

BLIND SOURCE SEPARATION USING HELLINGER DIVERGENCE AND COPULAS

AMAL OURDOU*, ABDELGHANI GHAZDALI AND ABDELMOUTALIB METRANE

Abstract. Whenever there is a mixture of signals of any type, *e.g.* sounds, images or any other form of source signals, Blind Source Separation (BSS) is the method utilized to separate these signals from the observations. The separation is done without any prior knowledge about the mixing process nor the source signals. In literature multiple algorithms have been deployed for this particular problem, however most of them depends on Independent Component Analysis (ICA) and its variations assuming the statistical independence of the sources. In this paper, we develop a new algorithm improving the separation quality for both independent and dependent sources. Our algorithm used copulas to accurately model the dependency structure and the Hellinger divergence as a distance measure since it can converge faster and it is robust against noisy source signals. Many simulations were conducted for various samples of sources to illustrate the superiority of our approach compared to other methods.

Mathematics Subject Classification. 62H05, 92C55.

Received December 27, 2021. Accepted July 28, 2022.

1. INTRODUCTION

Blind Source Separation is the technique used to extract sources, from observations of their mixtures without the knowledge of the original signals or the mixing process. Signal processing and Machine learning communities have widely explored the challenges in BSS during the last three decades. It was first used for the cocktail party problem, where the aim was to separate the sound signals of each person's speech. Then exploited in other scientific fields such as signal processing, image processing, medical signal processing, artificial neural networks, statistics, and information theory, speech recognition systems, telecommunications.

BSS is an ill posed problem, therefore, in literature, various assumptions on the sources has been made to enable the separation of the observed mixtures. For the linear and static mixing environment, Independent Component Analysis (ICA) [9] is used. It considers the sources to be mutually independent and non-Gaussian. Under these assumptions, the source signals can be estimated by optimizing a cost function. Numerous variations of ICA were introduced to literature citing for examples, maximizing likelihood [5], minimizing the mutual information [13, 29], minimizing the criteria of ϕ -divergences [14, 18], the second or higher order statistics [7, 28], etc. A good overview on the problem can be found in [10].

Keywords. Blind source separation, Hellinger divergence, copulas, dependent sources, noise-contaminated sources.

LIPIM, ENSA Khouribga, Sultan Moulay Slimane University, Khouribga, Morocco.

*Corresponding author: amal.ourdou@gmail.com

In [17, 18, 22] a new BSS algorithm was proposed to overcome the drawbacks of ICA techniques. This algorithm uses Copula to accurately model the dependency structure between the source components, hence omitting the mutual independence assumption. In this paper we make use of this copula and focus on the Hellinger divergence between the copula densities as our cost function to minimize, due to its efficiency and robustness in improving the results even for noisy data [4, 23], moreover one of its main characteristics is its rapid convergence compared to any other divergences.

This paper is organized as follows, section “Blind Source Separation” gives an overview on the BSS principle and model, then a review on copula in section “What are copulas?”. After that in section “Hellinger divergence and copula” we present our cost function as the Hellinger divergence between the copula densities. We then introduce our new approach detailing separately the independent and the dependent cases, in section “The proposed approach”. Then the comparison between our approach and various other methods is made, illustrating that the superiority of our approach in section “Simulation results”. Finally, we conclude the paper and give some further research directions.

2. BLIND SOURCE SEPARATION

The linear BSS problem states that the n unknown source components $\mathbf{s}(t) \in \mathbb{R}^n$, are blindly mixed together through a matrix \mathbf{M} containing the mixing coefficients. In vector notations

$$\mathbf{x}(t) := \mathbf{M}\mathbf{s}(t) + \mathbf{b}(t), \quad t \in \mathbb{R}, \quad (2.1)$$

where $\mathbf{x}(t) \in \mathbb{R}^p$ are the observations, and $\mathbf{b}(t)$ is an additive noise. In our work we consider the determine case where the number of sources is the same as the number of the observations *e.g.* $p = n$ and that the additive noise is omitted using a pre-processing technique [14]. The new BSS model is as follows:

$$\mathbf{x}(t) := \mathbf{M}\mathbf{s}(t), \quad \forall t \in \mathbb{R}, \quad (2.2)$$

having only the observed signals $\mathbf{x}(t)$, and no prior knowledge about the mixing process, the BSS solution searches for the optimum $p \times p$ un-mixing matrix \mathbf{B} , which gives the recovered sources

$$\mathbf{y}(t) := \mathbf{B}\mathbf{x}(t), \quad \forall t \in \mathbb{R}. \quad (2.3)$$

Where $\mathbf{y}(t)$ is the estimated sources that would be similar to the wanted sources $\mathbf{s}(t)$ if the un-mixing matrix \mathbf{B} is as close as possible to \mathbf{M}^{-1} .

3. WHAT ARE COPULAS?

Copulas has become very popular recently as a method to model the dependency structure of random variables. We can define copula as the function that helps us to connect univariate marginal distributions to a joint multivariate distribution function with a specific form of dependency. The Sklar’s theorem [32], which is the fundamental theorem for copulas, affirm the copula function’s existence, which it is of the form:

$$F(\mathbf{Z}) = \mathbb{C}_{\mathbf{Z}}(F_1(Z_1), \dots, F_p(Z_p)), \forall \mathbf{Z} := (Z_1, \dots, Z_p)^\top \in \mathbb{R}^p. \quad (3.1)$$

Where F is an p -dimensional distribution function with marginals F_1, \dots, F_p . $\mathbb{C}_{\mathbf{Z}}(\cdot)$ is the copula function which is also a joint distribution function on $[0, 1]^p$ in itself, with uniform margins. We have the following:

$$\begin{aligned} \forall \mathbf{u} := (u_1, \dots, u_p)^\top \in [0, 1]^p, \quad \mathbb{C}_{\mathbf{Z}}(\mathbf{u}) = \\ \mathbb{P}(F_1(Z_1) \leq u_1, \dots, F_p(Z_p) \leq u_p). \end{aligned}$$

If F_1, \dots, F_p are all continuous, then $\mathbb{C}_{\mathbf{Z}}(\cdot)$ is unique. In the opposite direction, consider a copula, $\mathbb{C}_{\mathbf{Z}}(\cdot)$, and univariate distribution functions, F_1, \dots, F_n . Then F as defined in (3.1) is a joint multivariate distribution function with marginals F_1, \dots, F_n .

TABLE 1. Examples of semiparametric copulas.

Family	$\mathbb{C}(u_1, u_2, \dots, u_p, \theta)$	Θ	θ_0
AMH	$\frac{\prod_{i=1}^p u_i}{1 - \theta \left(\prod_{i=1}^p (1 - u_i) \right)}$	$[-1, 1]$	0
Clayton	$\max \left[\left(\sum_{i=1}^p u_i^{-\theta} - p + 1 \right), 0 \right]^{-\frac{1}{\theta}}$	$[-1, +\infty[\setminus \{0\}]$	0
Frank	$-\frac{1}{\theta} \ln \left(1 + \frac{\prod_{i=1}^p (e^{-\theta u_i} - 1)}{(e^{-\theta} - 1)^{p-1}} \right)$	$\mathbb{R} \setminus \{0\}$	0

For the case, where the components of a random vector variable, $\mathbf{Z} := (Z_1, \dots, Z_p)^\top \in \mathbb{R}^p$ are statistically independent, we have the copula of independence denoted $\mathbb{C}_\Pi(\cdot)$ of the form:

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

If the copula has a density, then it is obtained in the following manner as

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

Using the last formula we can obtain the density of the copula of independence as follows:

$$c_\Pi(\mathbf{u}) := 1_{[0,1]^p}(\mathbf{u}), \quad \forall \mathbf{u} \in [0, 1]^p. \quad (3.2)$$

For the the random vector $\mathbf{Z} := (Z_1, \dots, Z_p)^\top$, let $f_\mathbf{Z}(\cdot)$ be its probability density if it exists, and $f_1(\cdot), \dots, f_p(\cdot) \in \mathbb{R}^p$ the marginal probability densities of Z_1, \dots, Z_p respectively. We can obtain the following relation after some uncomplicated computations

$$f_\mathbf{Z}(\mathbf{Z}) = \left(\prod_{j=1}^p f_j(Z_j) \right) c_\mathbf{Z}(F_1(Z_1), \dots, F_p(Z_p)). \quad (3.3)$$

Numerous models for copulas have been proposed in the literature. Semi-parametric copula models class are the most popular for modeling and estimating the structure of dependency. For this class the parametric copulas $\mathbb{C}(\cdot, \boldsymbol{\theta})$ is indexed by a parameter $\theta \in \Theta \subset \mathbb{R}^d$, with a non-parametric margins.

