Learning and Nonlinear Models - Revista da Sociedade Brasileira de Redes Neurais (SBRN), Vol. 5, No. 2, pp. 99-120, 2007

NONLINEAR INDEPENDENT COMPONENT ANALYSIS: THEORETICAL REVIEW AND APPLICATIONS

Eduardo F. Simas Filho[1,2], José Manoel de Seixas[1]

[1] Signal Processing Laboratory, COPPE/Poli/UFRJ, Rio de Janeiro, Brazil
[2] Federal Center for Technological Education, Simões Filho, Bahia, Brazil.
E-mails: esimas@lps.ufrj.br, seixas@lps.ufrj.br

Abstract – This paper reviews the Nonlinear Independent Components Analysis and its applications to blind source separation. An overview of the main statistical principles that guide the search for the independent components is formulated. The uniqueness of solution and some algorithms for estimating the nonlinear independent components are discussed. Experimental results using a synthetic database are used for performance comparison. A practical application in experimental high-energy physics is also presented.

Index Terms – Nonlinear ICA, Neural Networks, Blind Source Separation, Nonlinear Mixtures, Signal Detection.

Introduction

The basic linear Independent Component Analysis (ICA) model considers that the set of N measured signals $\mathbf{x} = [x_1, ..., x_N]^T$ is generated by a linear combination of unknown sources $\mathbf{s} = [s_1, ..., s_N]^T$:

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

where **A** is the N×N mixing matrix [1] (the N-dimensional vectors **s** and **x** denote, respectively, single observations of the source and measured signals, this notation will be used in the rest of the paper). Formulated this way, ICA is also referred to as Blind Source Separation (BSS) method [2] and its purpose is to estimate the original source signals ***s*** using only the observed (mixed) data ***x***. A solution can be obtained if one can find the inverse of the mixing matrix $\mathbf{B} = \mathbf{A}^{-1}$:

$$\mathbf{s} = \mathbf{B}\mathbf{x} \tag{2}$$

A general principle for estimating the matrix **B** can be found by considering that the original source signals are statistically independent. There are many mathematical methods for calculating the coefficients of matrix **B**. The nonlinear decorrelation and the maximally nongaussianity are the most applied ones [3]. There are some indeterminacies in the linear ICA model, the order of extraction of the independent components can change and scalar multipliers (positive or negative) may be modifying the estimated components. Fortunately, these limitations are insignificant in most applications [1].

Several good performance algorithms have been proposed for solving the linear ICA/BSS problem such as JADE (where search for independence is performed by cumulant matrix diagonalization) [4], Nonlinear Decorrelation (explores higher-order statistics, and thus independence, through nonlinear transformations) [5] and FastICA (fixed point algorithm that uses statistical and information theoretic measures in the search for independent sources) [6]. In [7], robust solutions for the noisy ICA problem ($\mathbf{x}=\mathbf{A}\mathbf{s}+\mathbf{n}$, where **n** is a random noise vector) were proposed. The linear ICA model has been applied successfully in a large number of signal processing tasks like noise removal [8, 9], passive sonar signal separation [10], telecommunications [11], feature extraction in biomedicine [12], face recognition [13, 14] and experimental particle physics [15]. Unfortunately, in problems where there is some sort of nonlinear phenomenon during the signal mixing process, linear ICA model may present poor results [16].

In a more general formulation, the nonlinear independent component analysis (NLICA) model considers that the measured signals **x** are formed by a nonlinear instantaneous mixing model:

$$\mathbf{x} = F(\mathbf{s}) \tag{3}$$

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 $G : R^N \rightarrow R^N$:

$$\mathbf{y} = G(\mathbf{x}) \tag{4}$$

so that the components of **y** are statistically independent. If $G = F^{-1}$ the sources are perfectly recovered (and so $\mathbf{s} = \mathbf{y}$) [17].

A characteristic of the NLICA problem is that the solutions are nonunique [17]. If **x** and **y** are independent random variables, it is easy to prove that *f(***x***)* and *g(***y***)*, 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 (NLBSS) is a more restrictive problem as its purpose is to estimate the original source signals from their nonlinear mixed version. Nonlinear BSS cannot be achieved without some prior information on the mixing model or sources. A complete investigation on the uniqueness of nonlinear ICA solutions can be found in [18]. NLICA algorithms have recently been applied in different problems such as speech processing [19] and image denoising [20].

Generally, the number of parameters to be estimated in a nonlinear ICA model increases when compared to the linear case. In view of this, the NLICA algorithms present higher computational complexity and consequently slower convergence. In the BSS problem, the accuracy of the estimated nonlinearly-mixed sources depends on the assumed mixing model, and thus the algorithm is usually chosen based on prior information concerning the mixing process. Considering this, the application of a nonlinear ICA algorithm is only justified if the estimation accuracy increases (compared to a linear algorithm) and there are no severe restrictions for the processing time (the developing phase may be executed offline).

Among the NLICA algorithms proposed in the literature, we can mention a class of methods that impose structural constraints to the mixing model, guaranteeing that the estimated nonlinear independent components are equal to the sources (unless by scaling and ordering indeterminacies). Another popular approach of direct implementation is to estimate the nonlinear mixing mapping through self-organizing neural networks. There is also a method closely related to the NLICA problem, which introduces nonlinear transformations by clustering the dataset into groups of similar characteristics, after that, linear ICA is applied to data belonging to each cluster producing independent components. More details concerning these algorithms will be provided in the following Sections.

This paper is divided as it follows. Section 1 provides a detailed description of the statistical independence measures often used for nonlinear ICA. In Section 2, it is presented a review on the uniqueness of the NLICA solution. Successful separation algorithms for nonlinear BSS and ICA are derived in Section 3. Experimental results obtained from synthesized and practical application signals are illustrated in Section 4. Conclusions are derived in Section 5.

1- Statistical Independence

Considering two random vectors **x** and **y**, they are statistically independent if and only if [21]:

$$p_{x,y}(\mathbf{x},\mathbf{y}) = p_x(\mathbf{x})p_y(\mathbf{y}) \tag{5}$$

where $p_{x,y}(\mathbf{x}, \mathbf{y})$, $p_x(\mathbf{x})$ and $p_y(\mathbf{y})$ are, respectively, the joint and marginal probability density functions (pdf) of **x** and **y**. Equivalent condition is obtained if for all absolutely integrable functions $g(\mathbf{x})$ and $h(\mathbf{y})$ the expression on Equation 6 holds:

$$E\{g(\mathbf{x})h(\mathbf{y})\} = E\{g(\mathbf{x})\}E\{h(\mathbf{y})\} \tag{6}$$

where *E{.}* is the expectation operator [1].

In typical blind signal processing problems, there is very little information on the source signals and so the pdfs estimation is a very difficult task, which can be avoided using Equation 6. Another principle that can be used to estimate the dependence of variables comes from the central limit theorem [22]: "The sum of two (independent) random variables is always closest to a Gaussian distribution than the original variable distributions". Thus, the independent components can be obtained through maximization of non-gaussianity [1].

Through one of the principles described above, in ICA/BSS algorithms, some mathematical properties are applied to estimate the independence of components, the most frequently used ones are detailed in the following subsections.

1.1- Nonlinear Decorrelation

Independent variables are uncorrelated, however, the reciprocal is not always true. Linear correlation is verified by second order statistics, while independence needs higher order information. In the nonlinear decorrelation methods, nonlinear functions introduce high-order statistics, making it possible the search for independent components.

As mentioned in Equation 6, two random variables are statistical independent if they are nonlinearly uncorrelated. As it is not possible to check all integrable functions *g(.)* and *h(.)*, estimates of the independent components are obtained while guaranteeing nonlinear decorrelation between a finite set of nonlinear functions [1].

For example, a well known linear 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.

Learning and Nonlinear Models - Revista da Sociedade Brasileira de Redes Neurais (SBRN), Vol. 5, No. 2, pp. 99-120, 2007

1.2- High-order Cumulants

The gaussianity (and consequently the statistical dependence) of a random variable can be measured by higher order cumulants. Considering a random vector **x**, the moment $\boldsymbol{\alpha}_k$ and central moment $\boldsymbol{\mu}_k$ of order k are defined by [23]:

$$\boldsymbol{\alpha}_k = E\{\mathbf{x}^k\} = \int_{-\infty}^{\infty} \mathbf{x}^k p_x(\mathbf{x}) d\mathbf{x} \qquad (7)$$

$$\boldsymbol{\mu}_k = E\{(\mathbf{x}-\boldsymbol{\alpha}_1)^k\} = \int_{-\infty}^{\infty} (\mathbf{x}-\boldsymbol{\alpha}_1)^k p_x(\mathbf{x}) d\mathbf{x} \qquad (8)$$

where $\boldsymbol{\alpha}_1 = \mathbf{m}_x$ is the mean vector of **x**. If **x** is zero mean ($\|\mathbf{m}_x\|=0$), than for all k holds: $\boldsymbol{\alpha}_k = \boldsymbol{\mu}_k$.

The cumulant κ_k of order k is defined as a function of the moments [23]. For a zero mean random variable x, the first four cumulants are:

$$\kappa_1 = 0;\ \ \kappa_2 = E\{x^2\} = \alpha_2;\ \ \kappa_3 = E\{x^3\} = \alpha_3; \qquad (9)$$

$$\kappa_4 = E\{x^4\} - 3[E\{x^2\}]^2 = \alpha_4 - 3\alpha_2^{2}$$

The third and fourth order cumulants are called respectively skewness (κ_3) and kurtosis (κ_4) [24]. Cumulants of order higher than four are rarely applied in practical ICA/BSS problems. Some interesting properties of cumulants are:

$$\begin{aligned} \kappa_k(x+y) &= \kappa_k(x) + \kappa_k(y) \\ \kappa_k(x) &= 0, \text{ for } k > 2 \text{ if } x \text{ is Gaussian} \end{aligned} \qquad (10)$$

Considering this, cumulants of order higher than two shall be applied to estimate data gaussianity. 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) [23]. Kurtosis can be easily computed from data substituting expectations in Equation 9 by sample means. One disadvantage is that κ_4 can be seriously influenced by outliers (observations that are numerically distant from the rest of the data), in extreme situations the kurtosis value may be dominated by a small number of points [24]. Some studies are being conduced with the purpose of obtaining robust estimation of high-order cumulants, specially the kurtosis [25].

1.3- Information Theoretic Contrasts

Alternative gaussianity measures can be obtained from information theory [26]. These parameters are usually more robust to outliers than cumulant based ones [2].

Negentropy of a random variable x is calculated through [26]:

$$J(x) = H(x_{gauss}) - H(x) \qquad (11)$$

where $H(.)$ is the entropy, and x_{gauss} is a Gaussian random variable with the same mean and variance of x. Entropy is one of the basic concepts of information theory and can be interpreted as the level of information contained in a random variable. Entropy $H(x)$ can also be viewed as the minimum code length needed to represent the variable x. Considering a discrete random variable x, **entropy** is defined as [27]:

$$H(x) = \sum_i P(x=a_i)\log P(x=a_i) \qquad (12)$$

where a_i are the possible values assumed by the variable x, and $P(x = a_i)$ is the probability that $x = a_i$.

An important result is that the Gaussian variable has maximum entropy between the variables of same variance [1]. So both entropy and negentropy can be used as gaussianity measures. The advantage of $J(x)$, compared to $H(x)$, is that it is always non-negative and zero when x is Gaussian. A problem with the computation of both $J(.)$ and $H(.)$ in blind signal processing is the pdf estimation (see Eq. 12). To avoid this, approximations using high order cumulants or non-polinomial functions shall be applied [1, 28].

