Learning and Nonlinear Models (L&NLM) – Journal of the Brazilian Neural Network Society, Vol. 10, Iss.1, pp. 51-69, 2012.

INDEPENDENT COMPONENT ANALYSIS AND BLIND SIGNAL SEPARATION: THEORY, ALGORITHMS AND APPLICATIONS

Eduardo F. Simas Filho[1,2], José M. de Seixas[1], Natanael N. Moura[1], Diego B. Haddad[1,3], José M. Faier[1], Maria C. S. Albuquerque[4]

[1] COPPE/Poli/Federal University of Rio de Janeiro, Brazil
[2] DEE/Poli/Federal University of Bahia, Brazil
[3] Federal Center of Technological Education, Rio de Janeiro, Brazil.
[4] Federal Institute of Education, Science and Technology, Bahia, Brazil.

E-mails: esimas@lps.ufrj.br, seixas@lps.ufrj.br, natmoura@lps.ufrj.br, diego@pads.ufrj.br, faier@lps.ufrj.br, cleaalbuquerque@ifba.edu.br

Abstract – This paper reviews Independent Components Analysis (ICA) and Blind Signal Separation (BSS) problems. An overview on the main statistical principles that guide the search for the independent components is formulated, methods for blind signal separation that require both high-order and second-order statistics are also illustrated. Some of the most successful algorithms for both ICA and BSS are derived. Experimental applications in different signal processing tasks such as passive sonar, nondestructive ultrasound inspection and electrical-load time series are presented.

Index Terms – ICA, Blind Source Separation, Signal processing, Feature extraction.

Introduction

In several multidimensional signal processing problems, it is desired to find a data transformation so that their essential structure becomes somehow more accessible [1]. Usually there is not much information available and the search for the desired data representation is performed through unsupervised learning. Among the linear techniques that searches for data transformation we can mention Principal Component Analysis (PCA) [2], Factor Analysis (FA) [3], Independent Component Analysis (ICA)[1] and Blind Signal Separation (BSS) [4].

The ICA model (in its linear and instantaneous version) considers that a set of N measured (observed) signals $\boldsymbol{x} = [x_1, ..., x_N]^T$ is generated by a linear combination of N unknown sources $\boldsymbol{s} = [s_1, ..., s_N]^T$:

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \tag{1}$$

where $\boldsymbol{A}$ is the $N\times N$ mixing matrix [1] (the N-dimensional vectors $\boldsymbol{s}$ and $\boldsymbol{x}$ denote, respectively, single observations of the source and measured signals, this notation will be used in the rest of the work). The purpose of ICA is to obtain an estimate $\boldsymbol{y}$ of the original source signals $\boldsymbol{s}$ using only the observed (mixed) data $\boldsymbol{x}$. For this, an inverse linear model is assumed:

$$\mathbf{y} = \mathbf{W}\mathbf{x}. \tag{2}$$

A general principle for estimating the de-mixing matrix $\mathbf{W}$ can be found by considering that the components of $\boldsymbol{y} = [y_1, ..., y_N]^T$ are statistically independent (or as independent as possible). There are many mathematical methods used to search for statistical independence. Among the most applied ones (in the ICA context) we can mention the nonlinear decorrelation and the maximally nongaussianity [5]. In the search for independent components, high-order statistics (HOS) information is usually required. There are some indeterminacies in the ICA model: the order of extraction of the independent components can change and unknown scalar multipliers (positive or negative) may be modifying the estimated components. Fortunately these limitations are insignificant in most applications [1].

The ICA model, as formulated in Equations 1 and 2, is also referred to as a Blind Signal Separation (BSS) method. Although ICA and BSS are closely related methods, blind separation can sometimes be achieved using only second-order statistics (SOS) information. This is the case when the source signals present temporal structure [4]. There exist also applications of ICA that do not assume a mixing model. For example, in feature extraction problems the independent components are estimated to reveal underlying characteristics of the data, instead of separating unknown source signals.

In the last decades, several algorithms have been proposed for solving the ICA problem, see for reference [6, 7, 8]. The ICA and BSS models have been applied successfully to different signal processing tasks like noise removal [9, 10, 11], passive sonar signal separation [12], telecommunications [13], biomedicine [14], face recognition [15, 16] and experimental high-energy physics [17]. A typical signal separation application that motivated theoretical and experimental developments in ICA/BSS research fields is the so-called cocktail-party problem. As illustrated in Figure 1, considering that in a room there are two speakers (sources) talking simultaneously, and two microphones positioned in different locations, the recorded signals are a linear combination of the sources (considering a simplified model where noise or sound propagation multi-path are both not considered). A human being has the ability to focus attention to one person and diminish surrounding noise. However, this is not a trivial task for an automatic signal processing system.

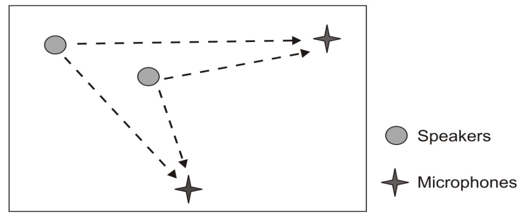


Figure 1 – Cocktail-party problem, the sound propagation paths are illustrated in dashed lines.

A simple experimental example of ICA application (in this case through FastICA algorithm [8]) for the separation of two linearly mixed sinusoidal signals is illustrated in Figure 2. The problem here is to recover the sources in Figure 2-a using only the observed data (Figure 2-b). As it can be seen, the estimated signals (see Figure 2-c) are very similar to the source signals (only their sign and scaling factors are modified).

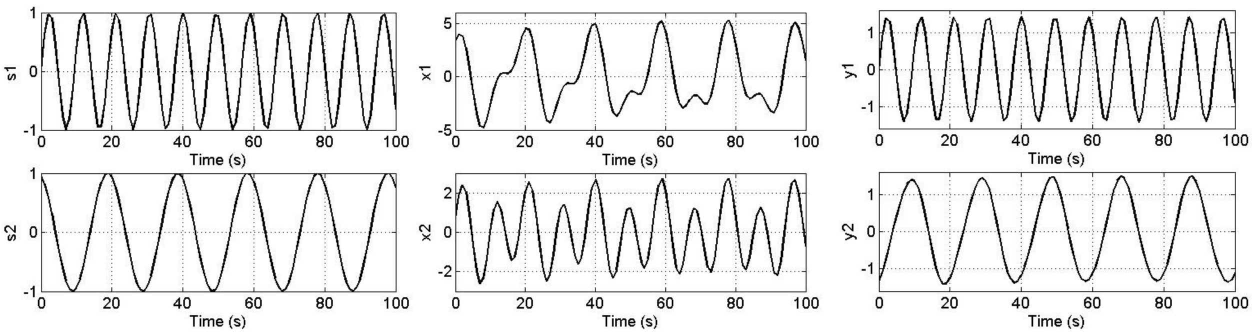


Figure 2 – Application of ICA for blind signal separation, sources (a), mixed (b) and blind recovered (c) signals.

This paper is divided as it follows. Section 1 comprises a detailed description of the statistical independence measures often used to derive ICA algorithms learning rules. Section 2 details some techniques often used as a pre-processing for ICA. In Section 3, successful separation algorithms for both ICA and BSS are derived. Extensions of the ICA model are described in Section 4 and experimental results obtained from synthesized and practical application signals are illustrated in Section 5. Conclusions are derived in Section 6.

1- Statistical Independence

If two random variables y_1 and y_2 are statistically independent, the following condition (necessary and sufficient) holds [18]:

$$p_{y_1,y_2}(y_1, y_2) = p_{y_1}(y_1)p_{y_2}(y_2), \tag{3}$$

where $p_{y1,y2}(y_1, y_2)$, $p_{y1}(y_1)$ and $p_{y2}(y_2)$ are, respectively, the joint and marginal probability density functions (pdf) of y_1 and y_2. Equivalent condition is obtained if, for all absolutely integrable functions $g(\cdot)$ and $h(\cdot)$, the expression on Equation 4 holds:

$$E\{g(y_1)h(y_2)\} = E\{g(y_1)\}E\{h(y_2)\} \tag{4}$$

where $E\{\cdot\}$ is the expectation operator [1]. Very little information on the source signals statistics is available in typical blind signal processing problems and so the pdfs estimation (required in Equation 3) is a very difficult task, which can be avoided using Equation 4. Different mathematical principles are used in ICA framework to determine whether random variables are statistically independent. In the following sub-sections some of these principles will be derived.

1.1- Non-gaussianity Leads to Independency

The ICA/BSS model described in Equation 1 can be re-written as:

$$x_i = \sum_{j=1}^{N} a_{ij} s_j \qquad i=1,..,N. \tag{5}$$

Considering the central limit theorem [19], which may be stated as: "The sum of two (independent) random variables is always closest to a Gaussian distribution than the original variables", and Equation 5, one can see that the observed signals x_i are formed by an averaged summation of the sources s_i. Thus, x_i are closer to Gaussian-distributed variables than s_i. In other words, the independent components can be obtained through maximization of non-gaussianity [1]. The non-gaussianity principle cannot be applied if one or more source signals are Gaussian distributed.

1.2- Nonlinear Decorrelation

As defined in Equation 4, one way to verify statistical independence is through nonlinear decorrelation. Linear correlation is verified by second-order statistics, while independence requires higher-order information, which in nonlinear decorrelation ICA methods is introduced through nonlinear functions.

By definition (see Equation 4), it is necessary to verify the correlation between all possible combinations of nonlinear functions in order to guarantee statistical independence between two random variables [1]. In practice, estimates of the independent components are obtained while guaranteeing decorrelation between a finite set of nonlinear functions. For example, the ICA algorithm, proposed by Cichocki and Unbehauen in [5], searches for independent components while providing decorrelation between a hyperbolic tangent and a polynomial function applied to the input signals.

1.3- High-order Cumulants

Moments and cumulants are statistical descriptors used to characterize the nature of a random variable distribution. Considering a random variable y, the moment α_k and central moment μ_k of order k are defined by [20]:

$$\alpha_k = E\{y^k\} = \int_{-\infty}^{\infty} y^k p_y(y)dy\,, \tag{6}$$

$$\mu_k = E\{(y-\alpha_1)^k\} = \int_{-\infty}^{\infty} (y-\alpha_1)^k p_y(y)dy\,. \tag{7}$$

The first moment $\alpha_1 = m_y$ is the mean of y and the second central moment μ_2 is the variance. If the random variable y is zero mean (or if the mean is removed: $y \leftarrow y - m_y$), than for all k holds: $\alpha_k = \mu_k$.

