{P}_1(\mathbf{s}(t)) = \mathbb{P}(\mathbf{s}(t) \mid r(t) = 1)$. At time t, according also to Equations (25) and (23), the distribution of $\mathbf{s}(t)$ is given by:

$$
\eta \mathbb{P}_0(\mathbf{s}(t)) + (1 - \eta)\mathbb{P}_1(\mathbf{s}(t))
$$

We have the following lemma:

Lemma 4. *Let $\mathbf{s}$ be a random variable (or random vector) with probability distribution given by*

$$
\mathbf{s} \sim \eta \mathbb{P}_0(\mathbf{s}) + (1 - \eta)\mathbb{P}_1(\mathbf{s})
$$

Let $\hat{r} \in \{0, 1\}$ be a binary random variable with probability distribution such that:

$$
\mathbb{P}(\hat{r} = 0 \mid \mathbf{s}) = \frac{\eta \mathbb{P}_0(\mathbf{s})}{\eta \mathbb{P}_0(\mathbf{s}) + (1 - \eta)\mathbb{P}_1(\mathbf{s})} \tag{26}
$$

$$
\mathbb{P}(\hat{r} = 1 \mid \mathbf{s}) = \frac{(1 - \eta)\mathbb{P}_1(\mathbf{s})}{\eta \mathbb{P}_0(\mathbf{s}) + (1 - \eta)\mathbb{P}_1(\mathbf{s})} \tag{27}
$$

Then, the conditional distribution $\mathbb{P}(\mathbf{s} \mid \hat{r} = 0)$ is $\mathbb{P}_0(\mathbf{s})$.

Proof. One can write the joint probability distribution $\mathbb{P}(\mathbf{s}, \hat{r}) = \eta \delta(\hat{r} = 0)\mathbb{P}_0(\mathbf{s}) + (1 - \eta)\delta(\hat{r} = 1)\mathbb{P}_1(\mathbf{s})$ and the result follows by conditioning. □

Lemma 4 relates our algorithm to the accept-reject algorithm for random variable generation. Indeed, drawing $\hat{r}$ and conditioning on $\hat{r} = 0$ to get $\mathbb{P}(\mathbf{s} \mid \hat{r} = 0)$ is performed in our algorithm by drawing $\hat{r}^{[q]}(t)$ and keeping only those samples $\mathbf{s}(t)$ of $\mathbf{S}$, where $\hat{r}^{[q]}(t) = 0$, that is, it amounts to rejecting those samples where $\hat{r}^{[q]}(t) = 1$. In other terms, through the ICE algorithm, the target

distribution $\mathbb{P}_0(\mathbf{s})$ is generated using the instrumental distribution $g(\mathbf{s}) = \eta\mathbb{P}_0(\mathbf{s}) + (1 - \eta)\mathbb{P}_1(\mathbf{s})$. Samples from this instrumental distribution $g(\mathbf{s})$ are given by the data themselves. By doing so, we obtain a set of data following the distribution $\mathbb{P}_0(\mathbf{s})$. Since $\mathbb{P}_0(\mathbf{s}(t)) = \mathbb{P}(\mathbf{s}(t) \mid r(t) = 0)$ and according to Assumption A4., this is precisely the distribution under which the ICA algorithm is applicable.

Remark 3. It is known that the probability distribution of $\mathbf{s}$ in Lemma 4 can be seen as the marginal of $(r, \mathbf{s})$, where r is a Bernoulli process and the conditional laws of $\mathbf{s}$ knowing r are given by $\mathbb{P}_0$ and $\mathbb{P}_1$. However, $\hat{r}$ is drawn independently of r and in particular, $\hat{r}$ is different from r. It means that in our algorithm, the original latent process $\mathbf{r}$ and the ICE sampling $\hat{\mathbf{r}}^{[q]}$ may be quite different, although the selected samples are distributed following $\mathbb{P}_0$. This will be illustrated in Section 7.1.2.

6.2.3 Step 3

A complete data estimator of the parameter η is given by the empirical frequency $\hat{\eta} = \frac{1}{T}\sum_{t=1}^{T}\delta(r(t) = 0)$. Equation (20) then yields the following update rule for the parameter η:

$$\hat{\eta}^{[q+1]} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{P}\left(r(t) = 0 \mid \mathbf{x}(t), \hat{\theta}^{[q]}\right) \qquad (28)$$

6.3 Markov latent process

We here detail the three steps of our algorithm in the case where $\mathbf{r}$ is a Markov process.

6.3.1 Step 1

Here, we assume in the ICE part of the procedure that $\mathbf{r}$ is a Markov process. Then, the posterior probability $\mathbb{P}(\mathbf{r} \mid \mathbf{X}; \hat{\theta}^{[q]})$ is Markov, and its transitions can be calculated according to the forward-backward or Baum-Welch algorithm. At each step q of the algorithm, $\hat{\mathbf{r}}^{[q]}$ is then generated according to $\mathbb{P}(\mathbf{r} \mid \mathbf{X}; \hat{\theta}^{[q]})$ as a non-stationary Markov chain. It is out of the scope to further detail these well-known procedures and the reader is referred to [29,30,35] and references therein for more explanations.

