Signal Processing 104 (2014) 319–324

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Fast communication

New blind source separation method of independent/dependent sources

A. Keziou [a,*], H. Fenniri [b], A. Ghazdali [c], E. Moreau [d,e]

[a] Laboratoire de Mathématiques de Reims EA 4535, Université de Reims Champagne-Ardenne, France
[b] CReSTIC, Université de Reims Champagne-Ardenne, France
[c] LAMAI, FST, Université Cadi-Ayyad, Morocco
[d] ENSAM, LSIS, UMR CNRS 7296, Université du Sud-Toulon-Var, France
[e] Aix Marseille Université, France

ARTICLE INFO

Article history:
Received 24 January 2014
Received in revised form
4 April 2014
Accepted 19 April 2014
Available online 30 April 2014

Keywords:
Blind source separation
Modified Kullback–Leibler divergence between copulas
Mutual information

ABSTRACT

We introduce a new blind source separation approach, based on modified Kullback–Leibler divergence between copula densities, for both independent and dependent source component signals. In the classical case of independent source components, the proposed method generalizes the mutual information (between probability densities) procedure. Moreover, it has the great advantage to be naturally extensible to separate mixtures of dependent source components. Simulation results are presented showing the convergence and the efficiency of the proposed algorithms.

1. Introduction

Blind source separation (BSS) is an instrumental problem in signal processing which has been addressed in the last three decades. We consider an instantaneous linear mixture described by

$$\mathbf{x}(t) := \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \in \mathbb{R}^p, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{p \times p}$ is an unknown non-singular mixing matrix, $\mathbf{s}(t) := (s_1(t), \ldots, s_p(t))^\top$ is the unknown vector of source signals to be estimated from $\mathbf{x}(t) := (x_1(t), \ldots, x_p(t))^\top$, the vector of observed signals. The number of sources and the number of observations, for the present work, are assumed to be equal. The presence of additive noise $\mathbf{n}(t)$ within the mixing model complicates significantly the BSS problem. It is reduced by applying some form of preprocessing such as denoising the observed signals through regularization approach, see e.g. [15]. The goal is to estimate the vector source signals $\mathbf{s}(t)$ using only the observed signals $\mathbf{x}(t)$. The estimate $\mathbf{y}(t)$ of the source signals $\mathbf{s}(t)$ can be written as

$$\mathbf{y}(t) = \mathbf{B}\mathbf{x}(t), \tag{2}$$

where $\mathbf{B} \in \mathbb{R}^{p \times p}$ is the de-mixing matrix. The question is how to obtain the de-mixing matrix $\widehat{\mathbf{B}}$ which has to be close to the ideal solution $\mathbf{A}^{-1}$, using only the observed signals $\mathbf{x}(t)$? It is well known, by Darmois theorem, that if the source components are mutually independent and at most one component is Gaussian, a consistent estimate $\widehat{\mathbf{B}}$ of $\mathbf{A}^{-1}$ (up to scale and permutation indeterminacies of rows) is the one that makes the components of the vector $\mathbf{y}(t)$ independent; see e.g. [6]. The corresponding signals $\widehat{\mathbf{y}}(t) := \widehat{\mathbf{B}}\mathbf{x}(t)$ are the estimate of the source signals $\mathbf{s}(t)$. Under the above hypotheses, many procedures have been proposed in the literature. Some of these procedures use second or higher order statistics, see e.g. [13,2] and the references therein, others consist of optimizing (on the

* Corresponding author.
E-mail addresses: amor.keziou@univ-reims.fr (A. Keziou), hassan.fenniri@univ-reims.fr (H. Fenniri), a.ghazdali@gmail.com (A. Ghazdali), moreau@univ-tln.fr (E. Moreau).

http://dx.doi.org/10.1016/j.sigpro.2014.04.017
0165-1684/

de-mixing matrix space) an estimate of some measure of dependency structure of the components of the vector $\mathbf{y}(t)$. As measures of dependence used in BSS, we find in the literature the criterion of mutual information (MI) [14,7], the criteria of α, β and Renyi's-divergences [5,19], and the criteria of ϕ-divergences [15]. The procedures based on minimizing estimates of MI are considered as the most efficient, since this criterion can be estimated efficiently, other procedures using divergences lead to robust method for appropriate choice of divergence criterion [15]. In this paper, we will focus on the criterion of MI (called also modified Kullback–Leibler divergence), viewed as measure of difference between copula densities, and we will use it to propose a new BSS approach that applies in both cases of independent or dependent source components. In the following, we will show that the mutual information of a random vector $\mathbf{Y} := (Y_1, \ldots, Y_p)^\top$ can be written as the modified Kullback–Leibler divergence (KL_m-divergence) between the copula of independence and the copula of the vector. Then, we propose a separation procedure based on minimizing an appropriate estimate of KL_m-divergence between the copula density of independence and the copula density of the vector. This approach applies in the standard case, and we will show that the proposed criterion can be naturally extended to separate mixture of dependent source components. The proposed approach can be adapted also to separate complex-valued signals. In all the sequel, we assume that at most one source is Gaussian, and we will treat separately the case of independent source components, and then the case of dependent source components. Chen et al. [3] proposed a BSS algorithm (for independent source components) based on minimizing a distance between the parameter of the copula of the estimated source and the value of the parameter corresponding to copula of independence. Ma and Sun [9] proposed a different criterion combining the MI between probability densities and Shannon entropy of semiparametric models of copulas.