Cumulants are an alternative to moments and in some cases provide simpler theoretical treatment. The cumulant κ_k of order k is defined as a function of the moments [20]. For a zero mean random variable y, the first four cumulants are:

$$\kappa_1 = \mathbf{0}\,;\;\; \kappa_2 = E\{y^2\} = \alpha_2\,;\;\; \kappa_3 = E\{y^3\} = \alpha_3\,; \tag{8}$$
$$\kappa_4 = E\{y^4\} - 3[E\{y^2\}]^2 = \alpha_4 - 3\alpha_2^2$$

The third and fourth order cumulants are called respectively skewness (κ_3) and kurtosis (κ_4) [21]. Cumulants of order higher than four are rarely applied in practical ICA/BSS problems. Cumulants can be easily estimated from data substituting expectations in Equation 7 by sample means. Some interesting properties of cumulants are:

$$\begin{gathered}\kappa_k(y_1 + y_2) = \kappa_k(y_1) + \kappa_k(y_2) \\ \kappa_k(y) = 0, \text{ for } k > 2 \text{ if } y \text{ is Gaussian}\end{gathered} \tag{9}$$

Considering this, the gaussianity of a random variable can be estimated by cumulants of order higher than two. The skewness value, for example, is related to pdf symmetry ($\kappa_3=0$ indicates symmetry). Spanning the interval $[-2, \infty)$, kurtosis is zero for a Gaussian variable. Negative values indicate sub-gaussianity (pdf flatter than Gaussian) and positive values super-gaussianity (pdf sharper than Gaussian) [20]. One disadvantage is that higher-order cumulants can be seriously influenced by outliers (observations that are numerically distant from the rest of the data). Thus, in extreme situations the kurtosis value may be dominated by a small number of points [21]. Some studies have been conduced with the purpose of obtaining robust estimation of high order cumulants [22].

1.4- Information Theory Contrasts

Information theory [23] deals with the quantification and description of information contained in a random variable. Some of the parameters used in information theory are usually applied to the search for independent components. A key parameter in information theory is **entropy**, which for a discrete random variable y is defined as [24]:

$$H(y) = \sum_i P(y = a_i) \log_b P(y = a_i) \tag{10}$$

where a_i are the possible values assumed by the variable $\mathbf{y}$, and $P(y = a_i)$ is the probability that $y = a_i$. There are a number of possible choices for the logarithm basis b, one commonly used is $b=10$, in this case, the entropy is measured in "digits". Intuitively, entropy quantifies the "uncertainty" of a random variable. Entropy is also expressed by the number of bits (code length) needed to represent the variable.

Negentropy is another information theory parameter and is computed through [23]:

$$J(y) = H(y_{gauss}) - H(y) \tag{11}$$

where y_{gauss} is a Gaussian random variable with the same mean and variance of y.

Both entropy and negentropy can be used as non-gaussianity measures because the Gaussian variable has maximum entropy between the variables of same variance [1]. The advantage of $J(y)$ is that it is always non-negative and zero when y is Gaussian. Considering a blind signal processing application, a problem with the computation of $J(\cdot)$ and $H(\cdot)$ is the pdf estimation (see Eq. 10). To avoid this, approximations using high-order cumulants or non-polynomial functions shall be applied [1, 25]. Commonly used negentropy approximations will be derived in the next sub-section.

Another statistical independence measure can be obtained through mutual information. **Mutual Information** measures the information that can be obtained about one variable by observing another. Considering this definition, if two random variables are independent, their mutual information is zero. The mutual information $I\ (y_1\ ,\ y_2\ ,\ ...,\ y_m\ )$ between m random variables $\mathbf{y} = [y_1\ ,\ y_2\ ,\ ...,\ y_m\ ]$ is obtained through [3]:

$$I(y_1,...,y_m) = \sum_{i=1}^{m} H(y_i) - H(\mathbf{y}) \tag{12}$$

Therefore, minimization of mutual information leads to statistical independence [23].

The **Kullback-Leibler (KL) divergence** is a parameter used to compare two different distributions and may be defined through:

$$C_{KL}(Q,P) = \int Q_y(y) \log \frac{Q_y(y)}{P_y(y)} dy \tag{13}$$

The Kullback-Leibler divergence is always nonnegative with minimum value zero when both densities are the same. If one pdf is Gaussian, maximizing C_{KL} is equivalent to maximize non-gaussianity.

1.4.1- Approximations of the Negentropy

As from equation (11), negentropy computation requires pdf estimation, which is not always computationally affordable in blind signal processing problems. Considering this some approximations for $J(\cdot)$ will be presented here.

Negentropy can be estimated through higher-order moments as follows [26]:

$$J(y) \approx \frac{1}{12} E\{y^3\}^2 + \frac{1}{48} kurt(y)^2 \tag{14}$$

To avoid problems by using the kurtosis as a cost function (considering that this operator is not robust to outliers), approximations based on the maximum-entropy principle shall be applied, as proposed in [27]:

$$J(y) \propto [E\{G(y)\} - E\{G(\nu)\}]^2 \tag{15}$$

where $\boldsymbol{G}$ is a non-quadratic function. If G does not grow too fast, more robust estimators are obtained. Some recommended choices are [27]:

$$G_1(u) = \frac{1}{a_1} \log \cosh a_1 u \quad , \quad G_2(u) = -\exp(-u^2/2) \tag{16}$$

a_1 is some suitable constant (usually $1 < a_1 < 2$).

2- Pre-processing for ICA/BSS

The standard ICA model assumes a mixing system where the number of sources and observed signals are the same. Additive noise is not considered either. Some signal processing procedures, which will be described in the following sub-section, are usually applied as pre-processing steps for ICA algorithms in order to improve their performance.

2.1- Principal Component Analysis

Principal Component Analysis (PCA) [2] is a statistical signal processing technique, also known as Karhunen-Loeve transform, which searches for a space where the projections $y_i=\boldsymbol{b}_i\, x_i$ of a zero mean random vector $\boldsymbol{x}$ ($E\{\boldsymbol{x}\} = 0$) in the direction $\boldsymbol{b}_i$ are non-correlated (second-order statistics is explored) and have maximum variance (i.e. composing an orthonormal basis). The principal components are ordered by their energy (variance), which is usually concentrated on a small number of components.

The first direction $\boldsymbol{b}_1$, can be computed through the maximization of [1]:

$$F_1^{PCA}(\mathbf{b}_1) = E\{z_i^2\} = E\{(\mathbf{b}_1\mathbf{x})^2\} = \mathbf{b}_1^T\mathbf{C}_x\mathbf{b}_1 \tag{17}$$

where $\boldsymbol{C}_x$ is the covariance matrix of $\boldsymbol{x}$ and $\|\boldsymbol{b}_1\|=\boldsymbol{1}$.

PCA can also be defined as a mean square error (MSE) compression of $\boldsymbol{x}$. The projection of $\boldsymbol{x}$ into the subspace spanned by the basis vectors $\boldsymbol{b}_i$ is $\sum_{i=1}^{m}(\mathbf{b}_i^T\mathbf{x})\mathbf{b}_i$, and the MSE criterion that shall be minimized by the orthonormal basis becomes:

$$F_{MSE}^{PCA} = E\{\| \mathbf{x} - \sum_{i=1}^{m}(\mathbf{b}_i^T\mathbf{x})\mathbf{b}_i \|^2\} \tag{18}$$

PCA for the vector $\boldsymbol{x}$ is equivalent to the eigenvalue decomposition of the matrix $\boldsymbol{C}_x$ [2]. Estimation of the principal components can also be performed through neural network models [28, 29].

PCA transformation is very useful for ICA as it eliminates second-order dependencies (correlation) between the signals, facilitating the search for independence. After PCA, the signals present zero mean and may have unit variance through a whitening process. Whitening is a linear transformation of the data $\boldsymbol{x}$ into $\boldsymbol{z} = \boldsymbol{Qx}$, such that the covariance matrix of $\mathbf{z}$ is the unity matrix:

$$\mathbf{E}\{\mathbf{zz}^{\mathbf{T}}\} = \mathbf{I} \tag{19}$$

Most ICA algorithms perform better over pre-whitened data; moreover, some ICA routines require this transformation in order to obtain accurate estimates of the independent components.

PCA is also applied for dimension reduction in over-determined (when there exist more sensors than sources) multidimensional problems. Measured data can be projected linearly into a subspace so that the maximum amount of information (in least square sense) is preserved if only the more energetic components are retained. If the number of independent components N is known, so the N more energetic principal components shall be preserved. Dimension reduction is also useful to reduce noise level and prevent over-learning.

2.2- Principal Discriminating Components

Considering a supervised classification problem, signal compaction can be performed using information on the target classes. The Principal Components for Discrimination (PCD) algorithm was initially proposed in [30] and uses target information to obtain a transformation of the input signals that maximizes discrimination efficiency (class separation) and data compaction rate simultaneously. PCD was successfully applied as a preprocessing step for ICA in classification problem in [31, 32].

The PCD can be performed through a MLP neural network [28], which is trained to maximize class discrimination. For simplicity, considering a binary discrimination process, a network with a single hidden neuron, extracts the first discriminating component z_1 (see Figure 3-a), which is defined as:

$$z_1 = \mathbf{d}_1 \times \mathbf{x}^T \tag{20}$$

where $\mathbf{d}_1=[d_{1,1}, d_{2,1}, \ldots, d_{N,1}]$ are the hidden neuron synaptic weights. The training procedure may be the traditional error back-propagation [28], using distinct target values for input signals belonging to different classes. The estimation of the next principal discriminating components is performed through the following procedure: 1. Add a new hidden neuron; 2. Keeping fixed only the previously estimated input layer weights (dashed lines in Figure 3-b), restart the training procedure; 3. Evaluate the discrimination efficiency; 4. Continue this procedure until there is no significant discrimination efficiency improvement by adding new hidden neurons.

The estimated PCDs are computed through:

$$\mathbf{z} = \begin{pmatrix} d_{1,1} & d_{2,1} & \cdots & d_{N,1} \\ d_{1,2} & d_{2,2} & \cdots & d_{N,2} \\ \vdots & \vdots & \vdots & \vdots \\ d_{1,j} & d_{2,j} & \cdots & d_{N,j} \end{pmatrix} \times \mathbf{x}^T \tag{21}$$

or, alternatively, in a more compact representation: $\mathbf{z}=\mathbf{D}\mathbf{x}^T$.

