Analysis of contrast functions in a genetic algorithm for post-nonlinear blind source separation

Fernando Rojas Ruiz, Carlos G. Puntonet and Ignacio Rojas Ruiz

Departamento de Arquitectura y Tecnología de Computadores. E.T.S.I. Informática. University of Granada, 18071, Spain.

Abstract. The task of recovering a set of unknown sources from another set of mixtures directly observable and little more information about the way they were mixed is called the blind source separation problem. If the assumption in order to obtain the original sources is their statistical independence, then ICA (Independent Component Analysis) maybe the technique to recover the signals. In this contribution, we propose and analyze three evaluation functions (contrast functions in Independent Component Analysis terminology) for the use in a genetic algorithm (PNL-GABSS, *Post-NonLinear Genetic Algorithm for Blind Source Separation*) which solves source separation in nonlinear mixtures, assuming the post-nonlinear mixture model. A thorough analysis of the performance of the chosen contrast functions is made by means of ANOVA (*Analysis of Variance*), showing the validity of the three approaches.

1   Introduction

The problem of *Blind Separation of Sources, BSS* is that of recovering a set of non-observable source signals from another set of related and observable "mixtures", without having any *a priori* knowledge neither about the sources, nor the way they were mixed. In many real world 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. If the assumption in order to obtain the original sources is their statistical independence, then ICA (*Independent Component Analysis*) can be applied. ICA is a method for finding underlying factors or components from multidimensional or multivariate statistical data [1].

This technique deals with the problem of transforming a set of observation patterns *x*, whose components are not statistically independent from one another, into a set of patterns $y = F(x)$ whose components are statistically independent from each other. In linear ICA, which is the most extensively studied case, the transformation F is restricted to being linear. Nonlinear ICA allows F to be nonlinear.

The nonlinear separation of sources has been addressed in [2], [3] and [4]. However, nonlinear ICA, is rather unconstrained, and normally demands additional information to make the estimations coincide with the original sources. Various dependence measures have been proposed for this problem: Burel [2] proposes a

quadratic "error" between probability densities function, Marques and Almeida [5] suggest a contrast function based on moments of all orders. Finally, other authors (Taleb and Jutten [6], Rojas et al [7]) proposed several contrast functions which approximate the mutual information of the estimated components.

In this contribution, we propose the use of a genetic algorithm to solve post-nonlinear blind source separation. Neural network approaches have the drawback of possibly being trapped into near-optimal solutions in situations where the search space presents many local minima. As an alternative, genetic algorithms deal simultaneously with multiple solutions, not a single solution, and also include random elements, which help to avoid getting trapped into sub-optimal solutions.

2   The Post-Non-Linear Mixture Model

A fundamental difficulty in nonlinear ICA is that is highly nonunique without some extra constraints. Therefore finding independent components does not lead us necessarily to the original sources [4]. Blind source separation in the nonlinear case is, in general, impossible. Taleb and Jutten [6], added some extra constraints to the nonlinear mixture so that the nonlinearities are independently applied in each channel after a linear mixture. In this way, the indeterminacies are the same as for the basic linear instantaneous mixing model: invertible scalings and permutations.

The mixture model can be described by the following equation:

$$x(t) = f(A \cdot s(t)) . \tag{1}$$

The unmixing stage, which will be performed by the algorithm here proposed is expressed by Equation (2):

$$y(t) = W \cdot g(x(t)) \tag{2}$$

3   Genetic Algorithm for Blind Source Separation

Genetic Algorithms (GAs) are nowadays one of the most popular stochastic optimization techniques. 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 proposed algorithm will be based on the estimation of mutual information, value which cancels out when the signals involved are independent. Mutual information between the elements of a multidimensional variable y is defined as:

$$I(y_1, y_2,...,y_n) = \sum_{i=1}^{n} H(y_i) - H(y_1, y_2,...,y_n) . \tag{3}$$

In order to exactly compute mutual information, we need also to calculate entropies, which likewise require to know the analytical expression of the probability density function (PDF) which is generally not available in practical applications of speech processing. Thus, we propose several evaluation functions (or contrast function in BSS terminology) that approximate mutual information. A *contrast*

function, $\Psi(\cdot)$, is any non-linear function which is invariant to permutation and scaling matrices, and attains its minimum value in correspondence of the mutual independence among the output components.

Independently of the selected contrast function, the operation of the basic genetic algorithm is invariant, needing the following features to be completely characterized:

- **Encoding Scheme.** The genes will represent the coefficients of the odd polynomials which approximate the family of nonlinearities *g* (see Eq. (2)). The linear part will be approximated by a well-known method such as JADE [8].
- **Initialization Procedure.** The polynomial coefficients which form part of the chromosome are randomly initialized.
- **Fitness Function.** The key point in the performance of a GA is the definition of the fitness function. In this case, the fitness function that we want maximize will be precisely the inverse of the approximation of mutual information given in Eq. (3):

$$Fitness(y) = \frac{1}{I(y)} = \frac{1}{\sum_{i=1}^{p} H(y_i) - H(y_1, y_2, ..., y_p)} \tag{4}$$

- **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" [10]. This operator 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.
- **Parameter Set.** Population size, number of generations, probability of mutation and crossover and other parameters relative to the genetic algorithm operation were chosen depending on the characteristics of the mixing problem.