In Table 1 we recall a description of three models of copula: Clayton [8], Ali-Mikhail-Haq (AMH) [1] and Frank [15], which were used in Section “Simulation results” in our simulation study. We provide the respective parameter space Θ for each model and the parameter θ corresponding to the independence hypothesis of margins denoted θ_0 , in other words

$$\begin{aligned} \mathbb{C}(u_1, \dots, u_p; \boldsymbol{\theta}_0) &= \mathbb{C}_\Pi(u_1, \dots, u_p) \\ &:= \prod_{i=1}^p u_i, \quad \forall (u_1, \dots, u_p)^\top \in [0, 1]^p. \end{aligned} \quad (3.4)$$

For a better understanding on the widely used semi-parametric copulas, one may refer to [21, 25].

In the following lines, we outline briefly one of the copula model selection procedures and the method of estimating the parameter θ from the data. For a random vector $\mathbf{Z} \in \mathbb{R}^p$, let's assume that a training sample of Z is available, that is, we dispose of i.i.d. realizations $\mathbf{z}(1), \dots, \mathbf{z}(N)$ of \mathbf{Z} .

The objective is to select the “best” copula model from the data, among a list of candidate models, that models the dependence structure of the components \mathbf{Z} , and to estimate the parameter θ of the model selected. Let $\{C_1(\cdot, \theta_1); \theta_1 \in \Theta_1 \subset \mathbb{R}^{d_1}\}, \dots, \{C_K(\cdot, \theta_K); \theta_K \in \Theta_K \subset \mathbb{R}^{d_K}\}$ be a list of candidate copula models. The selection of models can be done using the Bayesian information criterion (BIC) [30], resulting from the semi-parametric log-likelihood, see *e.g.* [16] and [33]. Denote by $c_k(\cdot, \theta_k)$ the density of the copula $\mathbb{C}_k(\cdot, \theta_k)$, for all k .

The BIC, of a given model k , is defined by

$$BIC(k) = -2 \sup_{\theta_k \in \Theta_k} \sum_{n=1}^N \log c_k \left(\widehat{F}_{Z_1}(\mathbf{z}_1(n)), \dots, \widehat{F}_{Z_p}(\mathbf{z}_p(n)), \theta_k \right) + d_k / \log(N),$$

The ideal model is the one which minimizes the BIC values, namely, the density copula model $\{c_{k^*}(\cdot, \theta_{k^*}); \theta_{k^*} \in \Theta_{k^*} \subset \mathbb{R}^{d_{k^*}}\}$ where

$$k^* = \arg \min_{k \in \{1, \dots, K\}} BIC(k).$$

Denote, simply, $\{c(\cdot, \theta); \theta \in \Theta \subset \mathbb{R}^d\}$ a selected model according to the above procedure. The parameter θ of the copula model in question can be estimated by maximizing the semi-parametric log-likelihood

$$\widehat{\theta} = \arg \sup_{\theta \in \Theta} \sum_{n=1}^N \log c \left(\widehat{F}_{Z_1}(\mathbf{z}_1(n)), \dots, \widehat{F}_{Z_p}(\mathbf{z}_p(n)), \theta \right).$$

4. HELLINGER DIVERGENCE AND COPULA

One of the important issues in many applications of probability theory is finding an appropriate measure of distance between two probability distributions. A number of divergence measures for this purpose have been studied. In this paper, we singled out Hellinger divergence [11] as the measure of instantaneous information because it improves the maximum likelihood in terms of efficiency-robustness for noisy data. It also converges faster than other divergences, see, *e.g.*, [4, 20].

The Hellinger distance denoted H between two probability density functions is defined through

$$H(q, p) := \int_{\mathbb{R}^M} 2 \left(\sqrt{\frac{q(\mathbf{t})}{p(\mathbf{t})}} - 1 \right)^2 p(\mathbf{t}) d\mathbf{t}. \quad (4.1)$$

where p and q are two probabilities on \mathbb{R}^M and q is absolutely continuous with respect to p . Note that the function $q \rightarrow H(q, p)$ is convex and non-negative, for any given probability p . Furthermore, we have the following fundamental property which was proved in [12]:

$$H(q, p) = 0 \text{ iff } q = p.$$

The Hellinger distance H between the joint density $f_{\mathbf{Y}}(\cdot)$ of the random vector $\mathbf{Y} := (Y_1, \dots, Y_p)^\top \in \mathbb{R}^p$, $p \geq 1$, and the product of the the marginal densities f_{Y_i} of the components Y_i , $i \in \{1, \dots, p\}$, is

given by

$$\begin{aligned}
 H\left(\prod_{i=1}^p f_{Y_i}(y_i), f_{\mathbf{Y}}(\mathbf{y})\right) &:= \int_{\mathbb{R}^M} 2 \left(\sqrt{\frac{\prod_{i=1}^p f_{Y_i}(y_i)}{f_{\mathbf{Y}}(\mathbf{y})}} - 1 \right)^2 \\
 &\quad f_{\mathbf{Y}}(\mathbf{y}) dy_1, \dots, dy_p, \\
 &:= \mathbb{E} \left[2 \left(\sqrt{\frac{\prod_{i=1}^p f_{Y_i}(y_i)}{f_{\mathbf{Y}}(\mathbf{y})}} - 1 \right)^2 \right].
 \end{aligned} \tag{4.2}$$

where \mathbb{E} is the mathematical expectation.

Note that $H\left(\prod_{i=1}^p f_{Y_i}, f_{\mathbf{Y}}\right)$ is non-negative and reaches its minimum value zero only when the components of the random vector \mathbf{Y} are statistically independent, in other words: $\prod_{i=1}^p f_{Y_i}(\cdot) = f_{\mathbf{Y}}(\cdot)$.

From equation (4.2) and using formula (3.3), the hellinger distance $H\left(\prod_{i=1}^p f_{Y_i}, f_{\mathbf{Y}}\right)$ can be reformulated as follows

$$\begin{aligned}
 H\left(\prod_{i=1}^p f_{Y_i}, f_{\mathbf{Y}}\right) &:= \int_{[0,1]^p} 2 \left(\sqrt{\frac{1}{c_{\mathbf{Y}}(\mathbf{u})}} - 1 \right)^2 c_{\mathbf{Y}}(\mathbf{u}) d\mathbf{u}, \\
 &:= \mathbb{E} \left[2 \left(\sqrt{\frac{1}{c_{\mathbf{Y}}(\mathbf{u})}} - 1 \right)^2 \right].
 \end{aligned} \tag{4.3}$$

This last equation implies that the Hellinger distance between the product of the marginal densities and the joint density of the random vector \mathbf{Y} can be also defined as the Hellinger distance between the copula density of independence c_{Π} , and copula density $c_{\mathbf{Y}}$ of the random vector \mathbf{Y}

$$H\left(\prod_{i=1}^p f_{Y_i}, f_{\mathbf{Y}}\right) := H(c_{\Pi}, c_{\mathbf{Y}}). \tag{4.4}$$

5. THE PROPOSED APPROACH

Before going into details let us first present the discrete version of the BSS problem. Considering the source signals $\mathbf{s}(n)$, $n = 1, \dots, N$ as N copies of the random source vector \mathbf{S} the equation (2.2) will take the following form:

$$\mathbf{X} := \mathbf{M}\mathbf{S}, \quad n = 1, \dots, N. \tag{5.1}$$

Hence, $\mathbf{y}(n) := \mathbf{B}\mathbf{x}(n)$, $n = 1, \dots, N$ is, N copies of the random source vector $\mathbf{Y} := \mathbf{B}\mathbf{X}$.

