A Canonical Genetic Algorithm for Blind Inversion of Linear Channels

Fernando Rojas[1], Jordi Solé-Casals[2], Enric Monte-Moreno[3], Carlos G. Puntonet[1]
and Alberto Prieto[1]

[1] Computer Architecture and Technology Department, University of Granada, 18071, Spain
{frojas, carlos}@atc.ugr.es, aprieto@ugr.es
http://www.atc.ugr.es/
[2] Signal Processing Group, University of Vic (Spain)
jordi.sole@uvic.es
http://www.uvic.es/
[3] TALP Research Center, Polytechnic University of Catalonia (Spain)
enric@gps.tcs.upc.es
http://gps-tsc.upc.es/veu/personal/enric/enric.html

Abstract. It is well known the relationship between source separation and blind deconvolution: If a filtered version of an unknown i.i.d. signal is observed, temporal independence between samples can be used to retrieve the original signal, in the same manner as spatial independence is used for source separation. In this paper we propose the use of a Genetic Algorithm (GA) to blindly invert linear channels. The use of GA is justified in the case of small number of samples, where other gradient-like methods fails because of poor estimation of statistics.

1 Introduction

The problem of source separation may be formulated as the recovering of a set of unknown independent signals from the observations of mixtures of them without any prior information about either the sources or the mixture [1, 2]. The strategy used in this kind of problems is based on obtaining signals as independent as possible at the output of the system. In the bibliography multiple algorithms are proposed for solving the problem of source separation in instantaneous linear mixtures, from neural networks based methods [3], cumulants or moments methods [4, 5], geometric methods [6] or information theoretic methods [7]. In real world situations, however, the majority of mixtures can not be modeled as instantaneous and/or linear. This is the case of convolutive mixtures, where the effect of channel from source to sensor is modeled by a filter [8].

A particular case of blind separation is the case of blind deconvolution, which is presented in figure 1. Development of this framework is presented in the following section.

input output
s(t) e(t) y(t)
h w

unknown observation
(convolution system) (deconvolution system)
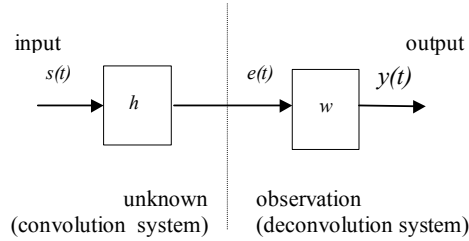
Fig. 1. Block diagram of the convolution system and blind deconvolution system. Both filter h and signal s(t) on the convolution process are unknown.

This paper proposes the use of Genetic Algorithms (GA) for the blind inversion of the (linear) channel. The theoretic framework for using source separation techniques in the case of blind deconvolution is presented in [9]. There, a quasi-nonparametric gradient approach is used, minimizing the mutual information of the output as a cost function to deal with the problem. A parametric approach can be found in [10]. The aim of the paper is to present a different optimization procedure to solve the problem, even if a small number of samples is available. In this case, gradient-like algorithms fail because of poor estimation of statistics. Our method is shown to be useful in this case, where other methods can not be used. This paper is organized as follows. Section 2 describes the linear model and presents the basic equations. Section 3 explains the Genetic Algorithm for blind deconvolution. Finally, section 4 presents the experiments showing the performance of the method.

2 Model and system equations

2.1 Model

We suppose that the input of the system $S=\{s(t)\}$ is an unknown non-Gaussian independent and identically distributed (i.i.d.) process, and that subsystem h is a linear filter, unknown and invertible. We would like to estimate $s(t)$ by only observing the system output. This implies the blind estimation of the inverse structure composed of a linear filter w. Let **s** and **e** be the vectors of infinite dimension, whose t-th entries are $s(t)$ or $e(t)$, respectively. The unknown input-output transfer can be written as:

$$e = \mathrm{H}s \tag{1}$$

where:

$$\mathbf{H} = \begin{pmatrix} \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & h(t+1) & h(t) & h(t-1) & \cdots \\ \cdots & h(t+2) & h(t+1) & h(t) & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix} \tag{2}$$

is an infinite dimension Toeplitz matrix which represents the action of the filter h to the signal $s(t)$. The matrix $\mathbf{H}$ is non-singular provided that the filter h is invertible, i.e. satisfies $h^{-1}(t)*h(t) = h(t)*h^{-1}(t) = \delta(t)$, where $\delta(t)$ is the Dirac impulse. The infinite dimension of vectors and matrix is due to the lack of assumption on the filter order. If the filter h is a finite impulse response (FIR) filter of order N_h, the matrix dimension can be reduced to the size N_h. In practice, because infinite-dimension equations are not tractable, we have to choose a pertinent (finite) value for N_h.

2.2 Summary of equations

From figure 1, we can write the mutual information of the output of the filter w using the notion of entropy rates of stochastic processes as:

$$I(Y) = \lim_{T \to \infty} \frac{1}{2T+1} \left\{ \sum_{t=-T}^{T} H(y(t)) - H(y_{-T}, ..., y_T) \right\} = H(y(\tau)) - H(Y) \tag{3}$$

where τ is arbitrary due to the stationary assumption. The input signal $S=\{s(t)\}$ is an unknown non-Gaussian i.i.d. process, $Y=\{y(t)\}$ is the output process and y denotes a vector of infinite dimension whose t-th entry is $y(t)$. We shall notice that $I(Y)$ is always positive and vanishes when Y is i.i.d.

