MATHEMATICAL MODELING AND COMPUTING, Vol. 8, No. 4, pp. 761–769 (2021)

Robust approach for blind separation of noisy mixtures of independent and dependent sources

Ourdou A.[1], Ghazdali A.[1], Laghrib A.[2], Metrane A.[1]

[1] *LIPIM, ENSA Khouribga, Sultan Moulay Slimane University, Khouribga, Morocco*
[2] *LMA, FST Beni-Mellal, Sultan Moulay Slimane University, Beni-Mellal, Morocco*

(Received 23 May 2021; Accepted 7 June 2021)

In this paper, a new Blind Source Separation (BSS) method that handles mixtures of noisy independent/dependent sources is introduced. We achieve that by minimizing a criterion that fuses a separating part, based on Kullback–Leibler divergence for either dependent or independent sources, with a regularization part that employs the bilateral total variation (BTV) for the purpose of denoising the observations. The proposed algorithm utilizes a primal-dual algorithm to remove the noise, while a gradient descent method is implemented to retrieve the signal sources. Our algorithm has shown its effectiveness and efficiency and also surpassed the standard existing BSS algorithms.

Keywords: *blind source separation, noisy mixtures, dependent sources, bilateral total variation, Kullback–Leibler divergence.*

2010 MSC: 62-XX, 62H05 **DOI:** 10.23939/mmc2021.04.761

1. Introduction

Blind Source Separation (BSS) aims to recover unknown original sources given only the observed mixtures without any previous knowledge about the source and the mixing process. In literature, various techniques were introduced under assumptions of mutual independence between the sources and only one of the components could have Gaussian distribution, for instance, [1–9]. However, the issue of BSS is much more complex with dependent components [10–14]. We have:

$$\overline{\boldsymbol{o}}(t) := \boldsymbol{M}[\boldsymbol{s}(t)] + \boldsymbol{\nu}(t) \in \mathbb{R}^p, \tag{1}$$

where $\overline{\boldsymbol{o}}(t) \in \mathbb{R}^p$ are the mixed signals, $\boldsymbol{M}[\cdot]$ is the unknown mixing operator, $\boldsymbol{s}(t) \in \mathbb{R}^q$ is the sources vector, $\boldsymbol{\nu}(t) \in \mathbb{R}^p$ is an additive noise vector, and $t \in [0, T]$ is the sample index. The model (1) can be viewed as follows: $\overline{\boldsymbol{o}}(t) = \boldsymbol{o}(t) + \boldsymbol{\nu}(t)$, where $\boldsymbol{o}(t) := \boldsymbol{M}\,\boldsymbol{s}(t)$ denotes the unknown mixed vector in a noise free environment. In our study, we assume, the number of the signals is equal to the observations ($p = q$), the independence between the signals source $\boldsymbol{s}$ and the noise $\boldsymbol{\nu}$, and the linearity of the mixtures. By using the model (1), the estimate of signals source is: $\overline{\boldsymbol{z}}(t) := \boldsymbol{W}\,\overline{\boldsymbol{o}}(t)$, where $\boldsymbol{W}$ denotes the de-mixing matrix and $\overline{\boldsymbol{z}}(t) \in \mathbb{R}^p$ is the approximate of $\boldsymbol{s}(t)$ in the noisy environment. Nonetheless, $\overline{\boldsymbol{z}}(\cdot)$ does not present the best estimation for the source signals $\boldsymbol{s}(\cdot)$. As a matter of fact, the recovered signals can be formed as: $\overline{\boldsymbol{z}}(t) := \boldsymbol{W}\,\overline{\boldsymbol{o}}(t) = \boldsymbol{W}\,\boldsymbol{o}(t) + \boldsymbol{W}\,\boldsymbol{\nu}(t) =: \boldsymbol{z}(t) + \overline{\boldsymbol{\nu}}(t)$. The noisy restored source $\overline{\boldsymbol{z}}(t)$ is the sum of $\boldsymbol{z}(t) := \boldsymbol{W}\,\boldsymbol{o}(t)$ the source signals estimate in a free-noise environment, and $\overline{\boldsymbol{\nu}}(t) := \boldsymbol{W}\,\boldsymbol{\nu}(t)$ is the unknown product of the de-mixing matrix and the noise. Several algorithms have been developed in recent decades to treat the noisy BSS problem, for example, [15–17].

For a higher quality estimate of the mixing matrix as well as the sources, [18] introduced a new BSS technique for noisy mixtures of dependent or independent sources by fusing the estimation of Kullback–Leibler divergence between the copula densities and the total variation (TV) regularization, however, using total variation in order to remove noise has its deficiencies such as sensitivity to noise and easiness to blur. To overcome those issues, we propose a progression of the stated work by introducing a method based on the bilateral total variation regularization (BTV) which is generated from the bilateral filter [19]. This filter eliminates the noise completely although retaining the edge information then we minimize the Kullback-Liebler divergence between the copula densities. Next, this paper is organized as follows: Section 2 presents a short resume about copulas and some of their fundamental properties. In Section 3, we go into more details about our approach. Section 4 states how to implement the proposed methodology utilizing both statistical and numerical skills. In section 5 numerical results are given to illustrate the efficiency and robustness of our proposed approach and finally a conclusion is given.

