Journal of Numerical Mathematics and Stochastics, **8** (1) : 01-08, 2016 © ***JNM@S***
http://www.jnmas.org/jnmas8-1.pdf
Euclidean Press, LLC
Online: ISSN 2151-2302

Blind Source Separation Based on Wavelet and Spearman’s Rho

S. SOUALHI, Z. MOKHTARI, and A. BOUSSAAD

Laboratory of Applied Mathematics, University of Biskra, BP 145, Algeria, E-mail: zouhir_m@yahoo.fr

Abstract. *The problem of blind source separation consists of estimating a set of source signals when only instantaneous linear mixtures are observed. In many cases, the information concerning the source signals and mixing matrix are not available. Through the strong assumption that the sources are mutually independent and of linearity of mixtures, the mixing matrix can be estimated to reveal the source signals. In this note, to minimize the associated computational work, the authors propose a novel source separation technique, based on the wavelet transform. To ensure the previously mentioned independence, the Spearman’s rho is employed as a criterion to be minimized within a new " genetics" algorithm. Some simulations are reported to illustrate the good performance of this algorithm.*

Key words : Blind Source Separation, Wavelet Transform, Spearman’s Rho, Image Encryption.

AMS Subject Classifications : 65C50, 65C99

1. Introduction

Blind sources separation (BSS) has been among the essential topics in the development of signal processing (see for example [2, 3, 6,8]). We consider in this note the simplest case of BSS when N sequences $X_1(t),\dots,X_N(t)$ are observed. Each one of these sequences is a linear combination of N independent unknown sequences $S_1(t),\dots,S_N(t)$. Here we can write $X(t) = MS(t)$ where $X(t)$ and $S(t)$ denote the vectors of components $X_1(t),\dots,X_N(t)$ and $S_1(t),\dots,S_N(t)$ respectively, while M is a square matrix that is called the mixing matrix.

The posing problem is to recover the unknown sources $S_1(t),\dots,S_N(t)$ from the observations, without any priori knowledge of their probabilistic structure. It is only assumed that the sources are mutually independent. The first solution of this problem proposed in [9], was based on cancellation of higher order moments. However, it has been proved [6, 7] that such an algorithm can diverge if the sources do not have an even probability density function.

To circumvent such a difficulty, other criteria have been used by several researchers, which are based on minimization of various cost functions, such as the sum of squares of forth-order cumulants [10, 4], or some contrast functions [1, 5] have been employed. Other authors related this problem of *BSS* to independent component analysis (*ICA*) which was introduced by Comon in [5], and later improved by Pham in [11].

Given a random vector X with a probability distribution P_x , the *ICA* problem is to find a transformation square matrix B such as the components of transformed vector BX are as independent as possible, if $X = AS$ with S having independent components, then $B = A^{-1}$ (where A is the mixing matrix), then B is a solution to the *ICA* problem. In this paper we propose a new genetics algorithm for blind source separation based on the discrete wavelet transform (DWT). A method that exploits the fundamental characteristic of this transform, which is namely the preservation of the signal shape in the approximation sub-band of the wavelet domain, and employs the Spearman's rho as a measure of dependence between the random variables. In particular, the Spearman's rho represents, when using the genetics algorithm, the criterion to be minimized. Finally, in section 4, some simulations are performed to illustrate the good performance of this method.

2. Spearman's Rho

Spearman's rho, being a measure of dependence between random variables, is an essential parameter of this work. In particular, we shall use the estimator of multivariate Spearman's rho introduced by F. Schmid in [12]. As in his paper, we shall estimate the Spearman's rho through the copula function.

At this point, particularly for this estimation of Spearman's rho, some essential notation needs to be fixed. Let $X_1,\dots,X_d$ be the set of d random variables with joint distribution function:
$F(x) = P(X_1 \le x_1, X_2 \le x_2, \dots, X_d \le x_d)$, where $x = (x_1,x_2,\dots,x_d) \in \mathbb{R}^d$ and marginal function $F_i(x) = P(X_i \le x)$ for $x \in \mathbb{R}^d$ and $i = 1,2,\dots,d$. If not stated otherwise, we will assume that the F_i are continuous functions. Thus, Sklar's theorem states that there exists a unique copula $C : [0,1]^d \to [0,1]$ such that $F(x) = C(F_1(x_1),\dots,F_d(x_d))$ for all $x \in \mathbb{R}^d$.

