

# COMPETTIVE LEARNING, SIMULATED ANNEALING AND GENETIC

## ALGORITHMS FOR THE SEPARATION OF SOURCES

C. G. PUNTONET, A. MANSOUR, F.ROJAS,  
Dpto. Arquitectura y Tec. Computadores  
University of Granada  
Spain

### ABSTRACT

This paper presents a new adaptive procedure for the linear and non-linear separation of signals with non-uniform, symmetrical probability distributions, based on both simulated annealing (SA) and competitive learning (CL) methods by means of a neural network, considering the properties of the vectorial spaces of sources and mixtures, and using a multiple linearization in the mixture space. Also, the paper proposes the fusion of two important paradigms, Genetic Algorithms and the Blind Separation of Sources in Nonlinear Mixtures (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 seriously explored. However, in Nonlinear Mixtures, optimization of the system parameters and, especially, the search for invertible functions is very difficult due to 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. The main characteristics of the method are its simplicity and the rapid convergence experimentally validated by the separation of many kinds of signals, such as speech or biomedical data.

**KEY WORDS:** Independent Component Analysis, Separation of sources, Simulated annealing, Genetic algorithms, Neural networks.

### 1. INTRODUCTION

Blind Source Separation (BSS) consists in recovering unobserved signals from a known set of mixtures. 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 only. A typical example is speech recordings made in an acoustic environment in the presence of background noise and/or competing speakers. This general case is known as the *Cocktail Party Effect*, in reference to human's brain faculty of focusing in one single voice and ignoring other voices/sounds, which are produced simultaneously with similar amplitude in a noisy environment. Spatial differences between the sources highly increase this capacity. 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],[17],[19]. Clearly explain the nature of the problem, previous work, purpose, and contribution of the paper. In the framework of independent component analysis, ICA, many kinds of approaches have been presented concerning the blind separation of sources, with applications to real problems in areas such as communications, feature extraction, pattern recognition, data visualization, speech processing and biomedical signal analysis (EEG, MEG, fMRI, etc), considering the hypothesis that the medium where the sources have been mixed is linear, convolutive or non-linear. ICA is a linear transformation that seeks to minimize the mutual information of the transformed data,  $e(t)$ , the fundamental assumption being that individual components of the source vector,  $x(t)$ , are mutually independent and have, at most, one Gaussian distribution. The 'Infomax' or independent component analysis algorithm of Bell and Sejnowski [2] is an unsupervised neural network learning algorithm that can perform blind separation of input data into the linear sum of time-varying modulations of maximally independent component maps, providing a powerful method for exploratory analysis of functional magnetic resonance imaging (fMRI) data. Also using the maximization of the negentropy, ICA 'Infomax' algorithm is used for unsupervised exploratory data analysis and for general linear ICA applied to electroencephalograph (EEG) monitor output. Many solutions for blind separation of sources are based on estimating a separation matrix with algorithms, adaptive or not, that use higher-order statistics, including

minimization or cancellation of independence criteria by means of cost functions or a set of equations, in order to find a separation matrix [10]. ICA is a promising tool for the exploratory analysis of biomedical data. In this context, a generalized algorithm modified by a kernel-based density estimation procedure has been studied to separate EEG signals from tumour patients into spatially independent source signals, the algorithm allowing artifactual signals to be removed from the EEG by isolating brain-related signals into single ICA components. Using an adaptive geometry-dependent ICA algorithm, Puntinet et al. [14] demonstrated the possibility of separating biomedical sources, such as EEG signals, after analyzing only the observed mixing space, due to the almost symmetric probability distribution of the mixtures.

The general case of a Non-linear mixture of sources is shown in the following figure:

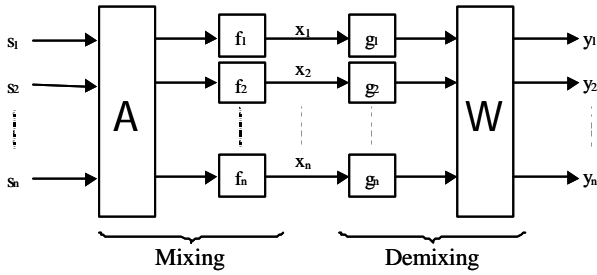


Fig. 1. Nonlinear mixing and demixing model.