2. Copulas

Copulas are functions that link a joint distribution function to its marginal functions. Let $\mathbf{S} := (S_1, \ldots, S_p)^\top \in \mathbb{R}^p$, be a random vector with $\mathbb{F}_{\mathbf{S}}(\mathbf{S}) := \mathbb{P}(S_1 \leqslant S_1, \ldots, S_p \leqslant S_p)$, the cumulative d.f. and $F_{S_j}(S_j) := \mathbb{P}(S_j \leqslant S_j), \forall j = 1, \ldots, p$, the marginal d.f's. The relationship between these two using a copula $\mathbb{C}_{\mathbf{S}}(\cdot)$, according to the theorem of Sklar [20] is: $\mathbb{F}_{\mathbf{S}}(\mathbf{S}) = \mathbb{C}_{\mathbf{S}}\left(F_{S_1}(S_1), \ldots, F_{S_p}(S_p)\right), \forall\ \mathbf{S} \in \mathbb{R}^p$ and for every vector $\mathbf{v} := (v_1, \ldots, v_p)^\top \in [0,1]^p$, we have $\mathbb{C}_{\mathbf{S}}(\mathbf{v}) = \mathbb{P}\left(F_{S_1}(S_1) \leqslant v_1, \ldots, F_{S_p}(S_p) \leqslant v_p\right)$, where its density is calculated as follows: $c_{\mathbf{S}}(\mathbf{v}) := \frac{\partial^p \mathbb{C}_{\mathbf{S}}(\mathbf{v})}{\partial v_1 \cdots \partial v_p}, \forall \mathbf{v} \in [0,1]^p$. The statistical independence of the components $S_1, \ldots, S_p$ is set iff: $\mathbb{C}_{\mathbf{S}}(\mathbf{v}) = \prod_{j=1}^p v_j =: \mathbb{C}_\Pi(\mathbf{v}), \forall \mathbf{v} \in [0,1]^p$, $\mathbb{C}_\Pi(\cdot)$ represent the copula of independence, consequently, the copula density of independence, $c_\Pi(\cdot)$, writes: $c_\Pi(\mathbf{v}) := \mathbf{1}_{[0,1]^p}(\mathbf{v})$. We consider $f_{\mathbf{S}}(\cdot)$, the probability density of the r.v. $\mathbf{S} := (S_1, \ldots, S_p)^\top$, if it exists, and, $f_{S_1}(\cdot), \ldots, f_{S_p}(\cdot) \in \mathbb{R}^p$, their marginal p.d's. By using the equations stated above, we can deduct the following:

$$f_{\mathbf{S}}(\mathbf{S}) = \left(\prod_{j=1}^p f_{S_j}(S_j)\right) c_{\mathbf{S}}\left(F_{S_1}(S_1), \ldots, F_{S_p}(S_p)\right). \tag{2}$$

In this paper, use the semiparametric copula models denoted $\mathbb{C}(\cdot, \boldsymbol{\theta})$, where $\theta \in \Theta \subset \mathbb{R}^d$, is the indexing parameter, with nonparametric margins, and we can estimate it by choosing the best estimate copula model, amid a large set of copula models. Given a model k, the Bayesian information criterion is: $BIC(k) = -2 \sup_{\theta_k \in \Theta_k} \sum_{n=1}^N \log c_k\big(\widehat{F}_{S_1}(s_1(n)), \ldots, \widehat{F}_{S_p}(s_p(n)), \theta_k\big) + d_k \log(N)$. The copula model that minimizes the BIC is the optimal model. We consider, $\{c(\cdot, \theta); \theta \in \Theta \subset \mathbb{R}^d\}$ the chosen model, the copula model parameter θ is then approximated by maximizing the semiparametric log-likelihood $\widehat{\theta} = \arg\sup_{\theta \in \Theta} \sum_{n=1}^N \log c\big(\widehat{F}_{S_1}(s_1(n)), \ldots, \widehat{F}_{S_p}(s_p(n)), \theta\big)$.

3. The proposed method

3.1. Denoising the observed data

To extract the noise-free signal $\boldsymbol{o}(t)$ from the noisy observation $\overline{\boldsymbol{o}}(t)$, with the smallest error, we use the mean squared error, by solving the least-square problem: $\inf_{\boldsymbol{o}} \frac{1}{2} \int_{[0,T]} |\overline{\boldsymbol{o}}(t) - \boldsymbol{o}(t)|^2 \mathrm{d}t$. We also consider the bilateral TV regularizer [21], due to its advantages, namely, the ability to smooth away the noise and small variation in a signal while preserving the major edges or discontinuity, and the ability of handling and removing high level noise unlike total variation. It is expressed as follows: $BTV(\boldsymbol{o}) := \sum_{j=-m}^m \alpha^{|j|} \|\boldsymbol{o} - \boldsymbol{G}^j \boldsymbol{o}\|_1$, where the matrix $\boldsymbol{G}^j$ entails a shift right of j samples and m is the spatial window size. To give a spacial decay effect, we deploy a weight $\alpha \in [0,1]$ to the sum of the regularization terms. Hence we obtain $\boldsymbol{o}$, by minimizing the following objective function

$$\inf_{\boldsymbol{o}} \left\{ \frac{1}{2} \int_{[0,T]} \|\overline{\boldsymbol{o}} - \boldsymbol{o}\|^2 \mathrm{d}t + \lambda \sum_{j=-m}^m \alpha^{|j|} \|\boldsymbol{o} - \boldsymbol{G}^j \boldsymbol{o}\|_1 \right\}, \tag{3}$$

where $\lambda > 0$ is the regularization parameter that measures the performance of the smoothing effect.