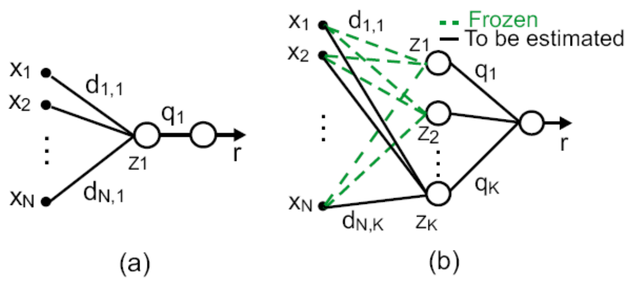


Figure 3 – Neural models to estimate (a) the first PCD and (b) the k-th PCD.

2.3- Noise Reduction

ICA/BSS estimation is severely degraded in presence of additive noise. Usual algorithms perform worse as the noise level increases and sometimes it is not possible to obtain meaningful results [33]. In practical applications some kind of noise is always present. A noisy ICA model can be described through:

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \mathbf{n} \tag{22}$$

where $\boldsymbol{n}$ is the random noise vector.

If some information on the random noise is available, such as frequency contents or statistical characteristics, it is strongly recommended that some appropriated signal processing procedure is performed to remove or at least reduce noise into acceptable levels.

Wavelet de-noising methods have been recently applied to improve the performance of ICA/BSS algorithms [33, 34]. Using discrete wavelet transform (DWT), a signal $\boldsymbol{y(t)}$ can be decomposed into both detail $\boldsymbol{D[k]}$ and approximation $\boldsymbol{H[k]}$ coefficients [35]. To achieve proper multi-resolution description, sequential decomposition levels shall be used (see Figure 4). The approximation signal is a smooth version of $\boldsymbol{y(t)}$, corresponding to a type of low-pass filtering, while $\boldsymbol{D[n]}$ carries high-frequency information. The de-noising algorithm used in this work basically consists on applying a threshold $\boldsymbol{T_m}$ to the detail coefficients of the decomposition level $\boldsymbol{m}$ such that:

$$D_m[k] = \begin{cases} \mathbf{0}, & D_m[k] \le T_m \\ D_m[k], & D_m[k] > T_m \end{cases} \tag{23}$$

The wavelet noise removal, different from a standard filtering, is able to eliminate high frequency components (that are usually produced by noise sources) in different frequency ranges (at each decomposition level).

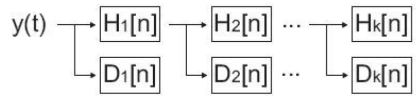


Figure 4 – Wavelet decomposition diagram.

3- Algorithms

Many algorithms have been proposed during the past years to solve the ICA/BSS problem. In this section, the attention was focused in some successful ICA algorithms such as Nonlinear Decorrelation [7], FastICA [8], JADE [6] and the Multiplicative Newton-like algorithm [36]. Approaches that explore second-order statistics (SOS) to perform blind source separation of temporary correlated signals [37] are also presented.

3.1- Nonlinear Decorrelation / Nonlinear PCA

As described through Equation 4, statistical independence can be achieved through nonlinear decorrelation. A problem with the direct application of this method is that, theoretically, all possible combinations of nonlinear functions need to be tested.

Approximations of the independent components are obtained by algorithms that perform nonlinear decorrelation between a finite number of functions. For example the Cichocki-Unbehauen algorithm [7] proposes a feedforward neural network, which is trained to estimate the demixing matrix $\boldsymbol{W}$ ($\boldsymbol{y}=\boldsymbol{Wx}$). The learning rule for $\boldsymbol{W}$ is [1,7]:

$$\Delta\mathbf{W} = \mu[\Lambda - f(\mathbf{y})g(\mathbf{y}^{\mathrm{T}})]\mathbf{W} \tag{24}$$

where $f(.)$ and $g(.)$ are nonlinear scalar functions (the authors have suggested a polynomial and a hyperbolic tangent), μ is the learning rate and $\boldsymbol{\Lambda}$ is a diagonal matrix (usually $\boldsymbol{\Lambda}=\boldsymbol{I}$). It is proved in [7] that if the learning rule in Equation 24 converges to a nonzero matrix $\boldsymbol{W}$, the outputs are nonlinearly decorrelated. To perform ICA using the algorithm proposed in equation 24, the signals must be pre-whitened.

The Nonlinear Principal Component Analysis (NLPCA) [38] is an extension of Principal Component Analysis (PCA), and thus can be defined through a modification of equation 17, deriving the following criterion to be minimized [39]:

$$J = E\{\| x - \sum_{i=1}^{m} g_i(w_i^T x) w_i \|^2\} \tag{25}$$

where $\boldsymbol{g}=[g_1,\ldots g_m]^T$ is a set of nonlinear functions (that needs to be specified a priori). The NLPCA criterion have been applied to estimate the ICA model in [39-41] and proved to be an efficient approach for this problem. Pre-whiten is also required here in order to NLPCA produce statistical independence. In [42], connections were established between NLPCA (for pre-whitened data) and other ICA criterion such as kurtosis maximization.

A nonlinear recursive least-squares (RLS) learning rule was proposed in [43] for ICA estimation through NLPCA. The method is a modification of the PAST algorithm [44] used for linear PCA. Considering that $\boldsymbol{z}$ is a whitened zero-mean variable ($\boldsymbol{y}=\boldsymbol{Wz}$), the updates for the RLS-NLPCA learning rule are [43]:

$$\mathbf{q}[k+1] = \mathbf{g}(\mathbf{W}[k]\mathbf{z}[k]) \tag{26}$$

$$\mathbf{h}[k+1] = \mathbf{P}[k]\mathbf{q}[k+1] \tag{27}$$

$$\mathbf{m}[k+1] = \mathbf{h}[k+1]/\beta + \mathbf{q}^T[k+1]\mathbf{h}[k+1] \tag{28}$$

$$\mathbf{P}[k+1] = \frac{1}{\beta}\mathbf{Tri}(\mathbf{P}[k] - \mathbf{m}[k+1]\mathbf{h}^T[k+1]) \tag{29}$$

$$\mathbf{r}[k+1] = \mathbf{z}[k+1] - \mathbf{W}^T[k]\mathbf{q}[k+1] \tag{30}$$

$$\mathbf{W}[k+1] = \mathbf{W}[k] + \mathbf{m}[k+1]\mathbf{r}^T[k+1] \tag{31}$$

The index k denotes the iteration step, the variables $\boldsymbol{q}$, $\boldsymbol{h}$, $\boldsymbol{m}$, $\boldsymbol{r}$ and $\boldsymbol{P}$ are internal to the algorithm and β is a forgetting constant. The matrix operator ***Tri****(.)* retains only the upper triangular part of a matrix, transposes it and copies to the lower triangular part, resulting in a symmetric matrix. As a recursive (on-line) algorithm, this method is able to track slow statistical variations of the data.

3.2- Fast Fixed-Point Algorithms (FastICA)

Considering the negentropy (*J(y)*) approximations of Equations 14 and 15, and the fact that minimizing *J(y)* leads to non-gaussianity maximization and thus statistical independence, the first choice for optimization algorithm for this criterion would be a gradient based one [45], as it is simple to implement and present low computational requirements. A limitation of such algorithms is that they present slow convergence and, moreover, a bad choice of the learning rate may destroy convergence [5].

The Fast Fixed-Point Algorithms so called FastICA were firstly proposed in [8] and were derived through approximate Newton-type iterations. Among the advantages of the method we can mention fast and more reliable convergence, computational simplicity and little memory requirements [5, 8].

Through some manipulations of Equation 15, the FastICA algorithm for estimation of one independent component is formulated as follows for pre-whitened data [5]:

1. Choose an initial (random) weight vector **w**;
2. Let $\mathbf{w}^{+} = E\{\mathbf{x}g(\mathbf{w}^{T}\mathbf{x})\} - E\{g'(\mathbf{w}^{T}\mathbf{x})\}\mathbf{w}$;
3. Let $\mathbf{w} = \mathbf{w}^{+} / \|\mathbf{w}^{+}\|$
4. If not converged, go back to step 2.

where *g(.)* in step 2 is the derivative of *G(.)* (see equations 14 and 15).

The authors suggest the use of one of the following nonlinear functions *G* (and their respective derivatives *g*) [8]:

$$G_1(u) = \frac{1}{a_1}\log\cosh a_1 u \quad , \quad g_1(u) = \tanh(a_1 u) \tag{32}$$

$$G_2(u) = -\frac{1}{a_2}\exp(-a_2 u^2/2) \quad , \quad g_2(u) = u\exp(-a_2 u^2/2) \tag{33}$$

$$G_3(u) = \frac{1}{4}u^4 \quad , \quad g_3(u) = u^3 \tag{34}$$

where $1 \le a_1 \le 2, \quad a_2 \approx 1$. The characteristics of the different contrast functions are [8]: G_1 is a good general purpose contrast function; G_2 are indicated when the independent components are highly super-Gaussian; and the use of G_3 is only justified for estimating sub-Gaussian components when there are no outliers.

In order to estimate several independent components, some deflationary orthogonalization using, for example, the Gram-Schmidt method [1], must be used. Modifications on the one unit FastICA algorithm which leads to estimation of several components simultaneously are available in [1, 5, 8].

3.3- Tensorial Methods

Tensors are considered as a higher dimensional generalization of matrices or linear operators [46]. Cumulant tensors are matrices containing the cross-cumulants. Considering this, the second order cumulant tensor is the covariance matrix (***C***) and the fourth-order tensor ($\mathbf{T}_4$) is formed by the fourth-order cross-cumulants $cum(x_i,x_j,x_k,x_l)$, that for zero mean random variables is defined as:

$$cum(x_i,x_j,x_k,x_l) = E\{x_i,x_j,x_k,x_l\} - E\{x_i,x_j\}E\{x_k,x_l\} - E\{x_i,x_k\}E\{x_j,x_l\} - E\{x_k,x_j\}E\{x_i,x_l\} \tag{35}$$

The fourth-order cumulant tensor $\mathbf{T}_4$ is a four-dimensional array, where each element is defined as $q_{ijkl}=cum(x_i,x_j,x_k,x_l)$, the indexes *i*, *j*, *k*, *l* vary from 1 to N (where N is the number of signals). The fourth-order cumulant tensor contains all fourth-order information of the data.

ICA tensorial methods are derived through a procedure analogous to diagonalization of covariance matrix ***C***, which produces signal decorrelation [2]. As $\mathbf{T}_4$ is a fourth-order counterpart of ***C***, independence can be achieved by diagonalizing $\mathbf{T}_4$, as for independent signals the only non-zero fourth-order cross-cumulants appears when $i=j=k=l$. Analogous to the second-order case, diagonalization of the fourth-order tensor can be achieved through eigenvalue decomposition (EVD).

