

NONLINEAR BLIND SOURCE SEPARATION USING GENETIC ALGORITHMS

F.Rojas¹, I.Rojas¹, R.M.Clemente², C.G.Puntonet¹

(1) Dept. of Computer Architecture and Technology, University of Granada (Spain)

(2) Area de Teoria de la Señal y Comunicaciones, University of Sevilla (Spain)

ABSTRACT

This article proposes the fusion of two important paradigms, Genetic Algorithms and the Blind Separation of Sources (GABSS). Although the topic of BSS, by means of various techniques, including ICA, PCA and neural networks, has been amply discussed in the literature, to date the possibility of using genetic algorithms has not been explored. However, in nonlinear mixtures, optimisation of the system parameters and, especially, the search for invertible functions is very difficult by the existence of many local minima. From experimental results, this paper demonstrates the possible benefits offered by GAs in combination with BSS, such as robustness against local minima, the parallel search for various solutions, and a high degree of flexibility in the evaluation function.

1. INTRODUCTION

The separation of independent sources from mixed observed data is a fundamental and challenging signal processing problem [2],[7],[14]. In many practical situations, one or more desired signals need to be recovered from the mixtures. A typical example is the case of speech recordings made in the presence of background noise and/or competing speakers. The source separation problem has been successfully studied for linear instantaneous mixtures [1],[4],[12],[14] and more recently, since 1990, for linear convolutive mixtures [10],[19],[21].

Even though the nonlinear mixing model is more realistic and practical, most existing algorithms for the BSS problem were developed for the linear model. Nevertheless, the linear mixing model may not be appropriate for some real environments. However, for nonlinear mixing models, many difficulties occur and neither the linear ICA nor the existing linear demixing methodologies are any longer applicable because of the complexity of nonlinear parameters.

Therefore, researchers have recently started addressing the blind source separation problem to nonlinear mixing models [8],[13],[17],[18],[20]. In [9],[13] the nonlinear components are extracted by a model-free methods using Kohonen's self-organizing-feature-map (SOFM). This approach suffers from the exponential growth of network complexity and the interpolation error that arises in recovering continuous sources. Burel [3] proposed a nonlinear mixing model using a two-layer perceptron for blind source separation, trained by the classical gradient descent method to minimize mutual information.

In [8] a new set of learning rules for the nonlinear mixing models based on the information maximization criterion is proposed. The mixing model is divided into a linear mixing part and a nonlinear

transfer channel, in which the nonlinear functions are approximated by parametric sigmoidal or by higher order polynomials.

More recently, Yang et al. [20] developed an information backpropagation algorithm for Burel's model by the natural gradient method, using special nonlinear mixtures in which the nonlinearity could be approximated by a two-layer perceptron. Other important work in source separation in post-nonlinear mixtures is presented in [17]. In this methodology, the estimation of nonlinear functions and of score functions is done using a multilayer perceptron with sigmoidal units trained by unsupervised learning.

A possible extension of ICA to the separation sources in a nonlinear mixture is to employ a nonlinear function to transform the mixture such that the new outputs become statistically independent after the transformation. However, this transformation is not unique and without limiting the function class for de-mixing transforms, this extension may give statistically independent output sources completely different from the original unknown sources. Although there exist many difficulties in this transformation, several nonlinear ICA algorithms have been proposed and developed .

Nevertheless, one of the greatest problems encountered with nonlinear mixing models is that the approximation of the nonlinear function by the various techniques (perceptron, sigmoidal, RBF, etc.) meets with a serious difficulty; there are many local minima in the search space of the solution parameters for adapting nonlinear functions. As shown in the experimental results section, output surface of a performance index based on the mutual information when the parameters of the nonlinear functions are modified presents multiple and severe local minima. Therefore, algorithms that are based on a gradient descent for the adaptation of these nonlinear-function parameters may become trapped within one such local minimum.

In this paper, we propose a new algorithm for the nonlinear mixing problem using flexible nonlinearities. This nonlinear function may be approximated by even n-th order polynomials. We develop an algorithm that makes use of the synergy between Genetic Algorithms and the Blind Separation of Sources (GABSS) for the optimization of the parameters that define the nonlinear functions. Simultaneously, a natural gradient descent method is applied to obtain the linear demixing matrix. Unlike many classical optimization techniques, GAs do not rely on computing local first or second order derivatives to guide the search algorithm; GAs is a more general and flexible method that is capable of searching wide solution spaces and avoiding local minima (i.e. it provides more possibilities of finding an optimal or near-optimal solution). GAs deal simultaneously with multiple

solutions, not a single solution, and also include random elements, which help to avoid getting trapped in sub-optimal solutions.