3.2. BSS-denoising procedure

Once having the noise-free observed signals, we can now estimate the source signals $\widehat{\boldsymbol{s}}(t) := \widehat{\boldsymbol{W}}\,\boldsymbol{o}(t)$. We do so by minimizing, in regards to $\boldsymbol{W}$ the following criterion:

$$\boldsymbol{W} \mapsto \mathcal{C}(\boldsymbol{W}) := \mathcal{C}_{\text{sep}}(\boldsymbol{W}) + \mathcal{C}_{\text{reg}}(\boldsymbol{z}), \tag{4}$$

where $\mathcal{C}_{\text{sep}}(\cdot)$ is the separating part, and $\mathcal{C}_{\text{reg}}(\boldsymbol{z}) := \frac{\gamma}{2T}\int_0^T \|\boldsymbol{z}(t) - \overline{\boldsymbol{z}}(t)\|^2\,dt + \mu\sum_{j=-m}^{m}\alpha^{|j|}\|\boldsymbol{z}(t) - \boldsymbol{G}^j\boldsymbol{z}(t)\|_1$, the regularization term, which is applicable for independent and dependent cases. γ, $\mu > 0$ present the regularization parameters. Let $f_{\boldsymbol{Z}}(\cdot)$ be the joint probability density of the r.v. $\boldsymbol{Z} \in \mathbb{R}^p$, and $f_{Z_1}, \dots, f_{Z_p}$ their marginal p.d's. The mutual information of $\boldsymbol{Z}$ writes then as: $\mathbb{MI}(\boldsymbol{Z}) := \int_{\mathbb{R}^p}\log\frac{f_{\boldsymbol{Z}}(\boldsymbol{z})}{\prod_{i=1}^p f_{Z_i}(z_i)}\,f_{\boldsymbol{Z}}(\boldsymbol{z})\,d\boldsymbol{z} = \mathbb{E}\left(\log\frac{f_{\boldsymbol{Z}}(\boldsymbol{z})}{\prod_{i=1}^p f_{Z_i}(Z_i)}\right)$. This $\mathbb{MI}$ is equal to the Kullback–Liebler divergence between $f_{\boldsymbol{Z}}(\cdot)$ and the product of the marginals and it is expressed as such: $\mathbb{MI}(\boldsymbol{Z}) = \mathbb{K}\left(f_{\boldsymbol{Z}}, \prod_{i=1}^p f_{Z_i}\right)$. and it is always positive and only reaches its minimum value zero iff $f_{\boldsymbol{Z}}(\cdot) = \prod_{i=1}^p f_{Z_i}(\cdot)$, i.e., the independence of the elements of $\boldsymbol{Z}$.

3.2.1. The case of independent source components

In this case, we use copula density of independence $c_{\prod}(\cdot)$, by applying the relation (2):

$$\begin{aligned}\mathbb{MI}(\boldsymbol{Z}) &= \int_{[0,1]^p}\log\left(c_{\boldsymbol{Z}}(\boldsymbol{v})\right)c_{\boldsymbol{Z}}(\boldsymbol{v})\,d\boldsymbol{u} \\ &= \int_{[0,1]^p}\log\left(\frac{c_{\boldsymbol{Z}}(\boldsymbol{v})}{c_{\prod}(\boldsymbol{v})}\right)c_{\boldsymbol{Z}}(\boldsymbol{v})\,d\boldsymbol{v} \\ &= \mathbb{E}\left(\log\frac{c_{\boldsymbol{Z}}(F_{Z_1}(Z_1), \dots, F_{Z_p}(Z_p))}{c_{\prod}(F_{Z_1}(Z_1), \dots, F_{Z_p}(Z_p))}\right) = \mathbb{K}\left(c_{\boldsymbol{Z}}, c_{\prod}\right),\end{aligned} \tag{5}$$

with $c_{\boldsymbol{Z}}(\cdot)$ the copula density of $\boldsymbol{Z}$. This measure is always positive, and only reaches the minimum, zero iff $c_{\boldsymbol{Z}}(\boldsymbol{v}) = c_{\prod}(\boldsymbol{v})$, $\forall\boldsymbol{v} \in [0,1]^p$, i.e., iff we reach the independence. Consequently, we have: $\mathcal{C}_{\text{sep}}(\boldsymbol{W}) := \mathcal{C}_{\text{sep}}^{\text{ind}}(\boldsymbol{W}) := \mathbb{K}\left(c_{\boldsymbol{Z}}, c_{\prod}\right)$. Hence $\boldsymbol{W} \mapsto \mathcal{C}_{\text{sep}}^{\text{ind}}(\boldsymbol{W})$ is non-negative and reaches its minimum value zero iff $\boldsymbol{W} = \boldsymbol{M}^{-1}$ with indeterminacies of scale and permutation.

3.2.2. The case of dependent source components case

Here, we assume that the dependency structure is modeled by a semiparametric copula density, $\{c_\theta(\cdot);\, \theta \in \Theta \subset \mathbb{R}^d\}$, with a multivariate parameter θ. In this case three approaches are used:
• **The model and the parameter are known**: In this case, we estimate $\boldsymbol{W}$ the de-mixing matrix in a straightforward way: $\boldsymbol{W} := \arg\inf_{\boldsymbol{W}}\mathcal{C}_{\text{sep}}^{\text{dep}}(\boldsymbol{W})$.
• **The model is known and the parameter is unknown**: Here the parameter θ of the dependency model is unknown, hence, we propose to estimate the separation matrix $\boldsymbol{W}$ by: $\boldsymbol{W} = \arg\inf_{\boldsymbol{W}}\inf_{\boldsymbol{\theta}\in\Theta}\mathcal{C}_{\text{sep}}^{\text{dep}}(\boldsymbol{W})$.
• **The model and the parameter are unknown**: Finally, when we have no knowledge about the parameter or the model, we consider T different models of copula densities of the source components, apply the approach described in the second case for each model, then take the model that minimizes the criterion, in other words: $\boldsymbol{W} = \arg\inf_W\inf_{\boldsymbol{\theta}\in\Theta^{k^*}} KL\big(\boldsymbol{c}_{\boldsymbol{s}_{\boldsymbol{\theta}\boldsymbol{k}^*}}, \boldsymbol{c}_{\boldsymbol{Z}}\big)$, where $k^* = \arg\min_{k=1\dots T}\inf_W\inf_{\boldsymbol{\theta_k}\in\Theta_k}\mathcal{C}_{\text{sep}}(\boldsymbol{W})$. We propose to estimate the separation criterion defined by:

$$\mathcal{C}_{\text{sep}}(\boldsymbol{W}) := \inf_{\theta\in\Theta}\,\mathbb{E}\left(\log\frac{c_{\boldsymbol{Z}}(F_{Z_1}(Z_1), \dots, F_{Z_p}(Z_p))}{c_\theta(F_{Z_1}(Z_1), \dots, F_{Z_p}(Z_p))}\right).$$