Using tensorial methods for ICA are theoretically simple, but computing EVD of four-dimensional matrices by ordinary algorithms requires a very large amount of memory and may be computationally prohibitive in some cases. In order to avoid this limitation, methods like FOBI (Fourth-Order Blind Identification) [47] and JADE (Joint Approximate Diagonalization of Eigenmatrices) [6] were proposed in the literature. The last one will be described in the next sub-section and is considered to be the most successful ICA tensorial method [1]. A connection between Tensorial methods and FastICA for kurtosis absolute value maximization is established in [1].

3.3.1- JADE

JADE algorithm proposes an approximate method for diagonalization of $\mathbf{T}_4$. Considering that the ICA model for pre-whitened data is satisfied:

$$\mathbf{z} = \mathbf{Bx} = \mathbf{BAs} = \mathbf{Ws} \ , \tag{36}$$

where $\mathbf{B}$ is the whitening matrix. In this case the cumulant tensor of $\mathbf{z}$ has a special structure and its eigenmatrices are described through [1]:

$$\mathbf{M} = \mathbf{w}_m \mathbf{w}_m^T \ , \tag{37}$$

where m=1,…,N and $\mathbf{w}_n$ are the rows of $\mathbf{W}$.

JADE algorithm uses the linear transformation F_{ij} applied to $\mathbf{M}$ (m_{kl} is an element of $\mathbf{M}$):

$$F_{ij}(\mathbf{M}) = \sum_{kl} m_{kl} \ \mathrm{cum}(x_i, x_j, x_k, x_l) \tag{38}$$

In JADE, a set of different matrices $\mathbf{M}_i$, i=1,…,k is taken and the purpose is to searches for a matrix $\mathbf{W}$ that makes $Q = \mathbf{W}F(\mathbf{M}_i)\mathbf{W}^T$ as diagonal as possible (one option for the $\mathbf{M}_i$ is to use the eigenmatrices of the cumulant tensor as they carry all the relevant information from the cumulants) [1]. The diagonality of matrix Q can be measured through the sum of the squares of off-diagonal elements.

3.4- Multiplicative Newton-Like Algorithm

A multiplicative ICA algorithm was proposed by Akuzawa and Murata in [48]. Using the kurtosis as cost function, this method applies second-order optimization (through the Newton method [32,45]) in the search for independent components (instead of first-order gradient iterations used in most of ICA algorithms).

This algorithm does not require pre-whitening and thus operates directly over the data. Some experimental results obtained in [49, 50] indicate that Akuzawa's algorithm overperforms both FastICA and JADE in the presence of Gaussian noise.

The purpose here (similar to other ICA algorithms) is to find a linear transformation C: $\mathbf{y}=\mathbf{Cx}$ (where $\mathbf{x} = [x_1, \ldots, x_N]^T$ and $\mathbf{y} = [y_1, \ldots, y_N]^T$), which, maximizes the independence between the components of $\mathbf{y}$. The following steps are executed during the iteration:

1. Chose C_o (initial de-mixing matrix) and Δ_o $(N \times N)$;
2. Evaluate the iteration step: $C_t = exp(\Delta_{t-1})C_{t-1}$;
3. Evaluate the cost function at C_t using a second-order expansion around C_{t-1};
4. Δ_t Is chosen as a saddle point of the cost function;
5. Back to step 2 until convergence.

More details on how the step 4 is executed can be found in [48]. Modifications on this method are proposed in [49] in order to reduce the computational cost by substituting the pure-Newton optimization by quasi-Newton iterations.

Comparing to FastICA or JADE, Akuzawa's algorithm is slower (even in its modified quasi-Newton version), and thus its application is not justified in noiseless mixtures. Otherwise, if high noise levels are present or when the problem is over-determined (i.e. number of sources is smaller than observations), the multiplicative algorithm shall present better separation performance.

3.5- Algorithms for Blind Source Separation of Time Structured Signals

The algorithms derived so far consider that the source signals are mutually independent random variables (no temporal structure was assumed). In many applications, however, the sources are time-dependent signals instead of random variables. In contrast to the standard ICA model, in which the samples have no particular order, when dealing with time structured signals, the samples cannot be shuffled as they present temporal structure. Additional information, obtained for example from autocovariances, can be applied to derive ICA algorithms that do not use the non-gaussianity principle and therefore can be applied to Gaussian signals (if they are correlated over time) [1].

One way to estimate the independent components of time series is to consider that data auto-covariances and cross-covariances are different from zero. All these statistics for a given time lag can be grouped in the time-lagged covariance matrix [51, 52]:

$$\mathbf{C}_\tau^x = E\left\{\mathbf{x}(t)\mathbf{x}(t-\tau)^T\right\} \tag{39}$$

where: $\mathbf{x}(t)=[x_1(t),\dots,x_k(t)]^T$ is the mixed signals vector (for zero delay) and $\mathbf{x}(t-\tau)$ is the mixed signal vector for the time delay τ ($\tau=1,2,3,\dots$).

Here, second order statistics (contained in the covariance matrix) is applied for blind source separation, instead of high-order information. The motivation for using covariances is that for two independent signals $\mathbf{y}_i(t)$ and $\mathbf{y}_j(t)$ the cross-covariance matrix is zero for all time delays:

$$\mathbf{C}_\tau^{yi,yj} = E\left\{\mathbf{y}_i(t)\mathbf{y}_j(t-\tau)^T\right\} = 0 \qquad \text{for } i \neq j \tag{40}$$

Considering now k time structured signals, it is equivalent to say that the covariance matrix of $\mathbf{y}(t)=[y_1(t),\dots,y_k(t)]^T$ is diagonal for all τ if the signals are independent (note that if the covariance matrix is zero only for $\tau=0$, the signals are said to be uncorrelated). The observed signals $\mathbf{x}(t)$ shall be made independent by searching for a matrix $\mathbf{W}$ ($\mathbf{y}(t)=\mathbf{W}\mathbf{x}(t)$) such that the estimated sources $\mathbf{y}(t)$ are uncorrelated for all time-delays.

There exists some algorithms in the literature that explore temporal structure of data in order to perform blind signal separation, among them we can mention AMUSE (Algorithm for Multiple Unknown Signals Extraction) [53] and SOBI (Second-Order Blind Identification) [54]. As it is not always possible (due to computational limitations) to verify all possible combinations of time-delays, practical algorithms usually verify a limited number of τ values.

AMUSE algorithm is based on the diagonalization of the covariance matrix for only one specific time delay τ_o. Considering that the observed signals $\mathbf{x}(t)$ are pre-whitened, generating the variables $\mathbf{z}(t)$, the de-mixing matrix $\mathbf{W}$ is obtained through eigenvalue decomposition of $\bar{C}_\tau^z = \frac{1}{2}\left[C_\tau^z + (C_\tau^z)^T\right]$, for $\tau = \tau_o$. The method can perform efficient signal separation only if the eigenvalues are distinct, which do not occurs frequently. This problem can be attenuated by searching for a proper time-delay which produces distinct eigenvalues.

An adaption of the eigenvalue decomposition method is proposed in SOBI algorithm [54] in order to allow simultaneous diagonalization of several matrices. Using this joint-diagonalization procedure, the separation matrix $\mathbf{W}$ is obtained considering several delayed covariance matrices and thus more robust results are obtained when compared to AMUSE.

More recently, some works [55, 56] had dedicated attention to second order signal separation, as this principle revealed to be very useful in the convolutive mixture case, which will be addressed in section 4.1.

4- Extensions to standard ICA/BSS models

In some practical applications, the standard ICA/BSS model (see Equations 1 and 2) do not describe properly the problem. For a more realistic application-dependent modeling, information on noise, multiple propagation paths or some sort of nonlinear shall be considered. In the following sub-sections extensions to the standard ICA/BSS models are illustrated.

4.1- Multi-channel Blind Deconvolution

The multi-channel convolutive blind source separation (CBSS) is also referred as a blind source separation of convolutive mixtures and consists of an extension of the basic instantaneous ICA model (see Eq. 1 and 2) in order to consider the existence of multiple propagation paths in the mixing system. This model better describes some acoustic, seismic, wireless communications and medical applications, where signal propagation is performed in a multipath environment. A M-tap mixing system can be described as:

$$x_p(n) = \sum_{n'=0}^{N}\sum_{k=0}^{M-1} h_{p,n'}(k)s_n(n-k) \tag{41}$$

where $h_{p,n'}(k)$ are attenuation coefficients of each path until the p-th sensor. The sources estimates can be expressed in terms of separation filters $w_{q,n'}$ whose aim is to invert the mixture system:

$$y_q(n) = \sum_{n'=0}^{N} \sum_{k=0}^{L-1} w_{q,n'}(k) x_{n'}(n-k) \tag{42}$$

The CBSS task slightly differs from the Multichannel Blind Deconvolution (BD) one. While BD algorithms aim to preserve the time structure of the signals, in CBSS methods, distorted versions of the sources are allowed. In digital communications, only BD techniques can be used. When the sources are acoustic signals, we can accept some distortion. Let $s_i(n)$ be the n-th sample of the i-th source and $y_j(n)$ the n-th sample of the j-th estimate. If y_j estimates the i-th source, the objective of BD and CBSS algorithms is to get [57]:

$$y_j^{BD}(n) = c_j s_i(n - \Delta_j) \tag{43}$$

$$y_j^{CBSS}(n) = \sum_{k=0}^{\infty} c_{j,k} s_i(n-k) \tag{44}$$

where c_j accounts for the *scaling* ambiguity and $c_{j,k}$ for the *filtering* ambiguity (for more details, see [58]). In this work, we do not focus in BD algorithms for convolutive mixtures, nor in channel identification techniques. In BSS methods there is a *scaling* ambiguity because of our ignorance about the dynamic level of the sources. As a result, our estimates, even in successful cases, are (approximately) scaled versions of the original sources. Our ignorance also implies that we cannot determine the order of the independent components and, for example, the third estimate can be a scaled version of the first source. The numbering of the sources is an arbitrary procedure, and the BSS algorithm cannot discover our labeling.

There are two main approaches for the CBSS problem: in the time [59-61] and in the frequency-domain [62-67]. Concisely, time-domain algorithms frequently use non-trivial extensions of ICA cost functions and they tend to present higher computational costs and less distortions and artifacts (estimation errors). A promising time-domain algorithm is the one presented in [59], which will be explained below.