2. SYSTEMS OF NONLINEAR MIXTURES

The task of blind signal separation (BSS) is that of recovering unknown source signals from sensor signals described by:

$$x(t) = F[As(t)] \quad (1)$$

where $x(t)=[x_1, x_2, \dots, x_n]^T$ is an available $n \times 1$ sensor vector, $s(t)=[s_1, s_2, \dots, s_n]^T$ is a $n \times 1$ unknown source vector having stochastic independent and zero-mean non-Gaussian elements $s_i(t)$, A is a $n \times n$ unknown full-rank and non singular mixing matrix, and $F=[f_1, f_2, \dots, f_n]^T$ are the set of invertible nonlinear transfer functions. The BSS problem consists in recovering the source vector $s(t)$ using only the observed data $x(t)$, the assumption of independence between the entries of the input vector $s(t)$ and possibly some a priori information about the probability distribution of the inputs. Statistical independence means that given one of the source signals, nothing can be estimated or predicted about any other source signal. If all the functions f_i are linear, (1) reduces to the linear mixing model. Even though the dimensions of x and s generally need not be equal, we make this assumption here for simplicity.

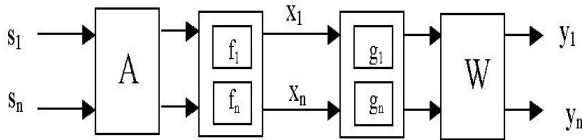


Figure 1: Nonlinear mixing and unmixing model.

Figure 1 shows that the mixing system is divided into two different phases: first, a linear mixing and second, for each channel i , a nonlinear transfer part. The unmixing system is the inverse; first we need to approximate the invert of the nonlinear function in each channel g_i , and then unmix the linear mixing by applying W to the output of the g_i nonlinear function.

$$y_i(t) = \sum_{j=1}^n w_{ij} g_j(x_j(t)) \quad (2)$$

In various approaches, the inverse function g_j is approximated by a sigmoidal transfer function, but because in certain situations where the a priori knowledge about the mixing model is not given by the human expert, a more flexible nonlinear transfer function based on a p -th order polynomial (odd polynomial) is used:

$$g_j(x_j) = \sum_{k=1}^p g_{jk} x_j^{2k-1} \quad (3)$$

Where $\vec{g}_j = [g_{j1}, \dots, g_{jP}]$ is a parameter vector to be determined. In this way, the output sources are calculated as:

$$y_i = \sum_{j=1}^n w_{ij} \sum_{k=1}^p g_{jk} x_j^{2k-1} \quad (4)$$

However, computation of the parameter vector \vec{g}_j is not easy, as it presents a problem with numerous local minima. Thus we require an algorithm that is capable of avoiding entrapment in such a minimum. As a solution to this first unmixing stage, we propose the hybridization of Genetic Algorithms. We have just used new meta-heuristics, as simulated annealing and genetic algorithms for the linear case [5],[15],[16], but in this paper we will focused in a more difficult problem as is the nonlinear ICA.

3. GENETIC ALGORITHMS AND BLIND SEPARATION IN NONLINEAR MIXTURES

3.1 Genetic Algorithms

GAs are currently one of the most popular stochastic optimisation techniques. They are inspired by natural genetics and the biological evolutionary process, and can be characterized by the following features [6],[11]:

- A scheme for encoding solutions to a problem in the form of a chromosome (chromosomal representation).
- An evaluation function which indicates the *fitness* of each chromosome relative to the others in the current set of chromosomes (referred to as population).
- An initialisation procedure for the population of chromosomes.
- Genetic operators which are used to manipulate the composition of the population.
- A set of parameters that provide the initial settings for the algorithm: the population size and probabilities employed by the genetic operators.

The GA evaluates a given population and generates a new one iteratively, with each successive population referred to as a generation. Given the current generation at iteration t , $G(t)$, the GA generates a new generation, $G(t+1)$, based on the previous generation, applying a set of genetic operations. The GA uses three basic operators to manipulate the genetic composition of a population: reproduction, crossover and mutation [6]. Reproduction consists in copying chromosomes according to their objective function (strings with higher evaluations will have more chances to survive). The crossover operator mixes the genes of two chromosomes selected in the phase of reproduction, in order to combine the features, especially their positive ones. Mutation is occasional, producing with low probability an alteration of some of the gene values in a chromosome (for example, in binary representation a 1 is changed into a 0 or vice versa). Mutation avoids the convergence to a population with a homogeneous gene pool and thus guarantees a certain variety of chromosomes.