6.3.2 Step 2

It is unchanged, although the interpretation as an accept-reject random variable generation does not hold any longer.

6.3.3 Step 3

From the values of $\mathbb{P}(r(t) \mid \mathbf{X}, \hat{\theta}^{[q]})$ and $\mathbb{P}(r(t), r(t + 1)) \mid \mathbf{X}, \hat{\theta}^{[q]})$, formulas similar to Equation (28) yield the estimated probabilities for $r(t)$ and the pairs $(r(t), r(t + 1))$. The transition probabilities are then obtained by the relation $\mathbb{P}(r(t + 1)) \mid r(t), \mathbf{X}, \hat{\theta}^{[q]}) = \frac{\mathbb{P}(r(t),r(t+1)) \mid \mathbf{X},\hat{\theta}^{[q]})}{\mathbb{P}(r(t) \mid \mathbf{X},\hat{\theta}^{[q]})}$. The reader is again referred to [29,30,35] for more details.

7 Simulations

We performed several simulations for different numbers of samples, respectively, $T = 1000$, $T = 5000$, $T = 10000$. We have set $q_{max} = 30$ in our algorithm: this value has been determined empirically by testing values of q_{max} up to 150. For values greater than $q_{max} = 30$, no significant quality improvement has been observed and this choice seems hence satisfactory as far as ICE convergence is concerned. Although highly interesting, a more detailed analysis of the convergence speed/rate of our algorithm is out of the scope of the article. The separation quality criterion considered is the mean square error (MSE): the provided value is a mean of the MSE over all sources. Additionally, we considered the empirical segmentation rate provided by the last sampling $\hat{\mathbf{r}}^{[q_{max}]}$ in Step 1 of the ICE procedure. We define this segmentation rate as follows:

$$\Upsilon_{ICE} = \frac{1}{T}\sum_{t=1}^{T}\delta\left(\hat{r}^{[q_{max}]}(t) = r(t)\right). \qquad (29)$$

This corresponds to the proportion of time samples which are correctly classified as corresponding either to independent ($r(t) = 0$) sources or to dependent ($r(t) = 1$) sources. In all cases, the source signals and the mixing matrix have been randomly generated.

The classical algorithm used in our simulations, which has been denoted ICA in Section 6 was CoM2 [25]. For comparison, we considered as a worst case reference the result when CoM2 is applied to the data, simply ignoring that the ICA assumptions are violated at some sample times. On the other hand, as an optimal best case reference, we considered the complete data situation, that is when $\mathbf{r}$ is available and the separation is performed based on $\mathbf{X}_0$.

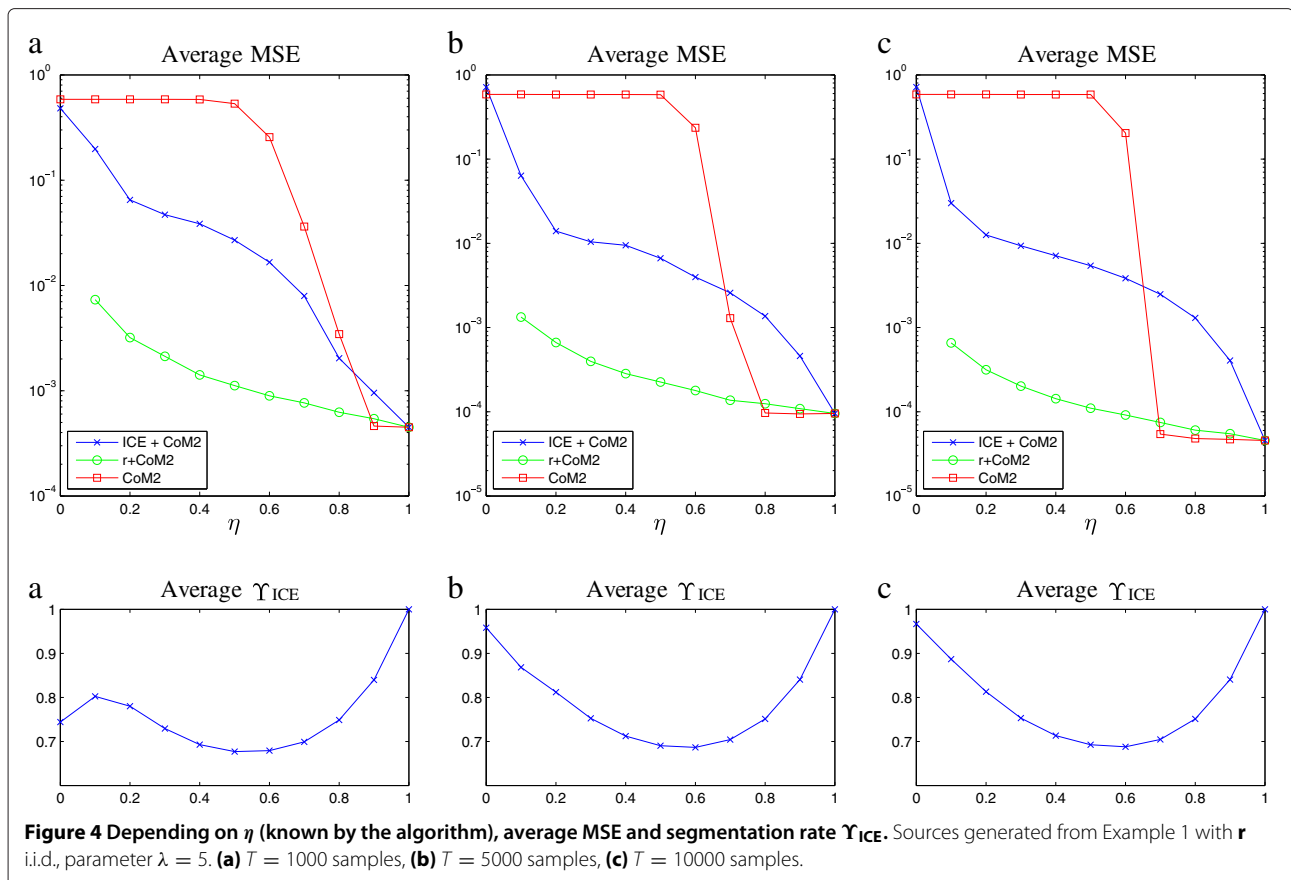In our simulations, we generated data according to the following models:

- sources satisfying **Example 1** in Section 4.2. We chose the process $a(t)$ as a mixture of Gaussians as defined in Equation (15), with $\mu = 0.5$. According to Section 3, the CoM2 algorithm, if applied on the dependent data only, is not successful in separating the sources.
- sources satisfying **Example 2** in Section 4.2. In this case, we choose $\lambda = \sqrt{2}$.

7.1 I.i.d. latent process with known η

We first provide some simulations in the case where $\mathbf{r}$ is an i.i.d. Bernoulli process with known η parameter (see Equation (24)).

7.1.1 Example 1

Separation results for sources generated according to Example 1 with $\mathbf{r}$ i.i.d. are gathered in Figures 4 and 5. Values in Figure 4 are average values over 1000

Figure 4 Depending on η (known by the algorithm), average MSE and segmentation rate Υ_{ICE}. Sources generated from Example 1 with **r**
i.i.d., parameter $\lambda = 5$. **(a)** $T = 1000$ samples, **(b)** $T = 5000$ samples, **(c)** $T = 10000$ samples.

Monte-Carlo realizations whereas values in Figure 5 are
the individual results for 100 Monte-Carlo realizations.

In Figure 4, the MSE values have been plotted depending
on $\eta \in [0, 1]$ for $T = 1000$ (a), $T = 5000$ (b), $T = 10000$
(c) samples. Note that for $\eta = 0$, all samples of the data
sources are dependent and the complete data separation is
therefore not applicable. Similarly, when setting η to zero
in the ICE algorithm, all samples are necessarily classi-
fied as dependent and our method is not applicable. For
this reason, when $\eta = 0$ has been used in the simulated
sources we have set η to 0.1 in the algorithm.

Naturally, the complete data case provides the best
results: when η tends to zero, the number of samples in $\mathbf{X}_0$
decreases, which explains the increasing MSE for smaller
η values. On the other hand, ignoring the dependence
provides results that are unacceptable for η smaller than
0.7, approximately. It seems however that CoM2 is able
to separate the particular sources considered for η greater
then 0.7, approximately: from the results in Section 3,
the existence of such a limit value above which CoM2
is successful seems quite natural. For η smaller than 0.7
the proposed method provides a very significant improve-
ment in terms of separation quality. The classification rate
Υ_{ICE} has also been plotted in Figure 4 and one can observe
that samples of dependent/independent sources are quite

well classified, which corroborates the good MSE values
obtained with our method.

Depending on the parameter λ, the source distribu-
tion assumed in the ICE estimation and described in
Section 5.3.1 is closer to or farther from the real source
distribution of Example 1, which has part of its samples
lying on the two bisectors (see the comparison between
Figure 1a and Figure 3a,b). In Figure 5, we tested the influ-
ence of the parameter λ for a fixed value of $\eta = 0.5$. The
MSE of 100 Monte-Carlo realizations have been plotted
for $T = 1000$ and $T = 10000$ samples. The correspond-
ing segmentation rate Υ_{ICE} has been plotted in the lower
part of the figures. First, it can be seen in Figure 5 that the
separation results of the Monte-Carlo realizations can be
clearly separated in two groups:

- for a minority of cases, the segmentation rate Υ_{ICE} is
  close to 0.5, meaning that the procedure completely
  fails to classify the latent process **r**. In such a case,
  the corresponding MSE is important (around 0.5)
  and separation is unsuccessful. This situation occurs
  more often for large values of λ: a greater value of λ
  consequently implies less robustness of our method.
- for most of the realizations, Υ_{ICE} is significantly
  greater than 0.5, indicating that the procedure
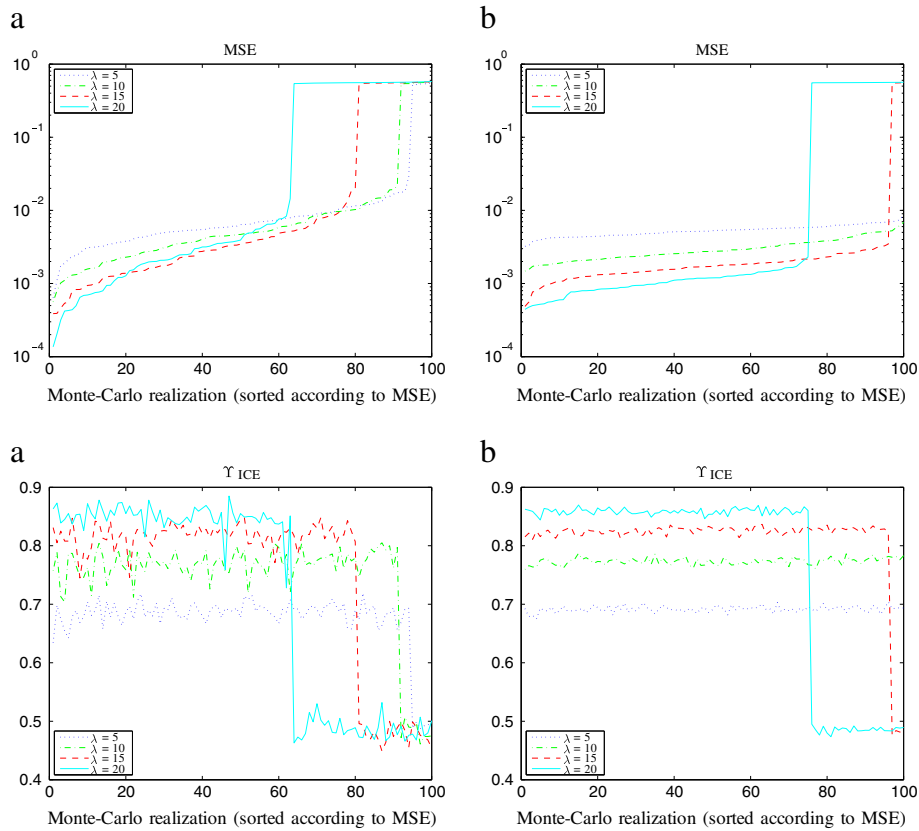
Figure 5 For different values of λ, MSE of 100 Monte-Carlo realizations sorted by increasing values and corresponding segmentation rate Υ_{ICE}**.** Sources generated from Example 1 with **r** i.i.d. $\eta = 0.5$ (known by the algorithm), **(a)** $T = 1000$ samples, **(b)** $T = 10000$ samples.

succeeds in classifying approximately the latent process **r**. In such a case, the corresponding MSE is low and the separation is successful. One can see that, in such a case where the separation is successful, a greater value of λ comes with a greater Υ_{ICE}, hence a better segmentation of **r** and a lower MSE, hence a better separation quality.