Although frequency-domains methods are faster than their time-domain counterparts, they often introduce artifacts in their estimates. They rely on convolution-product duality between time and frequency domains, which permit us to handle each frequency bin as an instantaneous mixture. This simple idea has three serious problems: i) in the STFT transform, the convolution must be circular and not linear, as usual; ii) the scaling ambiguity can change the power of each bin in a different manner (producing a significant distortion) and iii) the permutation ambiguity is able to exchange bins between estimates.

The first problem can be attenuated by zeroing (in the time-domain) the final part of each separation filter. This procedure can also alleviate the permutation problem. The second problem can be handled by the minimum distortion principle [68], while for the last one, an envelope correlation [62, 63], a direction of arrival [64, 65] or a hybrid approach [66] were proposed. One successful frequency-domain algorithm was presented in [67] and will be detailed later.

4.1.1 – Time-Domain CBSS

The method presented in [55] will be explained here. This method separates nonstationary and nonwhite sources (such as voice and music) using only second order statistics. Let L be the separation filters length, D the number of time-lags considered for the correlations ($1 \leq D \leq L$) and P ($P \geq D$) the length of output signals blocks used as a basis for the estimates of short-time correlations. The m-th block taken for processing can be placed in a matriz **U**:

$$\mathrm{U}_{\mathrm{p}}^{\mathrm{T}}(m) = \begin{bmatrix} x_p(mL) & x_p(mL-1) & \cdots & x_p(mL-L+1) \\ x_p(mL+1) & x_p(mL) & \ddots & x_p(mL-L+2) \\ \vdots & \ddots & \ddots & \vdots \\ x_p(mL+N-1) & \cdots & \cdots & x_p(mL-L+N) \end{bmatrix} \tag{45}$$

And the q-th output block can be written as:

$$\mathrm{Y}_q(m) = \begin{bmatrix} y_q(mL) & y_q(mL+1) & \cdots & y_q(mL+N-1) \\ \vdots & \ddots & \ddots & \vdots \\ y_q(mL-D+1) & y_q(mL-D+2) & \ldots & y_q(mL-D+N) \end{bmatrix} \tag{46}$$

So, defining:

$$\mathrm{X}_{\mathrm{p}}(m) = \begin{bmatrix} U_p(m) \\ U_p(m-1) \end{bmatrix} \tag{47}$$

$$\mathrm{W}_{pq}(m)=\begin{bmatrix} w_{p,q}(0) & w_{p,q}(1) & \cdots & w_{p,q}(L) & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & w_{p,q}(0) & w_{p,q}(1) & \cdots & w_{p,q}(L) & 0 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & 0 & \ddots & \ddots & 0 \\ 0 & \cdots & w_{p,q}(0) & w_{p,q}(1) & \cdots & w_{p,q}(L) & 0 & \cdots & 0 \end{bmatrix}_{Dx2L} \tag{48}$$

$$\mathrm{W}=\begin{bmatrix} \mathrm{W}_{11} & \cdots & \mathrm{W}_{1P} \\ \vdots & \ddots & \vdots \\ \mathrm{W}_{P1} & \cdots & \mathrm{W}_{PP} \end{bmatrix} \tag{49}$$

we can write $\mathrm{Y}_q(m)=\sum_{p=1}^{N}\mathrm{W}_{qp}\mathrm{X}_p(m)$ and $\mathrm{Y}(m)=\mathrm{WX}(m)$, where $\mathrm{X}(m)=\left[\mathrm{X}_1^T(m) \quad \cdots \quad \mathrm{X}_N^T(m)\right]^T$ and:

$$Y(m)=\begin{bmatrix} Y_1(m) \\ \vdots \\ Y_N(m) \end{bmatrix} \tag{50}$$

The defined cost function (a generalization of what was proposed in [69]) has roots in a general information-theoretical approach [70] and is written as:

$$\Im(m)=\sum_{i=0}^{\infty}\beta(i,m)\{\log[\det[\mathrm{bdiag}(Y(i)Y^T(i))]]-\log[\det[(Y(i)Y^T(i))]]\} \tag{51}$$

where $\beta(i,m)$, the learning factor, allows us to use on-line or off-line iterations (using only the blocks where $\beta(i,m)\neq 0$). The natural gradient of $\Im(m)$ w.r.t. **W** can be expressed as:

$$\nabla_{\mathrm{W}}^{\mathrm{NG}}\Im(m)=2\sum_{i=0}^{\infty}\beta(i,m)\mathrm{W}\left[Y(m)Y^T(m)\right]\left[\mathrm{bdiag}\left(Y(m)Y^T(m)\right)\right]^{-1} \tag{52}$$

where the bdiag(.) operator, acting on a partitioned block matrix consisted of several matrices, sets all submatrices on the off-diagonals to zero. For example, in $N = 2$ and determined case, we have:

$$\mathrm{bdiag}\left(\begin{bmatrix} \mathrm{Y}_1(m)\mathrm{Y}_1^{\mathrm{T}}(m) & \mathrm{Y}_1(m)\mathrm{Y}_2^{\mathrm{T}}(m) \\ \mathrm{Y}_2(m)\mathrm{Y}_1^{\mathrm{T}}(m) & \mathrm{Y}_2(m)\mathrm{Y}_2^{\mathrm{T}}(m) \end{bmatrix}\right)=\begin{bmatrix} \mathrm{Y}_1(m)\mathrm{Y}_1^{T}(m) & 0 \\ 0 & \mathrm{Y}_2(m)\mathrm{Y}_2^{\mathrm{T}}(m) \end{bmatrix} \tag{53}$$

The Sylvester matrix **W** update is $\mathrm{W}(m)=\mathrm{W}(m-1)+\mu\Delta\mathrm{W}(m)$. After updating, we get the first line of each $W_{pq}(m)$ matrix to reconstruct the $w_{p,q}$ filters and then we construct again $\mathrm{W}(m)$. There are other methods to obtain the filters from matrix $W(m)$ that can be more intuitive or more specific (case-sensitive); see [71] for more details. In order to reduce the computational cost, filter banks can be used [72].

4.1.2 – Frequency-Domain CBSS

The CBSS method described in [67] is fast and usually has good performance. It begins by windowing the mixtures, using the Hanning window of length *K* and shift length of *K*/4. Then, the STFT transform is applied and each bin becomes an instantaneous mixture. Let $W^{(k)}$ be the separation matrix of the *k*-th bin and $\mathrm{y}_i^{(k)}$ be the estimate of the *i*-th source at *k*-th bin.

To avoid the permutation problem, a multidimensional cost function is used. The update of each element of $W^{(k)}$ can be written as:

$$\Delta w_{ij}^{(k)}=\sum_{l=1}^{N}\left\{\delta(i-l)-E\left[\varphi^{(k)}\left(y_i^{(1)}\ldots y_i^{(K)}\right)y_l^{*(k)}\right]\right\}w_{lj}^{(k)} \tag{54}$$

where $E[\cdot]$ is the statistical average operator and $\varphi^{(k)}$ is a multidimensional *score function* defined as $\varphi^{(k)}\left(y_i^{(1)} \ldots y_i^{(K)}\right) = \frac{y_i^{(k)}}{\sqrt{\sum_{k=1}^{K}\left|y_i^{(k)}\right|^2}}$. This equation can be exposed in a matricial form, that leads to the well-known update:

$$W \leftarrow W + \mu\{I - E[\varphi(Y)Y^T]\}W \tag{55}$$

After convergence, we use the MD Principle [68] in each matrix $W^{(k)}$:

$$W^{(k)} \leftarrow \mathrm{diag}[W^{-1(k)}]W^{(k)} \tag{56}$$

where the diag operator zeroes the matrix off-diagonals elements. The MD Principle atenuates the scaling ambiguity. If we suppose that $W^{(k)} \approx D^{(k)}H^{-1(k)}$, where $D^{(k)}$ is a diagonal matrix that contains scaling coefficients, the MD Principle gives us $diag[W^{-1(k)}]W^{(k)} \approx diag[H^{(k)}]H^{-1(k)}$, where we have a reasonable scaling (not an arbitrary one). As a final procedure, we apply the inverse STFT-transform to get the estimates.

4.2- Nonlinear ICA/BSS

In problems where there exists some sort of nonlinear phenomena during the signal mixing process, the linear ICA model may present poor results [73]. A more general formulation considers that the measured signals $\boldsymbol{x}$ (where each sample are formed by a nonlinear instantaneous mixing model:

$$\mathbf{x} = F(\mathbf{s}) \tag{57}$$

where $F(.)$ is a $R^N \rightarrow R^N$ nonlinear mapping (the number of sources is assumed to be equal to the number of observed signals) and the purpose is to estimate an inverse transformation $G : R^N \rightarrow R^N$:

$$\mathbf{y} = G(\mathbf{x}) \tag{58}$$

so that the components of y are statistically independent. If $G = F^{-1}$ the sources are perfectly recovered [74].

A characteristic of the Nonlinear ICA problem is that the solutions are nonunique [74]. If y_1 and y_2 are independent random variables, it is easy to prove that $f(y_1)$ and $g(y_2)$, where $f(.)$ and $g(.)$ are differentiable functions, are also independent. So, it is clear that, without some restrictions, there is an infinite number of solutions for the inverse mapping G in a given application. The nonlinear blind source separation is a more restrictive problem as its purpose is to estimate the original source signals from their nonlinear mixed version. Nonlinear BSS cannot be achieve without some prior information on the mixing model or sources. A complete investigation on the uniqueness of nonlinear ICA solutions can be found in [75]. NLICA algorithms have been recently applied in different problems such as speech processing [76] and image denoising [77]. A complete review on nonlinear ICA/BSS theory, algorithms and applications can be found in [78].

5- Experimental Applications

In this Section are presented some experimental applications of ICA/BSS algorithms in different signal processing tasks such passive sonar signal detection, feature extraction in ultrasound inspection, information retrieval in time series, convolutive blind signal separation of music signals and NLICA for feature extraction in experimental high-energy physics.

5.1- Passive Sonar Signal Detection

In passive sonar [79, 80], a hydrophone array shall be used to examine the acoustic waves received from different directions. The acquired signals are used by experienced operators to verify if an important target is within the system reach. This procedure is time demanding and susceptible to human failure. A particular limitation in passive sonar systems is that the signals are immerse in huge background noise (generated from several underwater acoustic sources). Another problem that may appear is the cross-channel interference from different targets in adjacent directions (bearings) of the hydrophone array.

Signal processing techniques such as DEMON (Demodulation of Envelope Modulation on Noise) and LOFAR (Low Frequency Analysis and Recording) [79] are usually applied to enhance the signal of interest and reduce, as much as possible, the background noise. DEMON is a narrowband analysis which aims at identifying the number of shafts and shafts rotation frequency of the target propeller. Figure 5 shows a block diagram of the classical DEMON analysis.