The copula C is the joint distribution function of the random variables
$U_i = F_i(X_i), i = 1,2,\dots,d$ where $U_i \sim U[0,1]$.
Moreover
$C(u) = F(F_1^{-1}(u_1), F_2^{-1}(u_2),\dots,F_d^{-1}(u_d))$ for all $u \in [0,1]^d$,
where F^{-1} represents the generalized inverse of F such as
$F^{-1}(u) := inf\{x \in \mathbb{R} \cup \{\infty\} / F(x) \ge u\}, \qquad \forall u \in [0,1]$,
and
$F^{-1}(0) := sup\{x \in \mathbb{R} \cup \{-\infty\} / F(x) = 0\}$.

According to the detailed treatment of copulas, we can state some important results concerning them.

1. Every copula C is bounded in the following sense
   $W(u) \le C(u) \le M(u)$,
   such as
   $W(u) := max\{u_1 + u_2 + \dots + u_d - (d-1), 0\}$,
   and

$M(u) := min\{u_1, u_2, \dots, u_d\}$, for all $u \in [0,1]^d$,
where M and W are called the upper and lower frechet-hoeffding bounds, respectively.

2. An important copula is the independence copula $\prod(u) = \prod_{i=1}^{d}(u_i), u \in \mathbb{R}^d$, describing the dependence structure of stochastically independent random variables $X_1, X_2, \dots, X_d$.

Authors of [8] give for the Spearman's rho in the case of d-dimensional random vector X with copula C the expression

$$\rho = \frac{\int_{[0,1]^d} C(u)du - \int_{[0,1]^d} \prod(u)du}{\int_{[0,1]^d} M(u)du - \int_{[0,1]^d} \prod(u)du} = \frac{d+1}{2^d - (d+1)}(2^d \int_{[0,1]^d} C(u)du - 1).$$

Here ρ can be interpreted as the normalized average distance between the copula C and the independent copula $\prod(u)$. In the case of $d = 2$, with a simple calculation, we can obtain the remarkable results

$$\int_{[0,1]^2} M(u_1, u_2)du_1du_2 = \frac{1}{3},$$

and

$$\int_{[0,1]^2} \prod(u_1, u_2)du_1du_2 = \frac{1}{4}.$$

Then the formula for ρ can be rewritten as

$$\rho = 12\int_0^1\int_0^1 C(u_1, u_2)du_1du_2 - 3.$$

2.1. **Non parametric estimation**

The aim of this activity is to estimate Spearman's rho via the copula. Let $(X_k)_{k=1,n}$ be a random sample from a d-dimensional random vector X with joint distribution function F and copula C, which are completely unknown.The non parametric estimator of the marginal distribution functions is

$$\hat{F}_{i,n}(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{X_{ik}} \leq x, \qquad \forall x \in \mathbb{R}.$$

And

$$\hat{U}_{ik,n} := \hat{F}_{i,n}(X_{ik}), \qquad i = 1, \dots, d \quad k = 1, \dots, n.$$

Note that

$$\hat{U}_{ik} := \tfrac{1}{n}(Rank\ of\ (X_{ik})\ in\ (X_{i1}, \dots, X_{in})),$$

and the copula C can be estimated by the empirical copula, which is defined as

$$\hat{C}_n(u) = \frac{1}{n}\sum_{k=1}^{n}\prod_{i=1}^{d} 1_{\{\hat{U}_{ik,n} \leq u_i\}}, \forall u = (u_1, \dots, u_d) \in [0,1]^d.$$

Finally the estimator of ρ becomes

$$\hat{\rho} = h(d)(2^d \int_{[0,1]^d} \hat{C}_n(u)du - 1) \tag{1}$$

$$= h(d)(\frac{2^d}{n}\sum_{k=1}^{n}\prod_{i=1}^{d}(1 - \hat{U}_{ik,n}) - 1), \tag{2}$$

with

$$h(d) = \frac{d+1}{2^d - (d+1)}.$$

3. Proposed Algorithm

In this section, we explain our algorithm to achieve a fast separation of several unknown source signals. As stated earlier, the algorithm is based on the discrete wavelet transform *DWT*. The role of this transform is to estimate the inverse of the mixing matrix from the approximation sub-band of the wavelet domain. As for the specific criterion to be minimized, it is the absolute value of the Spearman's rho ($|\rho|$). The proposed genetics algorithm is essentially our tool for the minimization of this criterion, and can be analyzed according to the following steps:

- Step 1

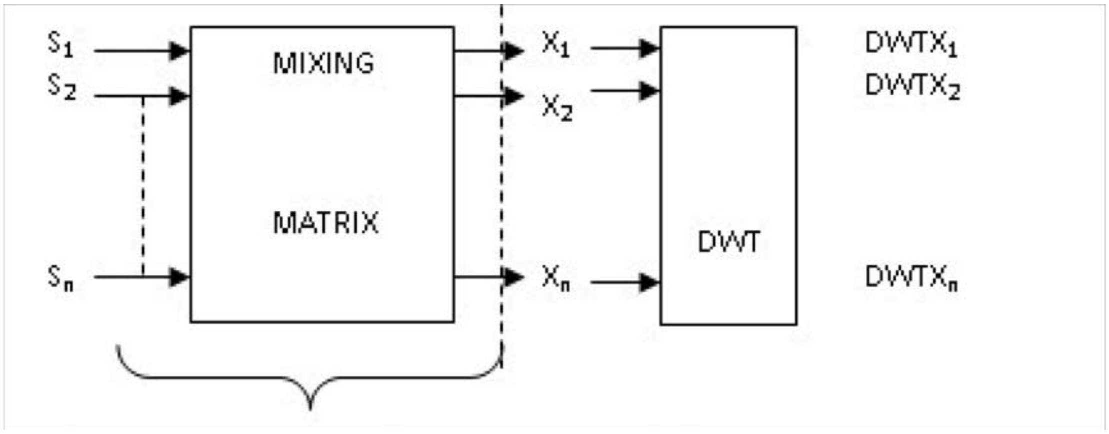

Figure 1: Decomposition of observed signals.

where

$$DWTX_i = [CA_i^3, CD_i^3, CD_i^2, CD_i^1], \qquad for \quad i = 1, \ldots, n,$$

with

CA_i^j : the approximation coefficient at level j,

CD_i^j : the detail coefficient at level j.

In this step we decompose each observed signal by the pyramidal digital wavelet transform, using the bi-orthogonal wavelet Bior(4.4) up to the level 3.

- Step 2

From the previous step we can formulate our objective function in the following way. Let *IM* be the inverse mixing matrix

$$IM = \begin{pmatrix} m_{11} & \ldots & m_{1n} \\ \ldots & \ldots & \ldots \\ m_{n1} & \ldots & m_{nn} \end{pmatrix},$$

where *IM* is an unknown square matrix and, $CA = [CA_i^3, CA_2^3, \ldots, CA_n^3]$ is the vector of the

approximation coefficients of wavelet level decompositions. The computational formula for the objective function is
$f(m) = |Rho(IM.CD)|$,
in which *Rho* is the Spearman's rho, $IM.CD$ is the simple product between the matrix IM and the vector *CD* and *m* is a vector of the variables with dimension $n \times n$.

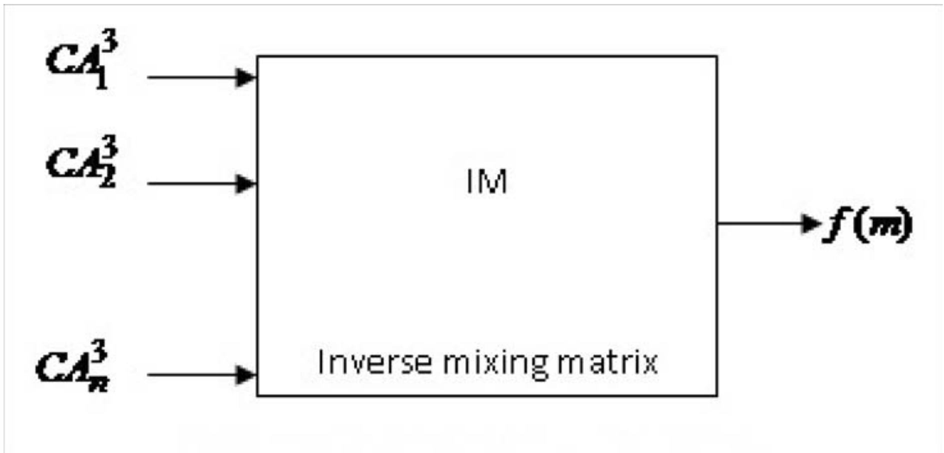

Figure 2: Formulation of the objective function.