In conclusion, the choice of λ should be a compromise between good separation quality (in case of success) and robustness in order to limit the rate of separation failure.
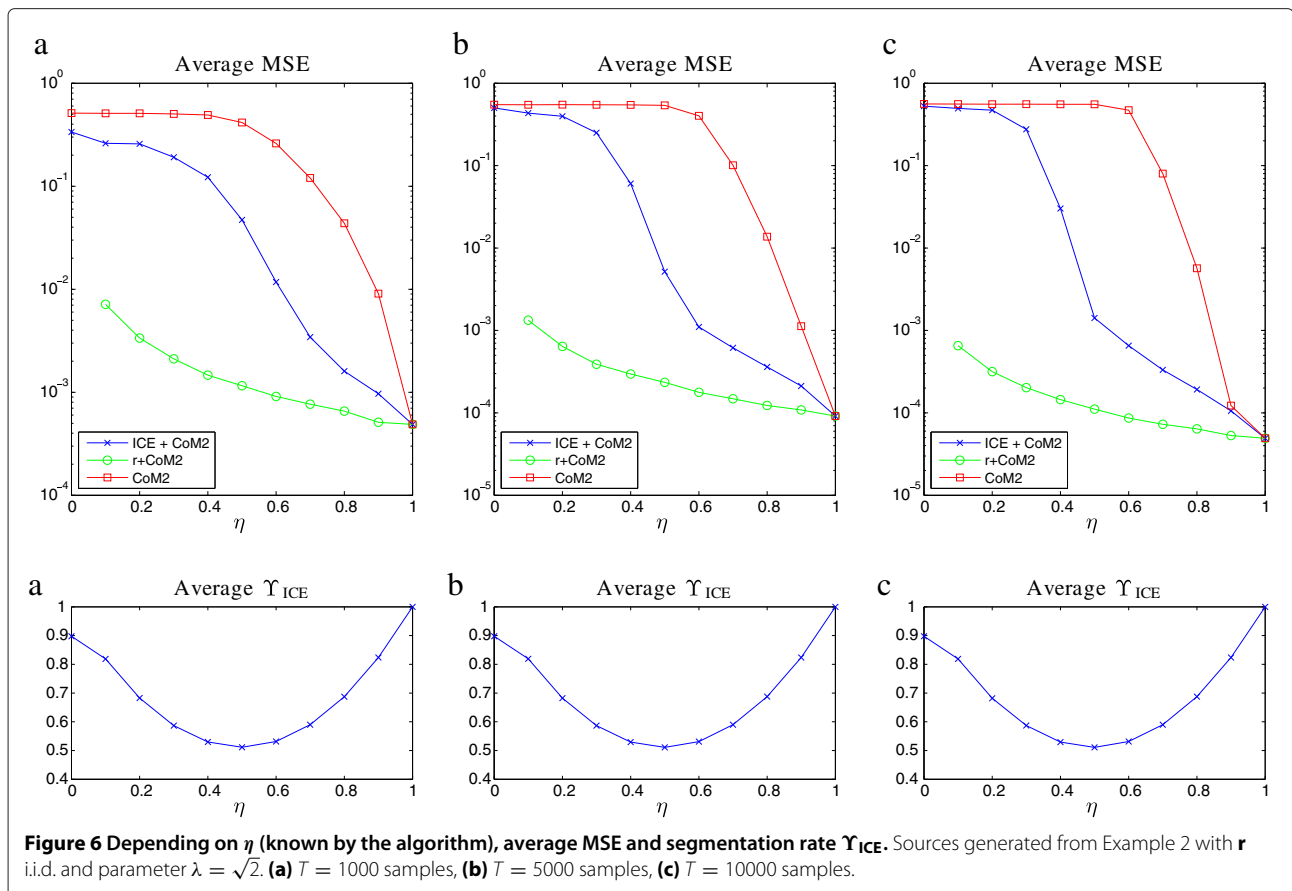
7.1.2 *Example 2*
The results for source signals following Example 2 (see Section 4.2.2 with $\lambda = \sqrt{2}$) are given in Figure 6. Similarly to the previous section, for $T = 1000$ (a), $T = 5000$ (b), $T = 10000$ (c) samples and depending on $\eta \in [0, 1]$ the MSE values have been plotted in the top graph and the segmentation rate Υ_{ICE} in the bottom graph.

As expected, the complete data estimation based on $\mathbf{X}_0$ of the separating matrix provides again the best results. In comparison with the case where the dependence of the sources is ignored, our method provides a very significant improvement, in particular for η lying between 0.4

and 0.8, approximately. Very interestingly, as witnessed in Figure 6, the good performance in terms of separation and MSE is obtained with very poor performance in terms of classification of the latent process **r**: note in particular that for $\eta = 0.5$, we have a low MSE whereas Υ_{ICE} is close to 0.5. The poor classification can be easily understood when comparing Figure 2b with Figure 1a. The important point in this observation is that a good classification of the latent process **r** is a sufficient but not a necessary condition for good ICA estimation. This is in contrast with the former example but is fully justified by the interpretation of Step 2 of the algorithm in Section 6.2. Indeed, according to Lemma 4, the ICE part of our algorithm selects a set of samples $\hat{\mathbf{X}}_0$ which is such that its distribution satisfies the usual ICA assumptions, although the classification may well be very poor.

7.2 I.i.d. latent process with unknown η
We now consider the case where **r** is i.i.d. but the parameter η in Equation (25) is unknown. In this case, η is estimated by the ICE algorithm. The results have been averaged over 1000 Monte-Carlo realizations and are gathered in Table 5 for both sources from Example 1

Figure 6 Depending on η (known by the algorithm), average MSE and segmentation rate Υ_{ICE}. Sources generated from Example 2 with **r** i.i.d. and parameter $\lambda = \sqrt{2}$. **(a)** $T = 1000$ samples, **(b)** $T = 5000$ samples, **(c)** $T = 10000$ samples.

and Example 2. In the case of Example 1, we have set the parameter $\lambda = 5$ in the assumed distribution of Section 5.3.1, whereas we have set $\lambda = \sqrt{2}$ in the case of Example 2. For comparison, we provide the MSE results when η is known and when the algorithm CoM2 is applied, just ignoring that the generated source signals are dependent.

One can see from the values in Table 5 that our separation method remains valid even in the case where η is estimated.

7.3 Markov latent process
As previously indicated, the ICE estimation algorithm is able to take into account the Markov dependence of **r**.

We have considered the importance of modeling **r** as a Markov or i.i.d. process.