Another statistical independence measure can be obtained through mutual information. The **Mutual Information** $I(x_1, x_2, ..., x_m)$ between m random variables $x_1, x_2, ..., x_m$ is obtained through equation 13 (where $\mathbf{x}=[x_1, x_2, ..., x_m]^T$) [3]:

Learning and Nonlinear Models - Revista da Sociedade Brasileira de Redes Neurais (SBRN), Vol. 5, No. 2, pp. 99-120, 2007

$$I(x_1,...,x_m) = \sum_{i=1}^{m} H(x_i) - H(\mathbf{x}) \tag{13}$$

It is proved elsewhere [26] that more efficient codes are obtained while using the set of variables **x** instead of the individual ones (and thus, $\sum_{i=1}^{m} H(x_i) > H(\mathbf{x}) \rightarrow I(x_1,...,x_m) > 0$), unless when the variables are independent (which implies in $I(x_1, x_2, ..., x_m)=0$). So, minimization of mutual information leads to statistical independence.

The **Kullback-Leiber (KL) divergence**, defined through Equation 14 [1]:

$$C_{KL}(Q,P) = \int Q_x(\mathbf{x}) \log \frac{Q_x(\mathbf{x})}{P_x(\mathbf{x})} dx \tag{14}$$

measures the distance between the two probability densities $P_x(\mathbf{x})$ and $Q_x(\mathbf{x})$, as it 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. The KL divergence is proved to be equivalent to mutual information [1].

2- Uniqueness of Nonlinear ICA/BSS Solution

In the nonlinear case, statistical independence is not enough to guarantee source separation. If two random variables *x* and *y* are independent and so $p_{x,y}(x, y) = p_x(x)p_y(y)$, for differentiable functions *f* and *g* it can be proved that [29]:

$$p_{f(x),g(y)}(x,y) = p_{f(x)}(x)p_{g(y)}(y) \tag{15}$$

and so the random variables $f(x)$ and $g(y)$ are also independent. This indeterminacy, different from the scaling and ordering of the linear case, is not acceptable in a source recovery problem. The nonlinear mapping that preserves independence is called trivial [17]. Examples of trivial mappings can be found in [16].

Some assumptions have been proposed to guarantee uniqueness of solutions for nonlinear ICA problems [18]:

- the problem dimension (number of components) equals two. So data can be considered as complex variables;
- the mixing function *F* is a conformal mapping and zero preserving ($f(0)= 0$). Conformal mapping is a one-to-one nonlinear mapping that locally preserves the coordinate orthogonality [30];
- the density functions of the independent components are limited to known values.

Another class of mappings that provides unique solution for the nonlinear ICA problem is obtained when structural constrains are imposed to the nonlinear mixing model.

The conclusion is that, to obtain source separation in nonlinear mixture, constraints on the mixing mapping *F* (see Equation 3) or sources must be applied.

3- Nonlinear ICA Algorithms

This section covers some successful separation algorithms, most of them based on neural network implementation.

3.1- Structural Constrained Algorithms

As mentioned in Section 2, information on the mixing model is needed to perform signal separation in the nonlinear ICA problem. In the following sub-sections some proposed structural models and their respective nonlinear BSS algorithms are detailed. More attention is focused on the Post-Nonlinear model, which is one of the most successful recovery architectures [29].

3.1.1- Post-nonlinear Mixtures

The post-nonlinear (PNL) model [29] assumes that the observed signals are generated by a linear mixing followed by component-wise nonlinearities (cross-channel nonlinearities are not allowed). The observed signals can be expressed as:

$$x_i = f_i(\mathbf{As}) \tag{16}$$

where the complete nonlinear mapping is $F(.)=[f_1(.),f_2(.),...,fn(.)]^T$. This model, although restrictive, can be applied to a great variety of practical problems, especially when the source signals propagate through a linear channel and the nonlinearities are present on the set of sensors.

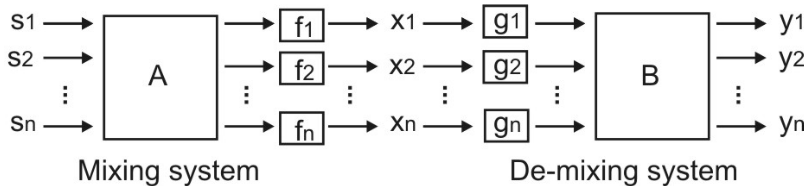


Figure 1 – Post-nonlinear Mixing / Demixing Model.

As illustrated in Figure 1, the recovery of the source signals is performed by an inverse model that comprises nonlinear (g_i) and linear (***B*** matrix) stages. The nonlinear section is usually estimated using neural network architectures like Multi-layer Perceptrons - MLP or Radial Basis Function - RBF, see [31] for detailed information on neural models. Different statistical independence measures are used in the training procedure. The linear part can be executed by any linear ICA/BSS method, see for example [2, 6]. The estimated sources are computed through:

$$\mathbf{y} = \mathbf{B}[\, g_1(x_1),..., g_N(x_N)]^T \tag{17}$$

The algorithm developed by Taleb and Jutten [29] is one of the earliest proposed in the literature for the post-nonlinear mixture problem. It is robust to variations in the source pdf, because it performs iterative estimation of the recovered signal statistics through the use of score function (ψ) computation (see Figure 2):

$$\psi = p'_{Yi}(u) / p_{Yi}(u) \tag{18}$$

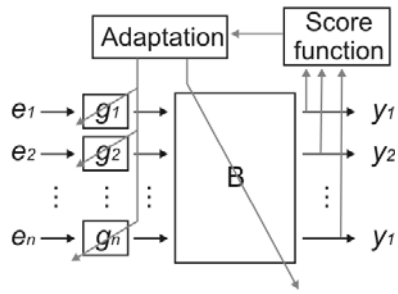


Figure 2 – Diagram of the Post-nonlinear algorithm proposed in [29].

Each nonlinear function block $g_k(u)$ (k=1,…,n) was modeled using MLP networks with linear output neuron:

$$g_k(u) = \sum_{h=1}^{N_H} \xi_j^h \sigma(\omega_j^h u - \eta_j^h) \tag{19}$$

where $u_j=g'_j(e_j)$, e_j are the observed signals (see Figure 2), $h = 1, ..., N_H$ is the hidden neurons index, and $j = 1, .., J$ is the observed signal (time) index. The Kullback-Lieber divergence is used to derive online learning rules for the nonlinear function estimation [29].

As there are lots of parameters to be adjusted in the inverse model and the optimization problem involves nonlinear functions, the algorithms some times suffer from local minima [17]. Different procedures were proposed in the literature in order to improve the neural network training performance in PNL mixing models. In [32, 33] a Genetic Algorithm [34] was used to perform a global search for the best set of parameters that maximize the independence measure. Different global optimization algorithms like Competitive Learning and Simulated Annealing were also tested and compared in [35]. The main problem concerning these approaches is that they present very high computational cost.

Alternative neural network architectures were also successfully applied in the PNL source separation problem. For example, in [36] Radial Basis Function (RBF) networks were used. In a different work [37], a separation algorithm using adaptive spline neural networks was proposed. A set of adaptively adjustable nonlinear series was used in [38] as activation functions for the hidden neurons of a feed-forward neural network based algorithm.

3.1.2- Different Structural Constrained Models

Some nonlinear BSS algorithms are based on models similar to the post-nonlinear one, although more general. In [39], for example, the mixing is considered to be performed by the following map:

$$\mathbf{x} = \mathbf{A}_2 f(\mathbf{A}_1 \mathbf{s}) \tag{20}$$

where $\mathbf{A_1}$ and $\mathbf{A_2}$ are square matrices and $f = [f_1, f_2, ..., f_N]^T$ are component-wise nonlinearities (see Figure 3). This model is sometimes called Post-Nonlinear-Linear (PNL-L) or Linear-Nonlinear-Linear (LNL).

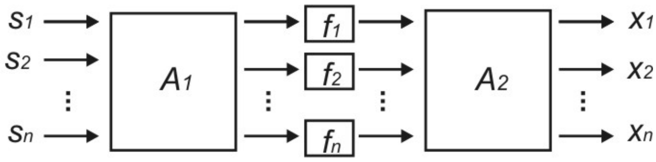


Figure 3 – Diagram of the Post-Nonlinear-Linear model [43].

Cross-channel nonlinearities are also not allowed in the model described in Equation 20, but a linear mixing (represented by $\mathbf{A_2}$) is performed after the nonlinear function blocks, providing a more general model when compared to post-nonlinear mixtures.

Some algorithms have been proposed to solve the PNL-L mixing problem. One of the earliest is described in [40] and uses a two hidden layers perceptron network trained through the back propagation of error functions derived from entropy or mutual information contrasts. A hybrid RBF-MLP neural model was applied in [41] to obtain an inverse of the PNL-L mapping. The use of this nontrivial architecture is justified in order to proper explore the good characteristics of each network model.

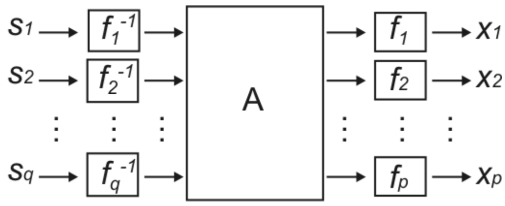


Figure 4 – Diagram of the mono-nonlinearity mixing model [42].

In [43] a different structural model is derived for the nonlinear BSS problem, called mono-nonlinearity mixing (see Figure 4). The observed signals are considered to be generated through Eq. 21. This model is said to be more general than the post-nonlinear, as each observed signal may be generated by different nonlinear functions of the sources and the number of observed signals p and sources q are not restricted to be the same..

$$\mathbf{x} = f(\mathbf{A} f^{-1}(\mathbf{s})) \tag{21}$$

The generality of the model proposed in Eq. 21 is stemmed from the theory of functional analysis [43] and it is proved that, this architecture can represent two-layer nonlinear mixing systems [42].

3.2- Mappings satisfying the Addition Theorem

Considering a special case where nonlinear mixtures can be reduced to linear through a simple mapping *H*, it is clear that signal separation can be achieved (through a linear ICA algorithm) if *H* is known a priori. It was demonstrated in [44] that a class of nonlinear mixtures satisfying an addition theorem (AT) can be mapped, using a simple transformation, into linear mixtures. An example of such mapping is [17]:

$$\begin{aligned} x_1 &= (s_1 + s_2)(1 - s_1 s_2)^{-1} \\ x_2 &= (s_1 - s_2)(1 + s_1 s_2)^{-1} \end{aligned} \tag{22}$$

Using the transformation $u_i = h(s_i) = \tan^{-1}(s_i)$, the mixing model, after some manipulation, reduces to:

$$x_1 = \tan(u_1 + u_2) \quad \text{and} \quad x_2 = \tan(u_1 - u_2) \tag{23}$$

now applying $h(.) = \tan^{-1}(.)$ to x_i:

$$\begin{aligned} v_1 &= \tan^{-1}(x_1) = (u_1 + u_2) \\ v_2 &= \tan^{-1}(x_2) = (u_1 - u_2) \end{aligned} \tag{24}$$

the variables v_i are a linear mixture of u_i . The original sources s_i are obtained through the following steps:

1. $v_i = h(x_i)$;
2. $\mathbf{u} = \mathbf{B}\mathbf{v}$, where $\mathbf{B}$ is obtained through any linear ICA algorithm;

3. $s_i = h^{-1}(u_i)$.

Although the class of AT models describes some reasonable nonlinear mixing, the structure of the system must be known a priori in order to be applied in practical source separation problems. Unfortunately, in most applications the system structure is unknown, and there may be no way of learning it [45]. Another limitation of this method is that the scale indeterminacy of the linear ICA estimation is transformed nonlinearly and, in some cases, this may produce severe distortion in the recovered nonlinear mixed signals.