5.1. A separation procedure for independent sources.

As shown in the previous section the Hellinger distance $H(c_{\Pi}, c_{\mathbf{Y}})$ between the copula density of independence and the copula density of the random variable \mathbf{Y} is always positive and only achieve its minimum zero if the components of \mathbf{Y} are statistically independent and the un-mixing matrix $\mathbf{B} = \mathbf{D}\mathbf{P}\mathbf{M}^{-1}$, where \mathbf{D} and \mathbf{P} are, a diagonal and permutation matrix respectively.

For a successful separation, the idea is to minimize an estimate $\hat{H}(\mathbf{c}_\Pi, \mathbf{c}_Y)$ constructed from the data $\mathbf{y}(1), \dots, \mathbf{y}(n)$. Therefore, the separation matrix is calculated in this fashion

$$\hat{\mathbf{B}} = \arg \min_{\mathbf{B}} \hat{H}(\mathbf{c}_\Pi, \mathbf{c}_Y). \quad (5.2)$$

That results in approximating the components $\hat{\mathbf{y}}(n) = \hat{\mathbf{B}} \mathbf{x}(n)$, $n = 1, \dots, N$. Considering equation (4.3), we introduce the following estimate of the distance $H(\mathbf{c}_\Pi, \mathbf{c}_Y)$ as

$$\hat{H}(\mathbf{c}_\Pi, \mathbf{c}_Y) := \frac{2}{N} \sum_{n=1}^N \left(\sqrt{\frac{1}{\hat{c}_Y(\hat{F}_{Y_1}(y_1(n)), \dots, \hat{F}_{Y_p}(y_p(n)))}} - 1 \right)^2 \quad (5.3)$$

where the kernel estimate of the copula density $c_Y(\cdot)$ is of the form

$$\hat{c}_Y(\mathbf{u}) := \frac{1}{N H_1 \dots H_p} \sum_{m=1}^N \prod_{j=1}^p k \left(\frac{\hat{F}_{Y_j}(y_j(m)) - u_j}{H_j} \right), \forall \mathbf{u} \in [0, 1]^p, \quad (5.4)$$

with $\hat{F}_{Y_j}(x)$, $j = 1, \dots, p$ the smoothed estimate of the marginal distribution functions $F_{Y_j}(x)$ for the random variable Y_j . For any real value $x \in \mathbb{R}$, $\hat{F}_{Y_j}(x)$ is defined by

$$\hat{F}_{Y_j}(x) := \frac{1}{N} \sum_{m=1}^N K \left(\frac{y_j(m) - x}{h_j} \right), \forall j = 1, \dots, p. \quad (5.5)$$

$K(\cdot)$ is a symmetric and centered probability density and the primitive of a kernel $k(\cdot)$ which we chose it to be the regular Gaussian density in this study. A more acceptable kernel choice $k(\cdot)$ that copes with the boundary effect can be rendered according to [26] to approximate the copula density.

H_1, \dots, H_p and h_1, \dots, h_p which are the bandwidth parameters seen in equations (5.4) and (5.5) are chosen with conform to the Silverman's thumb rule [31]. Hence, for all $j = 1, \dots, p$, we have:

$$\begin{cases} H_j = \left(\frac{4}{p+2} \right)^{\frac{1}{p+4}} N^{\frac{-1}{p+4}} \hat{\Sigma}_j, \\ h_j = \left(\frac{4}{3} \right)^{\frac{1}{5}} N^{\frac{-1}{5}} \hat{\sigma}_j, \end{cases} \quad (5.6)$$

where $\hat{\sigma}_j$ and $\hat{\Sigma}_j$ are, respectively, the empirical standard deviation of $y_j(1), \dots, y_j(N)$ and $\hat{F}_{Y_j}(y_j(1)), \dots, \hat{F}_{Y_j}(y_j(N))$.

We present a two-steps approach to estimate the separation matrix $\hat{\mathbf{B}}$. First, it is normal to start with a normalization stage, namely the spacial whitening, where we transform the data \mathbf{x} by a $p \times p$ -matrix \mathbf{W} such that

$$\mathbf{z}(n) = \mathbf{W} \mathbf{x}(n), \quad n = 1, \dots, N, \quad (5.7)$$

with

$$\mathcal{R}_Z = \mathbf{W} \mathcal{R}_X \mathbf{W}^T = \mathcal{I}_d \quad (5.8)$$

\mathcal{R}_Z and \mathcal{R}_X are the auto-covariances of \mathbf{Z} and \mathbf{X} successively. The spatial whitening of the observations consists in un-correlating the signals paired with a unit power constraint [3, 9].

The second step consists in applying series of Givens rotations, minimizing the estimate of the Hellinger distance. Let $\mathbf{U} \in \mathbb{R}^{p \times p}$ be a unitary matrix, satisfying $\mathbf{U} \mathbf{U}^T = \mathbf{I}_p$. This matrix can be written as $\mathbf{U}(\alpha) :=$

$\prod_{1 \leq i < l \leq p} G(i, k, \alpha_m)$, where $G(i, k, \alpha_m)$ is the $p \times p$ -matrix with the following inputs, for all $1 \leq j, l \leq p$,

$$G(i, k, \alpha_m)_{j,l} := \begin{cases} \cos(\alpha_m) & \text{if } j = i, l = i \text{ or } j = k, l = k; \\ \sin(\alpha_m) & \text{if } j = i, l = k; \\ -\sin(\alpha_m) & \text{if } j = k, l = i; \\ 1 & \text{if } j = l; \\ 0 & \text{else,} \end{cases} \quad (5.9)$$

the rotation angles $\alpha_m \in]-\pi/2, \pi/2[$, $m = 1, \dots, p(p-1)/2$, are the elements of the vector α .

The un-mixing matrix is written as follow: $\mathbf{B} = \mathbf{U}(\alpha)\mathbf{W}$ hence, the estimated sources take the upcoming form: $\mathbf{y}(n) = \mathbf{U}(\alpha)\mathbf{z}(n)$, $n = 1, \dots, N$. Accordingly, the estimate $\hat{H}(\mathbf{c}_\Pi, \mathbf{c}_\mathbf{Y})$ is a function of the parameter vector α which can be computed using a gradient descent algorithm by minimizing $\hat{\alpha} := \arg \min_{\alpha} \hat{H}(\mathbf{c}_\Pi, \mathbf{c}_\mathbf{Y})$ with respect to α . The un-mixing matrix is then estimated by

$$\hat{\mathbf{B}} = \mathbf{U}(\hat{\alpha})\mathbf{W}, \quad (5.10)$$

which results in approximating the source signals:

$$\hat{\mathbf{y}}(n) = \hat{\mathbf{B}} \mathbf{x}(n) = \mathbf{U}(\hat{\alpha})\mathbf{W} \mathbf{x}(n), \quad n = 1, \dots, N. \quad (5.11)$$

