Blind noisy mixture separation for independent/dependent sources through a regularized criterion on copulas

M. El Rhabi, H. Fenniri, A. Ghazdali, A. Hakim, A. Keziou

Blind noisy mixture separation for independent/dependent sources through a regularized criterion on copulas

M. El Rhabi[a], H. Fenniri[b], A. Ghazdali[c], A. Hakim[d], A. Keziou[e,*]

[a]*Ecole des Ponts ParisTech (ENPC), France*
[b]*CReSTIC, Université de Reims Champage-Ardenne, France*
[c]*LAMAI, FSTG, Université Cadi-Ayyad, Marrakech, Maroc*
[d]*LAMAI, FSTG, Université Cadi-Ayyad, Marrakech, Maroc*
[e]*LMR EA 4535 and ARC-Mathématiques CNRS 3399, Université de Reims Champagne-Ardenne, France*

Abstract

The paper introduces a new method for Blind Source Separation (BSS) in noisy instantaneous mixtures of both independent or dependent source component signals. This approach is based on the minimization of a regularized criterion. Precisely, it consists in combining the total variation method for denoising with the Kullback-Leibler divergence between copula densities. This latter takes advantage of the copula to model the structure of the dependence between signal components. The obtained algorithm achieves separation in a noisy context where standard BSS methods fail. The efficiency and robustness of the proposed approach are illustrated by numerical simulations.

Keywords: Blind source separation, Noisy instantaneous mixtures, Copulas, Total variation, Mutual information, Kullbak-Leibler divergence.

Contents

*Corresponding author
Email addresses:* `elrhabi@gmail.com` (M. El Rhabi), `hassan.fenniri@univ-reims.fr` (H. Fenniri), `a.ghazdali@gmail.com` (A. Ghazdali), `abdelilah.hakim@gmail.com` (A. Hakim), `amor.keziou@univ-reims.fr` (A. Keziou)

1. Introduction

The blind source separation (BSS) problem is a fundamental issue in applications of many different fields such as signal and image processing, biological and medical data analysis, communications, ... etc. The BSS aims to recover unknown source signals from a set of observations which are unknown mixtures of source signals. In order to separate the mixtures, different assumptions on the sources have to be made. In the literature, the most common assumptions are statistical independence of the source components and the condition that at most one of the components is gaussian. Under these assumptions, the BSS problem is linked to the well known problem of Independent Component Analysis (ICA), see for instance [1]. Using the above conditions, particularly, for free-noisy mixtures, many methods of BSS have been proposed in the

literature, see e.g. [2, 3, 4, 5]. Some of these procedures use second or higher order statistics [6], maximize likelihood [7], maximize nongaussianity [8], minimize the mutual information [9] or φ-divergences [10]. An interesting overview on the subject can be found in [11]. Recently, under the assumption that the source components are independent, based on estimation of mutual information and total variation (TV) regularization, [12] provided a new BSS algorithm to separate noisy mixtures of instantaneous or covolutive mixtures of independent source components. Further, it has been shown in [13] that, based on copula models, without the assumption of the independence of the source components, we can still identify both mixing matrix and sources uniquely (up to scale and permutation indeterminacies) of (free-noisy) mixtures of both independent and dependent source components. Motivated by various applications, we investigate, in the present paper, models of noisy linear instantaneous mixtures of independent/dependent sources, for which we propose, based on the previews works [12] and [13], a new BSS procedure. We successfully combined the two methods of [12] and [13] to obtain a new BSS algorithm able to separate noisy mixtures of independent/dependent source components. Our methodology consists in denoising the observed signals through the minimization of their total variation, and then minimizing Kullback-Leibler divergence, between copula densities, penalized by the total variation of the estimated source signals. The outline of this paper is organized as follows. In Section 2, we state the BSS problem in a noisy mixture context. In Section 3, we give a bref introduction on copulas and some of their main properties. In Section 4, we describe our approach. Section 5 presents how to implement the proposed approach using both numerical and statistical techniques. Section 6 gives some numerical results, illustrating the efficiency and robustness of the proposed method. Conclusions are drawn in Section 7.

2. The BSS model

2.1. BSS model in the noise-free case

BSS problem can be modeled as follows. Denoting $\boldsymbol{A}[\cdot]$ the (unknown) mixing operator, the relationship between the observed and source signals can be written as

$$\boldsymbol{x}(t) := \boldsymbol{A}[\boldsymbol{s}(t)] \in \mathbb{R}^p, \tag{1}$$

where $\boldsymbol{s}(t) \in \mathbb{R}^q$ is the unknown vector of source signals to be estimated, and $\boldsymbol{x}(t)$ represents the observed vector signal at time $t \in [0, T]$. The goal of BSS, is therefore to estimate the

unknown sources $\boldsymbol{s}(t)$ from the observed mixtures $\boldsymbol{x}(t)$. The estimation is performed with no prior information about either the sources or the mixing operator $\boldsymbol{A}[\cdot]$. Specific restrictions are made on both the mixing model and the source signals in order to limit the generality. We will restrict our self to the case where the number of source component and the number of observed mixture ones are equal $(p = q)$, and we assume, in the present paper, that the mixtures are linear and instantaneous, so that the mixing operator $\boldsymbol{A}$ can be considered as a $p \times p$ matrix. In this case, supposing in addition that $\boldsymbol{A}$ is invertible, the candidate estimates of the sources will be obviously of the form

$$\boldsymbol{y}(t) := \boldsymbol{B} \ \boldsymbol{x}(t) \in \mathbb{R}^p, \tag{2}$$

where $\boldsymbol{B} \in \mathbb{R}^{p \times p}$ represents an appropriate demixing matrix. In other words, the problem is to obtain an estimate, denote it $\widehat{\boldsymbol{B}}$, "closing" to the ideal solution $\boldsymbol{B} = \boldsymbol{A}^{-1}$, by the use of only the observation $\boldsymbol{x}(t)$, which leads to accurate estimation of the source $\boldsymbol{s}(t)$:

$$\widehat{\boldsymbol{s}}(t) := \widehat{\boldsymbol{B}} \ \boldsymbol{x}(t) \simeq \boldsymbol{s}(t). \tag{3}$$

2.2. BSS model for noisy mixtures

In the present paper, we will focus on the BSS problem for noisy linear instantaneous mixtures

$$\overline{\boldsymbol{x}}(t) := \boldsymbol{A} \ \boldsymbol{s}(t) + \boldsymbol{n}(t) \in \mathbb{R}^p, \ t \in [0, T], \tag{4}$$

where $\overline{\boldsymbol{x}}(t) \in \mathbb{R}^p$ is the vector of noisy observations, $\boldsymbol{s}(t) \in \mathbb{R}^p$ is the unknown vector of sources to be estimated, $\boldsymbol{n}(t) \in \mathbb{R}^p$ is the (unknown) noise, and $\boldsymbol{A} \in \mathbb{R}^{p \times p}$ is the unknown mixing matrix. We assume that the source signal $\boldsymbol{s}(t)$ and the noise $\boldsymbol{n}(t)$ are independent. The mixing model (4) can also be written as

$$\overline{\boldsymbol{x}}(t) = \boldsymbol{x}(t) + \boldsymbol{n}(t), \tag{5}$$

with

$$\boldsymbol{x}(t) := \boldsymbol{A} \ \boldsymbol{s}(t) \tag{6}$$

is the inaccessible noise-free mixed vector signals. The aim here, is to estimate the sources $\boldsymbol{s}(t)$ using only the observed noisy signal $\overline{\boldsymbol{x}}(t)$. The presence of noise $\boldsymbol{n}(t)$ within the mixing model (4), as well as the possible dependency structure of the source components, complicate significantly the BSS problem. Unlike the above case of free-noise mixtures in Subsection 2.1, the separating system of the form

$$\overline{\boldsymbol{y}}(t) := \boldsymbol{B} \ \overline{\boldsymbol{x}}(t) \tag{7}$$

does not provide a valid estimation for the source signals $\boldsymbol{s}(\cdot)$. In fact, the obtained signals according to the above system, can be written as

$$
\begin{aligned}
\overline{\boldsymbol{y}}(t) := \boldsymbol{B}\,\overline{\boldsymbol{x}}(t) \;&=\; \boldsymbol{B}\,\boldsymbol{x}(t) + \boldsymbol{B}\,\boldsymbol{n}(t) \\
&=:\; \boldsymbol{y}(t) + \overline{\boldsymbol{n}}(t),
\end{aligned} \tag{8}
$$

where $\overline{\boldsymbol{y}}(t) \in \mathbb{R}^p$ is the "noisy estimate" of $\boldsymbol{s}(t)$. That is the noisy estimated source $\overline{\boldsymbol{y}}(t)$ is the sum of

$$
\boldsymbol{y}(t) := \boldsymbol{B}\,\boldsymbol{x}(t) \tag{9}
$$

the "noisy-free" estimate of the sources, which is inaccessible, and the unknown "noise"

$$
\overline{\boldsymbol{n}}(t) := \boldsymbol{B}\,\boldsymbol{n}(t). \tag{10}
$$

Ideally, we would like to retrieve $\boldsymbol{y}(t)$ by denoising $\overline{\boldsymbol{y}}(t)$, but it is rather difficult since the noise $\overline{\boldsymbol{n}}(t)$ is unknown. During last years, several algorithms have been proposed to tackle the noisy BSS problem; In [14], the authors propose a two-step approach by combining the fraction allower order statistic for the mixing estimation and minimum entropy criterion for noise-free source component estimation. In [15], a whitening procedure is proposed to reduce the noise effect. At our knowledge, the noisy BSS problem for possibly dependent sources has not been considered in the literature. In [13], the authors proposed a new criterion to successfully separate noise-free mixtures of both independent and dependent sources, based on estimation of Kullback-Leibler divergence between copula densities. [16] provides a BSS procedure for noisy mixtures using TV-regularization of mutual information, which applies exclusively under the assumption that the source components are independent. Based on these two last works, the present paper introduces a new procedure for estimating sources in the context of noisy mixtures of possibly dependent source components. It proceeds on two stages : (i) denoising the observed signal $\overline{\boldsymbol{x}}(t)$ before demixing; (ii) a simultaneous BSS-denoisy procedure via minimizing a TV-regularized measure between copula densities.