## 2. HYBRIDATION OF CL, SA, AND GA.

### 2.1. SIMULATED ANNEALING

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

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

In different approaches, the inverse function  $g_j$  is approximated by a sigmoidal transfer function, but because of certain situations where the human expert does not give the a priori knowledge about the mixing model, a more flexible nonlinear transfer function based on even polynomial of  $P$ -th order is used:

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

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

$$y_i = \sum_{j=1}^n w_{ij} \sum_{k=1}^P g_{jk} x_j^{2k-1} \quad (3)$$

Nevertheless, computation of the parameter vector  $\overline{g_j}$  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 focus in a more difficult problem as is the nonlinear BSS. We propose an original method for independent component analysis and blind separation of sources that combines adaptive processing with a simulated annealing technique, and which is applied by normalizing the observed space,  $e(t)$ , in a set of concentric  $p$ -spheres in order to adaptively compute the slopes corresponding to the independent axes of the mixture distributions by means of an array of symmetrically distributed neurons in each dimension (Figure x). A preprocessing stage to normalize the observed space is followed by the processing or learning of the neurons, which estimate the high density regions in a way similar, but not identical to that of self organizing maps. A simulated annealing method provides a fast initial movement of the weights towards the independent components by generating random values of the weights and minimizing an energy function, this being a way of improving the performance by speeding up the convergence of the algorithm. The main process for competitive learning when a neuron approaches the density region, in a sphere  $\tilde{n}_k$  at time  $t$ , is given by:

$$w_i(\mathbf{r}_k, t+1) = w_i(\mathbf{r}_k, t) + \alpha(t) f(e(\mathbf{r}_k, t), w_i(\mathbf{r}_k, t)) \quad (4)$$

with  $\alpha(t)$  being a decreasing learning rate. Note that a variety of suitable functions,  $\alpha(t)$  and  $f()$ , can be used. In particular, a learning procedure that activates all the neurons at once is enabled by means of a factor,  $K_i(t)$ , that modulates competitive learning as in self-organizing systems, i.e.,

$$w_i(\mathbf{r}_k, t+1) = w_i(\mathbf{r}_k, t) + \alpha(t) \operatorname{sgn}[e(\mathbf{r}_k, t) - w_i(\mathbf{r}_k, t)] K_i(t) \quad (5)$$

$$K_i(t) = \exp(-h^1(t) \|w_i(\mathbf{r}_k, t) - w_i(\mathbf{r}_k, t)\|^2) \quad \forall i \in \{1, \dots, 2p\}$$

Simulated annealing is a stochastic algorithm that represents a fast solution to some combinatorial optimization problems. As an alternative to the competitive learning method described above, we first propose the use of stochastic learning, such as simulated annealing, in order to find a fast convergence of the weights around the maximum density points in the observation space  $e(t)$ . This technique is effective if the

chosen energy, or cost function,  $E_j$ , for the global system is appropriate. The procedure of simulated annealing is well known [16]. It is first necessary to generate random values of the weights and, secondly, to compute the associated energy of the system. This energy vanishes when the weights achieve a global minimum, the method thus allowing escape from local minima. For the problem of blind separation of sources we define an energy,  $E$ , related to the four-order statistics of the original  $p$  sources, due to the necessary hypothesis of statistical independence between them, as follows:

$$E = \sum_{i=1}^{p-1} \sum_{j=i+1}^p \langle cum_{22}^2(s_i(t), s_j(t)) \rangle \quad (6)$$

where  $cum_{22}^2(s_i(t), s_j(t))$  is the  $2 \times 2$  fourth-order cumulant of  $s_i(t)$  and  $s_j(t)$ , and  $\langle x(t) \rangle$  represents the expectation of  $x(t)$ . In spite of the fact that the technique presented in Section 2.2 is fast, the greater accuracy achieved by means of the competitive learning shown in Section 2.1 led us to consider a new approach. An alternative method for the adaptive computation of the  $\mathbf{W}\tilde{\mathbf{n}}_k$  matrix concerns the simultaneous use of the two methods, competitive learning and simulated annealing. Now, a proposed adaptive rule of the weights is the following:

$$\begin{aligned} W_{ijr_k}(t+1) = & W_{ijr_k}^s(t) \mathbf{b}(t) + \\ & + W_{ijr_k}^c(t) (1 - \mathbf{b}(t)) \end{aligned} \quad (7)$$

In Figure 2, we show the first simulation, that corresponds to the synthetic non-linear mixture presented by Lin and Cowan [7], for sharply peaked distributions, the original sources being digital 32-bit signals.