3.3- General Nonlinear Mixtures

If there is no constraint on sources or the mixing model, there is no guarantee that the obtained nonlinear independent components are related to the original sources (see Section 2). In view of this, nonlinear ICA algorithms are not able to perform blind source separation unless the uniqueness assumptions are satisfied. In the following, some popular nonlinear ICA algorithms are derived.

3.3.1- Self-Organizing Maps

One of the first attempts to perform nonlinear ICA was made through Self Organizing Map - SOM [46]. The Self Organizing Map is an unsupervised trained neural network that provides a topological organization of the input data set [47], transforming a k-dimensional continuous input space into a discrete characteristic map (generally bidimensional). Each neuron of the map is fully connected to all inputs. SOM compacts the information while preserving topological relations of the input data set. Self-organizing maps are widely applied in different signal processing tasks like fault diagnosis [48], image processing [49], control systems [50], robotics [51] and feature extraction [52]. An extensive review on SOM applications in engineering problems is presented in [53].

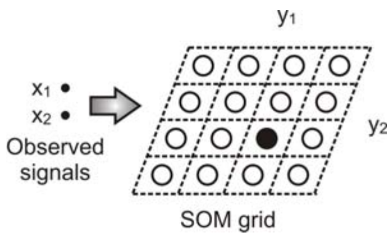


Figure 5 – SOM used for nonlinear ICA: for each input vector an independent component corresponds to the coordinates of the winner neuron (black dot).

It can be proved that the coordinates y_1 and y_2 of the winner neuron in the map (see Figure 5) are independent and roughly uniformly distributed [46]. To perform nonlinear ICA, SOM is trained using as inputs the observed signals and the coordinates of the winner vector correspond to the estimated independent components, which are assumed to have uniform pdf.

A disadvantage of the method is that the mapping is discrete and therefore some kind of regularization is needed to obtain continuous outputs. Another limitation is that the computational complexity increases very fast with the number of sources. To evaluate the computational cost, consider that N is the number of sources (and observed signals) and K is the number of desired quantization levels, thus the number of parameters Np of the network is given by:

$$Np = N \times K^N \tag{25}$$

Although very attractive, due to simple implementation, the use of SOM for nonlinear BSS requires that some conditions are satisfied in order to obtain good estimates of the original sources [17]:

- the size of the map (i.e. the number of neurons) must provide small quantization error;
- source signals are sub-Gaussian (the closer to uniform distribution the better are the results);
- the dimensionality of the problem (number of signals) shall be small;
- the mixing mapping presents mild nonlinearities.

Examples of the application of SOM for nonlinear ICA can be found in [46], where the method was firstly proposed and [20], where nonlinear ICA was used to remove multiplicative noise in images.

A similar method can be formulated through Generative Topographic Maps-GTM [54], providing a more solid theoretic foundation that can overcome some limitations of SOM. The GTM method closely resembles SOM in that it uses a discrete grid of points forming a regular array in the m-dimensional latent space (similar to SOM characteristic map). As in SOM, the dimension of the latent space is usually $m=2$ [1].

The advantage of GMT is that it can be modified to estimate nonlinearly mixed sources with virtually any probability distribution. Source variables are modeled as mixtures of Gaussian pdfs and the model parameters are determined by maximum likelihood using the expectation maximization (EM) algorithm [1]. GMT training procedure involves two steps, evaluation of the posteriori probabilities and adaption of the model parameters [54]. A nonlinear ICA method using GMT is proposed in [55]. For this model, the sources pdf must be known a priori and the correct separating mapping is considered to be the least complex one (given the sources pdf).

3.3.2- Bayesian Inference Methods

A nonlinear independent component analysis algorithm was proposed in [56] for general nonlinear mixtures, as described in:

$$\mathbf{x} = f(\mathbf{s}) + \mathbf{n} \tag{26}$$

where **n** is assumed to be a gaussian noise independent from the source signals.

The method uses MLP networks trained to estimate the nonlinear mapping. The sources are considered to be formed by a mixture of gaussian distributed signals. It can be proved [56] that given enough Gaussians in the mixture, virtually any distribution can be modeled with arbitrary accuracy. A variation of this method was applied previously to solve the linear ICA problem [57].

The Bayesian estimation method assigns posteriori probabilities to every nonlinear model that has possibly generated the measured data. The proposed method uses a technique called ensemble learning-EL [58], which is computationally more efficient and approximates the full Bayesian treatment. In EL, only the most probable subset of models is tested using a parametric approximation that is fitted to the posteriori probability [59]. The two layer MLP model is described as:

$$f(\mathbf{s}) = \mathbf{B}\varphi(\mathbf{A}\mathbf{s} + \mathbf{a}) + \mathbf{b} \tag{27}$$

for which, the parameters are: output layer weights **B**, output layer bias **b**, input layer weights **A**, input layer bias **a** and the nonlinear activation function $\varphi(.)$. For this, the MLP training procedure is not the traditional error back-propagation. For each parameter of the model, it is assigned a parametric probability distribution function and the purpose of the training process is to minimize the misfit between the exact posteriori distribution and the parametric approximation. Initial results obtained through the ensemble learning approaches are illustrated in [60] and [56].

As described in Equation 26, different from the standard nonlinear ICA formulation, the noise is modeled during the mixing process. Considering this, the authors refer to their algorithm as Nonlinear Independent Factor Analysis (NIFA) and it can be viewed as a nonlinear extension of Factor Analysis and Independent Factor Analysis [61]. The NIFA method is able to approximate pretty well the true sources in many cases, a great disadvantage is its high computational load [17].

Bayesian NLICA methods were proposed in [62] and [63]. Some tests were conduced in [64] to compare the experimental recovery performance of both Bayesian and structural constrained algorithms (PNL model) for nonlinear mixed signals. The main conclusions were:

- the PNL methods definitely perform better in classical PNL mixtures (same number of sources and observations and invertible nonlinearities);
- the performance of both methods can be improved by exploiting more mixtures than sources (especially in the noisy case);
- The main advantage of the Bayesian method is that no structural constraint is made on the nonlinear mixing model, producing more general algorithms. These methods are computationally more expensive and usually require several runs with different initializations (suffer from local minima problem).

3.3.3- Other Approaches for General Nonlinear Mixtures

In this section a brief review of some other methods proposed for solving the nonlinear ICA problem is presented. These algorithms usually apply nontrivial approaches to estimate the nonlinear independent components.

In [65], it was demonstrated that training low-complexity autoencoders may lead to nonlinear independent component analysis. The method is called LOCOCODE (Low Complexity Coding and Decoding). Low-complexity neural networks (not fully connected) are used to obtain descriptions of the inputs with as few and simple features as possible. Depending on statistical properties of the data, LOCOCODE may be able to approximate the nonlinear independent components. The method fails if high complexity transformations are to be estimated or a large number of sources are present. Some examples of low-complexity neural networks applications can be found in [66].

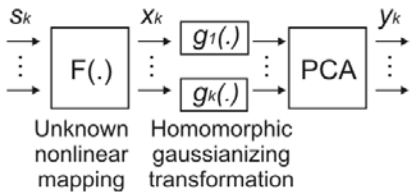


Figure 6: Diagram of the Homomorphic Nonlinear ICA method [67].

A novel technique for nonlinear ICA was proposed in [67]. Homomorphic transformations are applied to the observed signals (whether they are obtained from linear or nonlinear mixtures) in order to convert their marginal distributions to Gaussians (see Figure 6). The signals are further made orthogonal (independent) through principal components analysis (PCA) [68]. This is possible because for Gaussian variables decorrelation (performed by PCA) leads to independence.

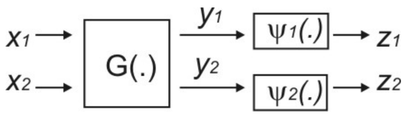


Figure 7: Diagram of the MISEP method [69].

Another approach to perform both linear and nonlinear ICA is the MISEP method, which is described in details in [69]. The MISEP algorithm uses mutual information contrast and it is considered an extension of the information maximization (INFOMAX) method [1]. A diagram of the MISEP model is illustrated in Figure 7, where y_i are the estimated independent components. The nonlinear functions ψ_i and the output variables z_i are only used in the training process. The nonlinearities ψ_i ideally should be the cumulative probability function of the corresponding y_i. If the aim is performing nonlinear ICA, the block labeled as G(.) shall model the existing nonlinear mapping.

3.4- Local ICA

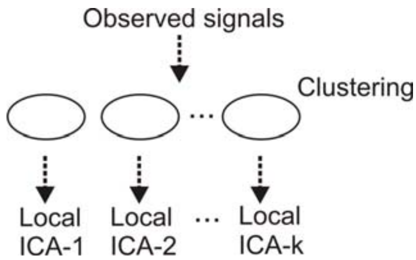


Figure 8: Local Independent Component Analysis based on Clustering.

If the ICA model is used for feature extraction instead of source separation, better description of the data set can be obtained while exploring local characteristics. Considering a high dimensional data set which presents severe variation on its statistics, the standard linear ICA model may not be able to reveal the underling structure of the data. In this situation, it is more reasonable to perform feature extraction (ICA estimation) from k subsets of the complete data set in which the elements belonging to the k-th subset present similar characteristics. This procedure leads to the Local ICA model.

As proposed in [70], a high dimensional data collection may be separated into clusters, through some clustering algorithm such as k-Means [71] or SOM [53], and linear independent components are then extracted from each subset (see Figure 8). Determining exactly the number of clusters that exists in a certain data set is not a simple task and usually requires some prior information. Fortunately, there are some criteria, such as the one proposed in [72], that can be used to choose the proper number of clusters in a blind signal processing application.

The clustering is responsible for an overall nonlinear representation, while linear ICA models estimated from each cluster describe local features of the data. Local ICA can be viewed as a compromise between linear and nonlinear ICA [17]. The purpose is to obtain better representation when compared to linear ICA, while avoiding computational problems of the nonlinear model [73]. In different Local ICA approaches the clusters may be overlapping, using, for example, fuzzy boundaries [74, 75] or non-overlapping [73, 76].

3.5- Extensions to the Standard Nonlinear Independent Component Analysis Paradigm

In many practical applications the instantaneous nonlinear independent component analysis model (see Equation 3) may not describe properly the mixing environment. Some modifications have recently been studied and consider additive noise [77, 78], convolutive nonlinear mixing models [79, 80] and the case when the number of sources and mixtures is different [81, 82].

An approach for nonlinear blind source separation of noisy signals was proposed in [78]. The sources are considered to be corrupted by additive noise (**n**) prior to the nonlinear mapping *F*, see Eq. 28. A radial basis function network (RBF) was applied to estimate the mixing model. The algorithm proposed therein uses as cost function for the RBF based on higher-order cumulant approximations of the mutual information criterion.

$$\mathbf{x} = F(\mathbf{s} + \mathbf{n}) \tag{28}$$

If the mixing process takes place within a multi-path environment, the observed signals are said to be generated through a convolutive model [79]. In this case, delayed versions of the sources are responsible for generating the mixed signals. A general nonlinear convolutive mixing model is described through Eq. 29:

$$\mathbf{x}[k] = F(\mathbf{s}[k], \mathbf{s}[k\text{-}1], \ldots \mathbf{s}[k\text{-}p]) \tag{29}$$

where k is the time index and p the number of time delays necessary to produce the observe d signals.