4. Discretization and Statistical estimation

4.1. Denoising the discrete observed data

To denoise the observations we use the primal dual algorithm [22, 23]. First, we set the following notations, $K = \lambda \sum_{j=-m}^{m} \alpha^{|j|}(I - \boldsymbol{G}^j)$, $\mathcal{F}(K\boldsymbol{o}) = \lambda \sum_{j=-m}^{m} \alpha^{|j|} \|\boldsymbol{o} - \boldsymbol{G}^j \boldsymbol{o}\|_1$ and $\mathcal{G}(\boldsymbol{o}) = \frac{1}{2}\int_{\mathbb{R}^p} |\boldsymbol{o}(t) - \bar{\boldsymbol{o}}(t)|^2 \mathrm{d}t$. The problem (3), is then of the form: $\inf_{\boldsymbol{o}}\{\mathcal{G}(\boldsymbol{o}) + \mathcal{F}(K\boldsymbol{o})\}$, where $\mathcal{F}$ and $\mathcal{G}$ are convex functions and K is the linear operator. The following primal-dual problem is obtained by using the saddle point problem: $\inf_{\boldsymbol{o}} \sup_{\boldsymbol{p}}\{\langle K\boldsymbol{o}, \boldsymbol{p}\rangle + \mathcal{G}(\boldsymbol{o}) + \mathcal{F}^*(\boldsymbol{p})\}$, where $\mathcal{F}^*(\boldsymbol{p}) = \sup_{\boldsymbol{p}}\langle \boldsymbol{p}, \boldsymbol{o}\rangle - \mathcal{F}(\boldsymbol{o})$, representing the dual of the function $\mathcal{F}$ and $\boldsymbol{p}$ present a dual variable with $\boldsymbol{p} \in \mathbb{R}^p$. After that, we verify: $\mathcal{F}^*(\boldsymbol{p}) = \delta_{\boldsymbol{P}}(\boldsymbol{p}) = 0$ if $\boldsymbol{p} \in \boldsymbol{P}$ and $= +\infty$ if $\boldsymbol{p} \notin \boldsymbol{P}$, where $\boldsymbol{P} = \{\boldsymbol{p} \colon \|\boldsymbol{p}\|_\infty \leqslant 1\}$, yet, one must determine the proximity operator functions $\mathcal{F}^*$ and $\mathcal{G}$ before continuing with the Primal–Dual algorithm. We have $(I + \sigma\partial\mathcal{F}^*)^{-1}(\boldsymbol{p}) = \Pi_{\boldsymbol{P}}(\boldsymbol{p})$, where $\Pi_{\boldsymbol{P}}(\boldsymbol{p}) = \frac{\boldsymbol{p}}{\max(||\boldsymbol{p}||_\infty, 1)}$, is a projection on $\boldsymbol{P}$ and $||\boldsymbol{p}||_\infty = \max_{i,j} |\boldsymbol{p}_{i,j}|$. Also by relying on the definition of the function $\mathcal{G}$, we have $(I + \tau\partial\mathcal{G})^{-1}(\boldsymbol{o}) = \frac{\boldsymbol{o} + \varepsilon\bar{\boldsymbol{o}}}{1+\varepsilon}$. The algorithm below, summarizes this step, with $K^\intercal$ is the adjoint of the operator K presented as such $K^\intercal = \lambda \sum_{j=-m}^{m} \alpha^{|j|}(I - \boldsymbol{G}^{-j})$.

Algorithm 1 The denoising step using Primal–Dual algorithm.

Data: $\bar{\boldsymbol{o}}$ the noisy observation.
Result: $\boldsymbol{y}$ the obtained noise-free signal.
Initialization: Given τ, $\sigma > 0$, $\eta \in [0, 1]$, $(\boldsymbol{p}^0, \boldsymbol{o}^0) \in \mathbb{R}^n \times \mathbb{R}^n$ and set $\boldsymbol{y}^0 = \boldsymbol{o}^0$.
Do:

$$\begin{aligned} \boldsymbol{p}^{n+1} &= (I + \sigma\partial\mathcal{F}^*)^{-1}(\boldsymbol{p} + \sigma K \boldsymbol{y}^n), \\ \boldsymbol{o}^{n+1} &= (I + \tau\partial\mathcal{G})^{-1}(\boldsymbol{o}^n - \tau K^\intercal \boldsymbol{p}^{n+1}), \\ \boldsymbol{y}^{n+1} &= \boldsymbol{o}^{n+1} + \eta(\boldsymbol{o}^{n+1} - \boldsymbol{o}^n). \end{aligned}$$

4.2. Statistical estimates of the separation terms

4.2.1. The case of independent source components

We estimate the principle (4) by

$$\boldsymbol{W} \mapsto \widehat{\mathcal{C}^{\mathrm{ind}}}(\boldsymbol{W}) := \widehat{\mathcal{C}^{\mathrm{ind}}_{\mathrm{sep}}}(\boldsymbol{W}) + \mathcal{C}_{\mathrm{reg,d}}(\boldsymbol{z}), \tag{6}$$