- Step 3

In this part of the algorithm, we estimate the inverse mixing matrix by the optimum m^* of the objective function which is computed by the "genetics" algorithm.

- Step4

In this step we estimate the source signals $\hat{S}_1, \hat{S}_2, \ldots, \hat{S}_n$ with the simple product between the observed signals $X_1, X_2, \ldots, X_n$ and the previous estimated mixing matrix $\widehat{IM}$

4. Computational Results

We present here some computational examples of signals, that were processed by the genetics algorithm, for the case of two and three source signals. In particular, we present some obtained graphical results, which demonstrate the effectiveness of the proposed method in the recovery of the source signals shape. Apparently, the source signals are recovered within expectations, and using some processing techniques after the separation operation, we have obtained the source signals almost exactly, as illustrated in the following Figures 3-5.

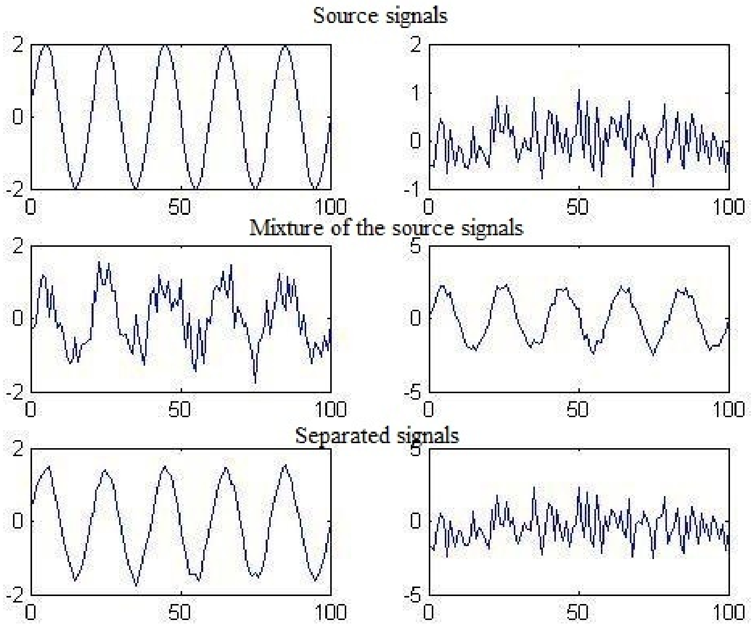

Figure 3: A sinusoidal signal with Gaussian noise.

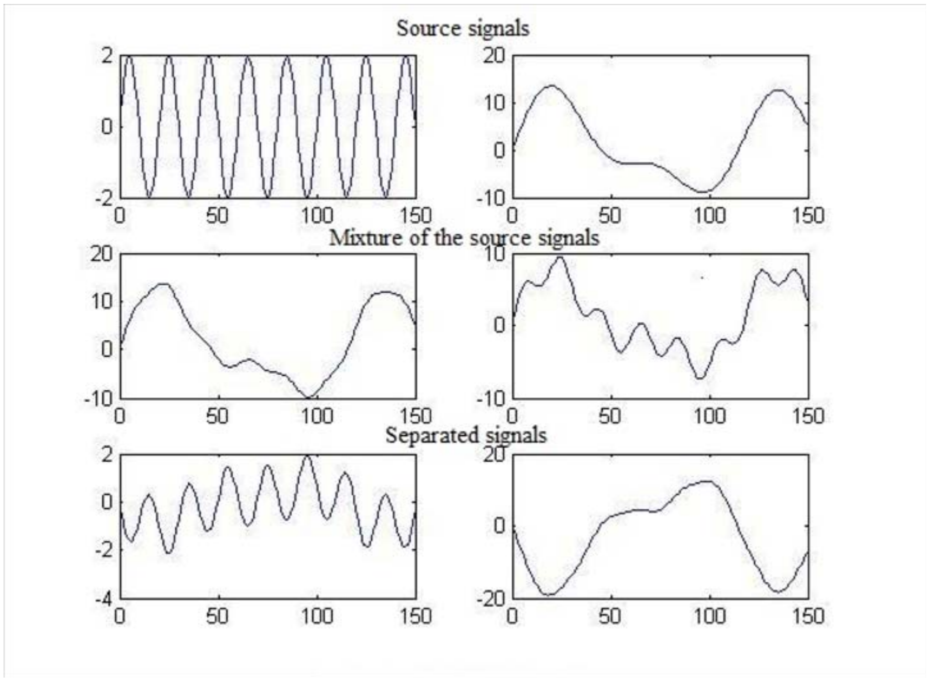

Figure 4: A two sources signal.