The mixed signal at time k ($\mathbf{x}[k]$) is considered to be generated by a nonlinear transformation *F(.)* applied to the source $\mathbf{s}[k]$ and to its p delayed versions ($\mathbf{s}[k\text{-}1]$, … $\mathbf{s}[k\text{-}p]$). As in the standard nonlinear ICA model, in the convolutive case the uniqueness assumptions (see Section 2) are also required to avoid multiple solutions. In [79], the PNL model was modified in order to incorporate a convolutive mixing block, as depicted in and Figure 9. A PNL structure was used to guarantee unique solution and a neural model, trained through KL divergence criterion, was applied to estimate the parameters of Eq. 30.

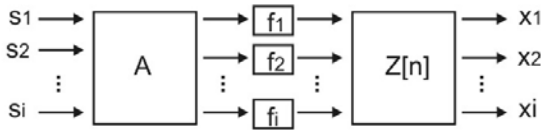


Figure 9: Convolutive PNL model as proposed in [79].

$$x_i = \sum_{n=1}^{p} \mathbf{z}_i[n] f_i(\mathbf{A}\mathbf{s}[k-n]) \tag{30}$$

As in the linear ICA model, most of NLICA algorithms assume that the number of sources N and observed signals M is the same. The underdetermined NL-BSS problem, when N>M, is a more complicated task, as the available information (number of observed signals) was diminished if compared to the determined case (N=M). In [81], through a Bayesian approach (see Section 3.3.2), a generalized Gaussian distribution model was applied to estimate sources pdf. Multilayer Perceptron neural networks were applied to approximate the nonlinear functions. The sources were approximately recovered in a three sources, two observed signals problem, indicating the algorithm's validity.

Another problem that has not been solved is how to optimally estimate the number of nonlinear independent components when there are more observations than sources N<M (the problem is overdetermined). In the linear case, PCA transformation [68] is usually applied for dimension reduction providing a measure of energy concentration that can be used to select the components that better describe de data set [1]. A modified PNL structure was proposed in [82], in which the observed signals are described through:

$$\mathbf{x} = \mathbf{B}F(\mathbf{A}\mathbf{s}) \tag{31}$$

where **A** and **B** are respectively NxN and MxN matrixes, *j*=1,.., N and *i*=1,…,M. In Eq. 31, **B** is responsible for increasing signal dimensionality. In the inverse model, see Eq. 32, a linear signal compaction algorithm (such as PCA) is used to estimate the compaction NxM matrix **W**. After that, a PNL is applied to the obtain the independent components **y**.

$$\mathbf{y} = \mathbf{D}G(\mathbf{W}\mathbf{x}) \tag{32}$$

4- Experimental Results

To evaluate the performance of some nonlinear ICA algorithms, five experimental tests were conduced using synthetic data. A practical application in experimental high energy physics is also described. In experiments 1 to 4, the purpose was to perform nonlinear blind signal separation. To evaluate the signal recovery performance, both the estimated components and the original sources were normalized in amplitude (using the maximum value as a normalization factor) and compared through mean square error computation:

$$\mathrm{MSE} = \frac{1}{Nk} \sum_{k=1}^{Nk} (s_n[k] - y_n[k])^2 \tag{33}$$

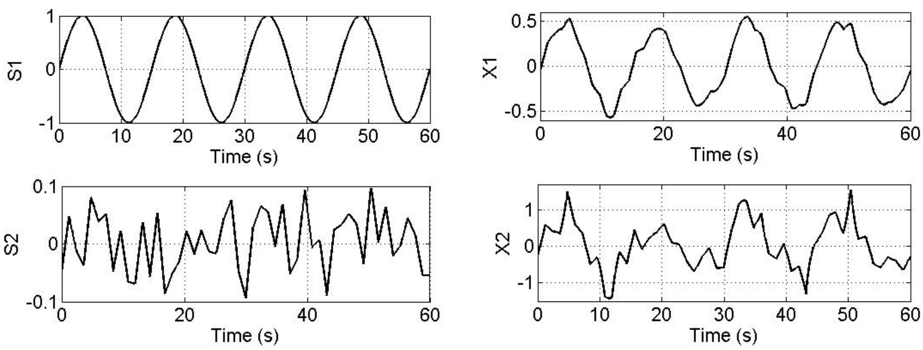
where s_n and y_n are respectively the normalized source and its corresponding estimated signal and k=1,…,Nk is the time index.

For the first four experiments, two algorithms were considered, one based on self-organizing maps (SOM) (see Section 3.3.1) and other that uses MLP neural networks to estimate the nonlinearities g_i of the inverse model of Post-Nonlinear Mixtures (PNL) (see Section 3.1.1), as proposed in [29]. These particular methods were chosen in order to provide a performance comparison between two of the most popular algorithms used for general (SOM) and structural constrained (PNL) nonlinear mixtures. The nonlinear methods were also compared to a linear ICA algorithm (FastICA) [1]. Section 4.6 provides a general discussion for the results of Experiments 1 to 4. Experiment 5 demonstrates the application of Local ICA to improve data representation, when compared to basic linear ICA. Finally, nonlinear BSS algorithms are applied in a complex real-world signal classification problem that concerns particle identification in high-energy signals.

4.1- Experiment 1: sub-gaussian sources and mild nonlinearity

In the first experiment, a sinusoidal signal (*ω=0.42 rad/s*) and a sub-gaussian random variable (*kurtosis = - 0.698*) formed the source vector **s**[k]. In this case the mixing map was realized by a mild nonlinearity: **x**[k] =**As**[k] + (**As**[k])^3 , where **A**=[0.6,0.2;0.1,0.9] (see Figure 10 for the original sources **s**[k] and mixed signals **x**[k]). The mixed signals present only small nonlinear distortion. Considering this, the performance of both SOM and PNL algorithms were compared to a linear ICA method (FastICA algorithm).

The signal recovery problem, in this example, can roughly be solved using the basic linear ICA model (see Figure 11-a). However, more accurate results are obtained through both SOM and PNL algorithms (see Figures 11-b and 11-c). A limitation of the SOM algorithm is that it tries to adapt an approximately uniform distribution to the source signals, even when it is not the case, which may result in distortion. Figure 12 illustrates this problem for a 20x20 map. The first signal (sinusoidal) was recovered with small distortion in its statistical characteristics (pdf and kurtosis), on the other hand, the subgaussian statistics was significantly modified, resulting in higher mean square error. The PNL algorithm proves to be robust to variations of the source pdf, executing iterative estimation of the score functions. This results in smaller reconstruction error (distortion).

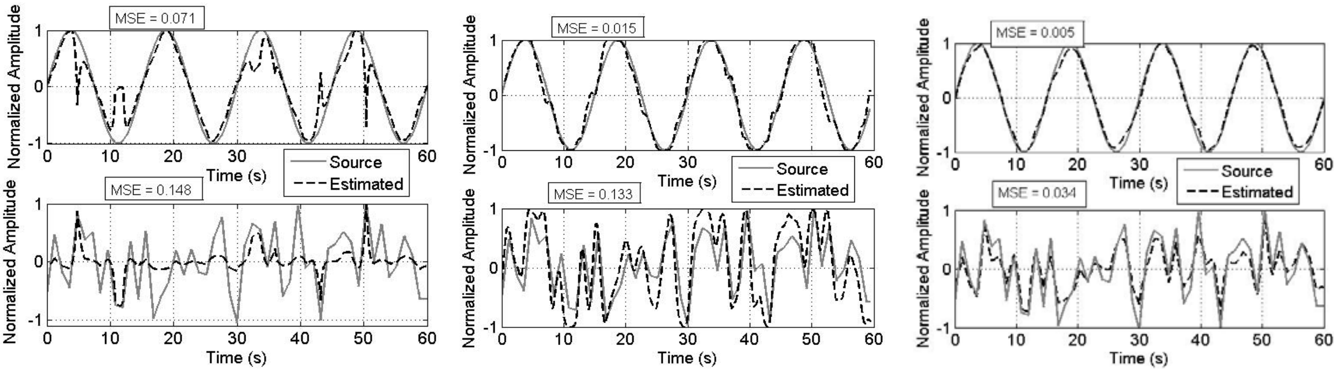


(a) (b)

Figure 10: Experiment 1 – (a) Original sources and (b) observed signals.

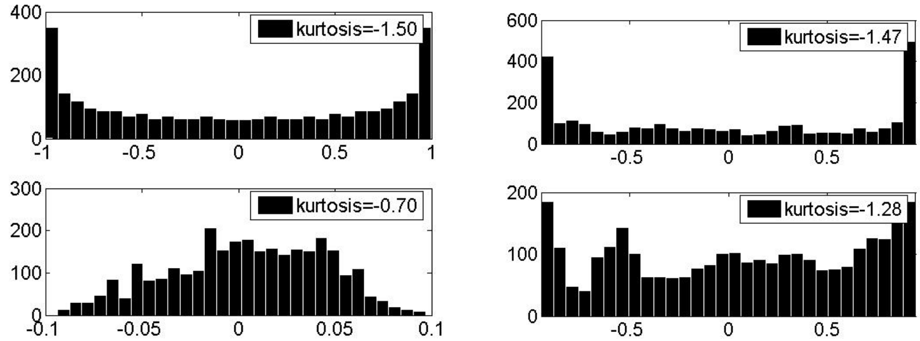
Another potential disadvantage of SOM is the high computational cost, which grows exponentially with the number of neurons (see Equation 25). In this work, only squared mapping was used to guarantee the same resolution (quantization interval) to both sources. The discrete SOM output signals were smoothed through a moving average filtering. In order to choose the number of neurons, the performance of different SOMs was evaluated through varying the size of the (squared) map (see Figure 13). For the sinusoidal signals (original and recovered), the figures of merit used here were the MSE and the Fast Fourier Transform (FFT) [83] amplitude coefficients. Considering these parameters, a 20×20 map may be the proper selection as it combines low MSE (for both components) and small frequency domain distortion (for the sinusoidal signal). Similar procedure was applied to the following experiments in order to choose the proper SOM dimensions for each problem.

Considering the MSE criterion, the PNL method achieved a better performance with respect to SOM. For the subgaussian signal, the result of SOM was slightly better than FastICA. However, SOM outperforms the linear method in sinusoidal signal recovery.



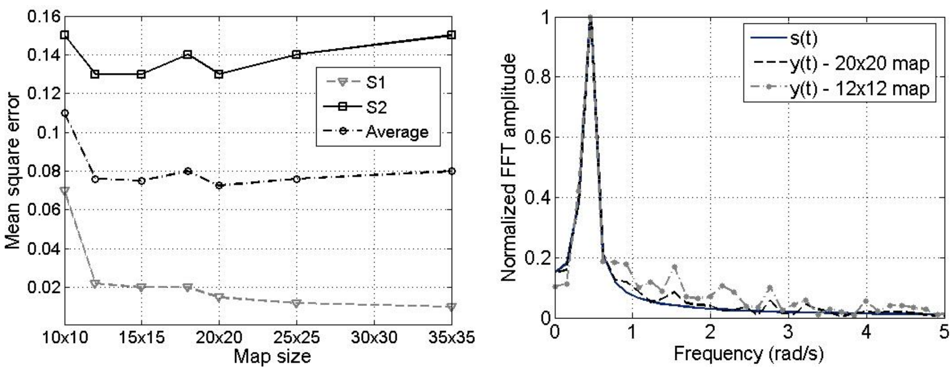
(a) (b) (c)

Figure 11: Experiment 1 - Estimated Sources through (a) Linear ICA, (b) SOM and (c) PNL.



(a) (b)

Figure 12: Experiment 1 - Distribution of original (a) sources and (b) recovered signals through SOM.