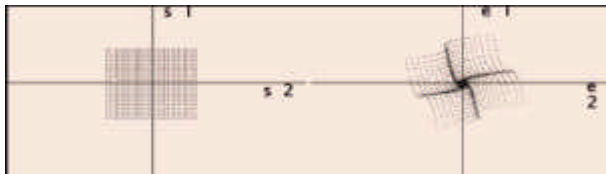


Figure 2. Non-linear mixture of  $p=2$  sources.

As shown in Figure 3, good estimation of the density distribution is obtained with 20000 samples, and using  $n=4$   $p$ -spheres ( $p=2$ ).

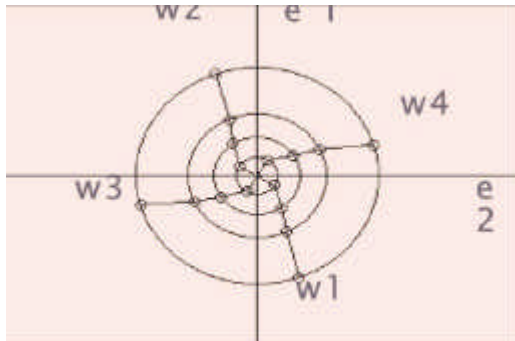


Figure 3. Network Estimation with SA and CL.

## 2.2 GENETIC ALGORITHMS

GAs are nowadays one of the most popular stochastic optimization techniques. They are inspired by the natural genetics and biological evolutionary process. The GA evaluates a 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 [5]. 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 the positive ones of them. Mutation is occasional; it produces with low probability, an alteration of some gene values in a chromosome (for example, in binary representation a 1 is changed into a 0 or vice versa).

To perform the GA, first is very important to define the fitness function (or contrast function in BSS context). This fitness function is constructed having in mind that the output sources must be independent from their nonlinear mixtures. For this purpose, we must utilize a measure of independence between random variables. Here, the mutual information is chosen as the measure of independence.

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). Unfortunately, regarding the separation of a nonlinear mixture, independence only 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. A similar index as proposed in [16] and [18], is used for the fitness function that approximates mutual information:

$$I(y) = -\log|W| - \sum_{i=1}^n E \left[ \sum_{k=1}^p (2k-1) g_{ik} x_i^{2k-2} \right] + \sum_{i=1}^n H(y_i) \quad (8)$$

Values near to zero of mutual information (8) between the  $y_i$  imply independence between those variables, being statically independent if  $I(y)=0$ .

In the above expression, the calculation of  $H(y_i)$  needs to approximate each marginal pdf of the output source vector  $y$ , 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  $y$  as:

$$H(y_i) \approx \frac{\log(2\pi e)}{2} - \frac{(k_3^i)^2}{2 \cdot 3!} - \frac{(k_4^i)^2}{2 \cdot 4!} + \frac{3}{8} (k_3^i)^2 k_4^i + \frac{1}{16} (k_4^i)^3 \quad (9)$$

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

The approximation of entropy (9) is only valid for uncorrelated random variables, being necessary to preprocess the mixed signals (*prewhitening*) before estimating its mutual information. Whitening or sphering of a mixture of signals consists in filtering the signals so that their covariances are zero (uncorrelatedness), their means are zero, and their variances equal unity.

The evaluation function that we compute will be the inverse of mutual information in (8), so that the objective of the GA will be maximizing the following function in order to increase statistical independence between variables:

$$eval\_function(y) = \frac{1}{I(y)} \quad (10)$$

There is a 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  $\bar{g}_j = [g_{j1}, \dots, g_{jp}]$ , 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 [18]:

$$\Delta W \propto h[I - \Phi(y)y^T]W \quad (11)$$

Where

$$\Phi(y) = F_1(k_3, k_4) \circ y^2 + F_2(k_3, k_4) \circ y^3$$

$$F_1(k_3, k_4) = -\frac{1}{2}k_3 + \frac{9}{4}k_3 \cdot k_4 \quad (12)$$

$$F_2(k_3, k_4) = -\frac{1}{6}k_4 + \frac{3}{2}k_3^2 + \frac{3}{4}k_4^2$$

And  $\circ$  denotes the Hadamard product of two vectors.

The typical genetic operators are crossover and mutation, that will be used for the manipulation of the current population in each iteration of the GA. The crossover operator is "Simple One-point Crossover". The mutation operator is "Non-Uniform Mutation" [11]. This operator presents the advantage when compared to the classical uniform mutation operator, of performing less significant changes to the genes of the chromosome as the number of generations grows. This property makes the exploration-exploitation trade-off be more favorable to exploration in the early stages of the algorithm, while exploitation takes more importance when the solution given by the GA is closer to the optimal solution.