The gradient in α of $H(\mathbf{c}_\Pi, \mathbf{c}_\mathbf{Y})$ can be calculated from the proper definitions of the estimates as follows:

$$\begin{aligned} \frac{d\hat{H}(\mathbf{c}_\Pi, \mathbf{c}_\mathbf{Y})}{d\alpha} &:= -\frac{2}{N} \sum_{n=1}^N \frac{d}{d\alpha} \left(\sqrt{\frac{1}{\mathbf{c}_\mathbf{Y}(\hat{F}_{Y_1}(y(n)), \dots, \hat{F}_{Y_p}(y_p(n)))}} - 1 \right)^2 \\ &:= -\frac{2}{N} \sum_{n=1}^N \left(\frac{1}{\mathbf{c}_\mathbf{Y}(\hat{F}_{Y_1}(y(n)), \dots, \hat{F}_{Y_p}(y_p(n)))} - \sqrt{\mathbf{c}_\mathbf{Y}(\hat{F}_{Y_1}(y(n)), \dots, \hat{F}_{Y_p}(y_p(n)))} \right) \\ &\quad \times \frac{\frac{d}{d\alpha} \mathbf{c}_\mathbf{Y}(\hat{F}_{Y_1}(y(n)), \dots, \hat{F}_{Y_p}(y_p(n)))}{\mathbf{c}_\mathbf{Y}(\hat{F}_{Y_1}(y(n)), \dots, \hat{F}_{Y_p}(y_p(n)))^2} \end{aligned} \quad (5.12)$$

where,

$$\begin{aligned} \frac{d\hat{\mathbf{c}}_\mathbf{Y}(\hat{F}_{Y_1}(y(n)), \dots, \hat{F}_{Y_p}(y_p(n)))}{d\alpha} &= \frac{1}{NH_1 \dots H_p} \sum_{m=1}^N \prod_{j=1, j \neq i}^p k \left(\frac{\hat{F}_{Y_j}(y_j(m)) - \hat{F}_{Y_j}(y_j(n))}{H_j} \right) \\ &\quad \times k' \left(\frac{\hat{F}_{Y_i}(y_i(m)) - \hat{F}_{Y_i}(y_i(n))}{H_i} \right) \frac{1}{H_i} \frac{d(\hat{F}_{Y_i}(y_i(m)) - \hat{F}_{Y_i}(y_i(n)))}{d\alpha}, \end{aligned} \quad (5.13)$$

with

$$\frac{d(\hat{F}_{Y_i}(y_i(m)))}{d\alpha} = \frac{1}{Nh_i} \sum_{n=1}^N k \left(\frac{y_i(n) - y_i(m)}{h_i} \right) \frac{d}{d\alpha} \left(\frac{y_i(n) - y_i(m)}{h_i} \right), \quad (5.14)$$

with $\mathbf{y}(n) = \mathbf{U}(\alpha)\mathbf{z}(n)$, $n = 1, \dots, N$.

The following algorithm sums up the proposed approach for the separation of independent components:

Algorithm 1 The separation algorithm for independent source components.

Data: the observations $\mathbf{x}(n)$, $n = 1, \dots, N$.

Result: the estimated sources $\hat{\mathbf{y}}(n)$, $n = 1, \dots, N$.

Whitening and Initialization: $\mathbf{z}(n) := \mathbf{W}\mathbf{x}(n)$, $\hat{\mathbf{y}}_0(n) = \mathbf{U}(\hat{\alpha}_0)\mathbf{z}(n)$. Given $\varepsilon > 0$ and $\mu > 0$.

Do: • **Update** α and \mathbf{y}

$$\alpha_{k+1} = \alpha_k - \mu \frac{d\hat{H}(\mathbf{c}_{\Pi}, \mathbf{c}_{\mathbf{Y}})}{d\alpha}.$$

$$\mathbf{y}_{k+1}(n) = \mathbf{U}(\alpha_{k+1})\mathbf{z}(n), \quad n = 1, \dots, N.$$

• **Until** $\|\alpha_{k+1} - \alpha_k\| < \varepsilon$

$$\hat{\mathbf{y}}(n) = \mathbf{y}_{k+1}(n), \quad n = 1, \dots, N.$$

5.2. A separation procedure for dependent sources.

In this section we tackle the case of dependent source components, hence, we can't use the independent copula density as in the previous section. Denote by $c_{\mathcal{S}}(\cdot)$ the unknown semi-parametric copula density of \mathcal{S} . We assume that it belongs to a set of L candidate semi-parametric models, say,

$$M_l := \{c_l(\cdot; \theta_l); \theta_l \in \Theta_l \subset \mathbb{R}\}, \quad l = 1, \dots, L. \quad (5.15)$$

Table 1 gives some examples of semi-parametric copula density models. Each “semiparametric” model M_l for $l = 1, \dots, L$, satisfies the following identifiability condition: for any regular matrix \mathbf{Q} , if the copula density, of $\mathbf{Q}\mathcal{S}$, belongs to $\{c_{\theta_l}(\cdot); \theta_l \in \Theta_l \subset \mathbb{R}\}$, then $\mathbf{Q} = \mathbf{D}\mathbf{P}$, where \mathbf{D} is diagonal and \mathbf{P} is a permutation. To get the objective function for dependent sources all we have to do is to replace the copula density of the independence sources in 5.3 by the semi-parametric copula density $\{c_{\theta_l}(\cdot)\}$ [22]. The new objective function will be of the following form:

$$H(\mathbf{c}_{\theta_l}, \mathbf{c}_{\mathbf{Y}}) := \mathbb{E} \left[2 \left(\sqrt{\frac{\mathbf{c}_{\theta_l}(F_{Y_1}(Y_1), \dots, F_{Y_p}(Y_p))}{\mathbf{c}_{\mathbf{Y}}(F_{Y_1}(Y_1), \dots, F_{Y_p}(Y_p))}} - 1 \right)^2 \right], \quad (5.16)$$

this term is non-negative and achieves its minimum value zero iff $\mathbf{B} = \mathbf{M}^{-1}$ (up to scale and permutation indeterminacies). Therefore, we estimate the demixing matrix by

$$\hat{\mathbf{B}} := \arg \inf_{\alpha} \inf_{\theta_{l^*} \in \Theta_{l^*}} \hat{H}(\mathbf{c}_{\theta_l}, \mathbf{c}_{\mathbf{Y}}), \quad (5.17)$$

where

$$l^* = \arg \min_{l=1, \dots, L} \inf_{\alpha} \inf_{\theta_l \in \Theta_l} \hat{H}(\mathbf{c}_{\theta_l}, \mathbf{c}_{\mathbf{Y}}), \quad (5.18)$$

The copula density as well as the marginal distribution functions estimates are defined as before. The solution $\hat{\mathbf{B}}$ can be computed by a gradient descent algorithm with respect to both α and θ of the criterion function $(\alpha, \theta_l) \mapsto \hat{H}(\mathbf{c}_{\theta_l}, \mathbf{c}_{\mathbf{Y}})$ for each model and then choose the solution minimizing the criterion over all considered models. The calculations for the gradient of the Hellinger divergence is the same as the one stated in the case of independence (5.12). The Algorithm (2) summarizes the presented method.

Algorithm 2 The separation algorithm for dependent source components.

Data: the observations $\mathbf{x}(n)$, $n = 1, \dots, N$.

Result: the estimated sources $\hat{\mathbf{y}}(n)$, $n = 1, \dots, N$.

Whitening and Initialization: $\mathbf{z}(n) := \mathbf{W}\mathbf{x}(n)$, $\hat{\mathbf{y}}_0(n) = \mathbf{U}(\hat{\theta}_0)\mathbf{z}(n)$. Given $\varepsilon > 0$ and $\mu > 0$.