Given an optimisation problem, simple GAs encode the parameters concerned into finite bit strings (or real numbers), and

then run iteratively using genetic operators in a random way but based on the fitness function evolution to perform the optimisation task. To perform the GA, it is first very important to define the fitness function (or contrast function in the BSS context). This fitness function is constructed bearing in mind that the output sources must be independent of their nonlinear mixtures. For this purpose, we must utilize a measure of the independence between random variables. Here, the mutual information is chosen as the measure of independence.

3.2 Fitness function

Evaluation functions of many forms can be used in a GA, subject to the minimal requirement that the function can map the population into a partially ordered set. As stated, the evaluation function is independent of the GA (i.e., stochastic decision rules). K-L e Mutual Information are stronger than cumulants or moments, since they use the definition of independence completely. The problem is to estimate the p.d.f. of the sources; the estimation of cumulants or moments is easier, more precise and faster than the estimation of p.d.f. Unfortunately, regarding the separation of a nonlinear mixture, independence alone is not sufficient to perform blind recovery of the original signals. Some knowledge of the moments of the sources, in addition to the independence, is required. An index similar to that proposed in [18] and [20], is used for the fitness function:

$$I = -\log|W| - \sum_{i=1}^n E \left[\sum_{k=1}^P (2k-1) g_{ik} x_i^{2k-2} \right] + \sum_{i=1}^n H(y_i) \quad (5)$$

In the above expression, the calculation of $H(y_i)$ needs to approximate each marginal pdf of the output source vector y_i , which are unknown. One useful method is the application of the Gram-Charlier expansion, which only needs some moments of y_i as suggested by Amari et al. [1] to express each marginal pdf of y_i as:

$$H(y_i) \approx \frac{\log(2\pi e)}{2} - \frac{(k_3^i)^2}{2 \cdot 3!} - \frac{(k_4^i)^2}{2 \cdot 4!} + \frac{3}{8} (k_3^i)^2 k_4^i + \frac{1}{16} (k_4^i)^3 \quad (6)$$

Where $k_3^i = m_3^i$, and $k_4^i = m_4^i - 3$.

3.3 Synergy between Genetic Algorithms and Natural Gradient Descent

Given a combination of weights obtained by the Genetic Algorithms for the nonlinear functions expressed as $G=[g_1, \dots, g_n]$, where the parameter vector that defines each function g_j is expressed by $\vec{g}_j = [g_{j1}, \dots, g_{jP}]$, it is necessary to learn the elements of the linear unmixing matrix W to obtain the output sources y_j . For this task, we use the natural gradient descent method to derive the learning equation for W as proposed in [20]:

$$\Delta W \propto \eta [I - \Phi(y) y^T] W \quad (7)$$

Where

$$\begin{aligned} \Phi(y) &= F_1(k_3, k_4) \circ y^2 + F_2(k_3, k_4) \circ y^3 \\ F_1(k_3, k_4) &= -\frac{1}{2} k_3 + \frac{9}{4} k_3 \cdot k_4 \\ F_2(k_3, k_4) &= -\frac{1}{6} k_4 + \frac{3}{2} k_3^2 + \frac{3}{4} k_4^2 \end{aligned} \quad (8)$$

And \circ denotes the Hadamard product of two vectors.

4. EXPERIMENTAL RESULTS

To provide an experimental demonstration of the validity of GABSS, we use a system of three sources. The independent sources are:

$$s(t) = \begin{bmatrix} \text{sign}(\cos(2\pi 155t)), \\ \sin(2\pi 300t + 6 \cos(2\pi 60t)), \text{rand}(t) \end{bmatrix} \quad (9)$$

Where $\text{rand}(t)$ is a random source uniformly distributed in $[-1, 1]$. These signals are first linearly mixed with a 3x3 mixture matrix

$$A = \begin{bmatrix} 0.6420 & 0.3016 & -0.3863 \\ -0.4347 & 0.82243 & -0.3150 \\ 0.3543 & 0.3589 & 0.942 \end{bmatrix} \quad (10)$$

The nonlinear distortions are selected as:

1. $f_1(x) = \text{Tanh}(x)$
2. $f_2(x) = 3/100 * x^5 - 33/200 x^3 + 4/5x$
3. $f_3(x) = 9/2500 * x^5 - 83/2000 x^3 + x/2$