3. Bref recall on Copula

When modeling multivariate distributions, one has to take into account the effects of the marginal distributions as well as the dependence between them. This can be achieved by using the copula approach, which allows to deal with the margins and the dependency structure

separately. It was introduced by [17] as a function which couples a joint distribution function with its univariate margins. Precisely, it can be presented as follows. Consider any random vector

$$\mathbf{Y} := (Y_1, \ldots, Y_p)^\top \in \mathbb{R}^p, \ p \geq 1,$$

with joint distribution function (d.f.)

$$\mathbb{F}_{\mathbf{Y}}(\cdot) : \mathbf{y} \in \mathbb{R}^p \mapsto \mathbb{F}_{\mathbf{Y}}(\mathbf{y}) := \mathbb{F}_{\mathbf{Y}}(y_1, \ldots, y_p) := \mathbb{P}(Y_1 \leq y_1, \ldots, Y_p \leq y_p),$$

and continuous marginal d.f.'s

$$F_{Y_j}(\cdot) : y_j \in \mathbb{R} \mapsto F_{Y_j}(y_j) := \mathbb{P}(Y_j \leq y_j), \ \forall j = 1, \ldots, p.$$

The characterization theorem of Sklar [17] shows that there exists a unique p-variate function $\mathbb{C}_{\mathbf{Y}}(\cdot) : [0, 1]^p \mapsto [0, 1]$, such that,

$$\mathbb{F}_{\mathbf{Y}}(\mathbf{y}) = \mathbb{C}_{\mathbf{Y}}(F_{Y_1}(y_1), \ldots, F_{Y_p}(y_p)), \forall \mathbf{y} := (y_1, \ldots, y_p)^\top \in \mathbb{R}^p.$$

The function $\mathbb{C}_{\mathbf{Y}}(\cdot)$ is called a copula and it is in itself a joint d.f. on $[0, 1]^p$ with uniform margins. We have

$$\forall \mathbf{u} := (u_1, \ldots, u_p)^\top \in [0, 1]^p, \ \mathbb{C}_{\mathbf{Y}}(\mathbf{u}) = \mathbb{P}\left(F_{Y_1}(Y_1) \leq u_1, \ldots, F_{Y_p}(Y_p) \leq u_p\right).$$

Conversely, for any marginal d.f.'s $F_1(\cdot), \ldots, F_p(\cdot)$, and any copula function $\mathbb{C}(\cdot)$, the function $\mathbb{C}(F_1(\cdot), \ldots, F_p(\cdot))$ is a multivariate d.f. on $\mathbb{R}^p$. On the other hand, since the marginal d.f.'s $F_{Y_j}(\cdot), j = 1, \ldots, p$, are assumed to be continuous, then the random variables $F_{Y_1}(Y_1), \ldots, F_{Y_p}(Y_p)$ are uniformly distributed on the interval $[0, 1]$. Moreover, we have that the components $Y_1, \ldots, Y_p$ are statistically independent if and only if (iff) the copula $\mathbb{C}_{\mathbf{Y}}(\cdot)$ of the random vector $\mathbf{Y} := (Y_1, \ldots, Y_p)^\top \in \mathbb{R}^p$ writes

$$\mathbb{C}_{\mathbf{Y}}(\mathbf{u}) = \prod_{j=1}^{p} u_j =: \mathbb{C}_{\Pi}(\mathbf{u}), \ \forall \mathbf{u} \in [0, 1]^p,$$

which will be denote by $\mathbb{C}_{\Pi}(\cdot)$; It is called the copula of independence. Define, when it exists, the copula density

$$c_{\mathbf{Y}}(\mathbf{u}) := \frac{\partial^p \mathbb{C}_{\mathbf{Y}}(\mathbf{u})}{\partial u_1 \cdots \partial u_p}, \ \forall \mathbf{u} \in [0, 1]^p.$$

Hence, the copula density of independence, denote it by $c_\Pi(\cdot)$, is the function taking the value 1 on $[0,1]^p$ and zero otherwise, namely,

$$c_\Pi(\mathbf{u}) := \mathbb{1}_{[0,1]^p}(\mathbf{u}), \quad \forall \mathbf{u} \in [0,1]^p. \tag{11}$$

Let $f_{\mathbf{Y}}(\cdot)$, if it exists, be the probability density of the random vector $\mathbf{Y} := (Y_1, \ldots, Y_p)^\top$, and, respectively, $f_{Y_1}(\cdot), \ldots, f_{Y_p}(\cdot)$, the marginal probability densities of the random variables $Y_1, \ldots, Y_p$. Then, a straightforward computation shows that, for all vector $\mathbf{y} := (y_1, \ldots, y_p)^\top \in \mathbb{R}^p$, we have the relation

$$f_{\mathbf{Y}}(\mathbf{y}) = \prod_{j=1}^{p} f_{Y_j}(y_j) \; c_{\mathbf{Y}} \left(F_{Y_1}(y_1), \ldots, F_{Y_p}(y_p) \right), \tag{12}$$

As previously highlighted, copulas play an important role in the construction of multivariate d.f.'s. Therefore, several investigations have been carried out concerning the construction of different families of copulas and their properties. For more details on modeling theory as well as surveys of the commonly semiparametric used copulas, we can refer to the monographs by [18, 19].

4. A combined TV-Copula approach for BSS

Our TV-Copula approach for BSS proceeds in two steps. (i) uses the TV regularization technique for denoising the observed signals. (ii) separates the mixtures by minimizing penalized Kullback-Leibler divergence between copulas with a TV-regularization term. In the following, we describe these two steps.

4.1. Setp 1 : Denoising the observed signals

As in [12], we would obtain the denoisy mixture signal

$$\boldsymbol{x}(t) := (x_1(t), \ldots, x_p(t))^\top, \quad t \in [0, T], \tag{13}$$

from the (noisy) observed one $\overline{\boldsymbol{x}}(t)$, see (4), by means of the following variational minimization problem

$$x_i(\cdot) = \arg \min_{w_i \in \mathcal{X}_c} \left\{ \frac{1}{T} \int_0^T \frac{1}{2} \left(w_i(t) - \overline{x}_i(t) \right)^2 \, dt + \lambda \frac{1}{T} \int_0^T \phi \left(|w_i'(t)| \right) \, dt \right\}, \; i = 1, \ldots, p, \tag{14}$$

7

where $\lambda > 0$ is the parameter of penalization, $\phi(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ is a well chosen function, $w_i'(t)$ is the first derivative of the signal $w_i(t)$ with respect to time t, and $\mathcal{X}_c$ is some appropriate space of signals. In practice, in order to reduce the noise, the function $\phi(\cdot)$ is chosen to encourage smoothing in regions where the variations of the signal are weak, that is $|w_i'| \approx 0$, and to preserve discontinuities where $|w_i'|$ is strong. The total variation case corresponds to the choice $\phi(x) = x$ which will be adopted here. One can make other choices like $\phi(x) = \sqrt{1 + x^2}$. A natural space when dealing with the above continuous variational problem (14), with $\phi(x) = x$, is $\mathcal{X}_c = BV([0, T])$, the space of all real valued functions on the interval $[0, T]$ with bounded variation; see e.g. [20] and [21]. Here T is the observation time of the signals. The mathematical treatment of the above problem, in its continuous form, is beyond the scope of this paper. We will consider only the corresponding discrete version which will be described in Subsection 5.1 below.

4.2. Step 2 : BSS via minimizing KL-divergence between copulas and denoising

The aim of the following step is to construct an estimated source signal

$$\widehat{s}(t) := \widehat{B}\, x(t)$$

from the denoised signal $\boldsymbol{x}(t) := (x_1(t), \ldots, x_p(t))^\top$ obtained as solution of (14). Our approach consists in minimizing, with respect to $\boldsymbol{B}$ on the demixing matrices space, an estimate (to be defined below) of some criterion of the form

$$\boldsymbol{B} \mapsto \mathcal{J}(\boldsymbol{B}) := \mathcal{J}_{\mathrm{sep}}(\boldsymbol{B}) + \mathcal{J}_{\mathrm{reg}}(\boldsymbol{y}), \tag{15}$$

where

$$\boldsymbol{y}(t) := \boldsymbol{B}\, \boldsymbol{x}(t) \in \mathbb{R}^p,$$

and $\boldsymbol{x}(t) = (x_1(t), \ldots, x_p(t))^\top \in \mathbb{R}^p$ is obtained from (14); $\mathcal{J}_{\mathrm{sep}}(\cdot)$ is some separating criterion, while $\mathcal{J}_{\mathrm{reg}}(\boldsymbol{y})$ is a regularization term devoted to denoising. For the denoising phase, for both cases, of independent or dependent source components, we propose to use as regularization term the following one

$$\mathcal{J}_{\mathrm{reg}}(\boldsymbol{y}) := \gamma \, \frac{1}{2T} \int_0^T \|\boldsymbol{y}(t) - \overline{\boldsymbol{y}}(t)\|^2 \, dt + \mu \, \frac{1}{T} \int_0^T |\nabla \boldsymbol{y}(t)| \, dt, \; \gamma > 0, \; \mu > 0, \tag{16}$$

where $\overline{\boldsymbol{y}} = \boldsymbol{B}\,\overline{\boldsymbol{x}}$; the real numbers γ and μ are regularization parameters to be suitably chosen by the user. In the above display, $\|\cdot\|$ is used to denote the Euclidian norm on $\mathbb{R}^p$, and