The process **r** has been generated as a stationary Markov chain with transition matrix $\begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}$, in which case $\mathbb{P}(r(t) = 0) = \mathbb{P}(r(t) = 1) = \frac{1}{2}$. On the corresponding data set with **r** Markov, our separation method has been performed both in the case where **r** is modeled as i.i.d. and in the case where **r** is modeled as a Markov chain. In both cases, the parameters in η have been estimated from the incomplete data. Figure 7 provides the results for sources from Example 1 (with $\lambda = 15$ in the algorithm) and Figure 8 provides the results for sources from Example 2 (with $\lambda = \sqrt{2}$ in the algorithm). The separation results with complete data

Table 5 Average MSE

| Number of samples	$T = 1000$			$T = 5000$			$T = 10000$		
Parameter η	**0.3**	**0.5**	**0.7**	**0.3**	**0.5**	**0.7**	**0.3**	**0.5**	**0.7**
Example 1, η known	4.70e-2	2.70e-2	7.94e-3	1.04e-2	6.66e-3	2.59e-3	9.34e-3	5.42e-3	2.48e-3
Example 1, η estimated	4.35e-1	8.17e-2	9.46e-3	5.36e-1	1.59e-2	4.59e-3	5.73e-1	8.50e-3	4.38e-3
Example 1, CoM2	5.86e-1	5.33e-1	3.62e-2	5.86e-1	5.84e-1	1.29e-3	5.86e-1	5.85e-1	5.43e-5
Example 2, η known	1.91e-1	4.71e-2	3.43e-3	2.51e-1	5.16e-3	6.15e-4	2.76e-1	1.42e-3	3.33e-4
Example 2, η estimated	3.85e-1	6.66e-2	3.53e-3	5.43e-1	6.34e-3	6.60e-4	5.59e-1	1.77e-3	3.55e-4
Example 2, CoM2	5.04e-1	4.16e-1	1.20e-1	5.47e-1	5.38e-1	1.01e-1	5.57e-1	5.54e-1	7.99e-2

Sources generated from Example 1 ($\lambda = 5$) and Example 2 ($\lambda = \sqrt{2}$) with r i.i.d.
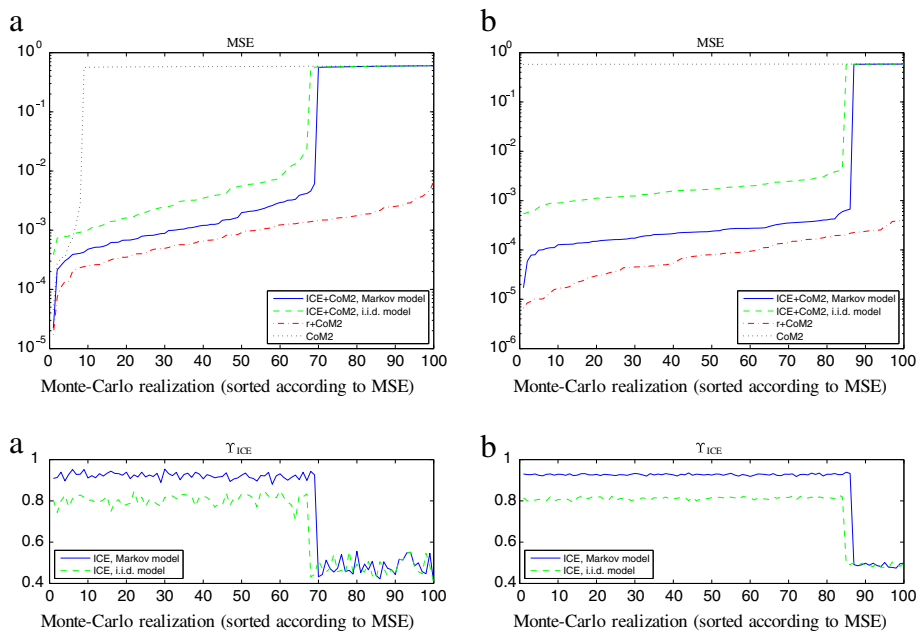
Figure 7 For an i.i.d. and a Markov model of **r**, MSE of 100 Monte-Carlo realizations sorted by increasing values and corresponding segmentation rate $\Upsilon_{\textbf{ICE}}$. Sources generated from Example 1 with **r** Markov, $\lambda = 15$ **(a)** $T = 1000$ samples, **(b)** $T = 10000$ samples.

and with CoM2 ignoring the dependence of the sources have also been plotted and, as expected, they correspond respectively to the best/worse MSE value. On the top plots of Figure 7, one can see that, in sources from Example 1, the performance is improved when taking into account