Do: • **Update** θ , α and \mathbf{y}

$$\theta_{k+1} = \theta_k - \nu \frac{d\hat{H}(\mathbf{c}_{\theta_k}, \mathbf{c}_{\mathbf{Y}})}{d\theta}.$$

$$\alpha_{k+1} = \alpha_k - \mu \frac{d\hat{H}(\mathbf{c}_{\theta_{k+1}}, \mathbf{c}_{\mathbf{Y}})}{d\alpha}.$$

$$\mathbf{y}_{k+1}(n) = \mathbf{U}(\alpha_{k+1})\mathbf{z}(n), \quad n = 1, \dots, N.$$

• **Until** $\|\theta_{k+1} - \theta_k\| < \varepsilon$

$$\hat{\mathbf{y}}(n) = \mathbf{y}_{k+1}(n), \quad n = 1, \dots, N.$$

6. SIMULATION RESULTS

In the following, we present the results of various simulations that were conducted to test our proposed approach and to better illustrate its performance. Our results will be compared with those obtained by [27](Copula-Alpha), [22](Copula-MI), [29](MI), [6] (JADE), [19] (FastICA), [24] (RADICAL) and [2] (InfoMax) under the same conditions.

In all the instances of our experiments the number of samples is $N = 3000$. The matrix used to mix the source components is $\mathbf{A} := [1 \ 0.7 \ 0.7 ; 0.7 \ 1 \ 0.7 ; 0.7 \ 0.7 \ 1]$, and $\mu = 0.1$ is the chosen gradient descent parameter. All simulations are iterated 80 times, and the accuracy of the estimated sources is calculated using the signal-to-noise-ratio criterion, which is defined by

$$SNR_i := 10 \log_{10} \frac{\sum_{n=1}^N s_i(n)^2}{\sum_{n=1}^N (\hat{s}_i(n) - s_i(n))^2}, \quad i = 1, 2, 3. \quad (6.1)$$

6.1. Independent source components

We consider in this experiment three mixed signals from two types of sample sources:

- Uniform i.i.d with independent components (see Fig. 1a).
- i.i.d sources with independent components drawn from the 4-ASK (Amplitude Shift Keying) alphabet (see Fig. 1b).

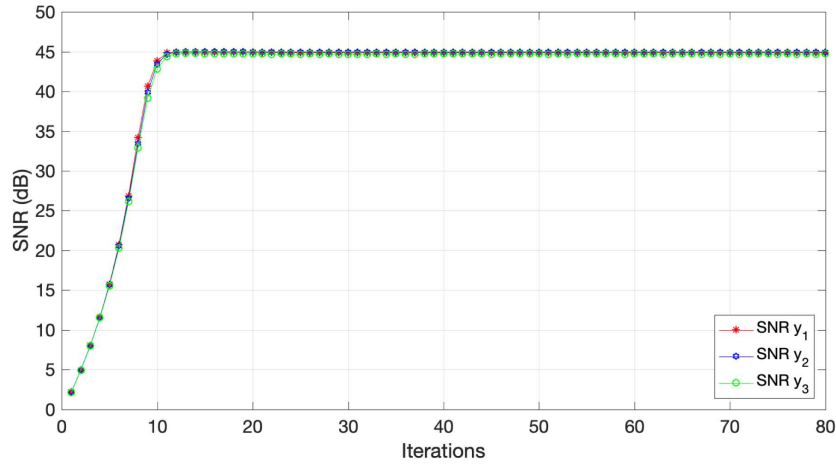
From Figures 1a and 1b, we observe that for both independent samples the SNR is close to 45dB which is considered highly satisfying for this classical case. In the other hand, Figures 2a and 2b present the criterion value *vs.* iterations. We can see that the separation is achieved when our criterion converges to its minimum value 0.

Table 2 illustrate the different SNR values of the sources, for our approach and other methods, we can see that the method proposed achieves the separation with similar accuracy with a slight improvement for the independent source components case and faster convergence compared to the latest Alpha-divergence method [27] when $\alpha = 0.5$.

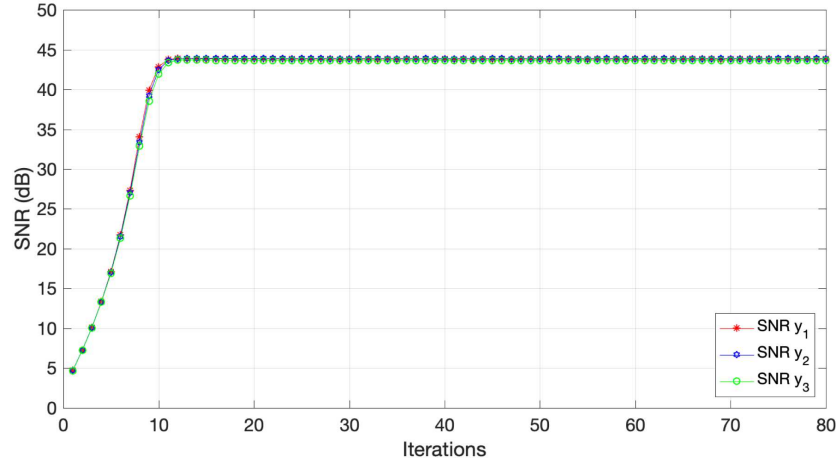
6.2. Dependent source components

Within this subsection we demonstrate the ability of the proposed approach (Algorithm 2 for dependent sources) to successfully separate mixtures of three dependent signals, we dealt with instantaneous mixtures of tree kinds of sample sources:

- i.i.d (with uniform marginals) sources with dependent components generated from AMH copula with $\theta = 0.75$.



(A)



(B)

FIGURE 1. Average output SNRs *versus* iteration number. (A) Uniform independent sources. (B) ASK independent sources.

- i.i.d (binary phase-shift keying(BPSK)) sources with dependent components generated from Clayton copula, with $\theta = 1.5$
- i.i.d (with uniform marginals) sources with dependent components generated from Frank copula, with $\theta = 2$.

In Figures 3a and 3b, we have shown the SNRs for dependent sources from Clayton and Frank copulas. From the simulation results it is noticeable that the proposed approach can separate the mixtures of dependent source components, with good performance.

Moreover, Figures 4a and 4b show the criterion value *versus* iterations for Clayton and Frank copulas. We can see that the separation is achieved when our criterion converges to its minimum value 0.

Table 3 exhibits the superiority of our proposed approach compared to [27](Copula-Alpha), [22](Copula-MI), [29](MI), [6] (JADE), [19] (FastICA), [24] (RADICAL) and [2] (InfoMax). with slight improvement compared to [27] when alpha is equal to 0.5, however our approach converges faster.