$|\nabla \boldsymbol{y}(t)| = \sum_{i=1}^{p} |y_i'(t)|$, for all t. The first term $\mathcal{J}_{\text{sep}}(\cdot)$, in (15), is a measure of dependence, or similarity between copula densities, and it depends obviously on the fact that the source components are independent or dependent. Hence, to describe the criterion $\mathcal{J}_{\text{sep}}(\cdot)$ hereafter, we will consider separately, the case where the source components are independent and the case where they are dependent. First, we will consider the following stochastic modeling of the analog signals $\boldsymbol{s}(t)$, $\boldsymbol{x}(t)$ and $\boldsymbol{y}(t) := \boldsymbol{B}\boldsymbol{x}(t)$, $t \in [0, T]$. We assume that the above last random processes are stationary, so that the corresponding sampled versions, with certain time period, say T_e,

$$\boldsymbol{s}(n) := \boldsymbol{s}(nT_e), \ \boldsymbol{x}(n) := \boldsymbol{x}(nT_e) \text{ and } \boldsymbol{y}(n) := \boldsymbol{y}(nT_e) = \boldsymbol{B}\,\boldsymbol{x}(n), \ n = 1, \ldots, N, \qquad (17)$$

can be considered as copies (realizations) of random vectors in $\mathbb{R}^p$, which will be denoted

$$\boldsymbol{S}, \quad \boldsymbol{X} \quad \text{and} \quad \boldsymbol{Y} := \boldsymbol{B}\boldsymbol{X}, \qquad (18)$$

respectively. It has been shown, in [13], for the noisy-free mixtures case, that if we dispose of some prior information about the copula density of the random source vector $\boldsymbol{S}$, we can detect, under some assumptions, both the mixing matrix and the sources uniquely (up to scale and permutation indeterminacies) for both cases of independent or dependent source components. Denote by

$$F_{\boldsymbol{Y}}(\cdot) : \ \boldsymbol{y} \in \mathbb{R}^p \mapsto F_{\boldsymbol{Y}}(\boldsymbol{y}) := F_{\boldsymbol{Y}}(y_1, \ldots, y_p) := \mathbb{P}(Y_1 \leq y_1, \ldots, Y_p \leq y_p), \qquad (19)$$

the joint distribution function of the random vector $\boldsymbol{Y} := (Y_1, \ldots, Y_p)^\top \in \mathbb{R}^p$, with continuous marginal distribution functions

$$F_{Y_i}(\cdot) : \ y_i \in \mathbb{R} \mapsto F_{Y_i}(y_i) := \mathbb{P}(Y_i \leq y_i), \ \forall i = 1, \ldots, p.$$

Let, if it exists, $f_{\boldsymbol{Y}}(\cdot)$ be the probability density of the random vector $\boldsymbol{Y}$, and, respectively, $f_{Y_1}, \ldots, f_{Y_p}$, the marginal probability densities of the components $Y_1, \ldots, Y_p$. The mutual information (MI) of $\boldsymbol{Y}$ is defined by

$$\mathbb{I}(\boldsymbol{Y}) := \int_{\mathbb{R}^p} \log \frac{f_{\boldsymbol{Y}}(\boldsymbol{y})}{\prod_{i=1}^{p} f_{Y_i}(y_i)} \, f_{\boldsymbol{Y}}(\boldsymbol{y}) \, d\boldsymbol{y} = \mathbb{E}\left(\log \frac{f_{\boldsymbol{Y}}(\boldsymbol{Y})}{\prod_{i=1}^{p} f_{Y_i}(Y_i)}\right) \qquad (20)$$

which equals the Kullback-Leibler divergence, denote it $\mathbb{K}(\cdot, \cdot)$, between the joint density $f_{\boldsymbol{Y}}(\cdot)$ and the product density $\prod_{i=1}^{p} f_{Y_i}(\cdot)$ of the margins $f_{Y_i}(\cdot)$, $i = 1 \ldots, p$, that is

$$\mathbb{I}(\boldsymbol{Y}) = \mathbb{K}\left(f_{\boldsymbol{Y}}, \prod_{i=1}^{p} f_{Y_i}\right). \qquad (21)$$

In (20), $\mathbb{E}(\cdot)$ is the mathematical expectation. Note also that the above criterion $\mathbb{I}(\boldsymbol{Y})$ is nonnegative and achieves its minimum value zero iff $f_{\boldsymbol{Y}}(\cdot) = \prod_{i=1}^{p} f_{Y_i}(\cdot)$, i.e., iff the components of the vector $\boldsymbol{Y}$ are statistically independent. We give now, the form of the separating term $\mathcal{J}_{\text{sep}}(\cdot)$ of the criterion (15).

4.2.1. The case of independent source components

Assume that the source components are independent. Using the relation (12), and applying the change variable formula for multiple integrals, we can show that $I(\boldsymbol{Y})$ can be written, in terms of copula densities, as follows

$$\mathbb{I}(\boldsymbol{Y}) = \int_{[0,1]^p} \log\left(c_{\boldsymbol{Y}}(\boldsymbol{u})\right) c_{\boldsymbol{Y}}(\boldsymbol{u})\, d\boldsymbol{u} \tag{22}$$

$$= \mathbb{E}\left(\log c_{\boldsymbol{Y}}\left(F_{Y_1}(Y_1), \ldots, F_{Y_p}(Y_p)\right)\right) =: -\mathbb{H}(c_{\boldsymbol{Y}}) \tag{23}$$

$$= \int_{[0,1]^p} \log\left(\frac{c_{\boldsymbol{Y}}(\boldsymbol{u})}{1}\right) c_{\boldsymbol{Y}}(\boldsymbol{u})\, d\boldsymbol{u} \tag{24}$$

$$= \int_{[0,1]^p} \log\left(\frac{c_{\boldsymbol{Y}}(\boldsymbol{u})}{c_{\Pi}(\boldsymbol{u})}\right) c_{\boldsymbol{Y}}(\boldsymbol{u})\, d\boldsymbol{u} \tag{25}$$

$$= \mathbb{E}\left(\log \frac{c_{\boldsymbol{Y}}(F_{Y_1}(Y_1), \ldots, F_{Y_p}(Y_p))}{c_{\Pi}(F_{Y_1}(Y_1), \ldots, F_{Y_p}(Y_p))}\right) = \mathbb{K}\left(c_{\boldsymbol{Y}}, c_{\Pi}\right), \tag{26}$$

where $c_{\boldsymbol{Y}}(\cdot)$ is the copula density of $\boldsymbol{Y}$, $c_{\Pi}(\cdot) := \mathbb{1}_{[0,1]^p}(\cdot)$ is the copula density of independence, and $\mathbb{E}(\cdot)$ the mathematical expectation operator. The above equations mean that the MI of a random vector $\boldsymbol{Y}$ can be seen as the opposite of the Shannon entropy $\mathbb{H}(\cdot)$ of the copula density $c_{\boldsymbol{Y}}(\cdot)$ of $\boldsymbol{Y}$, or as the Kullback-Leibler divergence between the copula density $c_{\boldsymbol{Y}}(\cdot)$ of $\boldsymbol{Y}$ and the copula density $c_{\Pi}(\cdot)$ of independence. Moreover, $\mathbb{I}(\boldsymbol{Y}) = \mathbb{K}\left(c_{\boldsymbol{Y}}, c_{\Pi}\right)$ is nonnegative and achieves its minimum value zero iff $c_{\boldsymbol{Y}}(\boldsymbol{u}) = c_{\Pi}(\boldsymbol{u})$, $\forall \boldsymbol{u} \in [0,1]^p$, namely, iff the components of the vector $\boldsymbol{Y}$ are independent. The separating term will be chosen then to be

$$\mathcal{J}_{\text{sep}} := \mathcal{J}_{\text{sep}}^{\text{ind}}(\boldsymbol{B}) := KL\left(c_{\boldsymbol{Y}}, c_{\Pi}\right) = \mathbb{E}\left(\log c_{\boldsymbol{Y}}\left(F_{Y_1}(Y_1), \ldots, F_{Y_p}(Y_p)\right)\right). \tag{27}$$

We have that the function $\boldsymbol{B} \mapsto \mathcal{J}_{\text{sep}}^{\text{ind}}(\boldsymbol{B})$ is nonnegative and attains its minimum value zero iff $\boldsymbol{B} = \boldsymbol{A}^{-1}$ (up to scale and permutation indeterminacies of rows).