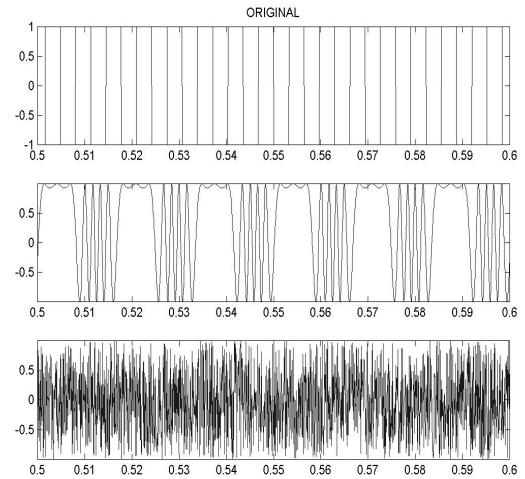


Figure 2: Original signals

The goal of the simulation was to analyse the behaviour of the GA and observe where the fitness function thus achieved is optimised; with this aim, therefore, we studied the mixing matrix obtained by the algorithm and the inverse function. When the number of generations reached a maximum value, the best individual from the population was selected and the estimated signals y were extracted, using the mixing matrix W , and the inverse function. Figure 2 represents the 1000 samples from the original signals. Figure 3 represents the mixed signals. As an example, Figure 4 shows the joint distribution of x_3 and x_2 mixed signal.

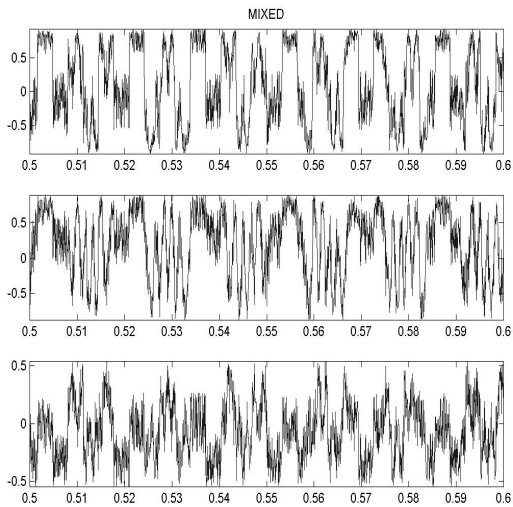


Figure 3: Mixed signal

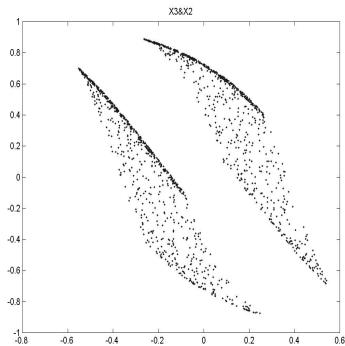


Figure 4: Detail example of a joint distribution of two mixed signal

Firstly, we demonstrate the existence of different local minima, for the simplest example in which only two parameters remain to be optimised. As an example, when the matrix W is equal to A^{-1} , and each of the functions g_j except one (for example, g_1) are optimally calculated, the surface of the error index is shown in Figure 5 (the index is normalized in the interval $[0,1]$). Even with only two parameters, the surface is bumpy and presents different local minima. The more parameters there are to be

optimised, the more complex is the hypersurface of the error index and the more local minima exist.

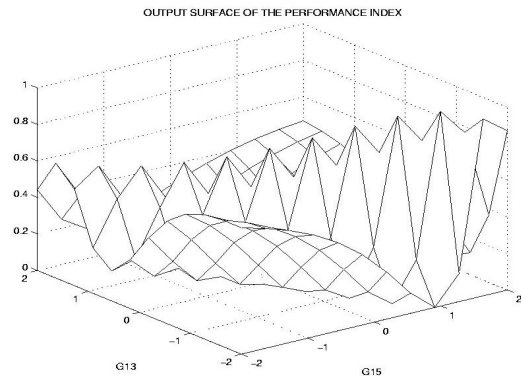


Figure 5: Output surface of the fitness function when all the parameters of the unmixing system are optimally calculated, except the two parameters of the function g_1