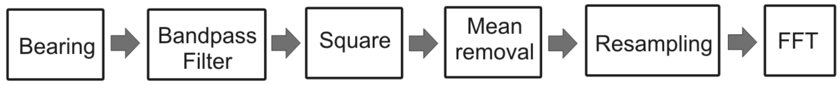
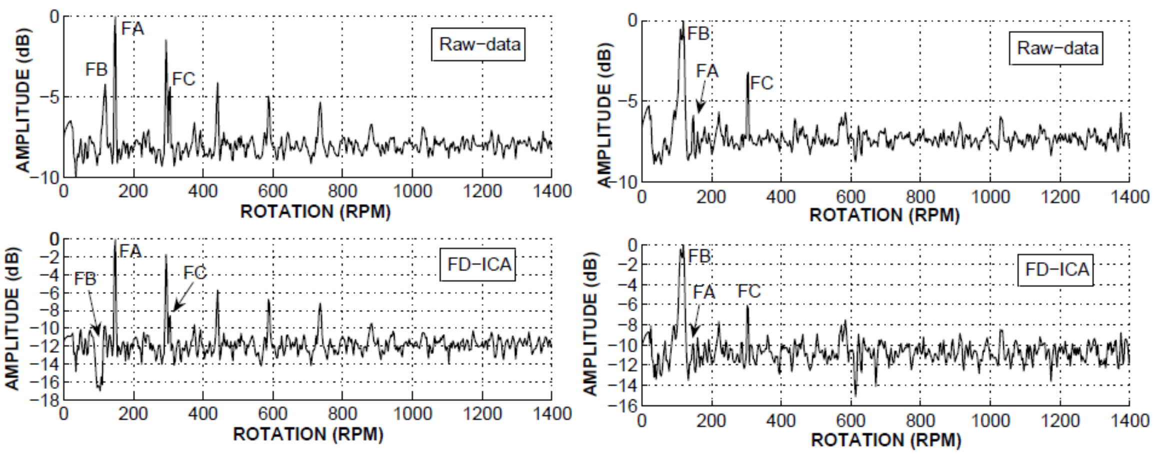


Figure 5 - Diagram of classical DEMON analysis.

In a given direction (bearing) the acoustic signal is bandpass filtered (selecting only the frequency range of interest to characterize the target propeller cavitation noise, which goes from hundreds to thousands of RPM). In the following, signal is squared (as in traditional AM demodulation) and TPSW [79] algorithm is applied to reduce the background noise. Finally, short-time FFT (Fast Fourier Transform) reveals frequency-domain information.

The work [81] illustrates how ICA can be applied to reduce both the cross-channel interference and the background noise in passive sonar signal processing. The DEMON plots for two adjacent bearings are illustrated in the top of Figures 6-a and 6-b, respectively for direction 190^o and 205^o. The main frequency components of 190° signal (FA=148 RPM and its multiples) are mixed together with information from the 205° direction (FB=119 RPM). The same problem is observed in the signal measured at bearing 205°. It was also observed that both signals (190° and 205°) are contaminated by FC=305 RPM, which is probably related to the submarine self-noise

ICA (through JADE algorithm) was applied in the frequency-domain to estimate the independent components which generated the DEMON signals at directions 190^o and 205^o. The independent DEMON plots are illustrated in the bottom of Figures 6-a and 6-b (respectively for directions 190^o and 205^o). It can be observed that cross-channel interference and the background noise level were reduced in approximately 4 dB and 5 dB, respectively.



(a) (b)

Figure 6 - DEMON plots for directions (a) 190^o and (b) 205^o. Top: measured data; Bottom: after ICA, extracted from [77].

5.2- Feature Extraction in Ultrasound Inspection of Pipeline Welded Joints

Non-destructive testing is used as an important tool to ensure industrial equipment reliability. Among the main available methods, ultrasound inspection is usually applied to identify weld defects. Pulse-echo ultrasound testing consists on using a transducer, coupled to the object to be analyzed, to emit an acoustic signal (in the ultrasound frequency range). The ultrasound signal propagates inside the object and reflects (producing echoes) every time it reaches interfaces, discontinuities and every variation present inside the material. These echoes are recorded as they reach back the transducers. A main limitation of this method is that the diagnosis is usually performed by experienced operators, by looking at the measured signals. This procedure is time consuming and susceptible to human failure.

An automatic neural diagnostic system for natural gas pipeline welded joints was proposed in [82]. The available signatures were analyzed by experience operators and labeled as defect (37 signals) and non-defect joints (58 signals). The ultrasound signatures were used (after a smoothing preprocessing step which includes low-pass filtering and down-sampling) to feed a supervised neural classifier (multi-layer perceptron architecture) [28]. In [83] it was shown that depending on underlying data characteristics, the application of ICA as a preprocessing step for classification problems may improve the discrimination efficiency. Considering this, as illustrated in Figure 7, ICA (through FastICA algorithm) was applied here to the ultrasound signatures aiming at extracting discriminant features of the data. As illustrated in Table 1, the neural network trained using the independent features produced higher discrimination efficiency (~+5%) if compared to a similar network trained directly with the smoothed ultrasound signatures. In this particular problem, the non-defect signatures were used as targets to be detected and false alarm means classifying a defect joint as non-defect one. Considering this, false alarm is not acceptable as it means that, in practice, a bad joint would be classified as a good one. So, the classifiers decision threshold were adjusted to produce the highest probability of detection (P_D) possible while maintaining zero probability of false alarm (P_F).

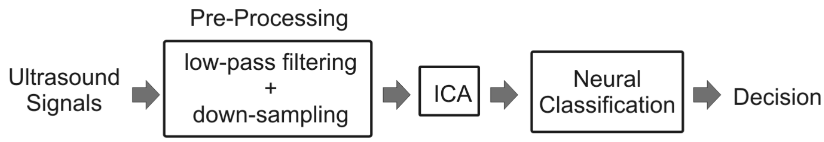


Figure 7 – Diagram of the signal processing chain.

Table 1 – Performance comparison between different discriminators.

	Neural Network	Neural Network + ICA
Highest P_D for P_F=0%	89.47 %	94.74 %

5.3- Information Retrieval in Time Series

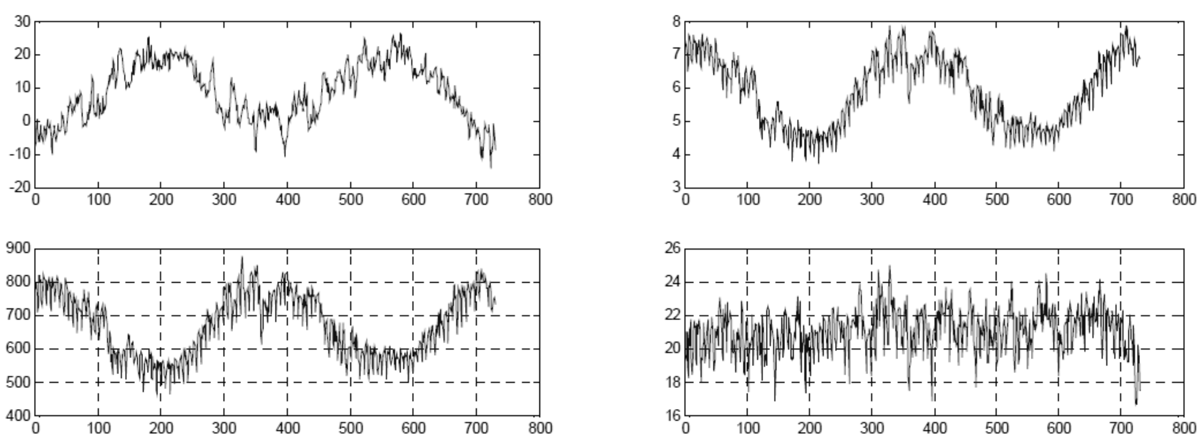
Parallel time series may present underlying common factors that can better describe the process represented by these series. In the work [84] ICA was applied to electrical load time series aiming at data quality monitoring. Electrical load and daily temperature data from an European energy provider (East-Slovakia Power Distribution Company) [85] were used. Electrical load data (in MW) was recorded in a 30 minutes period from January 1st 1997 to January 31st 1999 and the mean daily temperature information covers the same period.

The system proposed in [84] comprises a cascaded signal processing chain composed by ICA, a pre-processing block (which analyzes tendencies, periodicity and stationarity) and a neural network (to model the series and predict the future samples). The daily temperature and peak-load time series are illustrated in Figure 8-a (respectively at top and bottom). After applying ICA, the time-series illustrated in Figure 8-b were obtained. It was observed that the structure of the electrical load series was concentrated in a small number of components (most independent components are non-structured and model only noise), allowing a more accurate modeling by the neural networks.

The performance obtained through the data quality monitoring system with and without applying ICA (SOBI algorithm) was compared by computing the mean absolute percentage error (MAPE), which is defined as:

$$\mathbf{MAPE}\,(\%) = \frac{\mathbf{1}}{\mathbf{T}}\sum_{i=1}^{T}\left|\frac{\hat{x}_i - x_i}{x_i}\right| \times 100 \ , \tag{59}$$

where x_i , $\hat{x}_i$ and T are respectively the time signal, its prediction and the applied time-window. It was observed that the MAPE computed in the prediction of the daily peak-load time series was reduced from 7.03% (without ICA) to 2.46% (with ICA).



(a) (b)

Figure 8 – (a) Measured time-series and (b) independent components, daily temperatures (top) and peak-load (bottom), extracted from [80].

5.4- Multi-channel Blind Separation of Musical Signals

The performance evaluation of CBSS algorithms usually uses SIR (signal-interference ratio), SAR (signal-artifact ratio) and SDR (signal-distortion ratio) computations (the last two ones are more important when frequency-domain algorithms are used). For more details, see [86]. Using 24000 samples of two musical signals (sampled at 8 kHz) with mixture filters lengths equal to 8 and identical separation filters lenghts, 500 off-line iterations and $\beta(i,m)=10^{-1}/500$ (see Equation 49). Figure 9 ilustrates the SIR evolution obtained for a time-domain CBSS algorithm. An average SIR above 20 dB was obtained, a good result for convolutive separation.

In a different experiment, using two voice signals (a female and a male speaker) with sampling frequency of 8 kHz and (random) mixtures filters of length 32, we applied the Frequency-Domain algorithm of section 4.2.2 (with K = 128). The signals time length is 10 s. The performance evaluations parameters are presented in Table 1. The low values of SAR and SDR are typical of frequency-domain algorithms. Fig. 10 shows spectrograms of original sources, mixtures and estimates for 4 s of signals. Note that low frequency regions (where most of the energy is concentrated) reveals good separation of both sources. The high frequency regions were emphasized, because the MD Principle does not resolve completely the scaling ambiguity.