4.2.2. The case of dependent source components

In the case where the source components are dependent, we assume that we dispose of some prior information about the copula density of the random source vector $\boldsymbol{S}$. Note that this is possible

for many practical problems, it can be done, from learning samples of $\boldsymbol{S}$, by a model selection procedure, see e.g. [22], among semiparametric copula density models $\{c_\theta(\cdot); \theta \in \Theta \subset \mathbb{R}^d\}$, typically indexed by a multivariate parameter θ; see e.g. [19] and [18] for many examples of such models. The parameter θ can be estimated using maximum semiparametric likelihood, see e.g. [23]. Denote then by $\widehat{\theta}$, the obtained value of θ and $c_{\widehat{\theta}}(\cdot)$ the copula density modeling the dependency structure of the source components. Obviously, since the source components are assumed to be dependent, $c_{\widehat{\theta}}(\cdot)$ should be different from the density copula of independence $c_\Pi(\cdot)$. Hence, we naturally replace in (26), c_Π by $c_{\widehat{\theta}}$, and propose the following separating criterion $\mathcal{J}_{\mathrm{sep}}(\cdot)$

$$\mathcal{J}_{\mathrm{sep}}(\boldsymbol{B}) := \mathcal{J}_{\mathrm{sep}}^{\mathrm{dep}}(\boldsymbol{B}) := \int_{[0,1]^p} \log\left(\frac{c_{\boldsymbol{Y}}(\boldsymbol{u})}{c_{\widehat{\theta}}(\boldsymbol{u})}\right) c_{\boldsymbol{Y}}(\boldsymbol{u})\, \mathrm{d}\boldsymbol{u} := \mathbb{E}\left(\log \frac{c_{\boldsymbol{Y}}(F_{Y_1}(Y_1), \ldots, F_{Y_p}(Y_p))}{c_{\widehat{\theta}}(F_{Y_1}(Y_1), \ldots, F_{Y_p}(Y_p))}\right),$$
(28)

which equals $\mathbb{K}(c_{\boldsymbol{Y}}, c_{\widehat{\theta}})$, the Kullback-Leibler divergence between the copula densities $c_{\boldsymbol{Y}}(\cdot)$ and $c_{\widehat{\theta}}(\cdot)$. Moreover, we can show that the function $\boldsymbol{B} \mapsto \mathcal{J}_{\mathrm{sep}}^{\mathrm{dep}}(\boldsymbol{B})$ is nonnegative, and attains its minimum value iff $\boldsymbol{B} = \boldsymbol{A}^{-1}$ (up to scale and permutation indeterminacies of rows), provided that the copula density $c_{\widehat{\theta}}(\cdot)$ of the random vector source $\boldsymbol{S}$ satisfies the following assumption : for any regular matrix $\boldsymbol{M}$, if the copula density of the random vector $\boldsymbol{M}\boldsymbol{S}$ equals $c_{\widehat{\theta}}(\cdot)$, then $\boldsymbol{M} = \boldsymbol{D}\boldsymbol{P}$, for some diagonal matrix $\boldsymbol{D}$ and permutation matrix $\boldsymbol{P}$.

Remark 1. *The criterion (29) proposed in the above section supposes the knowledge of the copula density model $c_{\widehat{\theta}}(\cdot)$ (with known parameter $\widehat{\theta}$), or the availability of learning samples of the random vector source $\boldsymbol{S}$ by which we can obtain the copula density $c_{\widehat{\theta}}(\cdot)$ modeling the dependency structure of the source components, as described in Subsection 4.2.2. When only the model is known and the parameter is unknown, we can adapt the above criterion as follows. Denote by $\{c_\theta(\cdot); \theta \in \Theta\}$ the density copula model of the dependency structure of the sources. The criterion (29) can be in this case replaced by*

$$\mathcal{J}_{sep}(\boldsymbol{B}) := \inf_{\theta \in \Theta} \int_{[0,1]^p} \log\left(\frac{c_{\boldsymbol{Y}}(\boldsymbol{u})}{c_\theta(\boldsymbol{u})}\right) c_{\boldsymbol{Y}}(\boldsymbol{u})\, d\boldsymbol{u} := \inf_{\theta \in \Theta} \mathbb{E}\left(\log \frac{c_{\boldsymbol{Y}}(F_{Y_1}(Y_1), \ldots, F_{Y_p}(Y_p))}{c_\theta(F_{Y_1}(Y_1), \ldots, F_{Y_p}(Y_p))}\right). \quad (29)$$

In the more general case, when both the model and the parameter are unknown, the criterion may be adapted in a similar way as in our previous work [13] Section 5.3.

5. Discretization and Statistical estimation

In this section, we describe how to make operational the above methodology. Precisely, using both numerical and statistical techniques, we will describe hereafter the discrete versions of the regularization terms (14) and (16), as well as statistical estimates of the separation criterions (27) and (29). In all the sequel, analog signals are sampled with a certain period T_e. We need to define additional notation. To each scalar signal, say $u(t)$, $t \in [0, T]$, one associate a vector $\mathbb{U} := (u(1), \ldots, u(N))^\top \in \mathcal{X}_d := \mathbb{R}^N$ such that $u(k) := u(kT_e)$ for all $k = 1, \ldots, N$. The vector space $\mathcal{X}_d := \mathbb{R}^N$ will be equipped with the euclidian inner product

$$\langle \mathbb{U}, \mathbb{V} \rangle := \sum_{k=1}^{N} u(k)v(k),$$

for all $\mathbb{U}, \mathbb{V} \in \mathcal{X}_d$. The numerical first derivative of any $\mathbb{U} \in \mathcal{X}_d$, denoted $\mathbb{U}' := (u'(1), ..., u'(N))^\top$, belongs to $\mathcal{X}_d$, and is defined by

$$u'(k) := \frac{u(k+1) - u(k)}{T_e}, \quad \text{for all } k = 1, \ldots, N - 1, \tag{30}$$

and $u'(N) = 0$. Due to the discretization of the first derivatives and in order to avoid the *edge effect*, we define also the *backward derivative* of any $\mathbb{U} \in \mathcal{X}_d$, denoted $\mathbb{U}^* := (u^*(1), ..., u^*(N))^\top$, by

$$u^*(1) := \frac{u(1)}{T_e}, \ u^*(N) := -\frac{u(N-1)}{T_e}$$

and

$$u^*(k) := \frac{u(k) - u(k-1)}{T_e} \text{ for all } k = 2, \ldots, N - 1.$$

Here, the backward derivatives $(\cdot)^*$ is the discrete adjoint operator of $-(\cdot)'$. That is, for all $\mathbb{U}, \mathbb{V} \in \mathcal{X}_d$, we have

$$\langle \mathbb{U}^*, \mathbb{V} \rangle = -\langle \mathbb{U}, \mathbb{V}' \rangle. \tag{31}$$

All these definitions can be extended in a natural way to elements of the cartesian product space $\mathcal{X}^p = \mathbb{R}^p \times \mathbb{R}^N$. In the sequel, elements of the vector space $\mathcal{X}_d = \mathbb{R}^N$ will be denoted by double capital symbols, while the elements of the vector space $\mathcal{X}_d^p$ are denoted by bold double capital symbols, which can be considered as $p \times N$ dimension matrices.

5.1. Denoising the discrete observed signal

In this section, we show how to compute the denoisy observations

$$\mathbb{X} := \Big(\boldsymbol{x}(n) = (x_1(n), \ldots, x_p(n))^\top, n = 1, \ldots, N \Big)$$

from the observed noisy ones

$$\overline{\mathbb{X}} := \Big(\overline{\boldsymbol{x}}(n) = (\overline{x}_1(n), \ldots, \overline{x}_p(n))^\top, n = 1, \ldots, N \Big).$$

This will be obtained by solving the discrete version

$$x_i(\cdot) = \arg \min_{\mathbb{W}_i \in \mathcal{X}_d} \left\{ \frac{1}{N} \sum_{n=1}^{N} \frac{1}{2} (w_i(n) - \overline{x}_i(n))^2 + \lambda \frac{1}{N} \sum_{n=1}^{N} |w_i'(n)| \right\}, \, i = 1, \ldots, p, \qquad (32)$$

of the optimization problem (14). We start with the following proposition which gives a simple characterization of the solution; see [24] for a proof.

Proposition 5.1. *The discrete version (32) of the optimization problem (14) has a unique solution given by*

$$\mathbb{X} = \overline{\mathbb{X}} - \prod_{\lambda G} \overline{\mathbb{X}}, \qquad (33)$$

where $\prod_{\lambda G}$ is the euclidian orthogonal projection operator on the convex set λG with

$$G := \{\mathrm{V}^* / \mathrm{V} \in \mathcal{X}_d^p, \, |v_i(n)| \leq 1, \, \forall (i,n) \in \{1, \ldots, p\} \times \{1, \ldots, N\}\}. \qquad (34)$$

Thus, to compute $\mathbb{X}$, we are led to compute the projection operator $\prod_{\lambda G}$ on the convex set λG. For the denoising step, we propose then the algorithm 1 below; the reader may refer to [16] for more details, particularly, for a proof on convergence of this algorithm.

In all the sequel, the filtered observed signal $\mathbb{X}$ will be considered as the denoised version of the observed signal $\overline{\mathbb{X}}$.

5.2. BSS by minimizing an estimate of KL-divergence, between copulas, penalized by TV regularization

5.2.1. The case of independent source components

Recall that, when the source components are independent, the criterion function (15) is defined by

$$\boldsymbol{B} \mapsto \mathcal{J}^{\mathrm{ind}}(\boldsymbol{B}) := \mathcal{J}_{\mathrm{sep}}^{\mathrm{ind}}(\boldsymbol{B}) + \mathcal{J}_{\mathrm{reg}}(\boldsymbol{y}), \qquad (35)$$

Algorithm 1 The denoising step.

1: **Data:** $\overline{\mathbb{X}}$ the observed noised signal

2: **Result:** $\mathbb{X}$ the denoised signal

3: **Initialization:** $q = 0$, $\mathbb{X}^{(0)} = \overline{\mathbb{X}}$. Given $\epsilon > 0$, $\lambda > 0$ and $\rho > 0$ suitably chosen

4: **Do**: $q = q + 1$

5:      **for** $j = 1, \ldots, p$

6:         $\mathbb{X}_j^{(q)} = \prod_{\lambda G}(\mathbb{X}_j^{(q-1)} + 2\lambda\rho(\lambda(\mathbb{X}_j^{(q-1)})^* - \overline{\mathbb{X}}_j)'$

7:      **end**

8: **Until** $||\mathbb{X}^{(q)} - \mathbb{X}^{(q-1)}|| < \epsilon$

9: $\mathbb{X} = \mathbb{X}^{(q)}$.