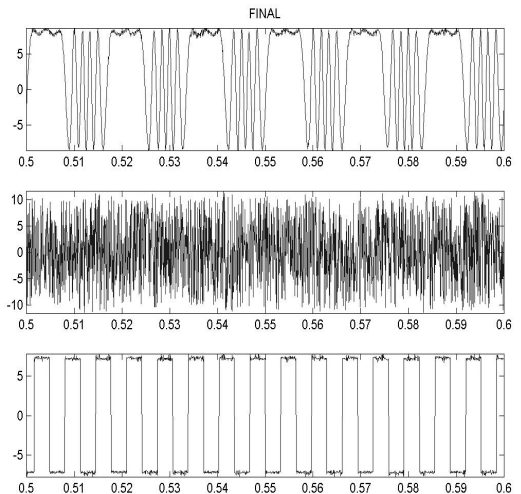


Figure 6: Obtained signal

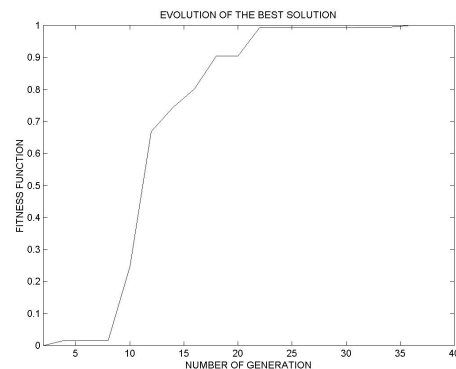


Figure 7: Evolution of the Best Fitness (Normalized fitness function)

Figure 6 shows the separated signal obtained with the proposed algorithm. Because these results are achieved directly by the presented methodology, there exist a different in the scaling factor in the order of the sources.

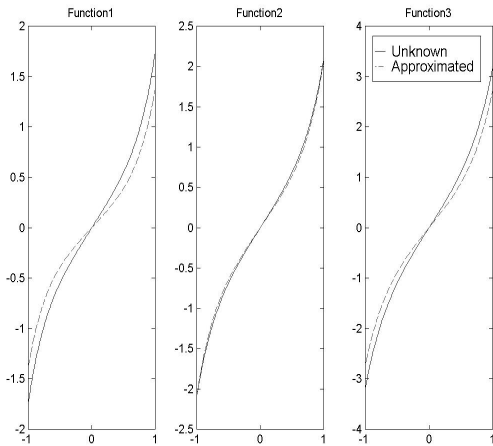


Figure 8: Comparison of the unknown f_i^{-1} and its approximation by g_i

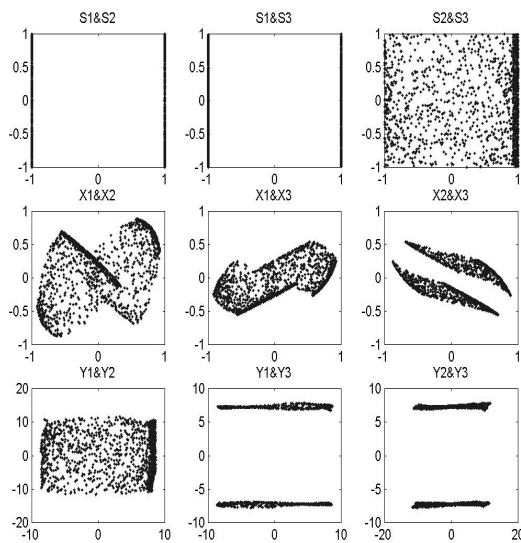


Figure 9: Representation of the joint distribution of the original (S), the mixed (X) and obtained (Y) sources (as there exists a permutation within the signals recovered, the joint representations are not ordered).

Figure 7 shows the development of the best fitness of the population within each of the generations, normalized in the interval $[0,1]$. It can be seen that, depending on the type of selection adopted and its parameters, the value of the optimum solution increases generation by generation until it stabilises at around the 35th generation. In the nonlinear transfer function g_i are presented, which successfully approximate the exact inverse

of f_i . Finally, Figure 9 gives the joint representation of the original, mixed and obtained signals. As there exists a permutation within the signals recovered, the joint representations are not ordered.

5. CONCLUSIONS

Many different approaches to the problem of blind separation of sources have been adopted by numerous researchers, using methods such as neural networks, artificial learning, higher order statistics, minimum mutual information, beam forming and adaptive noise cancellation, with various degrees of success being claimed. Despite the diversity of the approaches, the fundamental idea of the source signals being statistically independent remains the single most important assumption in most of these schemes. The neural network approach has the drawback that it may be trapped within local minima and therefore does not always guarantee optimal system performance. In blind source separation of nonlinear mixtures, the approximation of the nonlinear function by the various techniques (nonlinear ICA, neural network) meets with a serious difficulty; there are many local minima in the search space of the solution parameters for adapting nonlinear functions. For this reason, this article discusses a satisfactory application of genetic algorithms to the complex problem of the blind separation of sources. It is widely believed that the specific potential of Genetic or Evolutionary Algorithms originates from their parallel search by means of entire populations. In particular, the ability of escaping from local optima, an ability very unlikely to be observed in steepest-descent methods [6],[11]. Although to date, and to the best of the authors' knowledge, there is no mention in the literature of the synergy between BSS in nonlinear mixtures and GA, the article shows how GAs provide a tool that is perfectly valid as an approach to this problem.