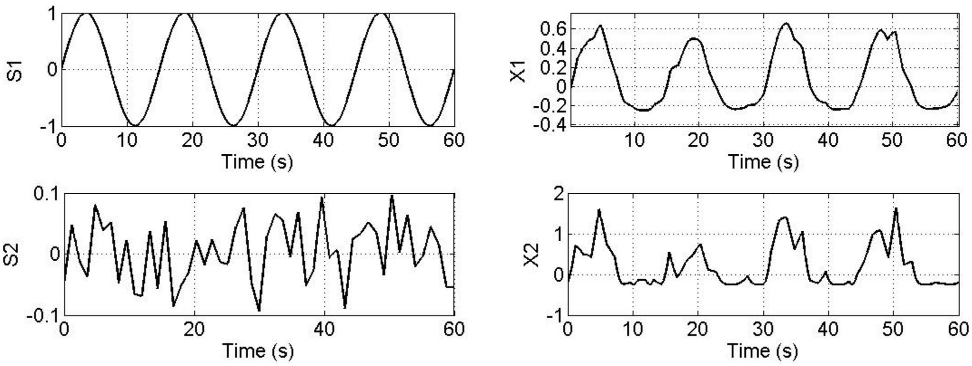


(a) (b)

Figure 13: Experiment 1 - (a) Mean square error and (b) FFT computed for different (squared) map dimensions.

4.2- Experiment 2: subgaussian sources and stronger nonlinearity

The same source signals used in Experiment 1 were now mixed using a stronger nonlinear condition: $\mathbf{x}[k]=\mathbf{A}\mathbf{s}[k] + (\mathbf{A}\mathbf{s}[k])^2$ (see Figure 14). The square function produces a more severe nonlinear distortion, if compared to the previous nonlinear mapping, as it eliminates negative values. Thus the complexity of the source recovery task increases [17].



(a) (b)

Figure 14: Experiment 2 – (a) Original source and (b) nonlinearly mixed signals.

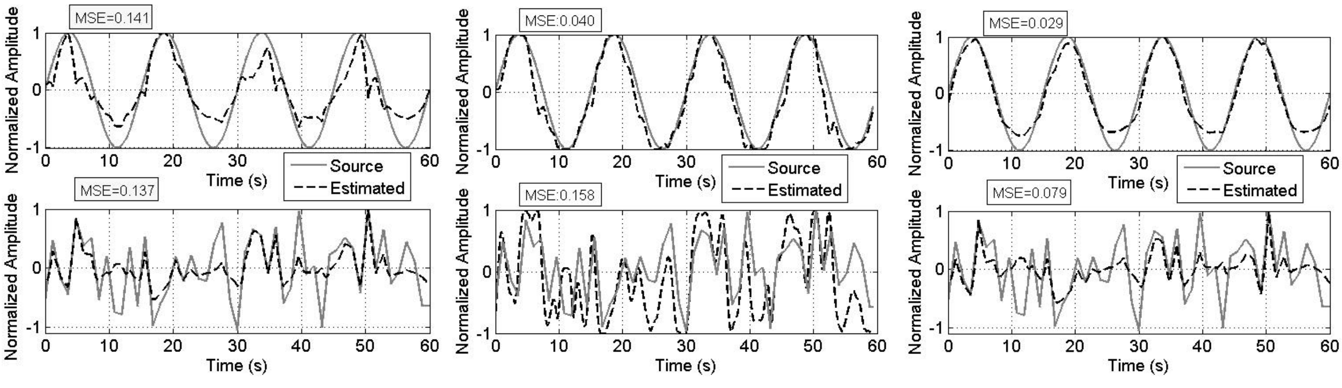
Once again SOM, PNL and FastICA algorithms were used to estimate the independent components. Figure 15 illustrates the performance of each method. As it can be depicted from Figure 15-a, for this problem the linear algorithm could not recover properly the original sources. The SOM method was able to recover the subgaussian source preserving its shape, but distorting its amplitude (see Figure 15-b), which results in a similar MSE value when compared to FastICA. Considering the sinusoidal signal, SOM method presented good performance. PNL algorithm estimated both signals with similar good accuracy. The source separation methods presented poorer results with respect to the previous experiment, probably due to the increase in the recovery task complexity.

4.3- Experiment 3: Gaussian and supergaussian sources

The sinusoidal signal was now mixed with approximately Gaussian (Experiment 3.1) and super-gaussian (Experiment 3.2) sources. The mixing mapping was as in Experiment 1: $\mathbf{x}[k] = \mathbf{As}[k] + (\mathbf{As}[k])^3$.

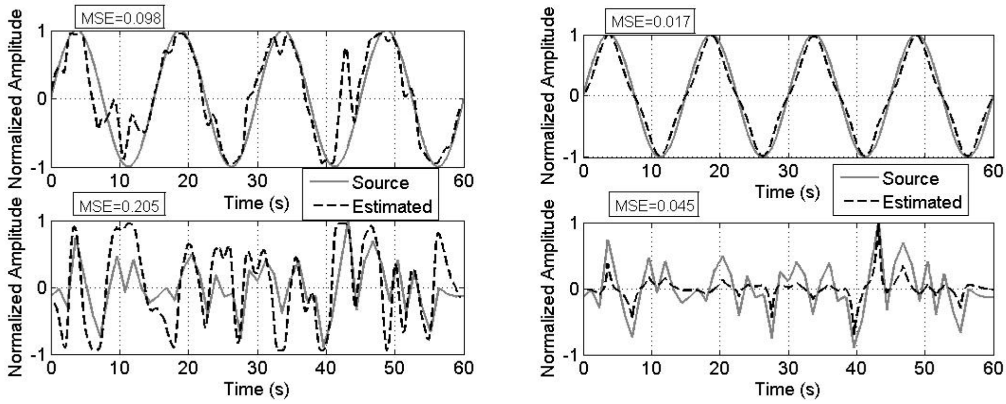
The aim of this test is to evaluate the nonlinear BSS algorithm sensitivity to sources statistics variations. As expected, the signal recovery performance was influenced by the source pdf while using SOM based nonlinear BSS (see both Figures 16 and 17). SOM produced severe distortion for the supergaussian source.

The PNL algorithm performs iterative estimation of the posteriori signal statistics, and so quite good approximations of the original sources were obtained. No significant performance degradation was observed for the PNL method with respect to the previous experiments.



(a) (b) (c)

Figure 15: Experiment 2 – Estimated sources through (a) FastICA, (b) SOM and (c) PNL.



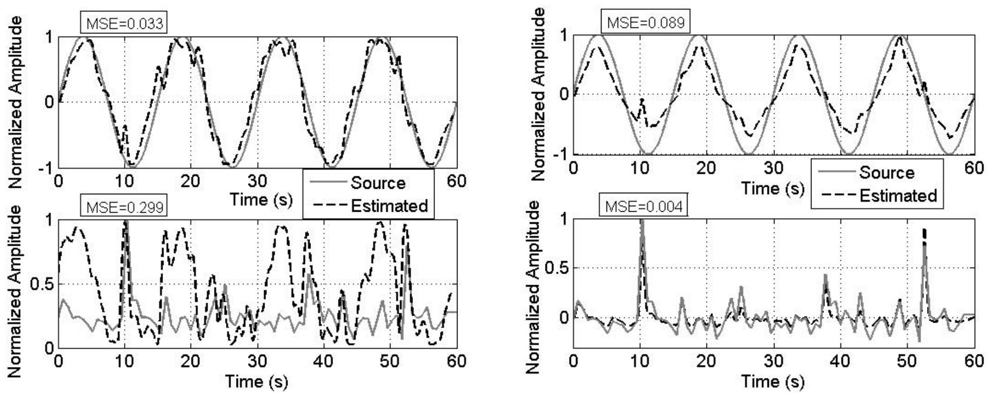
(a) (b)

Figure 16: Experiment 3.1 – Estimated sources through (a) SOM and (b) PNL.

4.4- Experiment 4: three subgaussian sources

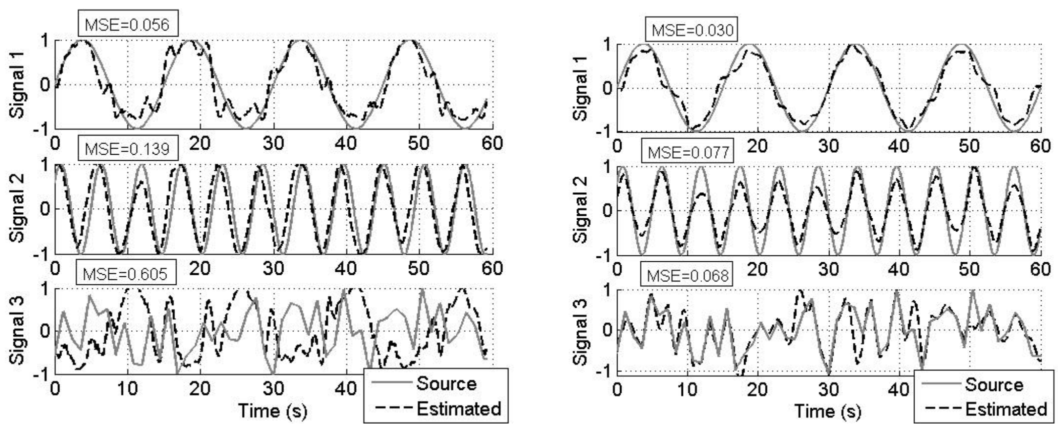
In this experiment the two source signals from Experiment 1 were mixed together with another sinusoidal component (*ω=1.13 rad/s*). The mixing function was again $\mathbf{x}[k]=\mathbf{A}\mathbf{s}[k] + (\mathbf{A}\mathbf{s}[k])^3$. The aim of this test is to evaluate the nonlinear BSS algorithms performance when a larger number of sources are involved in the mixing process.

It can be observed from Figure 18, that the recovery performance deteriorates for both methods. More severe distortions were observed while using SOM. Both sources 1 and 2 were relatively well recovered, but the algorithm has failed for the third source signal. The PNL algorithm also experienced a MSE increase with respect to the previous simulations, but the distortion might be considered acceptable.



(a) (c)

Figure 17: Experiment 3.2 – Estimated sources through (a) SOM and (b) PNL.

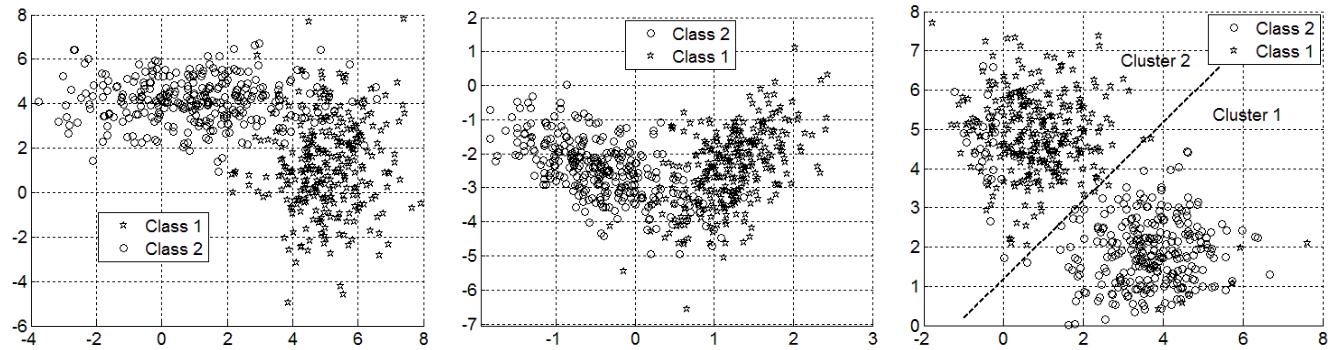


(a) (b)

Figure 18: Experiment 4 – Estimated sources through (a) SOM and (b) PNL.