In Figure 1 the flow chart for the genetic algorithm that solves blind source separation in post-nonlinear mixtures is shown.
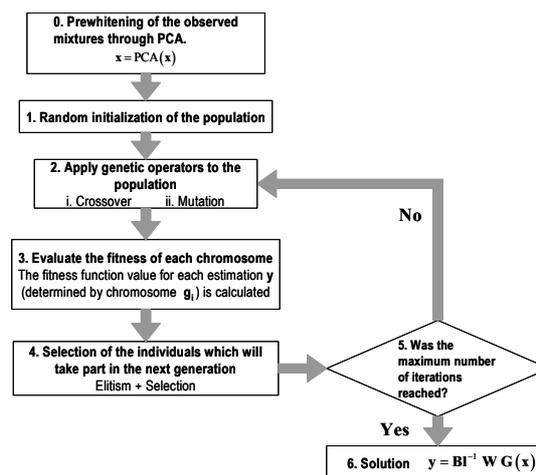
Fig. 1: Flow chart of the genetic algorithm for source separation of post-nonlinear mixtures (**Bl** is the whitening matrix).

4   Contrast Functions Proposed

In order to calculate the degree of independence of a set of variables precisely it is necessary to have the mathematical expression of their probability density functions. Actually, since that information is not typically available, is usual to utilize approaches of independence measures. In this research work, several functions have been evaluated, most of them based on the calculation of approaches of the mutual information or measures derived from it. One of the great advantages of genetic algorithms is its flexibility in the use of evaluation functions.

4.1   Cross-Cumulants Minimization.

It is possible to build a contrast function based on the simultaneous minimization of several cumulants, cross-cumulants 2-2, 3-1 and 1-3 to be precise. We can also add some penalization factors to the contrast functions, promoting solutions whose estimations are closed to zero mean and unit variance (C_1 and C_2, respectively).

$$\Psi_{CrossCumulants}(y_i, y_j) = \mathrm{cum}_{22}(y_i, y_j) + \mathrm{cum}_{31}(y_i, y_j) + \mathrm{cum}_{13}(y_i, y_j) + \alpha C_1 + \beta C_2 \qquad (5)$$

where α and β are weights (real numbers) for the penalization functions.

Unfortunately, this simple approach is only valid for separating two signals. Although it can be extended to three or more computing the contrast function by pairs, the computational cost would be exponentially increasing with the number of mixtures.

4.2   PDF Direct Approximation using histograms

We propose to approximate densities through the discretization of the estimated signals building histograms and then calculate their joint and marginal entropies. In this way, we define a number of bins *m* that covers the selected estimation space and then we calculate how many points of the signal fall in each of the bins ($B_i\ i = 1,...,m$). Finally, we easily approximate marginal entropies using the following formula:

$$H(y) = -\sum_{i=1}^{n} p(y_i) \log_2 p(y_i) \approx -\sum_{j=1}^{m} \frac{Card(B_j(y))}{n} \log_2 \frac{Card(B_j(y))}{n}. \qquad (6)$$

where *Card(B)* denotes cardinality of set B, *n* is the number of points of estimation *y*, and B_j is the set of points which fall in the jth bin.

The same method can be applied for computing the joint entropies of all the estimated signals:

$$H(y_1,...,y_p) = \sum_{i=1}^{p} H(y_i \mid y_{i-1},...,y_1) \approx -\sum_{i_1=1}^{m} \sum_{i_2=1}^{m} ... \sum_{i_n=1}^{m} \frac{\#B_{i_1 i_2...i_p}(y)}{n} \log_2 \frac{\#B_{i_1 i_2...i_p}(y)}{n}. \qquad (7)$$

where *p* is the number of components which need to be approximated.

Therefore, substituting entropies in Eq.(3) by approximations of Eqs. (6) and (7), we obtain a contrast function which will reach its minimum value when the estimations are independent.