---

where $\mathcal{J}_{\mathrm{sep}}^{\mathrm{ind}}(\boldsymbol{B})$ is given by, see (27),

$$\mathcal{J}_{\mathrm{sep}}^{\mathrm{ind}}(\boldsymbol{B}) := KL\left(c_{\boldsymbol{Y}}, c_{\Pi}\right) = \mathbb{E}\left(\log c_{\boldsymbol{Y}}\left(F_{Y_1}(Y_1), \ldots, F_{Y_p}(Y_p)\right)\right)$$

and

$$\mathcal{J}_{\mathrm{reg}}(\boldsymbol{y}) := \gamma \frac{1}{2T}\int_0^T \|\boldsymbol{y}(t) - \overline{\boldsymbol{y}}(t)\|^2 \, dt + \mu \frac{1}{T}\int_0^T |\nabla\boldsymbol{y}(t)| \, dt, \ \gamma > 0, \ \mu > 0,$$

with $\boldsymbol{y}(t) = \boldsymbol{B}\,\boldsymbol{x}(t)$ and $\overline{\boldsymbol{y}} = \boldsymbol{B}\,\overline{\boldsymbol{x}}(t)$. Using the stochastic modeling (17) and the relation (27), we propose to approximate the criterion (35) by

$$\boldsymbol{B} \mapsto \widehat{\mathcal{J}^{\mathrm{ind}}}(\boldsymbol{B}) := \widehat{\mathcal{J}_{\mathrm{sep}}^{\mathrm{ind}}}(\boldsymbol{B}) + \mathcal{J}_{\mathrm{reg,d}}(\boldsymbol{y}), \tag{36}$$

where

$$\mathcal{J}_{\mathrm{reg,d}}(\boldsymbol{y}) := \frac{\gamma}{2N}\sum_{i=1}^N \|\boldsymbol{y}(i) - \overline{\boldsymbol{y}}(i)\|^2 + \frac{\mu}{N}\sum_{i=1}^N |\nabla\boldsymbol{y}(i)| \tag{37}$$

is the discrete version of $\mathcal{J}_{\mathrm{reg}}(\boldsymbol{y})$, and $\boldsymbol{B} \mapsto \widehat{\mathcal{J}_{\mathrm{sep}}^{\mathrm{ind}}}(\boldsymbol{B})$ is the statistical estimate, of the separating criterion $\boldsymbol{B} \mapsto \mathcal{J}_{\mathrm{sep}}^{\mathrm{ind}}(\boldsymbol{B})$, which we define as follows

$$\boldsymbol{B} \mapsto \widehat{\mathcal{J}_{\mathrm{sep}}^{\mathrm{ind}}}(\boldsymbol{B}) := \frac{1}{N}\sum_{i=1}^N \log\left(\widehat{c}_Y(\widehat{F}_{Y_1}(y_1(i)), \ldots, \widehat{F}_{Y_p}(y_p(i)))\right), \tag{38}$$

where

$$\widehat{c}_Y(\widehat{F}_{Y_1}(y_1(i)), \ldots, \widehat{F}_{Y_p}(y_p(i))) := \frac{1}{NH_1\cdots H_p}\sum_{\ell=1}^N\prod_{j=1}^p k\left(\frac{\widehat{F}_{Y_j}(y_j(i)) - \widehat{F}_{Y_j}(y_j(\ell))}{H_j}\right),$$

14

is the kernel estimate of the copula density $c_{\boldsymbol{Y}}(\cdot)$, and $\widehat{F}_{Y_j}(\cdot)$, $\forall j = 1, \ldots, p$, is the estimate of the marginal distribution function $F_{Y_j}(\cdot)$ of the random variable Y_j, defined for any real number $r \in \mathbb{R}$, by

$$\widehat{F}_{Y_j}(r) := \frac{1}{N} \sum_{\ell=1}^{N} K\left(\frac{r - y_j(\ell)}{h_j}\right),$$

where $K(\cdot)$ is the primitive of a kernel $k(\cdot)$, a symmetric centered probability density. In our forthcoming simulation study, we will take as kernel $k(\cdot)$ a standard Gaussian probability density. A more appropriate choice of the kernel $k(\cdot)$, for estimating the copula density, can be done according to [25], which copes with the boundary effect. The bandwidth parameters $H_1, \ldots, H_p$ and $h_1, \ldots, h_p$ will be chosen according to Silverman's rule of thumb, see [26], i.e., for all $j = 1, \ldots, p$, we take $H_j = \left(\frac{4}{p+2}\right)^{\frac{1}{p+4}} N^{\frac{-1}{p+4}} \widehat{\Sigma}_j$, and $h_j = \left(\frac{4}{3}\right)^{\frac{1}{5}} N^{\frac{-1}{5}} \widehat{\sigma}_j$, where $\widehat{\Sigma}_j$ and $\widehat{\sigma}_j$ are, respectively, the empirical standard deviation of the data $\widehat{F}_{Y_j}(y_j(1)), \ldots, \widehat{F}_{Y_j}(y_j(N))$ and $y_j(1), \ldots, y_j(N)$. The source vector signal $\boldsymbol{s}(i), i = 1, \ldots, N$, will be then estimated by

$$\widehat{\boldsymbol{s}}(i) = \widehat{\boldsymbol{B}}\,\boldsymbol{x}(i), i = 1, \ldots, N,$$

where

$$\widehat{\boldsymbol{B}} := \arg\inf_{\boldsymbol{B}} \widehat{\mathcal{J}^{\mathrm{ind}}}(\boldsymbol{B}),$$

which can be computed using a gradient descent type algorithm. In fact, straightforward computation shows that, the gradient in $\boldsymbol{B}$ of the estimated criterion $\boldsymbol{B} \mapsto \widehat{\mathcal{J}^{\mathrm{ind}}}(\boldsymbol{B})$, can be written as $\dfrac{d\widehat{\mathcal{J}^{\mathrm{ind}}}(\boldsymbol{B})}{d\boldsymbol{B}} =$

$$\frac{1}{N} \sum_{i=1}^{N} \frac{\frac{d}{dB}\widehat{c}_{\boldsymbol{Y}}(\boldsymbol{u}(i))}{\widehat{c}_{\boldsymbol{Y}}(\boldsymbol{u}(i))} + \frac{\gamma}{N} \sum_{i=1}^{N} (\boldsymbol{y}(i) - \overline{\boldsymbol{y}}(i))(\boldsymbol{x}(i) - \overline{\boldsymbol{x}}(i))^{\top} + \frac{\mu}{N} \sum_{i=1}^{N} \left(\frac{\nabla \boldsymbol{y}(i)}{|\nabla \boldsymbol{y}(i)|}\right)^{*} \boldsymbol{x}(i)^{\top} \qquad (39)$$

where, $\dfrac{d}{d\boldsymbol{B}} := \left(\dfrac{\partial}{\partial \boldsymbol{B}_{l,j}}\right)$, $l, j = 1, \ldots, p$, $\boldsymbol{u}(i) := (\widehat{F}_{Y_1}(y_1(i)), \ldots, \widehat{F}_{Y_p}(y_p(i)))^{\top}$ and,

$$\frac{\partial \widehat{c}_{\boldsymbol{Y}}(\widehat{F}_{Y_1}(y(i)), \ldots, \widehat{F}_{Y_p}(y_p(i)))}{\partial \boldsymbol{B}_{l,j}} = \frac{1}{NH_1 \cdots H_p} \sum_{m=1}^{N} \prod_{j=1, j\neq i}^{p} k\left(\frac{\widehat{F}_{Y_j}(y_j(m)) - \widehat{F}_{Y_j}(y_j(i))}{H_j}\right)$$
$$\times k'\left(\frac{\widehat{F}_{Y_j}(y_j(m)) - \widehat{F}_{Y_j}(y_j(i))}{H_j}\right) \frac{1}{H_j} \frac{\partial(\widehat{F}_{Y_j}(y_j(m)) - \widehat{F}_{Y_j}(y_j(i)))}{\partial \boldsymbol{B}_{l,j}}, \qquad (40)$$

with

$$\frac{\partial(\widehat{F}_{Y_j}(y_j(m)))}{\partial \boldsymbol{B}_{l,j}} = \frac{1}{Nh_j}\sum_{i=1}^{N}k\left(\frac{y_j(i)-y_j(m)}{h_j}\right)(x_j(i)-x_j(m)). \tag{41}$$

We can then derive the following Algorithm 2.

Algorithm 2 BSS algorithm for independent source components.

Data: $\overline{\boldsymbol{x}}$ the observed signal

Result: $\widehat{\boldsymbol{s}}$ the estimated source signal

Initialization: Compute $\boldsymbol{x} = \overline{\boldsymbol{x}} - \prod_D \overline{\boldsymbol{x}}$ from Algorithm 1, $\boldsymbol{B}^{(0)} = \boldsymbol{I}_p$, $\boldsymbol{y}^{(0)} = \boldsymbol{B}^{(0)}\boldsymbol{x}$. Given $\varepsilon > 0$, $\nu > 0$.