4.5- Experiment 5: Local ICA

An application of Local ICA is demonstrated in this experiment for a binary decision problem. The purpose is to design a classifier system capable of separating the samples of two data classes with minimum error. Standard linear ICA has recently been applied for feature extraction (instead of blind signal separation) in classification problems [15]. For this, the estimated independent components were used to feed a classifier system (that can be implemented through neural architectures). In some cases, ICA transformation shall reveal underlying characteristics of data, producing discrimination performance improvement if compared to a classifier operating directly over the non-processed (non-independent) signals. When the dataset presents high variation in its statistical characteristics, the standard ICA estimation shall not present good results [17]. Considering this, if the available data set is clustered by similarity, a more accurate description can be obtained if the independent components are estimated for signals belonging to each cluster [70].



(a) (b) (c)

Figure 19: Experiment 5 - (a) The dataset, (b) linear and (c) local independent components.

In this experiment, the available bi-dimensional signals belonging to two distinct classes (see Figure 19-a) were pre-processed through both standard linear ICA and Local ICA. Neural classifiers perform class separation operating over the independent components. Figure 19-b illustrates the linear independent components, estimated through FastICA [3] in the feature extraction phase. The neural discriminator achieves 92% Class 1efficiency for a misclassification of 6% for Class 2. The neural classifier was implemented through MLP topology [31], traditional back-propagation training was used in a two layer network with hyperbolic tangents as neuron activation functions.

For Local ICA, the data set was clustered into two groups through k-means algorithm [71] and local independent components were extracted for each cluster (see Figure 19-c). MLP classifiers (similar to the ones used in the Linear ICA case) were supervised trained to maximize local discrimination. A considerable improvement was achieved in this configuration, as for 6% Class 2 misclassification, 97% of Class 1 was correctly identified (against 92% obtained from FastICA).

4.6- General Discussion

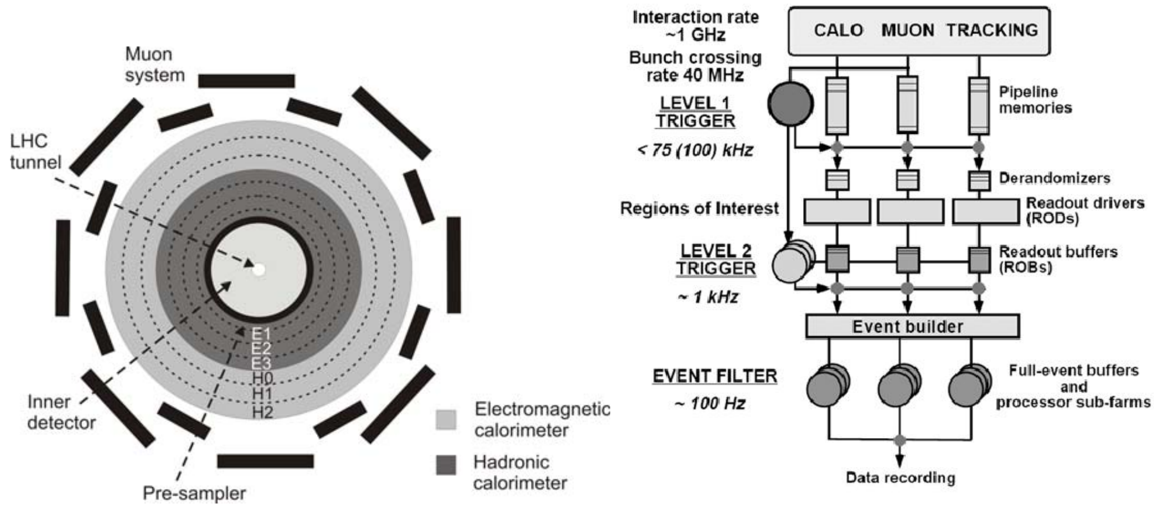
A comparison of the results obtained through the previous experiments (from 1 to 4) is shown in Table 1. In general, the PNL method obtained better results in terms of the mean square error (see equation 33), when compared to both SOM and FastICA. The linear method only achieved reasonable source recovery in the first experiment (due to mild nonlinear map). The source recovery performance obtained through SOM deteriorates when the original sources are not sub-Gaussian. If the problem dimensionality (the number of source signals) increases, the computational complexity grows exponentially, producing performance degradation in both nonlinear methods (SOM and PNL). Experiment 5 proved that, in some cases, local feature extraction describes the data set in a more meaningful way, providing valuable information for classification. The main indeterminacy of the Local ICA method is the appropriated number of clusters for each problem.

Table 1: Normalized mean-square error (MSE) for different experiments.

Experiment	1	2	3.1	3.2	4
Recov. Source	S_1 / S_2	S_1 / S_2	S_1 / S_2	S_1 / S_2	S_1 / S_2 / S_3
FastICA	0.071 / 0.148	0.141 / 0.137	0.153 / 0.129	0.115 / 0.133	0.352 / 0.247 / 0.230

SOM	0.015 / 0.133	0.040 / 0.158	0.098 / 0.205	0.033 / 0.299	0.056 / 0.139 / 0.605
PNL	0.005 / 0.034	0.029 / 0.079	0.017 / 0.045	0.089 / 0.004	0.030 / 0.077 / 0.068

4.7- Practical Application



(a) (b)

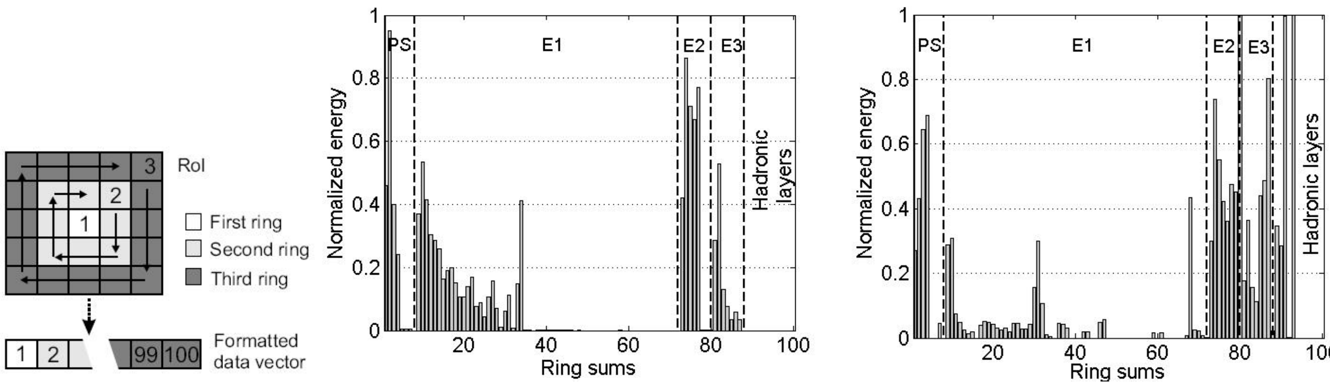
Figure 20: The ATLAS detector (a) and its triggering system (b).

Feature extraction procedures based on NLICA have recently been proposed in [52, 72, 84, 85] for the ATLAS detector [86] triggering system. The aim was to increase the particle discrimination efficiency of the online filtering task, which is performed in harsh conditions. ATLAS is one of the experiments of LHC (Large Hadron Collider) [87], the particle collider under construction at CERN (European Center for Nuclear Research). Placed at one of the LHC collision points, ATLAS has a cylindrical structure with the LHC beamline as the central axis (see Figure 20-a). Important information that guides the particle identification process is the energy deposition profile measured by the calorimeter system, one of the ATLAS subdetectors. The calorimeter system is segmented into four electromagnetic and three hadronic layers, producing more than 100,000 readout channels.

When operating at full capacity (high luminosity), LHC will produce $40x10^6$ bunch crosses per second. The total detector information will be near 60TB/s. Despite this very high event rate, the interesting channels will rarely occur. Therefore, a high-efficient filtering (triggering) system is required to guarantee that most of the background noise will be rejected and valuable information will not be lost. Considering this amount of data, the triggering task (see Figure 20-b) is performed online under short latency times and comprises three sequential filtering levels.

For LHC, interesting signatures can be found through decays that produce electrons as final state particles. Hadronic jets present energy deposition profiles similar to electrons (highly concentrated in the electromagnetic sections and almost no energy left in the hadronic layers), forming a huge background noise for the experiment. This practical, NLICA application, was conduced at the second trigger level using a dataset obtained through Monte Carlo simulator for proton-proton collisions [86]. The available signals comprise approximately 22,000 electron and 7,000 jet signatures. The particle discrimination procedure at LVL2 may be split into feature extraction, where relevant information is extracted from the measured signals, and hypothesis testing, where particle discrimination is performed. It is expected that, at high luminosity, 25,000 jets will reach the second-level trigger per second. In view of this, a slight improvement in background noise rejection, as for example one percent point, shall avoid the recording of 250 jets/second, providing cleaner data for offline analysis.

As proposed in [88], here the calorimeter information is formatted into concentric rings. In each calorimeter layer, the most energetic cell is considered as the first ring, and the next rings are sequentially formed around the first one (see Figure 21-a). The ring signals are obtained by summing the energy of the cells belonging to a given ring. This procedure makes the signals independent from the impact point in the detector. The ring energy is normalized within each layer. As a result, each signature becomes described by 100 rings. In practical calorimeter design, nonlinearities typically arise [89] and the independent features of the calorimeter signals may be better extracted by a nonlinear technique. Figures 21-b and 21-c depict typical electron and jet ring formatted signatures.



(a) (b) (c)

Figure 21: (a) Diagram of the ring mapping and signatures of typical (b) electron and (c) jet.

In this application, different NLICA algorithms were applied to extract, from the ring formatted data, relevant features for particle identification. The nonlinear independent components were used to feed MLP classifiers (see Figure 22). Three methods were used, Local ICA, SOM and PNL. In this particular problem, electron signatures represent the target signals to be detected and misclassified jets correspond to false alarm. The Receiver Operating Characteristic-ROC [90] was used as the figure of merit for the particle discrimination performance. The ROC shows how both detection (P_D) and false alarm (P_F) probabilities vary as the decision threshold changes. As the interesting events are very rare at LHC, high P_D is desired for the online triggering operation. Low P_F is also essential for the classifier design, as the huge background noise has to be rejected, as much as possible, to allow offline data analysis on clean data.

Figure 22: Block-diagram of the proposed NLICA based particle discriminators.

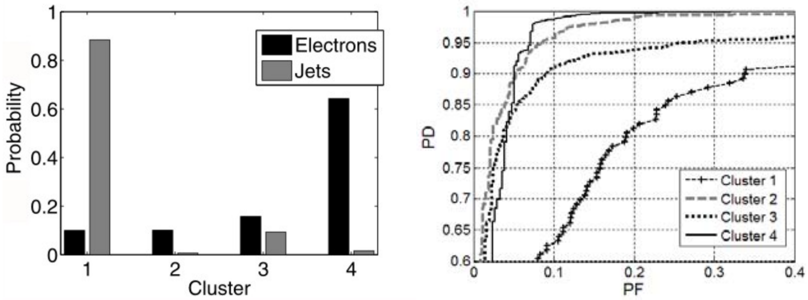
In [84], Local ICA (as described in Section 3.4) was applied to the ring formatted signals. The dataset was split into four groups using SOM for clustering. The concentrations of electrons and jets in each cluster are depicted in Figure 23-a. It can be observed that cluster 1 concentrates the majority of jets and cluster 4 most of the electron signatures. ICA (through FastICA algorithm) was applied to signals belonging to each cluster, and the estimated local independent components fed the input nodes of four neural classifiers. The ROC curves obtained from these discriminators are depicted in Figure 23-b. The main advantage of the local signal processing is that adjusting local thresholds allows the selection of different classes of particles, as for example, on offline analysis one may be interested only on electrons that present characteristics similar to typical jets (cluster 1). For different purposes, if the desired signatures are typical electrons cluster 4 threshold shall be adjusted.