4.3   Gram-Charlier Expansion for Mutual Information Approximation.

One useful method for mutual information approximation is the application of the Gram-Charlier expansion, which only needs some moments of $\mathbf{y}_i$ as suggested by Amari et al.[9] to express each marginal entropy of $\mathbf{y}$ as:

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

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

Applying Eq. (7) to the calculation of mutual information (Eq. (3)), we obtain a new contrast function:

$$\Psi_{IM-GramCharlier} = Est(I(\mathbf{y})) = Est\left(\sum_{i=1}^{n} H(y_i) - H(\mathbf{y})\right) =$$

$$= \sum_{i=1}^{n}\left[\frac{\log(2\pi e)}{2} - \frac{\left(k_3^i\right)^2}{2\cdot 3!} - \frac{\left(k_4^i\right)^2}{2\cdot 4!} + \frac{3}{8}\left(k_3^i\right)^2 k_4^i + \frac{1}{16}\left(k_4^i\right)^3\right] - \ldots \tag{9}$$

$$\ldots - \log\left|\det\left(\mathbf{W}\right)\right| - \sum_{i=1}^{n} E\left[\log\left|\sum_{k=1}^{P}(2k-1)p_{ik}x_i^{2k-2}\right|\right] - H(\mathbf{x})$$

5   Analysis of Variance of Experimental Results

Analysis of Variance (ANOVA) is a statistical method that yields values that can be tested to determine whether a significant relation exists between variables of interest. This method is widely used in industry to help identify the source of potential problems in the production process, and identify whether variation in measured output values is due to variability between various manufacturing processes, or within them.

In our case, we will use ANOVA in order to analyze the estimations in terms of their similarity to the original sources and testing which variables (factors) affect more to the performance of the proposed algorithm. It will be of special interest the influence of the contrast function chosen, so that we can determine which one works better in practice.

The simulation applies the following linear and non-linear transformation to a set of two voice signals of 10.165 samples each one:

$$A = \begin{bmatrix} 0.4891 & -0.1202 \\ -0.4641 & 0.8668 \end{bmatrix}, \ F = \begin{bmatrix} \tanh(x) \\ \tanh(x/2) \end{bmatrix} \tag{10}$$

In this case, ANOVA studies as the output variable the average *crosstalk* between estimations and the sources (lower the crosstalk, better the estimation). Table 1 briefly shows the different results obtained depending on the contrast function chosen for the genetic algorithm. As can be observed, each contrast function determines an homogenous group (groups of means within which there are no statistically significant differences) in the ANOVA.

From the results, it can be drawn that contrast function based on the Gram-Charlier expansion for mutual information approximation achieves better results. Also, from the ANOVA analysis is deduced that results also improve if the number of

samples increases, as it is normally expected. Other factors, as the crossover or mutation probabilities do not have a strong influence on the crosstalk results.

Contrast Function	Number of runs for each level	Average Crosstalk	Homogenous Groups		
$\Psi_{Histograms}$	1296	-11.3995 dB		X	
$\Psi_{CrossCumulants}$	1296	-10.201 dB			X
$\Psi_{IM-GramCharlier}$	1296	-15.6378 dB	X		

Table 1: Performance summary of the three contrast functions in the ANOVA.

6 Conclusions

This article discusses a satisfactory application of genetic algorithms to the complex problem of the blind separation of sources in post-nonlinear mixtures. Also, three different contrast functions are proposed in order to evaluate candidate solutions inside the genetic algorithm. The results are thoroughly analyzed by a powerful statistical tool as the ANOVA, showing that the function based on the mutual information approximation by means of the Gram-Charlier expansion achieves the best results in nonlinear source separation.

References

[1] A. Hyvärinen, J. Karhunen and E.Oja, *Independent Component Analysis*. J. Wiley-Interscience, New York, 2001.

[2] G.Burel, Blind separation of sources: A nonlinear neural algorithm, *Neural Networks*, vol.5, pp.937-947, 1992.

[3] T-W.Lee, B.Koehler and R.Orglmeister, Blind separation of nonlinear mixing models, In *IEEE NNSP*, pp.406-415, Florida, USA, 1997.

[4] A. Hyvärinen and P. Pajunen. Nonlinear Independent Component Analysis: Existence and Uniqueness results. *Neural Networks* 12(3): pp. 429-439, 1999.

[5] G. C. Marques and L. B. Almeida. An objective function for independence. In Proc. *International Conference on Neural Networks*, pp. 453–457, Washington DC, 1996.

[6] A.Taleb, C.Jutten, Source Separation in Post-Nonlinear Mixtures, *IEEE Transactions on Signal Processing*, vol.47 no.10, pp.2807-2820 (1999).

[7] F. Rojas, C.G. Puntonet, M. Rodríguez, I. Rojas, R. Martín-Clemente. Blind Source Separation in Post-Nonlinear Mixtures Using Competitive Learning, Simulated Annealing and a Genetic Algorithm. *IEEE Transactions on Systems, Man and Cybernetics*, Part C, Volumen: 34 (4), pp. 407 - 416, November 2004.

[8] J.F. Cardoso, A. Souloumiac, Blind beamforming for non Gaussian signals, *IEEE Proceedings-F*, vol. 140, no. 6, pp. 362-370, December 1993

[9] 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.

[10] Z. Michalewicz, Genetic Algorithms + Data Structures = Evolution Programs, Springer-Verlag, New York USA, Third Edition, (1999).