Abstract — This 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, the possibility of using genetic algorithms has not been explored thus far. 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.


I. 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 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].

A. Independent Component Analysis.

The blind separation of sources problem can be approached by a wider point of view by using Independent Component Analysis (ICA) [6]. ICA appeared as a generalization of the popular statistical technique Principal Component Analysis (PCA).

ICA goal is finding a linear transformation given by a matrix \( W \), so that the random variables \( y_i, i=1,...,n \) are as independent as possible in (1):

\[
y_i(t) = W \cdot x_i(t)
\]

B. Nonlinear BSS.

Nevertheless, the linear mixing model may not be appropriated for some real environment. Even though the nonlinear mixing model is more realistic and practical, most existing algorithms for the BSS problem were developed for the linear model. However, for nonlinear mixing models, many difficulties occur and both the linear ICA and the existing linear demixing methodologies are no longer applicable because the complexity of nonlinear parameters.

Therefore, researchers have recently started addressing the blind source separation problem to nonlinear mixing models[8],[13],[15],[16],[18]. In [9],[13] the nonlinear components are extracted by a model-free method, which used Kohonen’s self-organizing-feature-map (SOFM), but suffers from the exponential growth of network complexity and interpolation error in recovering continuous sources. Burel [3] proposed a nonlinear mixing model using a two-layer perceptron to the blind source separation problem, which is trained by the classical gradient descent method to minimize the mutual information.

In [8] a new set of learning rules for the nonlinear mixing models based on the information maximization criterion are 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. [18] developed and information backpropagation algorithm for Burel’s model by natural gradient method, using special nonlinear mixtures in which the nonlinearity could be approximated by a two-layer perceptron. Other important work in the source separation in post-nonlinear mixture is presented in [15]. In this methodology, the estimation of nonlinear functions and score functions is done using a multilayer perceptron with sigmoidal units trained by unsupervised learning.
In addition, an extension of ICA to the separation sources in 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 demixing transforms, this extension may give statistically independent output sources completely different to the original unknown sources. Although, to the best of our knowledge, there exist today 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 parameters solutions for adapting nonlinear functions. Output surface of a performance index based on the mutual information when the parameters of the nonlinear function 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 work, a new algorithm for the nonlinear mixing problem using flexible nonlinearities is proposed. This nonlinear function may be approximated by odd polynomials of n-th order. An algorithm that makes use of the synergy between Genetic Algorithms and the Blind Separation of Sources (GABSS) was developed 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.

II. NONLINEAR MIXTURE MODEL AND DEMIXING SYSTEM.

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

\[ x(t) = F[As(t)] \]  

(2)

where \( x(t)=[x_1,x_2,...,x_n]^T \) is an available \( nx1 \) sensor vector, \( s(t)=[s_1,s_2,...,s_n]^T \) is an \( nx1 \) unknown source vector having stochastic independent and zero-mean non-Gaussian elements \( s(t) \), A is a \( nxm \) unknown full-rank and non-singular mixing matrix, and \( F=[f_1,f_2,...,f_n]^T \) are the set of invertible nonlinear transfer function. 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 knowledge 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, (2) 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.

![Fig. 1. Nonlinear mixing and demixing model.](image)

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 \( g_i \) nonlinear function.

\[ y_j(t) = \sum_{j=1}^{n} w_{jj} g_j(x_j(t)) \]  

(3)

In different approaches, the inverse function \( g_i \) 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:

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

(4)

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

\[ y_j = \sum_{j=1}^{n} w_{jj} \sum_{k=1}^{P} g_{jk} x_j^{2k-1} \]  

(5)

Nevertheless, computation of the parameter vector \( g_j \) 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. For the nonlinear case, we have applied only genetic algorithms at this moment. Simulated annealing may lead also to good results. Nevertheless, in our opinion, the search space for the nonlinear situation is too wide for this technique to be employed and genetic algorithms have a higher capacity of parallelization.

III. APPLYING GENETIC ALGORITHMS TO NONLINEAR BSS

A. 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).

B. Evaluation Function based on Mutual Information.

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 prewhitening, 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 (6), so that the objective of the GA will be maximizing the following function in order to increase statistical independence between variables:

\[
\text{eval-function}(y) = \frac{1}{I(y)}
\]

C. Synergy between Genetic Algorithms and Natural Gradient descent

Given a combination of weights obtained by the genetic algorithms for the nonlinear functions expressed as \( G = [g_1, \ldots, g_n] \), where the parameter vector that defines each function \( g_j \) is expressed by \( g_j = [g_{j1}, \ldots, g_{jn}] \), it is necessary to learn the elements of the linear unmixing matrix \( W \) to obtain the output sources \( y \). For this task, we use the natural gradient descent method to derive the learning equation for \( W \) as proposed in [16] [18]:

\[
\Delta W \propto \eta \left[ I - \Phi(y)y^T \right] W
\]

Where

\[
\Phi(y) = F_1(k_1, k_4) \circ y^2 + F_2(k_3, k_4) \circ y^3
\]

\[
F_1(k_1, k_4) = -\frac{1}{2}k_1 + \frac{9}{4}k_1k_4
\]

\[
F_2(k_3, k_4) = -\frac{1}{6}k_1 + \frac{3}{2}k_3^2 + \frac{3}{4}k_4^2
\]

And \( \circ \) denotes the Hadamard product of two vectors.
D. Genetic Operators.

Typical crossover and mutation operators 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.

IV. SIMULATION RESULTS.

A. Execution of the algorithm.

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:

\[
s(t) = \begin{bmatrix} \sin(2\pi \cdot 300t + 6 \cdot \cos(2\pi \cdot 60t)) \\ \text{sign} \left( \sin(2\pi \cdot 200t) \right) \\ \text{rand}(t) \end{bmatrix}
\]  

(11)

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

\[
A = \begin{bmatrix} 0.6420 & 0.5016 & 0.4863 \\ 0.3347 & 0.82243 & -0.6150 \\ 0.3543 & -0.3589 & 0.542 \end{bmatrix}
\]  

(12)

The nonlinear distortion are selected as:

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

The goal of the simulation was to analyze the behavior of the GA and observe whether the fitness function thus achieved is optimized; 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 \( \hat{u} \) 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.

Figure 4 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.

Figure 5 compares the approximation of the functions \( g_i \) to the inverse of \( f_i \). Finally, Figure 6 shows the joint representation of the original, mixed and obtained signals.

Fig. 2. Original signals

Fig. 3. Mixed signals

Fig. 4. Obtained signals

Fig. 5. Comparison of the unknown \( f_i \) and its approximation by \( g_i \).
Fig. 6. Representation of the joint distribution of the original (S), mixed (X), and obtained (Y) signals.

B. Parameters of the GA.

In this practical application, the population size was $\text{population}_{\text{size}} = 20$ and the number of generations was $\text{generation}_{\text{number}} = 40$. Regarding genetic operators parameters, crossover probability per chromosome was $p_c = 0.8$ and mutation probability per gene was $p_m = 0.01$. These values were selected because of their better performance when compared with other combinations that were evaluated as well.

V. CONCLUSIONS

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 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.

Acknowledgements:
This work has been partially supported by the CICYT Spanish Projects TIC2001-2845, DPI2001-3219 and TIC2000-1348.

VI. REFERENCES