the Markovianity of **r**. More precisely, similarly to the simulations in Section 7.1.1 with $\lambda = 15$, the MSE values show a successful separation for a majority of realizations: in such a case, the Markovianity assumption clearly improves the MSE. However, the Markovianity
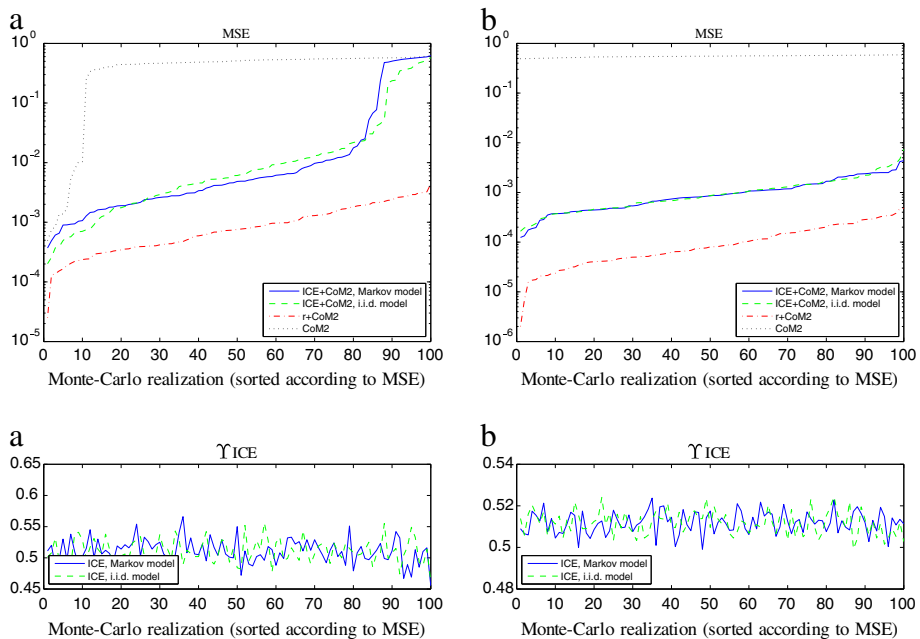
Figure 8 For an i.i.d. and a Markov model of **r**, MSE of 100 Monte-Carlo realizations sorted by increasing values and corresponding segmentation rate. Sources generated from Example 2 with **r** Markov, $\lambda = \sqrt{2}$. **(a)** $T = 1000$ samples, **(b)** $T = 10000$ samples.

property does not significantly improve the robustness of the method, since the number of realizations where separation has failed is approximately identical with a Markov and an i.i.d. model.

Interestingly, a different behavior can be seen in Figure 8 with data generated according to Example 2. Indeed, the performance has not improved when **r** has been modeled as a Markov chain. Intuitively, this may come from the fact that the two components of the probability mixture in the source distribution are harder to distinguish, as it has been illustrated in Figure 2.

8 Conclusion

In this article, we have first studied the behavior of kurtosis based contrast functions in specific cases of dependent sources. Observing that the criteria are polynomial functions of the parameters, we have been able to explicitly give the theoretical maxima as a function of a kurtosis value parameter $\kappa_4^{(a)}$. The behavior of the classical kurtosis contrast function can thus be understood depending on $\kappa_4^{(a)}$ and its restricted validity has been proved.

We have then introduced a model of dependent sources which consists in a probability mixture of two distributions. One component of the mixture satisfies the ICA assumption and, using an ICE estimation method, we have been able to exploit this information in order to perform separation. This example suggests that many more dependent sources might be separated if an adequate model of their distribution is provided. More generally, since the distribution model as a probability mixture may be an approximate one, an interesting problem would be to know to what extent a given distribution may be approximated by our proposed model in order to perform BSS. Finally, simulations have confirmed and validated our theoretical results.

Endnotes

[a] It rectifies an error in [22].
[b] One can show that the number of solutions of (10) is indeed finite. Note that one can also solve the polynomial equations, as has been done in the proof of Proposition 1.

Author details
[1] Institut Mines-Télécom/Télécom Sudparis, CNRS UMR 5157 SAMOVAR, 9 rue Charles Fourier, 91011 Évry Cedex, France. [2] GIPSA-Lab, UMR 5216, BP 46, 38402 Saint-Martin d'Hères Cedex, France.