Table 2 – Objective evaluations (in dB) of frequency-domain CBSS algorithm.

Measurement	Estimate 1	Estimate 2
SIR	10,33	13.22
SAR	3,90	5.49
SDR	2,70	4.64

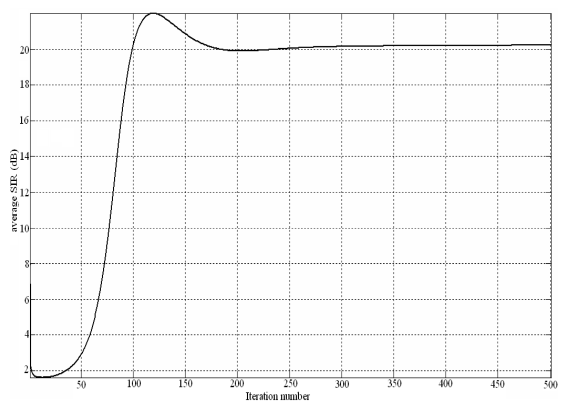


Figure 9 – Average SIR evolution (in dB) of time-domain CBSS algorithm.

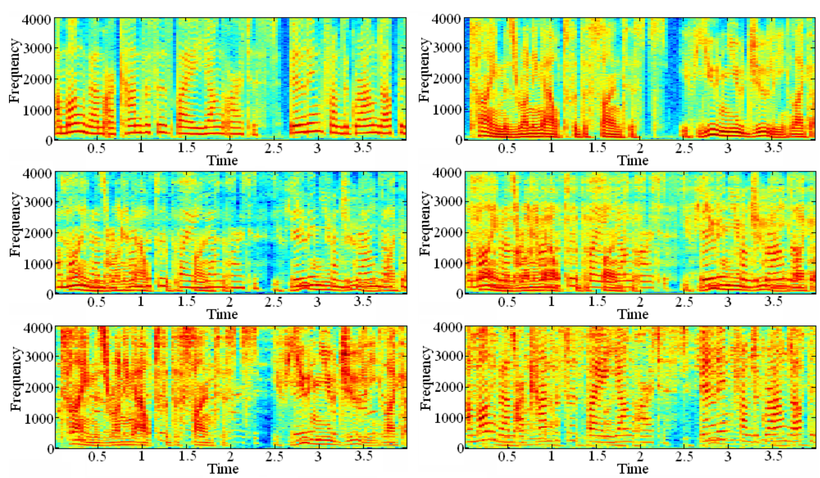


Figure 10 – Spectrograms of original sources (first line), mixtures (second line) and estimates (third line) of frequency-domain CBSS algorithm.

6- Conclusions

This paper describes the independent component analysis in its theoretical aspects, algorithms and applications. The statistical principles used to guide the search for independent components, such as non-gaussianity maximization and nonlinear decorrelation, are derived. The most commonly applied pre-processing steps (PCA and noise reduction) used to obtain more accurate estimates of the independent components are illustrated. Among the numerous algorithms proposed in the literature for solving the ICA problem, some popular ones (like FastICA and JADE) were derived in this paper. Extensions to the standard (linear and instantaneous) ICA model, such as the convolutive mixtures and nonlinear ICA paradigms are also discussed. The ICA model proved to be very useful in a great number of signal processing applications like blind signal separation, feature extraction, interference removal and information retrieval. Experimental results illustrate some of the benefits of applying ICA in challenging real-world applications such as passive sonar, ultrasound inspection, time-series data quality monitoring, blind separation of musical signals and high-energy physics.

Acknowledgements

The authors are thankful for the support provided by CNPq and FAPERJ (Brazil), and for the Brazilian Navy Research Institute (IPqM) for providing the data set used in this work.

References

[1] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Wiley, New York, 2001.
[2] I. T. Jolliffe, *Principal Component Analysis*. Springer, 2a ed., New York, 2002.
[3] H. H. Harman, *Modern Factor Analysis*. University of Chicago Press, 2nd edition, 1967.
[4] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*. Willey, New York, 2002.
[5] A. Hyvarinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Networks*, vol. 13, n. 4-5, pp. 411-430, 2000.
[6] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," *IEE Proceedings- F,* vol. 140, n. 6, pp. 362-370, November 1993.
[7] A. Cichocki and R. Unbehauen, "Robust neural networks with on-line learning for blind identification and blind separation of sources," *IEEE Transactions on Circuits and Systems-I*, vol. 43, n. 11, pp. 894-906, 1996.
[8] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626-634, 1999.
[9] A. Cichocki, S. Douglas, and S. Amari, "Robust techniques for independent component analysis with noisy data," *Neurocomputing*, vol. 22, n. 1-3, pp. 113-129, 1998.
[10] A. Budillon, F. Palmieri, and R. Varriale, "A hybrid method for blind signal de-noising via independent component analysis," *Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation* - ICA2000, pp. 145-150, Helsinki, Finland, 2000.
[11] H.-M. Park, S.-H. Oh, and S.-Y. Lee, "Adaptive noise cancelling based on independent component analysis," *Electronics Letters*, vol. 38, n 15, pp. 832-833, 2002.
[12] N. N. Moura, J. M. Seixas, W. S. Filho, and A. V. Greco, "Independent component analysis for optimal passive sonar signal detection," *Proceedings of the 7th International Conference on Intelligent Systems Design and Applications*, Rio de Janeiro, pp. 671-678, October 2007.
[13] L. Sarperi, X. Zhu, and A. K. Nandi, "Blind OFDM receiver based on independent component analysis for multiple-input multiple-output systems," *IEEE Transactions on Wireless Communications*, vol. 6, no. 11, pp. 4079-4089, 2007.
[14] J. Escudero, R. Hornero, D. Abasolo, A. Fernandez, and M. Lopez-Coronado, "Artifact removal in magneto-encephalogram background activity with independent component analysis," *IEEE Transactions on Biomedical Engineering*, vol. 54, n. 11, pp. 1965-1973, November 2007.
[15] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *IEEE Transactions on Neural Networks*, vol. 13, no. 6, pp. 1450-1464, 2002.
[16] K.-C. Kwak and W. Pedrycz, "Face recognition using an enhanced independent component analysis approach," *IEEE Transactions on Neural Networks*, vol. 18, n. 2, pp. 530-541, March 2007.
[17] E. F. Simas Filho, J. M. Seixas and L. P. Caloba. "Optimized Calorimeter Signal Compaction for an Independent Component based ATLAS Electron/Jet Second-Level Trigger", *Proceedings of Science*, vol. ACAT08, article 102, pp. 1-10, 2009.
[18] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York, 1991.
[19] M. Kendall and A. Stuart, *The Advanced Theory of Statistics*. 4th ed., Charles Griffing Company, London, 1977.
[20] M. R. Spiegel, J. J. Schiller, and R. A. Srinivasan, *Probability and Statistics*. 2 ed., McGraw-Hill, New York, 2000.
[21] T.-H. Kim and H. White, "On more robust estimation of skewness and kurtosis," *Finance Research Letters*, vol. 1, n. 1, pp. 56-73, 2004.
[22] M. Welling, "Robust higher-order statistics," *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics*, pp. 405-412, Barbados, January 2005.
[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, New York, 1991.
[24] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, n. 6, pp. 379-423, July 1948.
[25] A. Hyvarinen, "New approximations of diferencial entropy for independent component analysis and projection pursuit," *Advances in Neural Information Signal Processing*, no. 10, n. 1, pp. 273-279, 1998.