A SOM-based NLICA algorithm (see Section 3.3.1) was applied to the ring formatted calorimeter signals in [52]. The number of neurons was varied and higher discrimination efficiency was obtained for a 6x10 map. As a bi-dimensional neuron grid was used, the dataset was mapped into two nonlinear independent components. The nonlinear independent components joint pdf is illustrated in Figure 24-a for electrons and jets. It can be observed that most electron signatures were concentrated in one side of the y_1-y_2 plane and jets in the opposite region, facilitating the discrimination performance.

A PNL model was applied in [85] to estimate the nonlinear independent components. This structure seems to properly describe calorimeter data, as in this application, signal mixing is performed in an approximately linear medium and the nonlinearities are present in the sensors (cintilators, photo-multipliers and optical-electronic devices) [89]. The PNL model restricts the number of independent components to be equal to the number of observed signals (100 ring sums). As in our particular application a compromise between high discrimination efficiency and fast decision is required, it is very important to reduce data dimensionality taking care not to discard relevant features. For this, the relevance R_i of the i-th nonlinear independent component was estimated through $Ri = (1/N)\sum_k [y_k - y_{k,i(\mathbf{x}_k=\overline{\mathbf{x}_{k,i}})}]^2$ where N is the number of elements in the training set, y_k is the neural network output when vector $\mathbf{x}_k$ is present to the classifier. The relevance R_i is the mean square error in the classifier output when the i-th component of the input $\mathbf{x}_k$ is replaced by its mean in the training set ($\mathbf{x}_k = \overline{\mathbf{x}_{k,i}}$). Figure 24-b depicts the relevance of each component. Eliminating the 70 less relevant components, the classifier was re-trained, reducing the computational cost and improving discrimination efficiency.

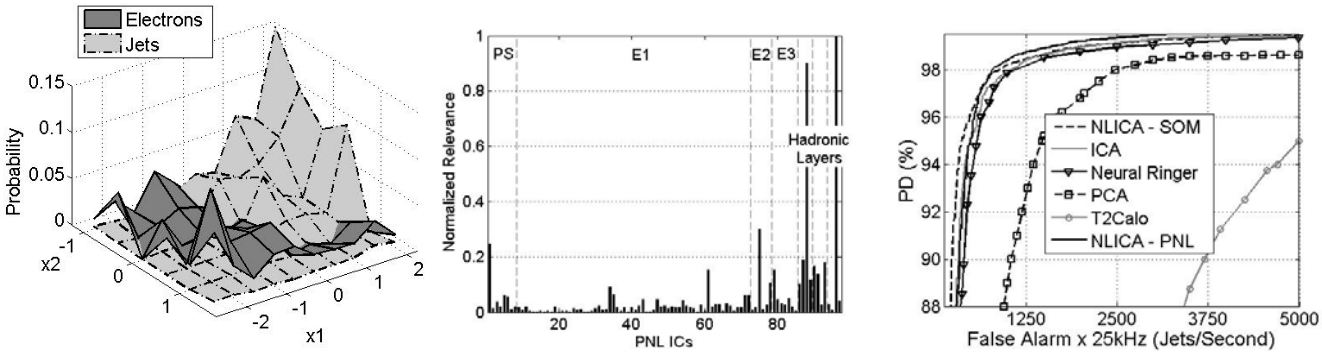
Considering the hypothesis testing procedure, neural classifiers were used in all three NLICA-based discriminators. The nonlinear independent components fed the input nodes of Multi-layer Perceptron (MLP) classifiers [31], trained to maximize particle discrimination. The networks comprised a single hidden layer and one output neuron. The number of hidden neurons was chosen after testing exhaustively the discrimination performance of a number of networks (varying the number of hidden neurons).

Different discriminators are compared in Figure 24-c. The baseline algorithm used in ATLAS for electron/jet discrimination (T2Calo) [91] extracts, directly from calorimeter measurements, high discriminating parameters that estimate the shape of the energy deposition profile. Thresholds on these parameters perform the particle discrimination. The Neural_Ringer [89] is an alternative method that is also implemented in the ATLAS software platform. Using a MLP neural classifier operating over the ring formatted signals, Neural_Ringer algorithm achieved better discrimination performance and similar computational cost when compared to T2Calo. Both segmented principal component analysis (PCA) and linear independent component analysis (ICA) were also applied, respectively in [92] and [15] for feature extraction and a MLP classifier performed the hypothesis testing over estimated principal or independent components. As illustrated in Figure24-c, the nonlinear ICA based classifiers, outperforms all other discriminators (note that Local ICA efficiency cannot be summarized in a single ROC as for each cluster one classifier was trained, see Figure 23-c). Table 2 provides a comparison between the proposed techniques. It can be observed from Figure 24-c and Table 2 that nonlinear ICA based discriminators present higher efficiency and among them, SOM method performs slightly better.



(a) (b)

Figure 23: Local ICA (a) cluster probabilities and (b) local ROC curves.



(a) (b) (c)

Figure 24: Nonlinear independent components (a) joint pdf (SOM) and (b) relevance (PNL) and (c) ROC curves.

Table 2: Performance comparison of different discriminators

Discrimination Technique	T2Calo	PCA	Neural_Ringer	ICA	Local ICA	NLICA PNL	NLICA SOM
PD (%) for PF=3%	81	88	95	96	97	98	98
False electrons (jets/second) for PF=95%	5000	1450	500	400	450	400	280

5- Conclusions

The nonlinear independent component analysis paradigm was discussed in this paper. It was demonstrated that, without proper regularization, there is an infinite number of solutions to the problem and thus the obtained independent components may not be related to the original source signals. Some special cases, where statistical independence leads to source separation were mentioned. Contrast functions that are usually applied for measuring statistical independence, in both linear and nonlinear cases, were exposed. Methods for nonlinear BSS, which impose constraints on the sources or the nonlinear mapping, were presented.

Experimental tests were conduced to evaluate the performance of some nonlinear ICA algorithms in different conditions. The SOM method, although simple in terms of implementation, has exhibited poorer signal recovery performance and higher computational cost when compared to PNL algorithm. A simple experiment using cluster based Local ICA illustrated the discrimination improvement that can be achieved while exploiting local features of a dataset. The efficiency of the nonlinear ICA model was also verified through a complex real world application, where discrimination performance was improved after estimation of the nonlinear independent components through the SOM, PNL and Local ICA methods.

A potential limitation that appears in nonlinear ICA algorithms is that there exist some model indeterminacies such as the ideal size of the SOM that better fits data, the number of hidden neurons in the MLP networks used for the nonlinear function approximation in the PNL model or the optimum number of clusters for a particular local ICA problem. These characteristics, combined with a computational cost increase, when compared to linear methods, sometimes shall prevent the application of nonlinear ICA, especially in high-dimensionality problems.

Acknowledgements

The authors are thankful for the support provided by CNPq, FINEP and FAPERJ (Brazil), CERN (Switzerland) and the European Union (through the HELEN project). The authors also thank the ATLAS Trigger/DAQ collaboration at CERN for providing the simulation calorimeter data set used in this work and for fruitful discussions concerning the nonlinear ICA application to the ATLAS detector online filtering system.

References

[1] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Wiley, 2001.

[2] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*. Willey, 2002.

[3] A. Hyvarinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Networks*, n. 13, pp. 411-430, 2000.

[4] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," *IEE Proceedings- F,* vol. 140, pp. 362-370, November 1993.

[5] 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*, November 1996.

[6] 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.

[7] A. Cichocki, S. Douglas, and S. Amari, "Robust techniques for independent component analysis with noisy data," *Neurocomputing*, vol. 22, pp. 113-129, 1998.

[8] 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*, pp. 145-150, Helsinki, Finland, June 2000.

[9] H.-M. Park, S.-H. Oh, and S.-Y. Lee, "Adaptive noise cancelling based on independent component analysis," *Electronics Letters*, vol. 38, pp. 832-833, Julu 2002.

[10] 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.

[11] 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.

[12] 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, pp. 1965-1973, November 2007.

[13] 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.

[14] K.-C. Kwak and W. Pedrycz, "Face recognition using an enhanced independent component analysis approach," *IEEE Transactions on Neural Networks*, vol. 18, pp. 530-541, March 2007.

[15] E. F. Simas Filho, L. P. Caloba, and J. M. Seixas, "Segmented independent component analysis for online filtering using highly segmented detectors," *Proceedings of the 7th International Conference on Intelligent Systems Design and Applications*, pp. 659-664, Rio de Janeiro, Brazil,October 2007.

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

[17] 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* (ICA2003), pp. 245-256, Nara, Japan, 2003.

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

[19] 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.