where $\mathcal{C}_{\mathrm{reg,d}}(\boldsymbol{z}) := \frac{\gamma}{2N}\sum_{i=1}^{N} |\overline{\boldsymbol{z}}(i) - \boldsymbol{z}(i)|^2 + \mu \sum_{i=1}^{N}\sum_{j=-m}^{m} \alpha^{|j|} |\boldsymbol{z}(n) - \boldsymbol{G}^j \boldsymbol{z}(n)|$, denotes the discrete version of $\mathcal{C}_{\mathrm{reg}}(\boldsymbol{z})$, and $\widehat{\mathcal{C}^{\mathrm{ind}}_{\mathrm{sep}}}(\boldsymbol{W})$, the statistical approximate of $\mathcal{C}^{\mathrm{ind}}_{\mathrm{sep}}(\boldsymbol{W})$, and is defined by: $\boldsymbol{W} \mapsto \widehat{\mathcal{C}^{\mathrm{ind}}_{\mathrm{sep}}}(\boldsymbol{W}) := \frac{1}{N}\sum_{i=1}^{N} \log\left(\widehat{c}_Z(\widehat{F}_{Z_1}(z_1(i)), \ldots, \widehat{F}_{Z_p}(z_p(i)))\right)$, where $\widehat{c}_Z\left(\widehat{F}_{Z_1}(z_1(i)), \ldots, \widehat{F}_{Z_p}(z_p(i))\right) := \frac{1}{NH_1 \cdots H_p}\sum_{\ell=1}^{N}\prod_{j=1}^{p} k\left(\frac{\widehat{F}_{Z_j}(z_j(i)) - \widehat{F}_{Z_j}(z_j(\ell))}{H_j}\right)$, denotes the kernel approximate of the copula density $c_{\boldsymbol{Z}}(\cdot)$, and $\widehat{F}_{Z_i}(\cdot)$, $\forall i = 1, \ldots, p$, are the smooth estimates of $F_{Z_i}(\cdot)$, and it is expressed as $\widehat{F}_{Z_i}(r) := \frac{1}{N}\sum_{\ell=1}^{N} K\left(\frac{r - z_i(\ell)}{h_i}\right)$, where $K(\cdot)$ denotes the kernel $k(\cdot)$'s primitive, which is a symmetric centered probability density. We picked up the standard Gaussian probability density as our kernel. Using the Silverman's rule of thumb [24], the bandwidth parameters are selected as follows: for all $i = 1, \ldots, p$, $H_i = \left(\frac{4}{p+2}\right)^{\frac{1}{p+4}} N^{\frac{-1}{p+4}} \widehat{\Sigma}_i$, and $h_i = \left(\frac{4}{3}\right)^{\frac{1}{5}} N^{-\frac{1}{5}} \widehat{\sigma}_i$, where $\widehat{\Sigma}_i$ and $\widehat{\sigma}_i$ are the empirical standard deviation of $\widehat{F}_{Z_i}(z_i(1)), \ldots, \widehat{F}_{Z_i}(z_i(N))$ and $z_i(1), \ldots, z_i(N)$, respectively. We can estimate the source signals by $\widehat{\boldsymbol{s}}(j) = \widehat{\boldsymbol{W}}\boldsymbol{o}(j), j = 1, \ldots, N$, where $\widehat{\boldsymbol{W}} := \arg\inf_{\boldsymbol{W}} \widehat{\mathcal{C}^{\mathrm{ind}}}(\boldsymbol{W})$, which can be estimated by deploying gradient descent algorithm. The gradient in $\boldsymbol{W}$ of $\boldsymbol{W} \mapsto \widehat{\mathcal{C}^{\text{ind}}}(\boldsymbol{W})$, is expressed as follows: $\frac{d\widehat{\mathcal{C}^{\text{ind}}}(\boldsymbol{W})}{d\boldsymbol{W}} = \frac{1}{N}\sum_{n=1}^{N} \frac{\frac{d}{d\boldsymbol{W}}\widehat{c}_{\boldsymbol{Z}}(\boldsymbol{v}(n))}{\widehat{c}_{\boldsymbol{Z}}(\boldsymbol{v}(n))} + \frac{\gamma}{N}\sum_{n=1}^{N}(\boldsymbol{z}(n) - \overline{\boldsymbol{z}}(n))(\boldsymbol{o}(n) - \overline{\boldsymbol{o}}(n))^{\top} + \frac{\mu}{N}\sum_{n=1}^{N}\sum_{j=-m}^{m}\alpha^{|j|}\left(I - \boldsymbol{G}^{-j}\right)\operatorname{sign}\left(\boldsymbol{z}(n) - \boldsymbol{G}^{j}\boldsymbol{z}(n)\right)$, where $\frac{d}{d\boldsymbol{W}} := \left(\frac{\partial}{\partial \boldsymbol{W}_{ij}}\right)_{ij}$, $i,j = 1,\ldots,p$, $\boldsymbol{v}(n) := (\widehat{F}_{Z_1}(z_1(n)),\ldots,\widehat{F}_{Z_p}(z_p(n)))^{\top}$ and, $\frac{\partial \widehat{c}_{\boldsymbol{Z}}(\widehat{F}_{Z_1}(z(n)),\ldots,\widehat{F}_{Z_p}(z_p(n)))}{\partial \boldsymbol{W}_{ij}} = \frac{1}{NH_1\cdots H_p}\sum_{m=1}^{N}\prod_{j=1,j\neq i}^{p} k\left(\frac{\widehat{F}_{Z_j}(z_j(m))-\widehat{F}_{Z_j}(z_j(n))}{H_j}\right) \times k'\left(\frac{\widehat{F}_{Z_i}(z_i(m))-\widehat{F}_{Z_i}(z_i(n))}{H_i}\right)\frac{1}{H_i}\frac{\partial(\widehat{F}_{Z_i}(z_i(m))-\widehat{F}_{Z_i}(z_i(n)))}{\partial \boldsymbol{W}_{ij}}$, with $\frac{\partial(\widehat{F}_{Z_i}(z_i(m))}{\partial \boldsymbol{W}_{ij}} = \frac{1}{Nh_i}\sum_{n=1}^{N} k\left(\frac{z_i(n)-z_i(m)}{h_i}\right)(o_j(n) - o_j(m))$. The following Algorithm presents the summary of the proposed approach:

Algorithm 2 BSS algorithm for separation of noisy independent source components.