2. Brief recalls on copulas

Consider a 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 marginals. We have for all $\mathbf{u} := (u_1, \ldots, u_p)^\top \in [0,1]^p$, $\mathbb{C}_\mathbf{Y}(\mathbf{u}) = \mathbb{P}(F_{Y_1}(Y_1) \leq u_1, \ldots, F_{Y_p}(Y_p) \leq u_p)$. Conversely, for any marginal d.f.'s $F_1(\cdot), \ldots, F_p(\cdot)$, and any copula function $\mathbb{C}(\cdot)$, the function $\mathbb{C}(F_1(y_1), \ldots, F_p(y_p))$ 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 $[0,1]$. So, if the components $Y_1, \ldots, Y_p$ are statistically independent, then the corresponding copula writes $\mathbb{C}_0(\mathbf{u}) := \prod_{j=1}^p u_j$, $\forall \mathbf{u} \in [0,1]^p$. It is called the copula of independence. Define, when it exists, the copula density (of the random vector $\mathbf{Y}$) $\mathbf{c}_\mathbf{Y}(\mathbf{u}) := (\partial^p / \partial u_1 \cdots \partial u_p) \mathbb{C}_\mathbf{Y}(\mathbf{u})$, $\forall \mathbf{u} \in [0,1]^p$. Hence, the copula density of independence $\mathbf{c}_0(\cdot)$ is the function taking the value 1 on $[0,1]^p$ and zero otherwise, namely,

$$\mathbf{c}_0(\mathbf{u}) := \mathbb{1}_{[0,1]^p}(\mathbf{u}), \quad \forall \mathbf{u} \in [0,1]^p. \tag{3}$$

Let $f_\mathbf{Y}(\cdot)$, if it exists, be the probability density on $\mathbb{R}^p$ 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 $\mathbf{y} \in \mathbb{R}^p$, we have

$$f_\mathbf{Y}(\mathbf{y}) = \prod_{j=1}^p f_{Y_j}(y_j) \mathbf{c}_\mathbf{Y}(\mathbf{u}), \tag{4}$$

where $\mathbf{u} := (u_1, \ldots, u_p)^\top := (F_{Y_1}(y_1), \ldots, F_{Y_p}(y_p))^\top$. In the monographs by [11,8], the reader may find detailed ingredients of the modeling theory as well as surveys of the commonly used semiparametric copulas.

3. Mutual information and copulas

The MI of a random vector $\mathbf{Y} := (Y_1, \ldots, Y_p)^\top \in \mathbb{R}^p$ is defined by

$$MI(\mathbf{Y}) := \int_{\mathbb{R}^p} -\log \frac{\prod_{j=1}^p f_{Y_j}(y_j)}{f_\mathbf{Y}(\mathbf{y})} f_\mathbf{Y}(\mathbf{y}) \, dy_1 \cdots dy_p. \tag{5}$$

It is called also the modified Kullback–Leibler divergence (KL_m-divergence) between the product of the marginal densities and the joint density of the vector. Note also that $MI(\mathbf{Y}) =: KL_m(\prod_{j=1}^p f_{Y_j}, f_\mathbf{Y})$ is nonnegative and achieves its minimum value zero if and only if (iff) $f_\mathbf{Y}(\cdot) = \prod_{j=1}^p f_{Y_j}(\cdot)$, i.e., iff the components of the random vector $\mathbf{Y}$ are statistically independent. An equivalent formula of (5) is

$$MI(\mathbf{Y}) := E\left(-\log \frac{\prod_{j=1}^p f_{Y_j}(Y_j)}{f_\mathbf{Y}(\mathbf{Y})} \right), \tag{6}$$

where $E(\cdot)$ is the mathematical expectation. Using the relation (4), and applying the change variable formula for multiple integrals, we can show that $MI(\mathbf{Y})$ can be written as

$$\begin{aligned} MI(\mathbf{Y}) &= \int_{[0,1]^p} -\log \left(\frac{1}{\mathbf{c}_\mathbf{Y}(\mathbf{u})} \right) \mathbf{c}_\mathbf{Y}(\mathbf{u}) \, d\mathbf{u} =: KL_m(\mathbf{c}_0, \mathbf{c}_\mathbf{Y}) \\ &= E(\log \mathbf{c}_\mathbf{Y}(F_{Y_1}(Y_1), \ldots, F_{Y_p}(Y_p))) =: -H(\mathbf{c}_\mathbf{Y}), \end{aligned}$$

where $H(\mathbf{c}_\mathbf{Y}) := \int_{[0,1]^p} -\log(\mathbf{c}_\mathbf{Y}(\mathbf{u})) \mathbf{c}_\mathbf{Y}(\mathbf{u}) \, d\mathbf{u}$ is the Shannon entropy of the copula density $\mathbf{c}_\mathbf{Y}(\cdot)$. The relation above means that the MI of the random vector $\mathbf{Y}$ can be seen as the KL_m-divergence between the copula density of independent $\mathbf{c}_0(\cdot)$, see (3), and the copula density $\mathbf{c}_\mathbf{Y}(\cdot)$ of the random vector $\mathbf{Y}$. We summarize the above results in the following proposition.

Proposition 1. *Let $\mathbf{Y} \in \mathbb{R}^p$ be any random vector with continuous marginal distribution functions. Then, the MI of $\mathbf{Y}$ can be written as the KL_m-divergence between the copula density $\mathbf{c}_0$ of independence and the copula density of the vector $\mathbf{Y}$:*

$$\begin{aligned} MI(\mathbf{Y}) &= \int_{[0,1]^p} -\log \left(\frac{1}{\mathbf{c}_\mathbf{Y}(\mathbf{u})} \right) \mathbf{c}_\mathbf{Y}(\mathbf{u}) \, d\mathbf{u} =: KL_m(\mathbf{c}_0, \mathbf{c}_\mathbf{Y}) \\ &= E(\log \mathbf{c}_\mathbf{Y}(F_{Y_1}(Y_1), \ldots, F_{Y_p}(Y_p))). \end{aligned} \tag{7}$$