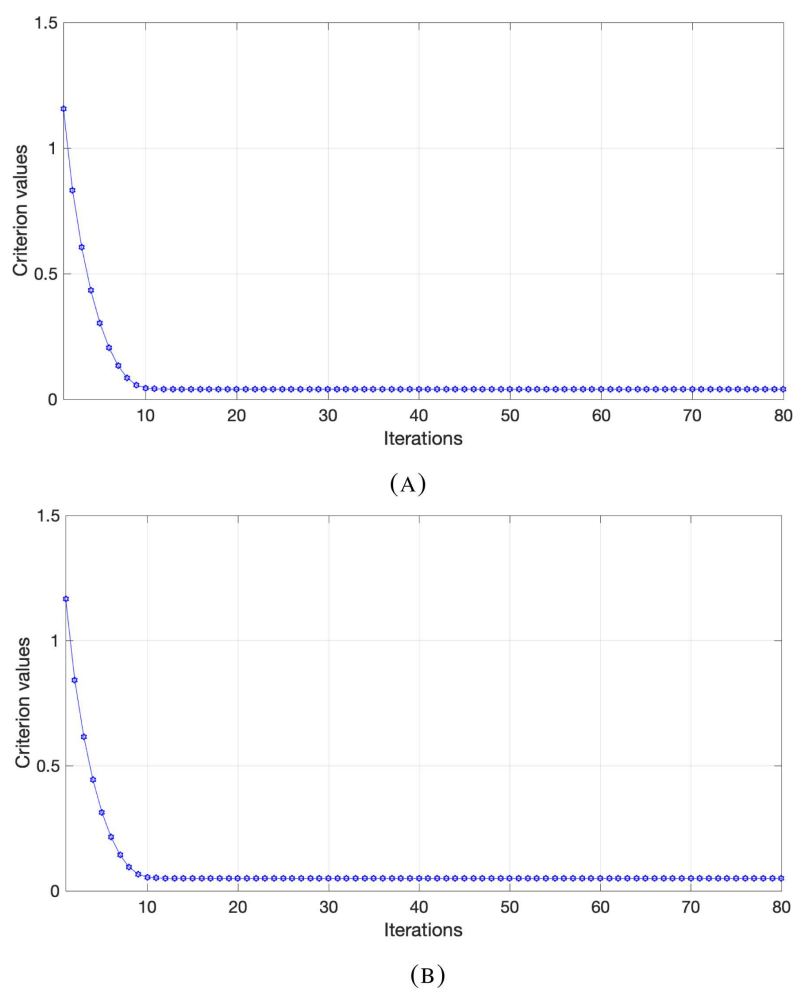
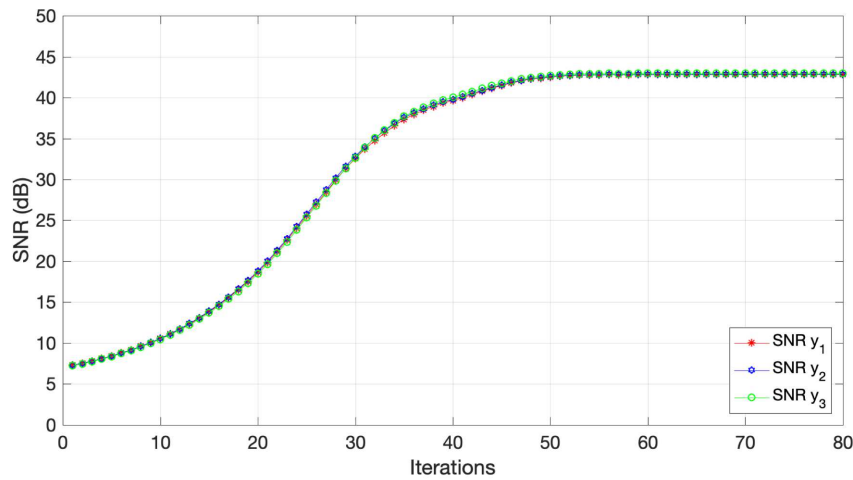


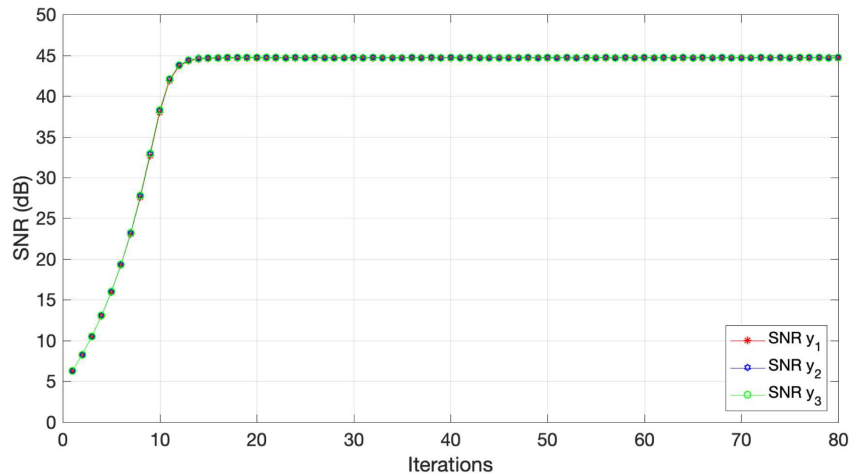
FIGURE 2. The criterion value *vs.* iterations. (A) Uniform independent sources. (B) ASK independent sources.

TABLE 2. Output SNR's for independent source components.

Type	Uniform			ASK		
Sources	S1	S2	S3	S1	S2	S3
Our method	46.7660	46.9336	46.8821	44.9767	44.7364	44.8914
Copula-Alpha	46.4311	46.3267	46.6112	44.9051	44.3076	44.5188
Copula-MI	45.8914	45.9560	45.7158	43.8195	43.8786	43.6588
MI	45.6757	45.6216	45.6142	43.1125	43.0654	43.0589
FastICA	43.8497	44.4492	40.1810	42.7949	42.0838	40.2107
JADE	45.5146	44.9522	44.9522	43.1671	43.1508	43.1889
RADICAL	44.9902	43.9986	44.7361	42.9023	42.4095	42.6272
InfoMax	45.3687	45.6243	44.8909	43.8846	43.3195	43.1528



(A)



(B)

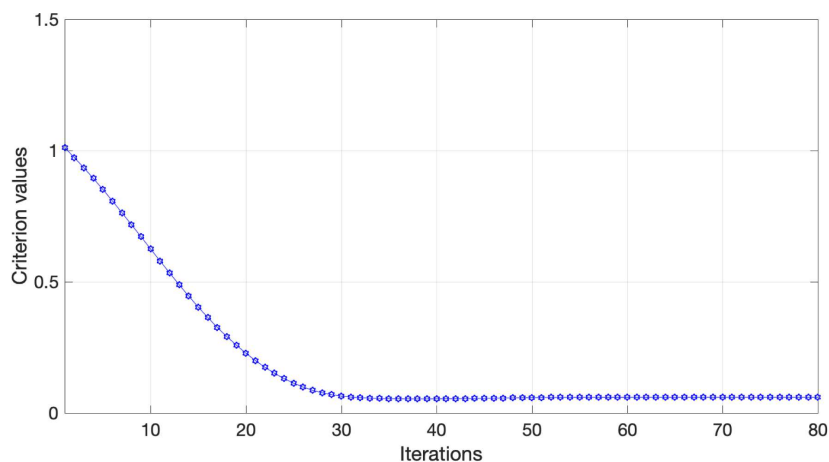
FIGURE 3. Average output SNRs *versus* iteration number. (A) Uniform dependent sources from Clayton-copula. (B) Uniform dependent sources from Frank-copula.

6.3. Noisy source components

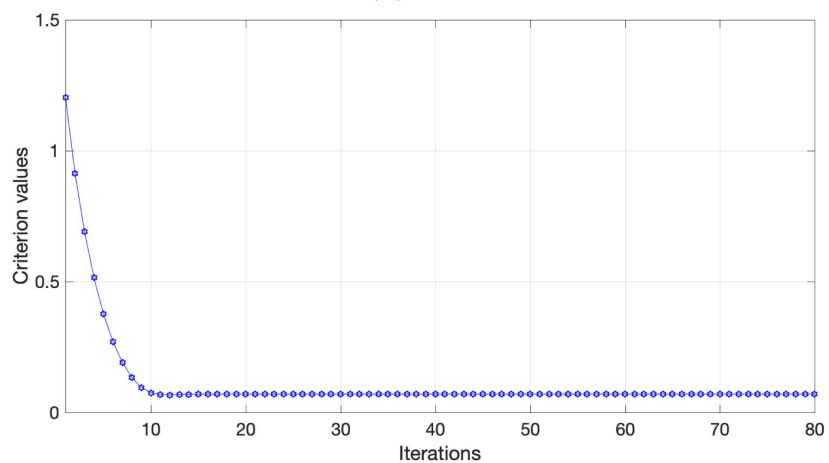
In this subsection we test the accuracy of our approach for noisy data. We work with the same source signals as above and the same conditions with an added white gaussian noise to the observed signals. We take $SNR = -25dB$.