Data: $\overline{\boldsymbol{o}}$ the noise free observations
Result: $\widehat{\boldsymbol{s}}$ the approximation of the source signal
Initialization: Calculate $\boldsymbol{o} = \overline{\boldsymbol{o}} - \prod_{\lambda G}\overline{\boldsymbol{o}}$
from Algorithm 1, $\boldsymbol{W}^{(0)} = \boldsymbol{I}_p$, $\boldsymbol{z}^{(0)} = \boldsymbol{W}^{(0)}\,\boldsymbol{o}$. Given $\varepsilon > 0$, $\nu > 0$.
Do: Update $\boldsymbol{W}$ and $\boldsymbol{z}$:

$$\boldsymbol{W}^{(q+1)} = \boldsymbol{W}^{(q)} - \nu\frac{d\widehat{\mathcal{C}^{\text{ind}}}(\boldsymbol{W})}{d\boldsymbol{W}},$$

$\boldsymbol{z}^{(q+1)} = \boldsymbol{W}^{(q+1)}\,\boldsymbol{o}$.
Until $||\boldsymbol{W}^{(q+1)} - \boldsymbol{W}^{(q)}|| < \varepsilon$,
$\widehat{\boldsymbol{s}} = \boldsymbol{z}^{(q+1)}$.

4.2.2. The case of dependent source components

The criterion (4) in this case, is given as follows

$$\mathcal{C}^{\text{dep}}(\cdot)\colon \boldsymbol{W} \mapsto \mathcal{C}^{\text{dep}}(\boldsymbol{W}) := \mathcal{C}^{\text{dep}}_{\text{sep}}(\boldsymbol{W}) + \mathcal{C}_{\text{reg}}(\boldsymbol{z}). \tag{7}$$

We suggest to estimate the principle (7) by: $\boldsymbol{W} \mapsto \widehat{\mathcal{C}^{\text{dep}}}(\boldsymbol{W}) := \widehat{\mathcal{C}^{\text{dep}}_{\text{sep}}}(\boldsymbol{W}) + \mathcal{C}_{\text{reg,d}}(\boldsymbol{z})$, where $\widehat{\mathcal{C}^{\text{dep}}_{\text{sep}}}(\boldsymbol{W})$ present the approximate, of $\mathcal{C}^{\text{dep}}_{\text{sep}}(\boldsymbol{W})$, and it writes: $\widehat{\mathcal{C}^{\text{dep}}_{\text{sep}}}(\boldsymbol{W}) := \frac{1}{N}\sum_{i=1}^{N}\log\left(\frac{\widehat{c}_{\boldsymbol{Z}}(\widehat{F}_{Z_1}(z(i)),\ldots,\widehat{F}_{Z_p}(z_p(i)))}{\widehat{c}_{\widehat{\theta}}(\widehat{F}_{Z_1}(z(i)),\ldots,\widehat{F}_{Z_p}(z_p(i)))}\right)$. The source signals are approximated as follows: $\widehat{\boldsymbol{s}}(i) = \widehat{\boldsymbol{W}}\boldsymbol{o}(i)$, $i = 1,\ldots,N$, where $\widehat{\boldsymbol{W}} := \arg\inf_{\boldsymbol{W}}\widehat{\mathcal{C}^{\text{dep}}}(\boldsymbol{W})$, which is estimated using the gradient descent algorithm. In fact, the gradient of the approximated criterion, with respect to $\boldsymbol{W}$, is calculated as follows $\frac{d\widehat{\mathcal{C}^{\text{dep}}}(\boldsymbol{W})}{d\boldsymbol{W}} = \frac{1}{N}\sum_{n=1}^{N}\left[\frac{\frac{d}{d\boldsymbol{W}}\widehat{c}_{\boldsymbol{Z}}(\boldsymbol{v}(n))}{\widehat{c}_{\boldsymbol{Z}}(\boldsymbol{v}(n))} - \frac{\frac{d}{d\boldsymbol{W}}\widehat{c}_{\widehat{\theta}}(\boldsymbol{v}(n))}{\widehat{c}_{\widehat{\theta}}(\boldsymbol{v}(n))}\right] + \frac{\gamma}{N}\sum_{n=1}^{N}(\boldsymbol{z}(n) - \overline{\boldsymbol{z}}(n))(\boldsymbol{o}(n) - \overline{\boldsymbol{o}}(n))^{\top} + \frac{\mu}{N}\sum_{n=1}^{N}\sum_{j=-m}^{m}\alpha^{|j|}\left(I - \boldsymbol{G}^{-j}\right)\operatorname{sign}(\boldsymbol{z}(n) - \boldsymbol{G}^{j}\boldsymbol{z}(n))$, where $\boldsymbol{v}(n) := \left(\widehat{F}_{Z_1}(z(n)),\ldots,\widehat{F}_{Z_p}(z_p(n))\right)$; the gradients $\frac{d}{d\boldsymbol{W}}\widehat{c}_{\boldsymbol{Z}}(\boldsymbol{v}(n))$ and $\frac{d}{d\boldsymbol{W}}\widehat{c}_{\widehat{\theta}}(\boldsymbol{v}(n))$ can be calculated in a similar manner as in Subsection 4.2.1. We summarize the approach above in the following Algorithm.