Do: Update $\boldsymbol{B}$ and $\boldsymbol{y}$:

$$\boldsymbol{B}^{(q+1)} = \boldsymbol{B}^{(q)} - \nu\frac{d\widehat{\mathcal{J}^{\mathrm{ind}}}(\boldsymbol{B})}{d\boldsymbol{B}}$$

$\boldsymbol{y}^{(q+1)} = \boldsymbol{B}^{(q+1)}\boldsymbol{x}$.

Until $||\boldsymbol{B}^{(q+1)} - \boldsymbol{B}^{(q)}|| < \epsilon$

$\widehat{\boldsymbol{s}} = \boldsymbol{y}^{(q+1)}$.

5.2.2. The case of dependent source components

Recall that, in the case where the source components are dependent, the criterion function (15) is defined by

$$\boldsymbol{B} \mapsto \mathcal{J}^{\mathrm{dep}}(\boldsymbol{B}) := \mathcal{J}_{\mathrm{sep}}^{\mathrm{dep}}(\boldsymbol{B}) + \mathcal{J}_{\mathrm{reg}}(\boldsymbol{y}), \tag{42}$$

where $\mathcal{J}^{\mathrm{dep}}(\boldsymbol{B})$ is defined by, see (29),

$$\mathcal{J}_{\mathrm{sep}}^{\mathrm{dep}}(\boldsymbol{B}) := \int_{[0,1]^p}\log\left(\frac{c_{\boldsymbol{Y}}(\boldsymbol{u})}{c_{\widehat{\theta}}(\boldsymbol{u})}\right)c_{\boldsymbol{Y}}(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} := \mathbb{E}\left(\log\frac{c_{\boldsymbol{Y}}(F_{Y_1}(Y_1),\ldots,F_{Y_p}(Y_p))}{c_{\widehat{\theta}}(F_{Y_1}(Y_1),\ldots,F_{Y_p}(Y_p))}\right).$$

and

$$\mathcal{J}_{\mathrm{reg}}(\boldsymbol{y}) := \gamma\frac{1}{2T}\int_0^T\|\boldsymbol{y}(t)-\overline{\boldsymbol{y}}(t)\|^2\,dt + \mu\frac{1}{T}\int_0^T|\nabla\boldsymbol{y}(t)|\,dt,\ \gamma > 0,\ \mu > 0.$$

Using the stochastic modeling (17) and the relation (29), we propose to approximate the criterion (42) by

$$\boldsymbol{B} \mapsto \widehat{\mathcal{J}^{\mathrm{dep}}}(\boldsymbol{B}) := \widehat{\mathcal{J}_{\mathrm{sep}}^{\mathrm{dep}}}(\boldsymbol{B}) + \mathcal{J}_{\mathrm{reg,d}}(\boldsymbol{y}), \tag{43}$$

where

$$\mathcal{J}_{\mathrm{reg,d}}(\boldsymbol{y}) := \frac{\gamma}{2N}\sum_{i=1}^{N}\|\boldsymbol{y}(i)-\overline{\boldsymbol{y}}(i)\|^2 + \frac{\mu}{N}\sum_{i=1}^{N}|\nabla\boldsymbol{y}(i)| \tag{44}$$

is the discrete version of $\mathcal{J}_{\text{reg}}(\boldsymbol{y})$, and $\boldsymbol{B} \mapsto \widehat{\mathcal{J}_{\text{sep}}^{\text{dep}}}(\boldsymbol{B})$ is the statistical estimate, of the criterion $\boldsymbol{B} \mapsto \mathcal{J}_{\text{sep}}^{\text{dep}}(\boldsymbol{B})$, which we define as follows

$$\boldsymbol{B} \mapsto \widehat{\mathcal{J}_{\text{sep}}^{\text{dep}}}(\boldsymbol{B}) := \frac{1}{N}\sum_{i=1}^{N} \log\left(\frac{\widehat{c}_{\boldsymbol{Y}}(\widehat{F}_{Y_1}(y(i)), \ldots, \widehat{F}_{Y_p}(y_p(i)))}{\widehat{c}_{\widehat{\theta}}(\widehat{F}_{Y_1}(y(i)), \ldots, \widehat{F}_{Y_p}(y_p(i)))}\right). \tag{45}$$

The source vector signal $\boldsymbol{s}(i), i = 1, \ldots, N$, will be then estimated by

$$\widehat{\boldsymbol{s}}(i) = \widehat{\boldsymbol{B}}\,\boldsymbol{x}(i), i = 1, \ldots, N,$$

where

$$\widehat{\boldsymbol{B}} := \arg\inf_{\boldsymbol{B}} \widehat{\mathcal{J}^{\text{dep}}}(\boldsymbol{B}),$$

which can be computed using a gradient descent type algorithm. In fact, straightforward computation shows that, the gradient in $\boldsymbol{B}$ of the estimated criterion $\boldsymbol{B} \mapsto \widehat{\mathcal{J}^{\text{dep}}}(\boldsymbol{B})$, can be written as $\dfrac{d\widehat{\mathcal{J}}_{\text{dep}}(\boldsymbol{B})}{d\boldsymbol{B}} =$

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\frac{d}{d\boldsymbol{B}}\widehat{c}_{\boldsymbol{Y}}(\boldsymbol{u}(i))}{\widehat{c}_{\boldsymbol{Y}}(\boldsymbol{u}(i))}\,\frac{\widehat{c}_{\widehat{\theta}}(\boldsymbol{u}(i))}{\frac{d}{d\boldsymbol{B}}\widehat{c}_{\widehat{\theta}}(\boldsymbol{u}(i))} + \frac{\gamma}{N}\sum_{i=1}^{N}(\boldsymbol{y}(i) - \overline{\boldsymbol{y}}(i))(\boldsymbol{x}(i) - \overline{\boldsymbol{x}}(i))^{\top} + \frac{\mu}{N}\sum_{i=1}^{N}\left(\frac{\nabla\boldsymbol{y}(i)}{|\nabla\boldsymbol{y}(i)|}\right)^{*}\boldsymbol{x}(i)^{\top}, \tag{46}$$

where $\boldsymbol{u}(i) := (\widehat{F}_{Y_1}(y(i)), \ldots, \widehat{F}_{Y_p}(y_p(i)))$; the gradients $\frac{d}{d\boldsymbol{B}}\widehat{c}_{\boldsymbol{Y}}(\boldsymbol{u}(i))$ and $\frac{d}{d\boldsymbol{B}}\widehat{c}_{\widehat{\theta}}(\boldsymbol{u}(i))$ can be explicitly computed in a similar way as in Subsection 5.2.1 above. We obtain then the following Algorithm 3.

Algorithm 3 BSS algorithm for dependent source components.

    Data: $\overline{\boldsymbol{x}}$ the observed signal

    Result: $\widehat{\boldsymbol{s}}$ the estimated source signal

    Initialization: Compute $\boldsymbol{x} = \overline{\boldsymbol{x}} - \prod_{D}\overline{\boldsymbol{x}}$ from Algorithm 1, $\boldsymbol{B}^{(0)} = \boldsymbol{I}_p$, $\boldsymbol{y}^{(0)} = \boldsymbol{B}^{(0)}\boldsymbol{x}$. Given $\varepsilon > 0, \nu > 0$.

    Do: Update $\boldsymbol{B}$ and $\boldsymbol{y}$:

        $\boldsymbol{B}^{(q+1)} = \boldsymbol{B}^{(q)} - \nu\dfrac{d\widehat{\mathcal{J}}_{\text{dep}}(\boldsymbol{B})}{d\boldsymbol{B}}$

        $\boldsymbol{y}^{(q+1)} = \boldsymbol{B}^{(q+1)}\boldsymbol{x}$.

    Until $||\boldsymbol{B}^{(q+1)} - \boldsymbol{B}^{(q)}|| < \epsilon$

        $\widehat{\boldsymbol{s}} = \boldsymbol{y}^{(q+1)}$.

---

6. Simulation results

In this section, we present simulation results illustrating the performance of our proposed method. For simplicity, we will deal with only the case of two mixtures/two sources. We will consider many examples of independent/dependent source component signals. The results will be compared with those of [16] under the same conditions. Recall that the separating algorithm proposed in [16] is based on minimizing a kernel type estimate of mutual information between probability densities, penalized by the same TV-regularization term. In all examples the number of samples is $N = 2000$. The two sources are mixed with the matrix $\boldsymbol{A} := [1\ 0.5; 0.5\ 1]$. A centered gaussian noise with standard deviation 0.01 was added to the normalized mixtures, so that the signal-to-noise ratio equals -20 dB. The gradient descent parameter is taken to be $\nu = 0.1$. In the denoising step, see Algorithm 1, we take $\rho = 0.1$, $\lambda = 0.01$, and $\epsilon = 0.001$. In the second step, see Algorithm 2 and Algorithm 3, we chose $\gamma = 0.001$ and $\mu = 0.01$. All simulations are repeated 100 times, and the accuracy of source estimation is evaluated through the following criterion, called again the signal-to-noise-ratio (SNR), defined by

$$SNR_i := 10 \log_{10} \frac{\sum_{n=1}^{N} s_i(n)^2}{\sum_{n=1}^{N} (\widehat{s}_i(n) - s_i(n))^2}, i = 1, 2. \tag{47}$$

6.1. Examples where the source components are independent

We consider two noisy mixed signals of two kinds of sample sources: uniform i.i.d with independent components (Fig. 1); i.i.d sources with independent components drawn from the 4-ASK (Amplitude Shift Keying) alphabets ($-3, -1, 1, 3$ with the same weights 0.25) at which was added a centered gaussian random variable with variance equals 0.25 (Fig. 2). We observe from Fig. 1 and Fig. 2, that our proposed method (Algorithm 2) gives accurate estimation of sources in the standard case of independent component sources, with performance around 27 dB in term of SNR. Fig. 3 and Fig. 4 shows the criterion values versus iterations. Observe that the criterion values $\widehat{\mathcal{J}^{\mathrm{Ind}}}(\cdot)$ converges to 0 when the separation is achieved. In Fig. 11 and Fig. 12, we compare Algorithm 2 with the MI-TV presented in [16], to separate the same instantaneous mixtures with independent sources. We can see, from figures 11 and 12, that the two approches are equivalent, in this standard case, as expected.