Figure 5a illustrate the SNR of the independent sources, it can be seen that the proposed approach is able to separate noisy independent sources with good performance, and Figure 5b shows that when the separation is achieved our criterion converges to its minimum 0. Figure 6a showcase the SNR of the dependent sources from Clayton copula, the proposed approach is able to separate even noisy dependent sources. Moreover Figure 6b shows that the criterion in this case also converges to its minimum 0.

Table 4 present the output SNR values of the estimated sources using our approach and the other methods, we can see that the approaches are equivalent, with superiority of our method, in case of noise-contaminated



(A)

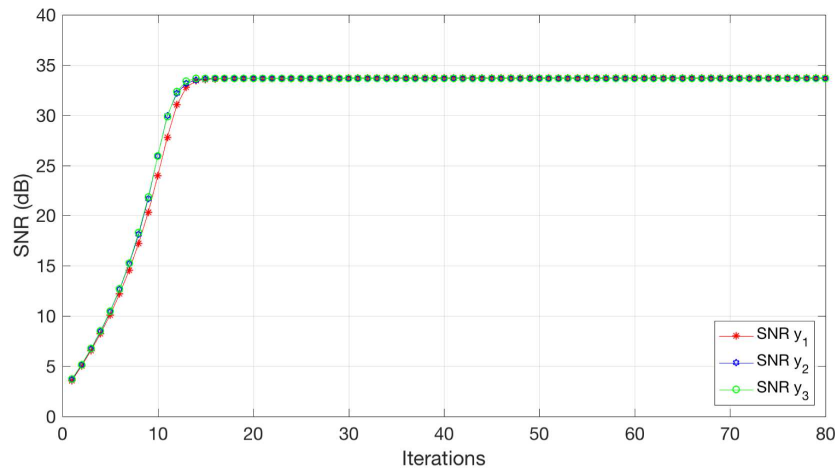


(B)

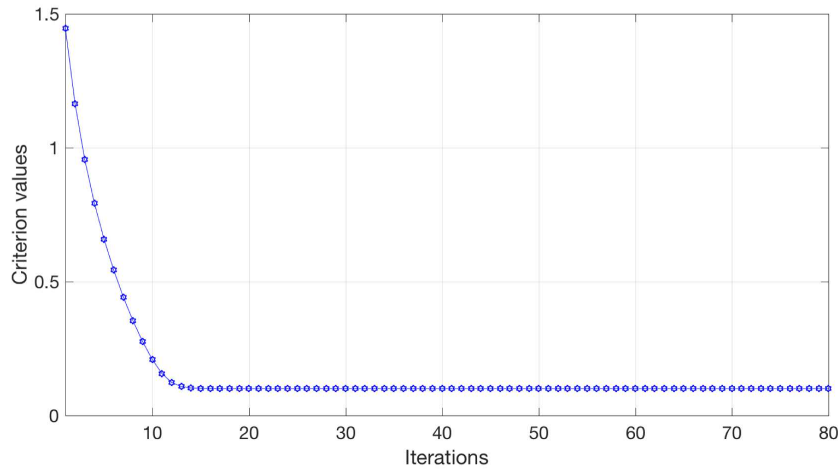
FIGURE 4. The criterion value *vs.* iterations. (A) Uniform dependent sources from Clayton-copula. (B) Uniform dependent sources from Frank-copula.

TABLE 3. Output SNR's for dependent source components.

Copulas	AMH			Clayton			Frank		
Sources	S1	S2	S3	S1	S2	S3	S1	S2	S3
Our method	45.2240	45.4681	45.3301	43.7845	43.8392	44.5118	46.0179	46.2626	46.1243
Copula-Alpha	44.7440	45.0932	45.3209	43.5692	42.9346	43.8850	45.4471	45.8412	46.0578
Copula-MI	44.2309	44.2773	44.2510	42.8152	42.8256	42.9534	44.6781	44.7247	44.6984
MI	15.3014	15.3221	15.3083	10.1302	10.3273	10.2757	18.9585	18.9875	18.9681
FastICA	41.1604	8.0131	8.0765	38.0900	5.9141	2.4933	42.9433	11.7862	8.7365
JADE	14.9945	14.1343	14.3148	9.7169	9.6885	10.7102	17.0637	17.2252	17.5806
RADICAL	9.3148	8.7314	6.7465	9.4081	8.7376	10.9933	17.8044	17.1656	17.4945
InfoMax	9.2981	9.2577	9.2871	8.2981	8.2577	8.2871	10.5136	10.3983	10.4529



(A)



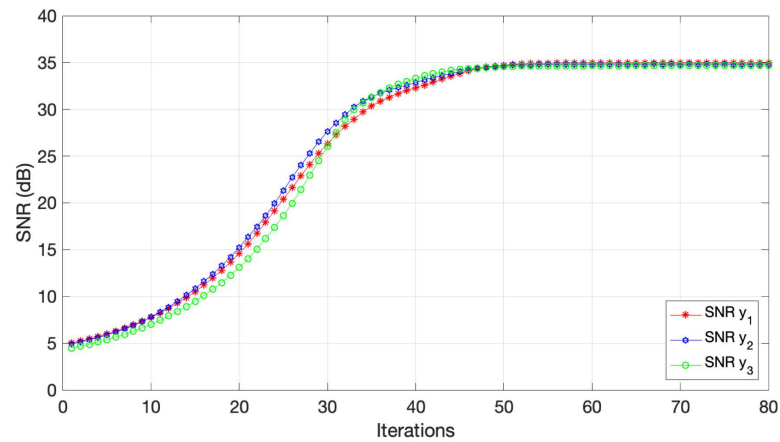
(B)

FIGURE 5. Uniform noisy independent sources. (A) Average output SNRs *versus* iterations. (B) The criterion value *vs.* iterations.

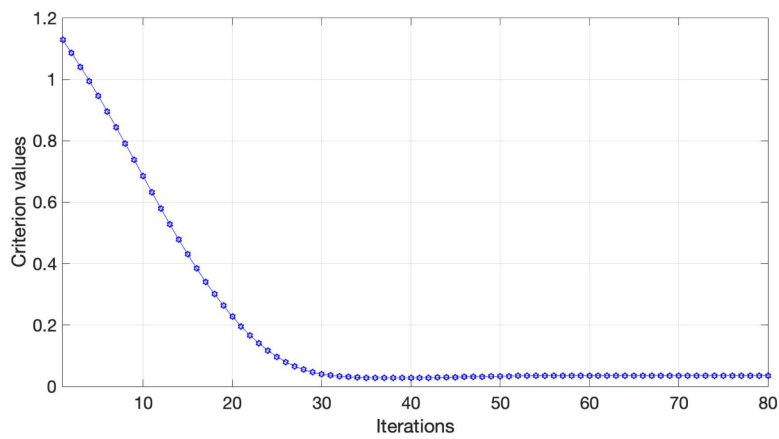
independent source components. On the other hand, our approach is apt to separate even noisy mixtures of dependent source components with higher accuracy.

7. CONCLUSION

We have presented a new BSS algorithm, that is able to separate instantaneous linear mixtures of both independent and dependent source components. Our approach proceeds in two steps: First a normalization stage with spatial whitening and then the application of Givens rotations, minimizing the estimate of the Hellinger distance. This divergence works better in presence of noise and it also converge faster than the usual Kullback-Leibler divergence as illustrated in section “Simulation results” for 3×3 mixture-sources, where the efficiency and the accuracy of the proposed algorithms is evaluated through the signal-to-noise-ratio criterion. It



(A)



(B)

FIGURE 6. Uniform dependent noisy sources from Clayton-copula. (A) Average output SNRs *versus* iterations. (B) The criterion value *vs.* iterations.