[26] M. Jones and R. Sibson. "What is projection pursuit". *Journal of the Royal Statistical Society, Sec.A*, vol. 150, n. 1, pp. 1-36., 1987.
[27] A. Hyvärinen. "New approximations of diferencial entropy for independent component analysis and projection pursuit". *Advances in Neural Information Processing Systems*, vol. 10, n. 1, pp.273-279. MIT Press, 1998.
[28] S. Haykin, *Neural Networks and Learning Machines*. Prentice Hall, New Jersey, 2008.
[29] J. B. O. Souza Filho,. L. P. Calôba, J. M. Seixas. "An accurate and fast neural method for PCA extraction". *International Joint Conference on Neural Networks*, 2003, Portland, Oregon, EUA. p. 797-802.
[30] L. P. Caloba, J. M. Seixas and F. S. Pereira. "Neural Discriminating Analysis for a Second-Level Trigger System". *Int. Conf. on Computing in High Energy Physics*, pp. 1-5, Rio de Janeiro, September, 1995.
[31] E. F. Simas Filho, J. M. de Seixas and L. P. Calôba. "Modified post-nonlinear ICA model for online neural discrimination". *Neurocomputing*, Vol. 73, n. 16-18, pp. 2820–2828 , 2010.
[32] E. F. Simas Filho, J. M. de Seixas and L. P. Calôba. "Optimized Calorimeter Signal Compaction for an Independent Component based ATLAS Electron/Jet Second-Level Trigger". *Proceedings of Science*, v. ACAT08, p. 1-10, 2009.
[33] A. Paraschiv-Ionescu, C. Jutten, K. Aminian, B. Najafi and Ph. Robert, "Source Separation in Strong Noisy Mixtures: a Study of Wavelet De-noising Pre-processing". *Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing*, vol. 2, pp. 1681-1684, Orlando, 2002.
[34] C. Sánchez et al. "Wavelet Denoising as Preporcessing Stage to Improve ICA Performance in Atrial Fibrillation Analysis". *Proceedings of the 6th International Conference on Independent Component Analysis and Blind Signal Separation*, pp. 486-494, 2006.
[35] S. Mallat. "*A Wavelet Tour of Signal Processing*" 3rd Ed., Academic Press, Burlington, US, 2008.
[36] T. Akuzawa. "Multiplicative Newton-like algorithm and independent component analysis". *Proceedings of the IEEE International Joint Conference on Neural Networks*, vol. 4, pp. 79 – 82, Como, Italy, 2000.
[37] A. Belouchrani, K. Abedi-Meraim, J. Cardoso, E. Moulines. "A Blind Source Separation Technique Using Second Order Statistics". *IEEE Transactions on Signal Processing*, vol. 45, n. 2, pp. 434-444, 1997.
[38] M. A. Kramer. "Nonlinear Principal Component Analysis Using Autoassociative Neural Networks". *AIChE Journal*, vol. 37, No. 2, pp. 233-243, February 1991.
[39] E. Oja. "The Nonlinear PCA Learning Rule in Independent Component Analysis". *Neurocomputing*, vol. 17, n. 1, pp. 25-45, 1997.
[40] J. Karhunen and J. Joutsensalo. "Representation and separation of signals using nonlinear PCA type learning. *Neural Networks*, vol. 7, n. 1, pp. 113-127, 1994.
[41] J. Karhunen, L. Wang, and R. Vigario. "Nonlinear PCA Type Approaches for Source Separation and Independent Component Analysis". *Proceedings of the International Conference on Neural Networks*, pp. 995-1000, Perth, 1995.
[42] P. Pajunen and J. Karhunen. "Least-squares methods for blind source separation based on nonlinear PCA". *International Journal of Neural Systems*. Vol. 8, n. 5-6, pp. 601-612, 1998.
[43] Karhunen, J., Pajunen, P., and Oja, E., The Nonlinear PCA Criterion in Blind Source Separation: Relations with Other Approaches. *Neurocomputing*, vol. 22, November 1998, pp. 5-20.
[44] B. Yang. "Projection approximation subspace tracking". *IEEE Transactions on Signal Processing*, vol. 46, n. 1, pp. 95-107, 1995.
[45] R. Fletcher. *Practical Methods of Optimization*, 2nd Ed. Wiley, New York, 2000.
[46] A. D. Michal, *Matrix and Tensor Calculus*, 1st Ed., Dover, West Sussex, UK, 2008.
[47] J.-F. Cardoso. "Source separation using higher order moments". *Proceedings of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'89)*, pages 2109–2112, Glasgow, UK, 1989.
[48] T. Akuzawa and N. Murata: "Multiplicative Nonholonomic Newton-like Algorithm". *Chaos, Solitons and Fractals*, Volume 12, Number 4, pp. 785-79, January, 2001.
[49] T. Akuzawa. "Extended Quasi-Newton Method for the ICA" *Proc. Int. Workshop Independent Component Anal. Blind Signal Separation* pp.521-525, Helsinki, 2000.
[50] W. M. Cuenca et al. "Independent Component Analysis for Blind Estimation of Gausian Noise in High Voltage Equipments Parcial Discharges", *Proceedings of the Brazilian Neural Networks Conference*, pp. 133-138, São Paulo, Brazil, 2003 (in Portuguese).
[51] P. Pajunen. "Blind Source Separation Using Algorithmic Information Theory", *Neurocomputing*, vol. 22, n. 1, pp. 35-48,1998.
[52] P. Pajunen. *Extensions of Linear Independent Component Analysis: Neural and Information-Theoretic Methods*. PhD Thesis, Helsink Univertity of Technology, 1998.
[53] L. Tong, R.-W. Liu, V.C. Soon, and Y.-F. Huang. "Indeterminacy and identifiability of blind identification". *IEEE Trans. on Circuits and Systems*, vol. 38, n. 5, pp. 499–509, 1991.
[54] A. Belouchrani, K. Abedi-Meraim, J. F. Cardoso, E. Moulines. "A Blind Source Separation Technique Using Second Order Statistics". *IEEE Transactions on Signal Processing*, Vol. 45, No. 2, pp. 434-444, 1997.
[55] H. Buchner, R. Aichner, and W. Kellerman, "A Generalization of Blind Source Separation Algorithms for Convolutive Mixtures Based on Second-Orders Statistics," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 120-134, 2005.
[56] Y. Xiang, S. Nahavandi, H. Trinh H. Zheng. "A new second-order method for blind signal separation from dynamic mixtures", *Computers and Electrical Engineering*, vol. 30, No 5, pp. 347-359, July, 2004.
[57] S. C. Douglas, H. Sawada, and S. Makino, "Natural Gradient Multichannel Blind Deconvolution and Speech Separation Using Causal FIR Filters," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 92-104, 2005.
[58] D. W. E. Schobben and P. C. W. Sommen, "On the Indeterminacies of convolutive blind signal separation based on second-order statistics", *Proceedings of the International Symposium of Signal Processing and its Applications*, pp. 215-218, 1999.
[59] J. Thomas, Y. Deville, and S. Hosseini, "Time-domain Fast Fixed-Point Algorithms for Convolutive ICA," *IEEE Signal Processing Letters,* vol. 13, no. 4, pp. 228-231, 2006.
[60] R. Aichner, H. Buchner, S. Araki, and S. Makino, "On-line Time-Domain Blind Source Separation of Nonstationary Convolved Signals," *Proceedings of the International Conference on Independent Component Analysis and Blind Signal Separation*, pp. 987-992, Nara, Japan, 2003.
[61] M. Joho, "Blind Signal Separation of Convolutive Mixtures: A Time-Domain Joint-diagonalization Approach," *Proceedings of the International Conference on Independent Component Analysis and Blind Signal Separation*, pp. 578-585, Granada, Spain, 2004.
[62] J. Anemüller and B. Kollmeier, "Amplitude Modulation Decorrelation for Convolutive Blind Source Separation," *Proceedings of the International Conference on Independent Component Analysis and Blind Signal Separation,* pp. 215-220, Helsinki, Finland, 2000.

[63] N. Murata, S. Ikeda e A. Ziehe, "An Approach to Blind Source Separation Based on Temporal Structure of Speech Signals," *Neurocomputing*, vol. 41, n. 1-4, pp. 1-24, 2001.

[64] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Evaluation of Blind Signal Separation Method using Directivity Pattern Under Reverberant Conditions," *Proceedings of the International Conference on Acoustic, Speech and Signal Processing,* pp. 3140-3143, Orlando, 2002.

[65] M. Z. Ikram and D. R. Morgan, "A Beamforming Approach to Permutation Alignment for Multichannel Frequency-Domain Blind Speech Separation," *Proceedings of the International Conference on Acoustic, Speech and Signal Processing*, pp. 881-884, Orlando, 2002.

[66] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A Robust and Precise Method for Solving the Permutation Problem of Frequency-domain Blind Source Separation," *IEEE Transactions on Speech and Audio Processing*, vol. 12, n. 5, pp. 530-538, 2004.

[67] T. Kim, H. T. Attias, S.-Y. Lee, and Te-Won Lee, "Blind Source Separation Exploiting Higher-Order Frequency Dependencies," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 1, pp. 70-79, 2007.

[68] K. Matsuoka, "Minimal Distortion Principle for Blind Source Separation," *Proceedings of SICE Annual Conference*, pp. 2138-2143, 2002.

[69] K. Matsuoka, M. Ohya, and M. Kawamoto, "A Neural Net for Blind Separation of Nonstationary Signals", *Neural Networks*, vol. 8, no. 3, pp. 411-419, 1995.

[70] H. Buchner, R. Aichner, and W. Kellerman, "Blind Source Separation for Convolutive Mixtures Exploiting Nongaussianity, Nonwhiteness, and Nonstationarity,". *Proc. Int. Workshop Acoustic Echo Noise Control (IWAENC)*, pp. 275-278, 2005.

[71] H. Buchner, R. Aichner, and W. Kellerman, "TRINICON-based Blind System Identification with application to Multiple-Source Localization and Separation", in *Blind Speech Separation* (S. Makino, T.-W. Lee and H. Sawada, editors), Springer, Berlin, 2007.

[72] M. R. Petraglia, P. B. Batalheiro, and D. B. Haddad, "Blind Source Separation for Convolutive Mixtures Using a Non-Uniform Oversampled Filter Bank", *Proceedings of the European Signal Processing Conference*, p. 1-5, 2008.

[73] C. Jutten, M. B. Zadeh, and S. Hosseini, "Three easy ways for separating nonlinear mixtures,"*Signal Processing*, vol. 88, n.2, pp. 217-229, 2004.

[74] C. Jutten and J. Karhunen, "Advances in nonlinear blind source separation," *Proceedings of the 4th Int. Symp. on Independent Component Analysis and Blind Signal Separation*, pp. 245-256, 2003.

[75] A. Hyvärinen and P. Pajunen, "Nonlinear independent component analysis: Existence and uniqueness results," *Neural Networks*, vol. 12, no. 3, pp. 429-439, 1999.

[76] F. Rojas, C. G. Puntonet, and I. Rojas, "Independent component analysis evolution based method for nonlinear speech processing," *Artificial Neural Nets Problem Solving Methods, PT II*, vol. 2687, pp. 679-686, 2003.

[77] M. Haritopoulos, H. Yin, and N. M. Allinson, "Image denoising using self-organizing map-based nonlinear independent component analysis," *Neural Networks*, vol. 15, n. 8-9, pp. 1085-1098, 2002.

[78] E. F. Simas Filho and J. M. de Seixas. "Nonlinear Independent Component Analysis: Theoretical Review and Applications". *Learning and Nonlinear Models*, vol. 5, n. 2, pp. 99-120, 2007.

[79] R. O. Nielsen. *Sonar Signal Processing*, Artech House Inc, Nortwood, MA, 1991.

[80] A.D. Waite. *Sonar for practicing Engineers*, Wiley, New York, 2003.

[81] N. N. Moura, E. F. Simas Filho and J. M. de Seixas. "Independent Component Analysis for Passive Sonar Signal Processing". In: *Advances in Sonar Signal Processing*, Ed. Sergio Rui Silva, pp. 91-110, I-Tech, Vienna, Austria, 2009.

[82] D. Ferreira, M. C. C. Albuquerque, E. F. Simas Filho. "Automatic Identification of Defects in Pipeline Welded Joints through Neural Networks". *Proceedings of the National Conference on Control and Automation*, pp. 1-5 Salvador, Brazil, October, 2009 (in Portuguese).

[83] V. Sanchez-Poblador, E. Monte-Moreno, and J. Solé-Casals. "ICA as a Preprocessing Technique for Classification" *C.G. Puntonet and A. Prieto (Eds.): ICA 2004, LNCS 3195*, pp. 1165–1172, 2004.

[84] J. M. Faier and J. M. de Seixas. "Data Quality Monitoring of Electrical Load Time-Series using Neural Networks and Independent Component Analysis". Proceedings *of the Brazilian Conference on Neural Networks / Computational Intelligence*, Ouro Preto, Brazil, October, 2009 (in Portuguese).

[85] European Network on Intelligent Technologies for Smart Adaptive Systems, dataset available at <http://neuron.tuke.sk/competition>, accessed on September, 2009.

[86] E. Vincent, R. Gribonval, and C. Févotte, "Performance Measurement in Blind Audio Source Separation," *IEEE Transactions on Audio, Speech, and Language Processing,* vol. 14, no. 4, 2006.