References

1. P Comon, C Jutten (eds), in *Handbook of Blind Source Separation, Independent Component Analysis and Applications* (Academic Press, Oxford, 2010)
2. TH Li, Finite-alphabet information and multivariate blind deconvolution and identification of linear systems. IEEE Trans. Inf. Theory. **49**, 330–337 (2003)
3. P Comon, Contrasts, independent component analysis, and blind deconvolution. Int. J. Adapt. Control Sig. Proc. **18**(3), 225–243 (2004). [hal-00542916]
4. A Hyvärinen, S Shimizu, in *Proc. International Conference on Artificial Neural Networks (ICANN)*. A quasi-stochastic gradient algorithm for variance-dependent component analysis (Athens, Greece, 2006), pp. 211–220
5. JF Cardoso, in *Proc. ICASSP*. Multidimensional independent component analysis (Seattle, 1998), pp. 1941–1944
6. A Quirós, SP Wilson, in *European Signal Processing Conference (EUSIPCO)*. Dependent Gaussian mixture models for source separation (Barcelona, Spain, 2011), pp. 1723–1727
7. M Almeida, R Vigário, JB Dias, in *European Signal Processing Conference (EUSIPCO)*. Phase locked matrix factorization (Barcelona, Spain, 2011), pp. 1728–1732
8. HW Gutch, J Krumsiek, FJ Theis, in *European Signal Processing Conference (EUSIPCO)*. An ISA algorithm with unkonwn group sizes identifies meaningful clusters in metabolomics data (Barcelona, Spain, 2011), pp. 1733–1737
9. K Nordhausen, H Oja, in *European Signal Processing Conference (EUSIPCO)*. Scatter matrices with independent block property and ISA (Barcelona, Spain, 2011), pp. 1738–1742
10. S Rafi, M Castella, W Pieczynski, in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*. An extension of the ICA model using latent variable (Prague, Czech Republic, 2011), pp. 3712–3715
11. FR Bach, MI Jordan, in *Uncertainty in Artificial Intelligence (UAI): Proceedings of the Eighteenth Conference*. Tree-dependent component analysis (San Francisco, 2002), pp. 36–44
12. A Hyvärinen, PO Hoyer, M Inki, Topographic independent component analysis. Neural Comput. **13**(7), 1527–1558 (2001)
13. CF Caiafa, EE Kuruoglu, AN Proto, in *International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (Maxent)*. A minimax entropy method for blind separation of dependent components in astrophysical images, (2006)
14. EE Kuruoglu, in *Proc. of LVA/ICA, Volume 6365 of* LNCS. Dependent component analysis for cosmology: a case study (St-Malo, France, 2010), pp. 538–545
15. T Kim, I Lee, TW Lee, in *Fortieth Asilomar Conference on Signal, Systems and Computers (ACSSC)*. Independent vector analysis: definition and algorithms (Pacific Grove, CA, 2006), pp. 1393–1396
16. F Kohl, G Wübbeler, D Kolossa, C Elster, M Bär, R Orglmeister, in *Proc. of ICA'09, Volume 5441 of* LNCS. Non-independent BSS: a model for evoked MEG signals with controllable dependencies (Paraty-RJ, Brazil, 2009), pp. 443–450
17. S Akaho, Conditionally independent component analysis for supervised feature extraction. Neurocomputing. **49**, 139–150 (2002)
18. P Loubaton, P Regalia, in *Proceedings of ICC*. Blind deconvolution of multivariate signals: a deflation approach (Geneva, Switzerland, 1993), pp. 1160–1164
19. N Delfosse, P Loubaton, Adaptive blind separation of independent sources: a deflation approach. Signal Process. **45**, 59–83 (1995)
20. P Comon, Blind identification and source separation in 2×3 under-determined mixtures. IEEE Trans. Signal Process. **52**, 11–22 (2004)
21. P Comon, O Grellier, in *Proc. of ICA'99*. Non-linear inversion of underdetermined mixtures (Aussois, France, 1999), pp. 461–465
22. M Castella, P Comon, in *Proc. of ICA'07, Volume 4666 of* LNCS. Blind separation of instantaneous mixtures of dependent sources (London, UK, 2007), pp. 9–16
23. M Castella, Inversion of polynomial systems and separation of nonlinear mixtures of finite-alphabet sources. IEEE Trans. Signal Process. **56**(8, Part 2), 3905–3917 (2008)
24. W Pieczynski, Statistical image segmentation. Mach. Graph. Vis. **1**(1/2), 262–268 (1992)
25. P Comon, Independent component analysis, a new concept. Signal Process. **36**(3), 287–314 (1994)
26. JF Cardoso, A Souloumiac, Blind beamforming for non gaussian signals. IEE- Proceedings-F. **140**(6), 362–370 (1993)

27. A Hyvärinen, E Oja, Independent component analysis: algorithms and applications. Neural Netw. **13**(4–5), 411–430 (2000)
28. C Lossen, Singular: a computer algebra system. Comput. Sci. Eng. **5**(4), 45–55 (2003)
29. P Lanchantin, J Lapuyade-Lahorgue, W Pieczynski, Unsupervised segmentation of randomly switching data hidden with non-Gaussian correlated noise. Signal Process. **91**, 163–175 (2011)
30. S Derrode, W Pieczynski, Signal and image segmentations using pairwise Markov chains. IEEE Trans. Signal Process. **52**(9), 2477–2489 (2004)
31. W Pieczynski, in *Stochastic Modeling Techniques and Data Analysis international conference (SMTDA)*. EM and ICE in hidden and triplet Markov models (Chania, Greece, p. 2010
32. A Peng, W Pieczynski, Adaptive mixture estimation and unsupervised local Bayesian image segmentation. Graphic. Models Image Process. **57**(5), 389–399 (1995)
33. J Delmas, An equivalence of the EM and ICE algorithm for exponential family. IEEE Trans. Signal Process. **45**(10), 2613–2615 (1997)
34. W Pieczynski, Sur la convergence de l'estimation conditionnelle itérative. Compte rendus de l'Académie des Sciences Mathématiques. **246**(7–8), 457–460 (2008)
35. PA Devijver, Baum's forward-backward algorithm revisited. Pattern Recogn. Lett. **3**, 369–373 (1985)