TABLE 4. Output SNR's for independent and dependent noisy source components.

Copulas Sources	Independence			Clayton			Frank		
	S1	S2	S3	S1	S2	S3	S1	S2	S3
Our method	36.1882	36.2384	36.0518	34.9658	34.8174	34.6611	36.2770	36.5339	36.3887
Copula-Alpha	35.8113	35.9858	36.0351	34.6367	33.5660	33.7636	35.6460	35.8031	36.1108
Copula-MI	31.0214	31.0657	30.9011	29.1377	29.1850	29.7672	30.4445	30.4273	30.4370
MI	30.4531	30.4178	30.4130	7.1605	7.0444	6.6586	12.2857	12.3780	12.1843
FastICA	30.2118	29.6796	28.2773	25.5767	3.0302	3.7841	28.5473	8.4895	6.6148
JADE	30.2331	30.2209	30.2495	6.9056	6.8649	7.1214	13.2814	13.3712	12.8820
RADICAL	30.3161	29.9472	30.1102	6.5064	7.4440	6.9503	12.4281	14.7506	12.5084
InfoMax	30.0881	30.6651	30.5402	5.3751	5.3627	5.3168	8.3977	8.2703	8.3137

should be noted that our proposed algorithms are more time-consuming compared to the classic ones, considering that we estimate both copulas density of the vector and the marginal distribution function of each component, however, ours gives better results especially for the dependent noisy source components case.

Acknowledgements. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

- [1] M.M. Ali, N.N. Mikhail and M.S. Haq, A class of bivariate distributions including the bivariate logistic. *J. Multivariate Anal.* **8** (1978) 405–412.
- [2] A.J. Bell and T.J. Sejnowski, An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* **7** (1995) 1129–1159.
- [3] A. Belouchrani, K. Abed Meraim, J.F. Cardoso and E. Moulines, A blind source separation technique using second-order statistics. *IEEE Trans. Signal Process.* **45** (1997) 434–444.
- [4] R. Beran, Minimum hellinger distance estimates for parametric models. *Ann. Stat.* **5** (1977) 445–463.
- [5] J.F. Cardoso, Blind signal separation: statistical principles. *Proc. IEEE* **86** (1998) 2009–2025.
- [6] J.F. Cardoso and A. Souloumiac, Blind signal beamforming for non gaussian signals. *Proc. IEE* **140** (1993) 362–370.
- [7] M. Castella, S. Rhioui, E. Moreau and J.C. Pesquet, Quadratic higher order criteria for iterative blind separation of a mimo convolutive mixture of sources. *IEEE Trans. Signal Process.* **55** (2007) 218–232.
- [8] D.G. Clayton, A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65** (1978) 141–151.
- [9] P. Comon, Independent component analysis, a new concept? *Signal Process.* **36** (1994) 287–314.
- [10] P. Comon and C. Jutten, Handbook of Blind Source Separation: Independent Component Analysis and Applications. Academic Press, Oxford (2010).
- [11] I. Csizsár, Eine informationstheoretische ungleichung und ihre anwendung auf beweis der ergodizitaet von markoffschen ketten. *Magyer. Tud. Akad. Mat. Kutat. In. Kol.* **8** (1964) 85–108.
- [12] I. Csizsár, Information-type measures of difference of probability distributions and indirect observation. *Stud. Sci. Math. Hung.* **2** (1967) 229–318.
- [13] M. El Rhabi, G. Gelle, H. Fenniri and G. Delaunay, A penalized mutual information criterion for blind separation of convolutive mixtures. *Signal Process.* **84** (2004) 1979–1984.
- [14] M. El Rhabi, H. Fenniri, A. Keziou and E. Moreau, A robust algorithm for convolutive blind source separation in presence of noise. *Signal Process.* **93** (2013) 818–827.
- [15] M.J. Frank, On the simultaneous associativity of $F(x, y)$ and $x + y - F(x, y)$. *Aequ. Math.* **19** (1979) 194–226.
- [16] C. Genest, K. Ghoudi and L.-P. Rivest, A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82** (1995) 543–552.
- [17] A. Ghazdali, A. Hakim, A. Laghrib, N. Mamouni and S. Raghay, A new method for the extraction of fetal ECG from the dependent abdominal signals using blind source separation and adaptive noise cancellation techniques. *Theor. Biol. Med. Model.* **12** (2015) 25.
- [18] A. Ghazdali, M. El Rhabi, H. Fenniri, A. Hakim and A. Keziou, Blind noisy mixture separation for independent/dependent sources through a regularized criterion on copulas. *Signal Process.* **131** (2017) 502–513.
- [19] A. Hyvärinen and E. Oja, A fast fixed-point algorithm for independent component analysis. *Neural Comput.* **9** (1997) 1483–1492.
- [20] R. Jiménez and Y. Shao, On robustness and efficiency of minimum divergence estimators. *Test* **10** (2001) 241–248.
- [21] H. Joe, Multivariate Models and Dependence Concepts. Chapman & Hall London, Boca Raton (2001).
- [22] A. Keziou, H. Fenniri, A. Ghazdali and E. Moreau, New blind source separation method of independent/dependent sources. *Signal Process.* **104** (2014) 319–324.
- [23] B.G. Lindsay, Efficiency versus robustness: the case for minimum Hellinger distance and related methods. *Ann. Stat.* **22** (1994) 1081–1114.
- [24] E.G. Miller and J.W. Fisher III, Independent components analysis by direct entropy minimization. California Univ Berkeley Dept of Electrical Engineering and Computer Sciences (2003).
- [25] R.B. Nelsen, An Introduction to Copulas. Springer, New Haven (2007).
- [26] M. Omelka, I. Gijbels and N. Veraverbeke, Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing. *Ann. Stat.* **37** (2009) 3023–3058.
- [27] A. Ourdou, A. Ghazdali, A. Laghrib and M. Abdelmoutalib, Blind separation of instantaneous mixtures of independent/dependent sources. *Circuits Syst. Signal Process.* **40** (2021) 1–24.
- [28] D.T. Pham, Blind separation of instantaneous mixture of sources based on order statistics. *IEEE Trans. Signal Process.* **2** (2000) 363–375.
- [29] D.T. Pham, Mutual information approach to blind separation of stationary sources. *IEEE Trans. Inf. Theory* **48** (2002) 1935–1946.

- [30] G. Schwarz, Estimating the dimension of a model. *Ann. Stat.* **6** (1978) 461–464.
- [31] B.W. Silverman, Density Estimation for Statistics and Data Analysis. CRC Press, New Haven (1986).
- [32] M. Sklar, Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **8** (1959) 229–231.
- [33] H. Tsukahara, Semiparametric estimation in copula models. *Can. J. Stat.* **33** (2005) 357–375.

Subscribe to Open (S2O)

A fair and sustainable open access model



This journal is currently published in open access under a Subscribe-to-Open model (S2O). S2O is a transformative model that aims to move subscription journals to open access. Open access is the free, immediate, online availability of research articles combined with the rights to use these articles fully in the digital environment. We are thankful to our subscribers and sponsors for making it possible to publish this journal in open access, free of charge for authors.

Please help to maintain this journal in open access!

Check that your library subscribes to the journal, or make a personal donation to the S2O programme, by contacting subscribers@edpsciences.org

More information, including a list of sponsors and a financial transparency report, available at: <https://www.edpsciences.org/en/maths-s2o-programme>