### 3. SIMULATION RESULTS.

To provide an experimental demonstration of the validity of GABSS, we will use a system of three sources. Two of the sources are sinusoidal, while the third is a random

signal, uniformly distributed in  $[1, 1]$  (uniform noise). The independent sources are:

$$s(t) = \begin{bmatrix} \sin(2\mathbf{p} \cdot 30t + 6 \cdot \cos(2\mathbf{p} \cdot 6t)), \\ \text{sign}[\sin(2\mathbf{p} \cdot 20t)], \\ \text{rand}(t) \end{bmatrix} \quad (13)$$

These signals are first linearly mixed with a 3x3 mixture matrix

$$A = \begin{bmatrix} 0.6420 & 0.5016 & 0.4863 \\ 0.3347 & 0.82243 & -0.6150 \\ 0.3543 & -0.3589 & 0.542 \end{bmatrix} \quad (14)$$

The nonlinear distortion are selected as:

1.  $f_1(x) = \text{Tanh}(x)$
2.  $f_2(x) = \text{Tanh}(0.8x)$
3.  $f_3(x) = \text{Tanh}(0.5x)$

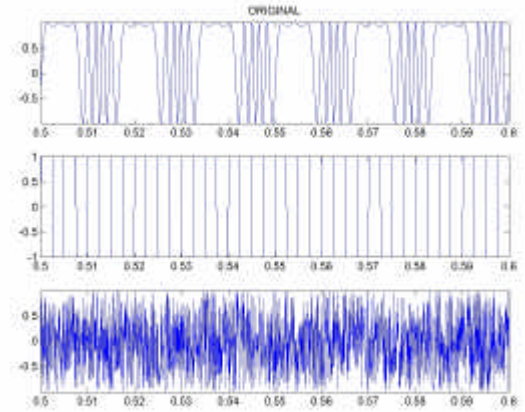


Fig. 4. Original signals

The goal of the simulation was to analyse the behaviour of the GA and observe whether 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  $u$  were extracted, using the mixing matrix  $W$ , and the inverse function.

Figure 4 represents the 1000 samples from the original signals.

Figure 5 represents the mixed signals.

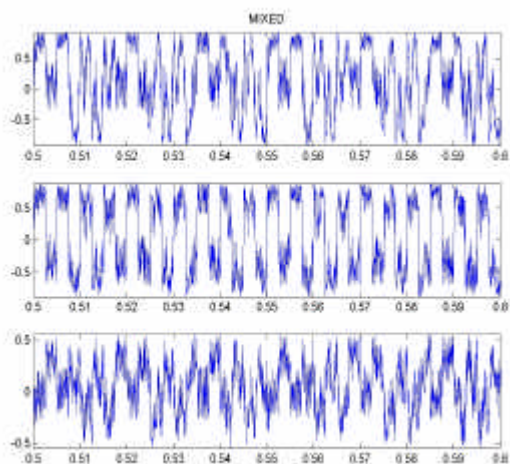


Fig. 5. Mixed signals

Figure 6 shows the separated signals obtained with the proposed algorithm. As it can be seen signals are very similar to the original ones, up to possible scaling factors and permutations of the sources.

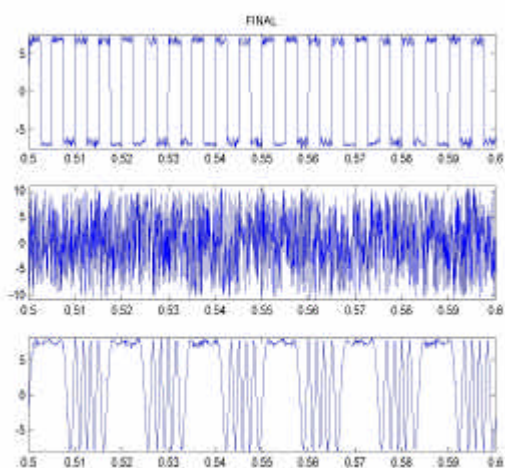


Fig. 6. Obtained signals

Figure 7 compares the approximation of the functions  $g_i$  to the inverse of  $f_i$ .

Figure 8 shows the joint representation of the original, mixed and obtained signals.