5. Simulation results

In order to test the performance of the suggested approach, various simulations were conducted on four signal types: uniform i.i.d sources with independent components, dependent uniform i.i.d sources from Clayton copula [25], with $\theta = 1.5$, dependent uniform i.i.d sources from FGM copula [26], with $\theta = 0.8$ and dependent uniform i.i.d sources from Gumbel copula [27], with $\theta = 2.5$. For each source we have $N = 3000$ samples, and a centered Gaussian noise with deviation equal to 0.01 added, to gain two different signal-to-noise ratio (SNR) values: -25 dB and -15 dB. Considering the mixing matrix $\boldsymbol{M} := [1\ 0.7\ 0.7;\ 0.7\ 1\ 0.7;\ 0.7\ 0.7\ 1]$, the gradient descent parameter $\nu = 0.1$, while for the denoising part, we adopted $\tau = 0.1$, $\sigma = 0.01$, $\eta = 0.01$, $\gamma = 0.01$, $\mu = 0.01$ and $\varepsilon = 0.001$. To calculate the

Algorithm 3 BSS algorithm for separating noisy dependent source components.

Data: $\overline{\boldsymbol{o}}$ the noise free observation
Result: $\widehat{\boldsymbol{s}}$ the approximation of source
Initialization: Calculate $\boldsymbol{o} = \overline{\boldsymbol{o}} - \prod_{\lambda G} \overline{\boldsymbol{o}}$
from Algorithm 1, $\boldsymbol{W}^{(0)} = \boldsymbol{I}_p$, $\boldsymbol{z}^{(0)} = \boldsymbol{W}^{(0)}\, \boldsymbol{o}$. Given $\varepsilon > 0$, $\nu > 0$.
Do: Update $\boldsymbol{W}$ and $\boldsymbol{z}$:

$$\boldsymbol{W}^{(q+1)} = \boldsymbol{W}^{(q)} - \nu \frac{d\widehat{\mathcal{C}^{\text{dep}}}(\boldsymbol{W})}{d\boldsymbol{W}},$$

$$\boldsymbol{z}^{(q+1)} = \boldsymbol{W}^{(q+1)}\boldsymbol{o}.$$

Until $||\boldsymbol{W}^{(q+1)} - \boldsymbol{W}^{(q)}|| < \varepsilon$,
$\widehat{\boldsymbol{s}} = \boldsymbol{z}^{(q+1)}$.

quality of the estimated sources, the SNR was used. We also compare our experimental result with the result of [18] and those of [17], [28] (JADE), [29] (FastICA) and [30] (RADICAL), penalized by the same BTV and TV-regularization versions.

Table 1 shows that our proposed approach is satisfying performance at different noise level in the independent components case.

Table 1. Output SNR's for independent source components.

Noise	-25			-15		
Sources	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
Our method	**36.547**	**36.720**	**36.314**	**28.332**	**28.181**	**28.356**
Copula-TV	34.451	34.795	34.541	26.924	26.815	26.723
MI-BTV	34.639	34.841	34.412	26.721	26.302	26.510
MI-TV	33.429	33.772	33.851	25.872	25.836	25.869
FastICA-BTV	33.703	33.936	30.424	26.521	25.697	24.729
FastICA-TV	32.534	32.913	29.183	26.901	25.348	24.199
JADE-BTV	34.634	34.147	33.881	26.289	26.644	26.781
JADE-TV	33.513	33.371	33.209	26.355	26.210	26.732
RADICAL-BTV	34.783	33.839	34.137	26.418	25.983	26.027
RADICAL-TV	34.168	32.953	34.243	25.852	25.939	25.527

For dependent sources, our approach scored the highest SNR values for the three samples, proving its superiority compared to the other methods. The comparison results are summarized in Table 2, Table 3 and Table 4.

Table 2. Output SNR's (dependent components generated from FGM copula).

Noise	-25			-15		
Sources	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
Our method	**36.012**	**36.391**	**36.143**	**28.295**	**28.720**	**28.281**
Copula-TV	34.429	34.290	34.133	26.248	26.410	26.234
MI-BTV	15.934	15.862	15.934	12.455	12.223	12.484
MI-TV	15.013	15.772	15.300	11.484	11.044	11.918
FastICA-BTV	34.254	10.723	8.126	24.099	9.546	8.135
FastICA-TV	34.343	10.291	7.281	24.911	9.240	7.691
JADE-BTV	14.198	14.926	14.419	13.307	13.442	12.497
JADE-TV	13.024	13.115	13.392	13.102	13.237	12.891
RADICAL-BTV	14.624	14.319	14.713	12.491	14.211	12.703
RADICAL-TV	13.833	13.713	14.290	13.661	14.218	12.329

Table 3. Output SNR's (dependent components generated from Clayton copula).

Noise	-25			-15		
Sources	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
Our method	**34.138**	**34.616**	**34.924**	**27.753**	**27.300**	**27.290**
Copula-TV	32.574	32.949	32.538	25.846	25.237	25.822
MI-BTV	9.710	9.669	9.761	7.823	8.896	8.210
MI-TV	9.379	9.801	9.476	7.484	8.226	7.847
FastICA-BTV	24.633	6.619	3.785	22.649	5.439	3.732
FastICA-TV	23.956	5.646	3.657	22.125	5.671	3.238
JADE-BTV	9.096	9.098	9.936	8.439	8.350	8.344
JADE-TV	8.600	8.585	8.898	8.040	8.166	8.506
RADICAL-BTV	8.903	8.460	9.726	8.582	8.092	8.945
RADICAL-TV	8.623	7.697	9.145	7.605	8.475	8.299

Table 4. Output SNR's (dependent components generated from Gumbel copula).

Noise	-25			-15		
Sources	**S1**	**S2**	**S3**	**S1**	**S2**	**S3**
Our method	**36.518**	**36.118**	**36.130**	**28.405**	**28.185**	**28.185**
Copula-TV	34.364	34.124	34.705	26.215	26.215	26.503
MI-BTV	15.674	15.371	15.576	12.161	12.424	12.120
MI-TV	15.709	15.608	15.214	11.757	11.867	11.621
FastICA-BTV	34.629	10.509	8.107	24.565	9.848	8.135
FastICA-TV	34.583	10.145	7.940	24.394	9.774	7.074
JADE-BTV	14.247	14.540	14.928	13.393	13.486	12.580
JADE-TV	13.844	13.311	13.989	13.257	13.260	12.806
RADICAL-BTV	14.170	14.389	14.432	12.346	14.283	12.920
RADICAL-TV	13.224	13.305	14.666	13.089	14.588	12.664

6. Conclusion

In this paper a new BSS approach is presented for noisy environments. This technique eliminates noise from instantaneous linear mixtures of independent and dependent source components and then separates the mixture. Our approach was tested and compared to existing methods and exhibited its superiority. In addition, the proposed framework can be enlarged to function in future interchanges with convolutive independent/dependent mixtures.