Acknowledgement

This work has been partially supported by the CICYT Spanish TIC98-0982.

REFERENCES

- [1] S-I. Amari, A.Cichocki, H.Yang, "A New Learning Algorithm for Blind Signal Separation", In Advances in Neural Information Processing Systems, vol.8, 1996.
- [2] A. Bell & T.J. Sejnowski: An Information-Maximization Approach to Blind Separation and Blind Deconvolution Neural Computation 1129-59 (1995).
- [3] G.Burel, "Blind separation of sources: A nonlinear neural algorithm", Neural Networks, vol.5, pp.937-947, 1992.
- [4] J.F.Cardoso, "Source separation using higher order moments", in Proc. ICASSP, Glasgow, U.K. May 1989, pp.2109-2212.
- [5] R.M.Clemente, CGPuntonet, J.I.Acha. "A conjugate gradient method and simulated annealing for blind separation of sources". Lecture Notes in Computer Science, Springer-Verlag, (IWANN-2001), Granada (Spain), 13-15 June, 2001.

- [6] D.E. Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning", AddisonWesley, Reading, MA, 1989.
- [7] A.Hyvärinen and E.Oja, A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9 (7), pp.1483-1492, 1997.
- [8] T-W.Lee, B.Koehler, R.Orglmeister, "Blind separation of nonlinear mixing models", In *IEEE NNSP*, pp.406-415, Florida, USA, 1997.
- [9] J.K.Lin, D.G.Grier, J.D.Cowan, "Source separation and density estimation by faithful equivariant SOM", in *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 1997, vol.9.
- [10] A.Mansour, C.Jutten, P.Loubaton, "Subspace method for blind separation of sources in convolutive mixtures", in *Proc. EUSIPCO*, Trieste, Italy, Sept.1996, pp.2081-2084.
- [11] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag, New York USA, Third Edition, 1999.
- [12] A.V. Oppenheim, E. Weinstein, K.C. Zangi, M. Feder, and D. Gauger. Single-sensor active noise cancellation. *IEEE Trans. on speech and audio processing*, 2(2):285-290, April 1994.
- [13] P.Pajunen, A.Hyvarinen, J.Karhunen, "Nonlinear blind source separation by self-organizing maps", in *Progress in Neural Information Processing: Proc. IONIP'96*, vol.2. New York, 1996, pp.1207-1210.
- [14] C.G.Puntonet, A.Prieto, "Neural net approach for blind separation of sources based on geometric properties", *Neurocomputing*, 18 (3), 1998, pp.141-164.
- [15] C.Puntonet, A.Mansour. "Blind Separation of Sources using density estimation and simulated annealing" *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences*, vol."E84-A", 2001.
- [16] I.Rojas, C.G.Puntonet, A.Cañas, B.Pino, J.Fernández "Genetic Algorithms for the Blind Separation of Sources", *IASTED International Conference on Artificial Intelligence and Applications (AIA2001)*, September, 2001, Marbella (Spain).
- [17] A.Taleb, C.Jutten, "Source Separation in Post-Nonlinear Mixtures", *IEEE Transactions on Signal Processing*, vol.47m no.10, pp.2807-2820 1999.
- [18] Y.Tan, J.Wang, J.M.Zurada, "Nonlinear Blind Source Separation Using a Radial Basis Function Network", *IEEE Trans. On Neural Networks*, vol.12, no.1, pp.124-134, 2001.
- [19] H.L.N.Thi, C.Jutten, "Blind sources separation for convolutive mixtures", *Signal Processing*, vol.45, pp.209-229, 1995.
- [20] H.H.Yang, S.Amari, A.Chichocki, "Information-theoretic approach to blind separation of sources in non-linear mixture", *Signal Processing*, vol.64, 1998, 291-300.
- [21] D.Yellin, E.Weinstein, "Multichannel signal separation: Methods and analysis", *IEEE Trans. Signal Processing*, vol.44, pp.106-118, Jan. 1996.