6.2. Examples where the source components are dependent

In this subsection, we show the capability of the proposed algorithm using BSS-copula-TV (Algorithm 3) to successfully separate two noisy mixed source signals with dependent components.

We dealt with noisy instantaneous mixtures of five kinds of sample sources: i.i.d. random vector sources (with uniform marginals in Fig. 5, and binary phase-shift keying (BPSK) marginals in Fig. 6), with dependent components generated from Ali-Mikhail-Haq (AMH) copula model where $\widehat{\theta} = 0.6$. i.i.d. vector sources (with uniform marginals in Fig. 7, and binary phase-shift keying (BPSK) marginals in Fig. 8) where the components are dependent, generated from Clayton copula with $\widehat{\theta} = 1$. i.i.d vector sources (with uniform marginals) with dependent components, generated from Fairlie-Gumbel-Morgenstern (FGM) copula model where $\widehat{\theta} = 0.8$ (Fig. 9). It can be seen from the results, in Fig. 5 to Fig. 9, that the proposed method is able to separate, with good performance (with level equivalent to that of independent source case, around 27 dB). Fig. 10 shows the convergence of the criterion $\widehat{\mathcal{J}}_{\mathrm{Dep}}(\cdot)$ to 0 when the separation is achieved for the last example. We compare, for the following two examples of dependent mixtures, the present approach (Algorithm 2) and the MI-TV of [16]. We consider i.i.d. vector sources (with uniform marginals in Fig. 14, and BPSK marginals in Fig. 15) where the component are dependent, generated from FGM copula (with $\widehat{\theta} = 0.8$). We can see, clearly, that separating dependent sources - with classical criterion devoted to the independent case - fails, and that our criterion is able to deal with the dependent case with good performance.

7. Conclusion

We have presented a new BSS approach in presence of noise, by minimizing a new regularized criterion. The approach is able to separate instantaneous linear mixtures of both independent and dependent source components. It proceeds in two steps: a preprocessing which consists in reducing the noise on the observations, and a second one minimizing a separation criterion based on Kullback-Leibler divergence between copula densities regularized through total variation. In Section 7, the accuracy and the consistency of the obtained algorithms are illustrated by simulation, for 2×2 mixture-source. It should be mentioned, however, that our proposed algorithms based on copula densities, rather than the classical ones based on probability densities, are more time consuming, since we estimate both copulas density of the vector and the marginal distribution function of each component. The present approach can be extended to deal with convolutive mixtures of independent/dependent sources, that will be addressed in future communications.

References

[1] P. Comon, Independent component analysis, a new concept?, Signal Processing 36 (3) (1994) 287 – 314, higher Order Statistics.

[2] A. Mansour, C. Jutten, A direct solution for blind separation of sources, IEEE Transactions on Signal Processing 44 (3) (1996) 746–748.

[3] A. Taleb, C. Jutten, Entropy optimization, Artificial Neural Networks - ICANN'97 (1997) 529–534.

[4] A. Hyvärinen, E. Oja, A fast fixed-point algorithm for independent component analysis, Neural Computation 9 (1997) 1483–1492.

[5] D. Pham, Blind separation of instantaneous mixture of sources based on order statistics, IEEE Transactions on Signal Processing 48 (2) (2000) 363–375.

[6] J.-C. Pesquet, E. Moreau, Cumulant-based independence measures for linear mixtures, IEEE Trans. Inform. Theory 47 (5) (2001) 1947–1956. doi:10.1109/18.930929.
URL http://dx.doi.org/10.1109/18.930929

[7] J.-F. Cardoso, Blind signal separation: statistical principles, Proceedings of the IEEE 86 (10) (1998) 2009–2025.

[8] M. Novey, T. Adali, Ica by maximization of nongaussianity using complex functions, Proc. MLSP.

[9] D. Pham, Mutual information approach to blind separation of stationary sources, IEEE Transactions on Information Theory 48 (7) (2002) 1935–1946.

[10] A. Keziou, H. Fenniri, M. Ould Mohamed, G. Delaunay, Séparations aveugle de sources par minimisation des α-divergences, in: XXIIe colloque GRETSI (traitement du signal et des images), Dijon (FRA), 8-11 septembre 2009, GRETSI, Groupe d'Etudes du Traitement du Signal et des Images, 2009.

[11] P. Comon, C. Jutten, Handbook of Blind Source Separation: Independent component analysis and applications, Academic press, 2010.

[12] M. El Rhabi, H. Fenniri, A. Keziou, E. Moreau, A robust algorithm for convolutive blind source separation in presence of noise, Signal Processing 93 (4) (2013) 818–827.

[13] A. Keziou, H. Fenniri, A. Ghazdali, E. Moreau, New blind source separation method of independent/dependent sources, Signal Processing 104 (2014) 319 – 324.

[14] M. Sahmoudi, H. Snoussi, M. G. Amin, Robust approach for blind source separation in non-gaussian noise environments, Proceedings of ISCCSP, Marrakesh, Morocco, IEEE/EURASIP.

[15] A. Belouchrani, A. Cichocki, Robust whitening procedure in blind source separation context, Electronics letters 36 (24) (2000) 2050–2051.

[16] M. El Rhabi, H. Fenniri, A. Keziou, E. Moreau, A robust algorithm for convolutive blind source separation in presence of noise, Signal Processing 93 (4) (2013) 818 – 827.

[17] A. Sklar, Fonctions de répartition à n dimensions et leurs marges, Publ. Inst. Statist. Univ. Paris 8 (1959) 229–231.

[18] R. B. Nelsen, An introduction to copulas, 2nd Edition, Springer Series in Statistics, Springer, New York, 2006.

[19] H. Joe, Multivariate models and dependence concepts, Vol. 73 of Monographs on Statistics and Applied Probability, Chapman & Hall, London, 1997. doi:10.1201/b13150.
URL http://dx.doi.org/10.1201/b13150

[20] L. Evans, R. Gariepy, Measure theory and fine properties of function, Studies in Advanced Mathematics, CRC Press, Boca Raton, FL.

[21] A. Cohen, D. Wolfgang, , I. Daubechies, R. DeVore, Harmonic analysis of the space bv, Rev. Mat. Iberoamericana 19(1) (2003) 235–263.

[22] X. Chen, Y. Fan, Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification, J. Econometrics 135 (1-2) (2006) 125–154.

[23] H. Tsukahara, Semiparametric estimation in copula models, Canad. J. Statist. 33 (3) (2005) 357–375. doi:10.1002/cjs.5540330304.
URL http://dx.doi.org/10.1002/cjs.5540330304

[24] A. Chambolle, An algorithm for total variation minimization and applications, Journal of Mathematical imaging and vision 20 (1-2) (2004) 89–97.

[25] M. Omelka, I. Gijbels, N. Veraverbeke, Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing, Ann. Statist. 37 (5B) (2009) 3023–3058.

[26] B. W. Silverman, Density estimation for statistics and data analysis, Monographs on Statistics and Applied Probability, Chapman & Hall, London, 1986.
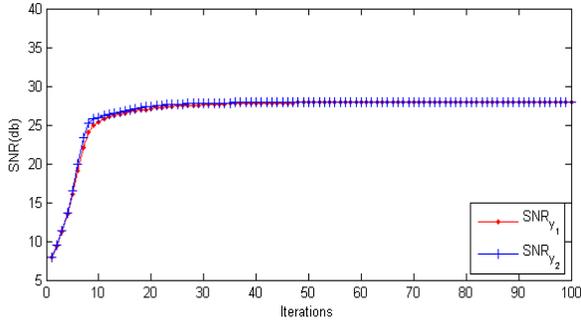
Figure 1: Average output SNRs versus iteration number : Uniform independent sources.
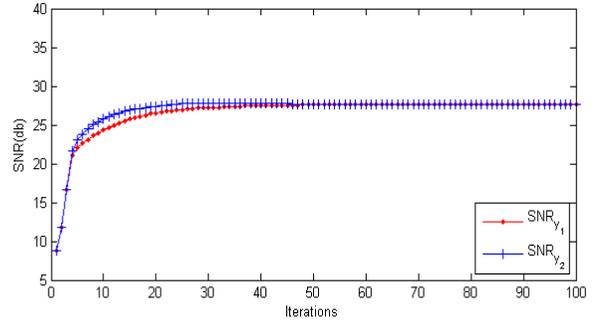
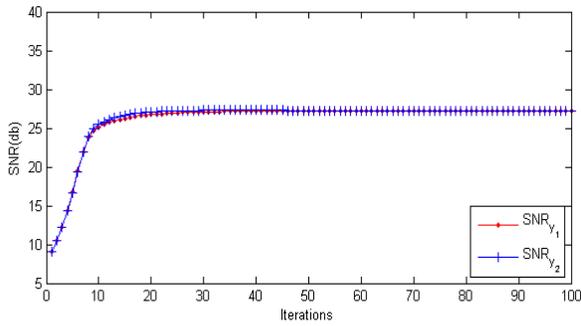