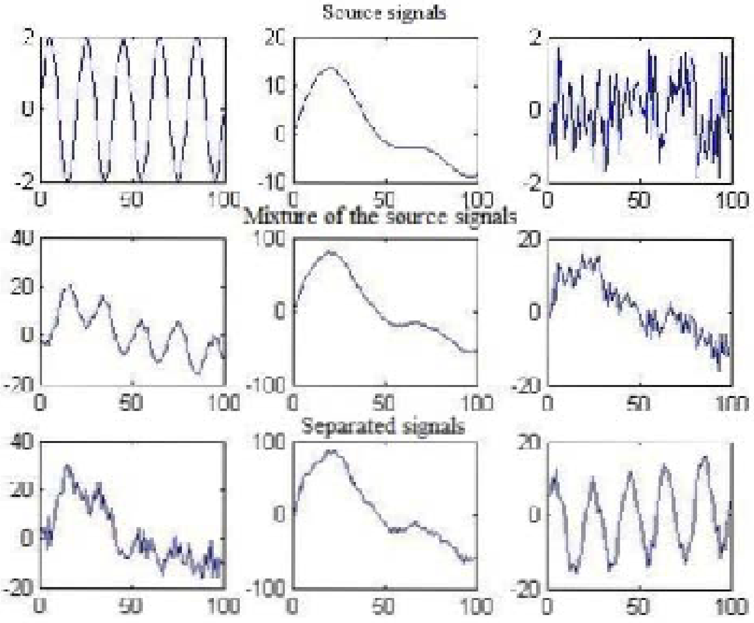

Figure 5: A three sources signal.

5. Conclusion

In this paper we have presented a new algorithm for blind source separation based on discrete wavelet transform with the aim to minimize the computational work. Particularly, we estimate the mixing matrix through the sub-band approximation coefficients in the wavelet domain. Minimization of the Spearman's rho serves in this algorithm as a measure of dependence between sources. The reported computational results demonstrate the effectiveness of this new algorithm for BSS.

Acknowledgments
We are grateful to an anonymous referee for some very useful remarks.

References

[1] J-F. Cardoso, Source separation using higher order moments, *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, *ICASSP-89*, IEEE, (1989), 2109-2112.

[2] J-F. Cardoso, Iterative technique for blind source separation using only fourth order cumulants, *Proceedings of EUSIPCO'***92**, (1992), 739-742.

[3] J-F. Cardoso, and B. H. Laheld, Equivariant adaptive source separation, *IEEE Transactions on Signal Processing* **44**(12), (1996), 3017-3030.

[4] P. Comon, Separation of sources using higher-order cumulants, *Proceedings of the 33rd Annual Techincal Symposium*, International Society for Optics and Photonics, (1989),170-183.

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

[6] P. Comon, C. Jutten, and J. Hérault, Blind separation of sources, Part II: Problems Statement, *Signal Processing* **24**, (1991), 11-20.

[7] J-C. Fort, Stabilité de l'algorithme de séparation de sources de Jutten et Hérault, *Traitement du Signal* **8**(1), (1991), 35-42.

[8] M. Gaeta, J-L. Lacoume, Source separation without a priori knowledge: the maximum likelihood solution, *Proceedings of EUSIPCO'***90**, Barcelona(Spain), (1990), 621-624.

[9] J. Hérault, C. Jutten, and B. Ans, Detection de grandeurs primitives dans un message composite par une architecture de calcul neuromimetique en apprentissage non supervise, *Dixieme Colloque sur le Traitement du Signal et ses Applications*, Nice, France, 1985.

[10] J. Lacoume, and P. Ruiz, Sources indentification: a solution based on the cumulants, *Fourth Annual ASSP Workshop on Spectrum Estimation and Modeling,* IEEE, (1988), 199-203.

[11] D. T. Pham, Blind separation of instantaneous mixture of sources via an independent component analysis, *IEEE Transactions on Signal Processing* **44**(11), (1996), 2768-2779.

[12] F. Schmid, and R. Schmidt, Multivariate extensions of Spearman's rho and related statistics, *Statistics & Probability Letters* **77**(4), (2007), 407-416.

Article history: Submitted February, 14, 2016; Revised April, 04, 2016; Accepted May, 27, 2016.