[20] M. Haritopoulos, H. Yin, and N. M. Allinson, "Image denoising using self-organizing map-based nonlinear independent component analysis," *Neural Networks*, pp. 1085-1098, 2002.
[21] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 1991.
[22] M. Kendall and A. Stuart, *The Advanced Theory of Statistics*. London: Charles Griffing and Company Limited, 4th ed., 1977.
[23] M. R. Spiegel, J. J. Schiller, and R. A. Srinivasan, *Probability and Statistics*. McGraw-Hill, 2 ed.,2000.
[24] T.-H. Kim and H. White, "On more robust estimation of skewness and kurtosis," *Finance Research Letters*, vol. 1, pp. 56-73, 2004.
[25] M. Welling, "Robust higher order statistics," *Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics* (AISTATS2005), pp. 405-412, Barbados, January 2005.
[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley, 1991.
[27] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, pp. 379-423, July 1948.
[28] A. Hyvarinen, "New approximations of diferencial entropy for independent component analysis and projection pursuit," *Advances in Neural Information Signal Processing*, no. 10, pp. 273-279, 1998.
[29] A. Taleb and C. Jutten, "Source separation in post-nonlinear mixtures," *IEEE Tran.s on Signal Processing*, no. 47, pp. 2807-2820, 1999.
[30] R. V. Churchill and J. W. Brown, *Complex Variables and Applications*. New York: McGraw-Hill, 5th ed., 1990.
[31] S. Haykin, *Neural Networks, Principles and Practice*. Bookman, 2001.
[32] Y. Tan and J. Wang, "Nonlinear blind source separation using higher order statistics and a genetic algorithm," *IEEE Transactions on Evolutionary Computation*, vol. 5, no. 6, pp. 600-612, 2001.
[33] S. Kai, W. Qi, and D. Mingli, "Approach to nonlinear blind source separation based on niche genetic algorithm," *Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications* (ISDA'06), Jinan, China, 2006.
[34] D. E. Goldberg, *Genetic algorithms in search, optimization and machine learning*. Addison-Wesley, 1989.
[35] F. Rojas, C. G. Puntonet, M. Rodríguez-Álvarez, I. Rojas and Martín-Clemente, "Blind source separation in post-nonlinear mixtures using competitive learning, simulated annealing, and a genetic algorithm," *IEEE Transactions on Ssystems, Man and Cybernetics – Part C: Applications and Reviews*, vol. 34, p. 407-416, 2004.
[36] Y. Tan and JunWang, "Nonlinear blind separation using an fbf network model," *Proceedings of the IEEE International Symposium on Circuits and Systems* (ISCAS00), vol. 3, pp. 634-637, Geneva, Switzerland, 2000.
[37] M. Solazzi, F. Piazza, and A. Uncini, "Nonlinear blind source separation by spline neural networks," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (ICASSP), vol. 5, pp. 2781-2784, Salt Lake City, US, 2001.
[38] P. Gao, L. Khor, W. Woo, and S. Dlay, "Two-stage series-based neural network approach to nonlinear independent component analysis," *Proceedings of the IEEE International Symposium on Circuits and Systems* (ISCAS), pp. 4559-4562, Island of Kos, Greece, 2006.
[39] M. Solazzi and A. Uncini, "Spline neural networks for blind separation of post-nonlinear-linear mixtures," *IEEE Transactions on Circuits and Systems - I: Regular Papers*, vol. 51, pp. 817-829, April 2004.
[40] H. Yang, S. I. Amari, and A. Cichocki, "Information backpropagation for blind separation of sources in nonlinear mixture," *Proceedings of the International Conference on Neural Networks*, vol. 4, pp. 2141-2146, Houston, US, June 1997.
[41] W. L. Woo and S. S. Dlay, "Nonlinear blind source separation using a hybrid rbf-fmlp network," *IEE Proceedings on Vision, Image and Signal Processing*, vol. 8, no. 2, pp. 173-183, 2005.
[42] P. Gao, W. Woo, and S. S. Dlay, "Neural network approaches to nonlinear blind source separation," *Proceedings of the Eighth International Symposium on Signal Processing and Its Applications*, vol. 1, pp. 78-81,Sidney, Australia, 2005.
[43] Griffel, D. H. *Applied Functional Analysis.* Wiley, 1984.
[44] J. Eriksson and V. Koivunen, "Blind identifiability of class of nonlinear instantaneous ica model," *Proceedings of XI European Signal Processing Conference*, pp. 7-10, Toulouse, France, September 2002.
[45] J. Eriksson and V. Koivunen, "Blind separation of a class of nonlinear ica models," *IEEE International Symposium on Circuits and Systems* (ISCAS05), vol. 6, pp. 5890-5893, Kobe, Japan, May 2005.
[46] P. Pajunen, A. Hyvarinen, and J. Karhunen, "Nonlinear blind source separation by self-organizing maps," *Proceedings of the International Conference on Neural Information Processing*, pp. 1207-1210, Hong Kong, 1996.
[47] T. Kohonen, "The self-organizing map," *Proceedings of the IEEE*, vol. 78, pp. 1464-1480, September 1990.
[48] T. Babnik, R. Aggarwal, and P. Moore, "Data mining on a transformer partial discharge data using the self-organizing map," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 14, pp. 444-452, April 2007.
[49] G. Cheng, J. Yang, K. Wang, and X. Wang, "Image color reduction based on self-organizing maps and growing self-organizing neural networks," *Sixth International Conference on Hybrid Intelligent Systems*, pp. 24-28, December 2006.
[50] Y. Zhao and J. A. Farrell, "Self-organizing approximation-based control for higher order systems," *IEEE Transactions on Neural Networks*, vol. 18, pp. 1220-1231, July 2007.
[51] H. Xiaoyuan, T. Ogura, A. Satou, and O. Hasegawa, "Developmental word acquisition and grammar learning by humanoid robots through a self-organizing incremental neural network," *IEEE Transactions on Systems, Man and Cybernetics, Part B*, vol. 37, pp. 1357-1372, October 2007.
[52] E. F. Simas Filho, L. P. Caloba, and J. M. Seixas, "Self-organized mapping of calorimetry information for high efficient online electron/jet identification in ATLAS," *Proceedings of Science*, PoS(ACAT2007)055, October 2007.
[53] T. Kohonen, E. Oja, O. Simula, A. Visa, and J. Kangas, "Engineering applications of the self-organizing map," *Proceedings of the IEEE*, vol. 84, pp. 1358-1384, October 1996.
[54] C. M. Bishop, M. Svensen, and C. K. I. Willians, "GTM: The generative topographic mapping," *Neural Computation*, vol. 10, pp. 215-234, 1998.
[55] P. Pajunen and J. Karhunen, "A maximum likelihood approach to nonlinear blind source separation,"in ICANN '97: *Proceedings of the 7th International Conference on Artificial Neural Networks*,(London, UK), pp. 541-546, Springer-Verlag, 1997.
[56] H. Lappalainen and A. Honkela, "Bayesian nonlinear independent component analysis by multilayer perceptrons," *Advances in Independent Component Analysis*, Springer-Verlag, pp. 93-121, 2000.

[57] H. Lappalainen, "Ensemble learning for independent component analysis," *Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation* - ICA1999, pp. 7-12, Aussois, France, 1999.

[58] H. Lappalainen and J. W. Miskin, "Ensemble learning," *Advances in Independent Component Analysis*, Springer-Verlag, pp. 75-92, 2000.

[59] H. Valpola, "Nonlinear independent component analysis using ensemble learning: Theory," *Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation*, pp. 351-356, Helsinki, Finland, 2000.

[60] H. Valpola, X. Giannakopoulos, A. Honkela, and J. Karhunen, "Nonlinear independent component analysis using ensemble learning: Experiments and discussion," *Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation* - ICA2000, pp. 351- 356, Helsinki, Finland, 2000.

[61] H. Attias, "Independent factor analysis," *Neural Computation*, vol. 11, pp. 803-851, May 1999.

[62] A. Honkela, "Approximating nonlinear transformations of probability distributions for nonlinear independent component analysis," *Proceedings of the IEEE International Joint Conference on Neural Networks*, pp. 2169-2174, Budapest, Hungary, 2004.

[63] A. Honkela, H. Valpola, A. Ilin, and J. Karhunen, "Blind separation of nonlinear mixtures by variational bayesian learning," *Digital Signal Processing*, 2007.

[64] A. Ilin, S. Achard, and C. Jutten, "Bayesian versus constrained structure approaches for source separation in post-nonlinear mixtures," *Proceedings of the IEEE International Joint Conference on Neural Networks*, vol. 3, pp. 2181-2186, Budapest, Hungary July 2004.

[65] S. Hochreiter and J. Schmidhuber, "Nonlinear ica through low-complexity autoencoders," *Proceedings of the 1999 IEEE International Symposium on Circuits and Systems* (ISCAS), vol. 5, pp. 53-56, Orlando, US, 1999.

[66] S. Hochreiter and J. Schmidhuber, "Feature extraction through lococode," *Neural Computation*, vol. 11, no. 3, 1999.

[67] D. Erdogmus, Y. N. Rao, and J. C. Principe, "Nonlinear independent component analysis by homomorphic transformation of the mixtures," *Proceedings of the IEEE International Joint Conference on Neural Networks*, vol. 1, pp. 52-55, Budapest, Hungary, 2004.

[68] I. T. Jolliffe, *Principal Component Analysis*. Springer, 2a ed., 2002.

[69] L. B. Almeida, "Linear and nonlinear ica based on mutual information: the misep method," *Signal Processing*, vol. 84, no. 2, pp. 231-245, 2004.

[70] J. Karhunen and S. Malaroiu, "Local independent component analysis using clustering," *Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation*, Aussois, France, 1999.

[71] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. Wiley, 2 ed., 2000.

[72] E. F. Simas Filho, J. M. Seixas and L. P. Caloba, "Segmented self-organized feature extraction for online filtering in a high event rate detector", *Proceedings of the European Conference on Signal Processing* (EUSIPCO), Lausanne, Switzerland, August, 2008.

[73] J. Karhunen, S. Malaroiu, and M. Ilmoniemi, "Local linear independent component analysis based on clustering," *International Journal of Neural Systems*, vol. 10, pp. 439-451, 2000.

[74] K. Honda, H. Ichihashi, M. Ohue, and K. Kitaguchi, "Extraction of local independent components using fuzzy clustering," *Proceedings of 6th International Conference on Neural Networks and Soft Computing*, pp. 837-842, Zakopane, Polland2000.

[75] T. Maenaka, K. Honda, and H. Ichihashi, "Local independent component analysis with fuzzy clustering and regression-principal component analysis," *Proceedings of IEEE International Conference on Fuzzy Systems*, pp. 857-862, Vancouver, Canada, 2006.

[76] F. Palmieri and A. Budillon, "Multi-class independent component analysis for rank deficient distributions," *Advances in Independent Component Analysis* ed. by M. Girolami, Springer-Verlag, pp. 145-160, 2000.

[77] S. Maeda and S. Ishii, "A noisy nonlinear independent component analysis," *IEEE Workshop on Machine Learning for Signal Processing*,, pp. 173-182, São Luis, Brazil, 2004.

[78] N. Zhang; X. Zhang; J. Lu; T. Yahagi, An approach for nonlinear blind source separation of signals with noise using neural networks and higher-order cumulants. *Proceedings of the IEEE International Symposium onCircuits and Systems* (ISCAS), Kobe, Japan, 23-26 May 2005, pp. 5726 – 5729.

[79] D. Vigliano, R. Parisi, and A. Uncini, "An information theoretic approach to a novel nonlinear independent component analysis paradigm," *Signal Processing*, vol. 85, pp. 997-1028, 2005.

[80] J. Zhang, W. L. Woo, and S. S. Dlay, "Expectation-maximisation approach to blind source separation of convolutive mixture," *IET Signal Processing*, vol. 1, no. 2, pp. 51-65, 2007.

[81] C. Wei, L. Khor, W. Woo, and S. Dlay, "A novel iterative conditional maximization method for post-nonlinear underdetermined blind source separation," *Proc. of the 15th Inter. Conference on Digital Signal Processing*, pp. 551-554, Cardiff, Wales, July 2007.

[82] E. F. Simas Filho, J. M. Seixas and L. P. Caloba, "Segmented overdetermined nonlinear independent component analysis for online neural filtering". *Proceedings of the 10th Brazilian Symposium on Artificial Neural Networks* (SBRN), Salvador, Brazil, October 2008 (to be presented).

[83] Paulo S. R. Diniz, Eduardo A. B. da Silva, Sergio L. Netto. *Digital Signal Processing: System Analysis and Design*, Cambridge University Press, UK, 2002.

[84] E. F. Simas Filho, J. M. Seixas and L. P. Caloba. "Local independent component analysis applied to highly segmented detectors", *Proceedings of the IEEE International Symposium on Circuits and Systems* (ISCAS), 18-21 May 2008, Seatle, US, pp: 3005-3008

[85] E. F. Simas Filho, J. M. Seixas and L. P. Caloba. "Nonlinear independent component analysis for online filtering is a high event rate detector", *Proceedings of the Brazilian Automatics Conference* (CBA), Juiz de Fora, Brazil, September, 2008 (In Portuguese).

[86] ATLAS-Collaboration, "The ATLAS experiment at the CERN LHC," *Journal of Instrumentation*, Vol. 3, S08003, 2008.

[87] M. Price, "The LHC project," *Nuclear Instruments and Methods in Physics Research*, pp. 46–61, 2002.

[88] A. Anjos, R. Torres, J. Seixas, B. Ferreira, and T. Xavier, "Neural triggering system operating on high resolution calorimetry information," *Nuclear Instruments and Methods in Physics Research*, no. 559, pp. 134-138, 2006.

[89] R. Wigmans. *Calorimetry: Energy Measurement in ParticlePhysics*. Clarendon Press, 2000.

[90] H. L. V. Trees, *Detection, Estimation, and Modulation Theory, Part I*. John Wiley and Sons, 2001.

[91] A. G. Mello, et al., "Overview of the high-level trigger electron and photon selection for the ATLAS experiment at the LHC," *IEEE Transactions on Nuclear Science*, vol. 53, no. 5, pp. 2839–2843, October 2006

[92] H. Lima Jr. and J. Seixas, “A segmented principal component analysis applied to calorimetry information at ATLAS,” *Nuclear Instruments and Methods in Physics Research*, no. 559, pp. 129–133, 2006.