Figure 2: Average output SNRs versus iteration number : ASK independent sources.
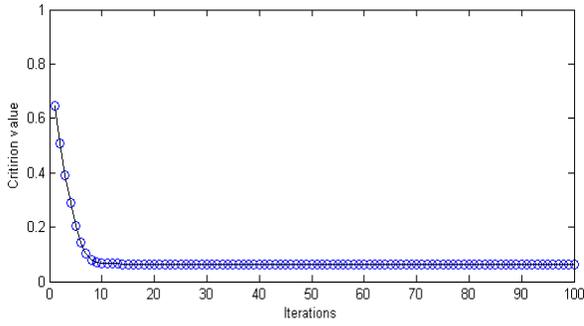
Figure 3: The criterion value vs iterations : uniform independent sources.
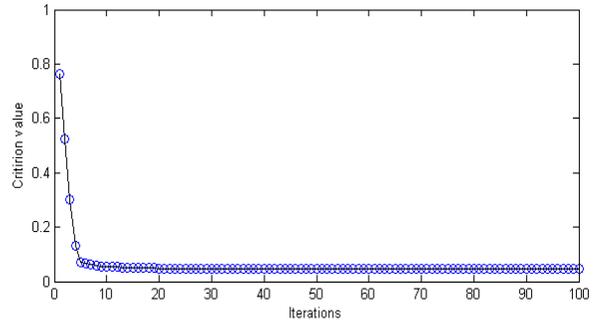
Figure 4: The criterion value vs iterations : Bpsk independent sources.
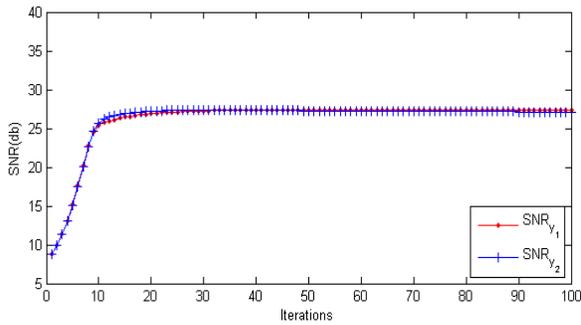
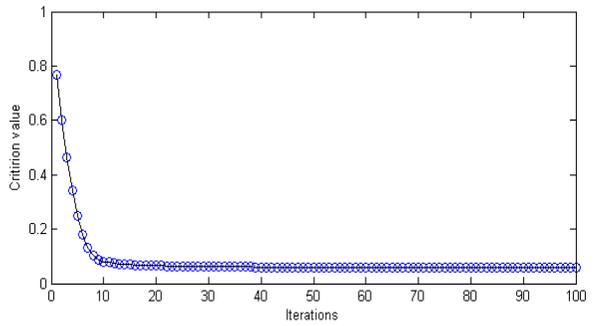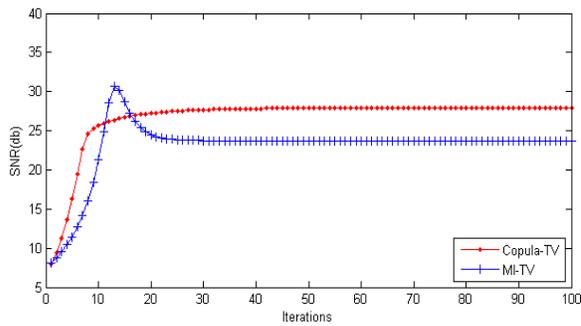Figure 5: Average output SNRs versus iteration number : Uniform dependent sources from AMH-copula.
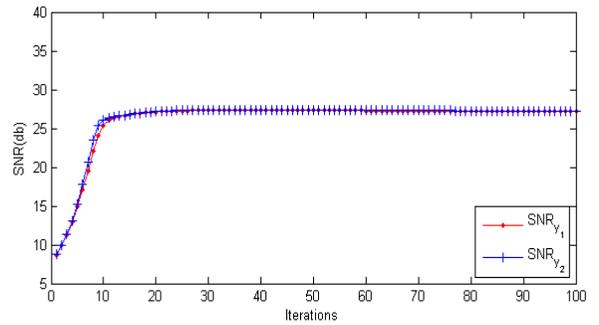
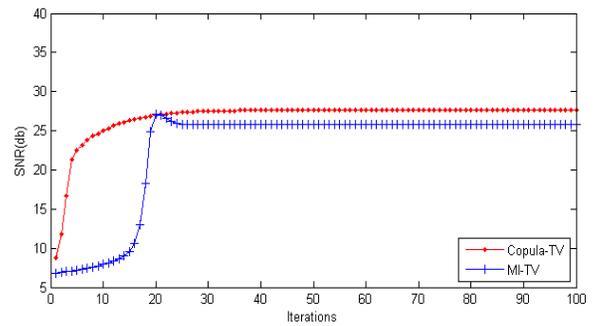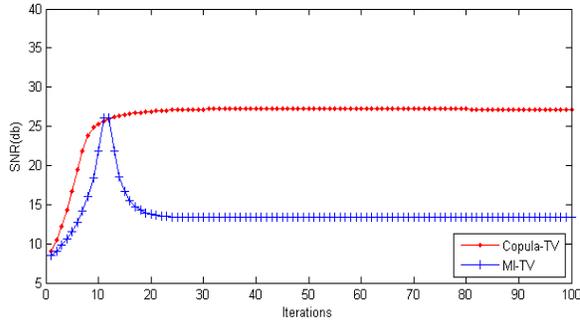Figure 6: Average output SNRs versus iteration number : Bpsk dependent sources from AMH-copula.
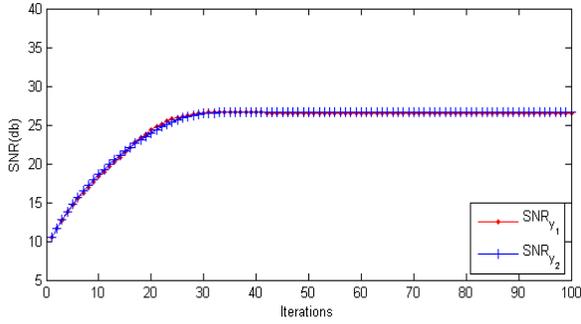
Figure 7: Average output SNRs versus iteration number : Uniform dependent sources from Clayton-copula.
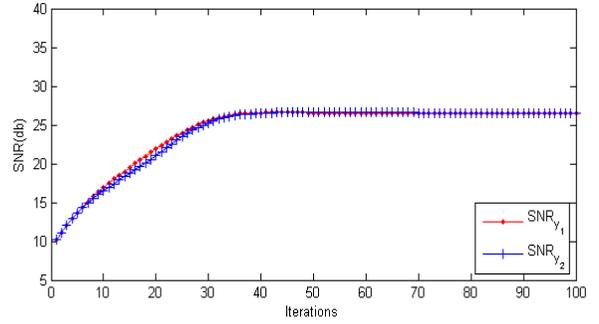
Figure 8: Average output SNRs versus iteration number : Bpsk dependent sources from Clayton-copula.

Figure 9: Average output SNRs versus iteration number : Uniform dependent sources from FGM-copula.

Figure 10: The criterion value vs iterations : Uniform dependent sources from FGM-copula.

Figure 11: Average output SNRs versus iteration number: Uniform independent sources.

Figure 12: Average output SNRs versus iteration number: Bpsk independent sources.

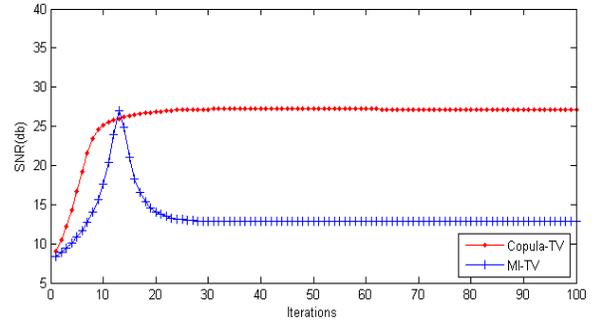Figure 13: Average output SNRs versus iteration number: Uniform dependent sources from FGM-copula.

Figure 14: Average output SNRs versus iteration number: Bpsk dependent sources from FGM-copula.
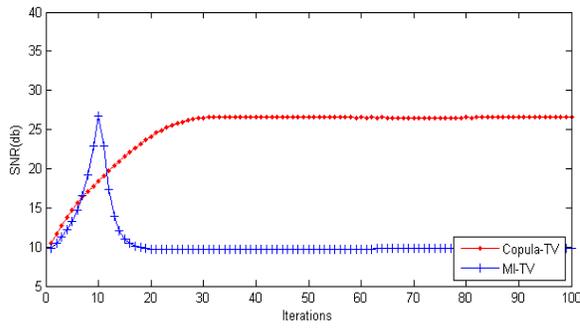

Figure 15: Average output SNRs versus iteration number: Uniform dependent sources from Clayton-copula.
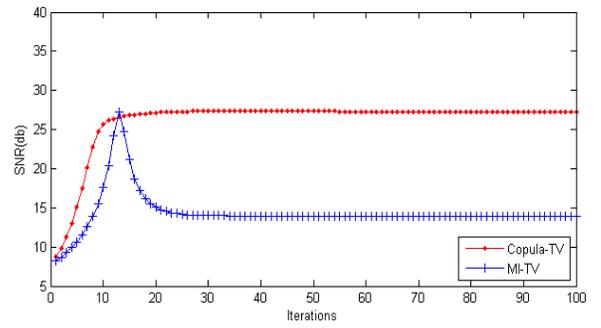

Figure 16: Average output SNRs versus iteration number: Bpsk dependent sources from AMH-copula.