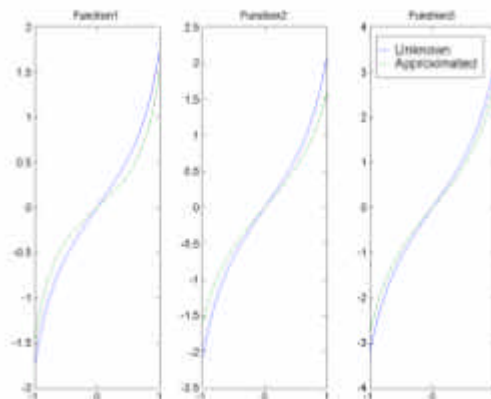


Fig. 7. Comparison of the unknown  $f_i^{-1}$  and its approximation by  $g_i$ .

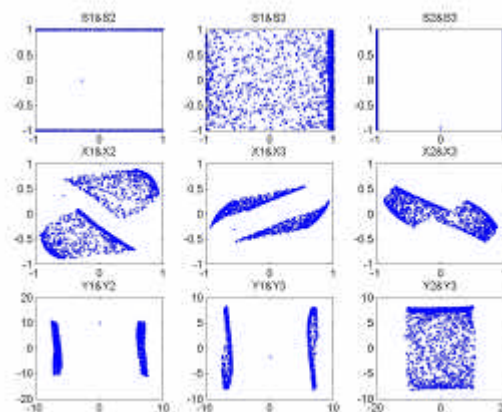


Fig. 8. Representation of the joint distribution of the original (S), mixed (X), and obtained (Y) signals.

In this practical application, the population size was  $population\ size = 20$  and the number of generations was  $generation\ number = 40$ . Regarding genetic operators parameters, crossover probability per chromosome was  $p_c = 0.8$  and mutation probability per gene was  $p_m = 0.01$ . As a special parameter for the non-uniform mutation operator  $b = 5$ .

#### 4. CONCLUSION

We have shown a new, powerful adaptive-geometric method based on competitive unsupervised learning and simulated annealing, which finds the distribution axes of the observed signals or independent components by means of a piecewise linearization in the mixture space, the use of simulated annealing in the optimization of a four-order statistical criterion being an experimental advance. The algorithm, in its current form, presents some drawbacks concerning the application of simulated annealing to a high number,  $p$ , of signals, and the complexity of the procedure  $O(2^p p^2 n)$  for the separation

of nonlinear mixtures, that also depends on the number,  $n$ , of  $p$ -spheres. In spite of these questions that remain open, the time convergence of the network is fast, even for more than two subgaussian or supergaussian signals, mainly due to the initial simulated annealing process that provides a good starting point with a low computation cost, and the accuracy of the network is adequate for the separation task, the competitive learning being very precise, as several experiments have corroborated.

Besides the study of noise, future work will concern the application of this method to independent component analysis with linear and nonlinear mixtures of biomedical signals, such as in Electroencephalograph and functional Magnetic Resonance Imaging, where the number of signals increases sharply, making simulated annealing suitable in a quantized high-dimensional space.

Many different approaches to the blind separation of sources problem have been adopted by numerous researchers, using 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 into local minima and therefore it does not always guarantee optimal system performance.

This article discusses, also, 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 is an ability very unlikely to be observed in steepest-descent methods. Although to date, and to the best of the authors' knowledge, there is no mention in the literature of this synergy between GAs and BSS in nonlinear mixtures, the article shows how GAs provide a tool that is perfectly valid as an approach to this problem.

## 5. ACKNOWLEDGEMENT

This work has been partially supported by the CICYT Spanish Project TIC2000-1348 and TIC2001-2845.

## 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] D.E. Goldberg, "*Genetic Algorithms in Search, Optimization and Machine Learning*", Addison-Wesley, Reading, MA, 1989.
- [6] Hyvärinen, J. Karhunen y E.Oja, *Independent Component Analysis*. J. Wiley-Interscience. 2001.
- [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.Orgmeister, "Blind separation of nonlinear mixing models", In *IEEE NNISP*, 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.Hyvärinen, 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] A.Taleb, C.Jutten, "Source Separation in Post-Nonlinear Mixtures", *IEEE Transactions on Signal Processing*, vol.47 no.10, pp.2807-2820 1999.
- [16] 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.
- [17] H.L.N.Thi, C.Jutten, "Blind sources separation for convolutive mixtures", *Signal Process.*, vol.45, pp.209-229, 1995.
- [18] 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.
- [19] D.Yellin, E.Weinstein, "Multichannel signal separation: Methods and analysis", *IEEE Trans. Signal Processing*, vol.44, pp.106-118, Jan. 1996.