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

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

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

[4] Belouchrani A., Abed-Meraim K., Cardoso J.-F., Moulines E. A blind source separation technique using second-order statistics. IEEE Transactions on Signal Processing. **45** (2), 434–444 (1997).

[5] Pesquet J.-C., Moreau E. Cumulant-based independence measures for linear mixtures. IEEE Trans. Inform. Theory. **47** (5), 1947–1956 (2001).

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

[7] Novey M., Adali T. ICA by maximization of nongaussianity using complex functions. 2005 IEEE Workshop on Machine Learning for Signal Processing. 21–26 (2005).

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

[9] Keziou A., Fenniri H., Ould Mohamed M., Delaunay G. Séparations aveugle de sources par minimisation des α-divergences. XXIIe colloque GRETSI, Dijon, 8–11 septembre 2009.

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

[11] Ghazdali A., Hakim A., Laghrib A., Mamouni N., Raghay S. A new method for the extraction of fetal ECG from the dependent abdominal signals using blind source separation and adaptive noise cancellation techniques. Theoretical Biology and Medical Modelling. **12**, Article number: 25 (2015).

[12] Mamouni N., Keziou A., Fenniri H., Ghazdali A., Hakim A. A new convolutive source separation approach for independent/dependent source components. Digital Signal Processing. **100**, 102701 (2020).

[13] Ourdou A., Ghazdali A., Laghrib A., Metrane A. Blind Separation of Instantaneous Mixtures of Independent/Dependent Sources. Circuits, Systems, and Signal Processing. **40**, 4428–4451 (2021).

[14] Ourdou A., Ghazdali A., Metrane A., Hakim M. Digital document image restoration using a blind source separation method based on copulas. In Journal of Physics: Conference Series. **1743**, 012034 (2021).

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

[16] Sahmoudi M., Snoussi H., Amin M. G. Robust approach for blind source separation in non-gaussian noise environments. Proceedings of ISCCSP, Marrakesh, Morocco, IEEE/EURASIP (2006).

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

[18] Ghazdali A., El Rhabi M., Fenniri H., Hakim A., Keziou A. Blind noisy mixture separation for independent/dependent sources through a regularized criterion on copulas. Signal Processing. **131**, 502–513 (2017).

[19] Tomasi C., Manduchi R. Bilateral Filtering for Gray and Color Images. Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271). 839–846 (1998).

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

[21] Farsiu S., Robinson D., Elad M., Milanfar P. Fast and Robust Multi-Frame Super-Resolution. IEEE Trans. on Image Processing. **13** (10), 1327–1344 (2003).

[22] El Mourabit I., El Rhabi M., Hakim A., Laghrib A., Moreau E. A new denoising model for multi-frame super-resolution image reconstruction. Signal Processing. **132**, 51–65 (2017).

[23] Afraites L., Hadri A., Laghrib A. A denoising model adapted for impulse and Gaussian noises using a constrained-PDE. Inverse Problems. **36** (2), 025006 (2019).

[24] Silverman B. W. Density estimation for statistics and data analysis. Monographs on Statistics and Applied Probability, Chapman & Hall, London (1986).

[25] Gumbel E. J. Bivariate exponential distributions. Journal of the American Statistical Association. **55** (292), 698–707 (1960).

[26] Morgenstern D. Einfache Beispiele zweidimensionaler Verteilungen. Mitteilungeblatt für mathematische statistik. **8**, 234–235 (1956).

[27] Clayton D. G. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika. **65** (1), 141–151 (1978).

[28] Cardoso J. F., Souloumiac A. Blind beamforming for non-gaussian signals. IEE Proceedings F (Radar and Signal Processing). **140** (6), 362–370 (1993).

[29] Hyvärinen A., Oja E. A fast fixed-point algorithm for independent component analysis. Neural Computation. **9** (7), 1483–1492 (1997).

[30] Miller E.-G., Fisher J.-W. III. Independent components analysis by direct entropy minimization, Tech. Rep. UCB/CSD-03-1221, University of California at Berkeley, January 2003.

Надійний підхід до сліпого розділення сумішей шумів незалежних і залежних джерел

Оурдоу А.[1], Газдалі А.[1], Лагріб А.[2], Метран А.[1]

[1] *LIPIM, ENSA Хурібга, Університет Султана Мулая Слімана, Хурібга, Марокко*
[2] *LMA, FST Бені-Меллал, Університет Султана Мулай Слімана, Бені-Меллал, Марокко*

У цій роботі представлено новий метод сліпого розділення джерел (СРД), який обробляє суміші шумів незалежних/залежних джерел. Це досягається мінімізацією критерію, що поєднує розділюючу частину (на основі розбіжності Кульбака–Лейблера для залежних або незалежних джерел) з частиною регуляризації, яка використовує двосторонню повну варіацію (ДПВ) з метою зниження шуму в спостереженнях. Запропонований алгоритм використовує алгоритм primal-dual для видалення шуму, тоді як метод градієнтного спуску реалізується для пошуку джерел сигналу. Представлений алгоритм довів свою ефективність та результативність, і навіть більше того, перевершив існуючі стандартні алгоритми СРД.

Ключові слова: *сліпе розділення джерел, суміші шумів, залежні джерела, двостороння загальна варіація, розбіжність Кульбака–Лейблера.*