After some algebra, Equation (2) can be rewritten as [10]:

$$I(Y) = H(y(\tau)) - \frac{1}{2\pi} \int_0^{2\pi} \log \left| \sum_{t=-\infty}^{+\infty} w(t) e^{-jt\theta} \right| d\theta - E[\mathcal{E}] \tag{4}$$

At this point we need to derive the optimization algorithm. One possibility is, for example, to use gradient-like algorithms, where the derivative of $I(Y)$ with respect to the coefficients of w filter is needed. In our system, a canonical genetic algorithm will be used, avoiding the calculus of hard statistics. The method is presented in next section.

3 Genetic Algorithm for Blind Deconvolution

## 3.1 Justifying the use of a GA for blind deconvolution

A genetic algorithm (GA hereinafter) is a search technique used in computer science to find approximate solutions to combinatorial optimization problems. GAs are a particular class of evolutionary algorithms that use techniques inspired by evolutionary biology such as inheritance, mutation, natural selection, and recombination (or crossover) [11].

The process of blind deconvolution can be handled by a genetic algorithm which evolves individuals corresponding to different inverse filters and evaluate the estimated solutions according to a measure of statistical independence. This is a problem of global optimization: minimizing or maximizing a real valued function $f(\vec{x})$ in the parameter space $\vec{x} \in P$. This particular type of problems is suitable to be solved by a genetic algorithm. GAs are designed to move the population away from local minima that a traditional hill climbing algorithm might get stuck in. They are also easily parallelizable and their evaluation function can be any that assigns to each individual a real value into a partially ordered set (poset). GAs have already been successfully applied to linear and post-nonlinear blind source separation [12].

3.2 GA characterization

The operation of the basic genetic algorithm needs the following features to be completely characterized:

− Encoding Scheme. The genes will represent the coefficients of the unknown deconvolution filter W (real coding). An initial decision must therefore be taken about the length of the inverse filter.
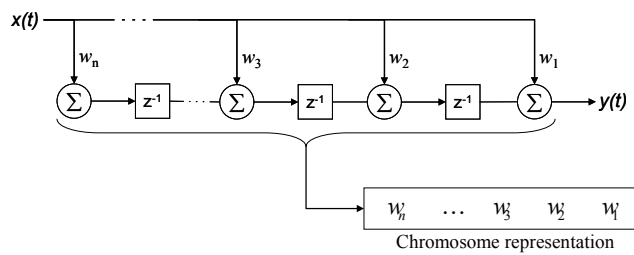
Fig. 2. Encoding scheme in a genetic algorithm for filter coefficients of linear blind deconvolution. The values of the variables stored in the chromosome are real numbers.

− Initialization Procedure. Coefficients of the deconvolution filter which form part of the chromosomes are randomly initialized.

- Fitness Function. The key point in the performance of a GA is the definition of the fitness function. In this case, we propose several fitness functions related with maximizing independence between the values of the estimated signal *y(t)*. Maximizing kurtosis absolute value is proposed as a first approach, followed by two cumulant-based expressions. Further details will be given in Section 3.2.1.

- 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 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. Generally a population of 80-100 individuals was used, stopping criteria was set between 60-100 iterations, crossover probability is 0.8 per chromosome and mutation probability is typically set between 0.05 and 0.08 per gene.

3.2.1 Evaluation functions proposed

One of the most remarkable advantages of genetic algorithms is its great flexibility for the application of new evaluation functions, being the only requirement that the evaluation function assigns a real value to each individual into a partially ordered set. Therefore, the evaluation function is extremely modular and independent from the rest of the GA. This ability will allow us to decide which evaluation function performs better in each situation. Generally, we will look for an evaluation function which gives higher scores for those chromosomes representing estimations which maximize statistical independence.

- Measuring nongaussianity by kurtosis. Absolute value of the kurtosis has been extensively used as a measure of nongaussianity in finding independent components [13]. Kurtosis is simple to compute. In this paper we propose using the absolute value of the normalized kurtosis as the first evaluation function:

$$\left| kurt(x) \right| = \left| \frac{\mathrm{E}(x^4)}{\mathrm{E}(x^2)^2} - 3 \right| \tag{5}$$

The evaluation function is directly derived from (5):

$$\mathrm{eval}_{\mathrm{Kurt}}(w) = \left| kurt(y) \right| \tag{6}$$

where y is the signal obtained after applying the filter w to the observation x.

- Measuring mutual information through approximation by cumulants. As it is well-known, kurtosis is very sensitive to outliers. Therefore we propose a different evaluation function. Using Edgewoth expansion, an approximation of mutual in-

formation can be reached using cumulants (higher-order statistics), as proposed in [14]:

$$I(y) = C - \left(\frac{1}{48} \sum_{i=1}^{n} \left[4\kappa_3(y_i)^2 + \kappa_4(y_i)^2 + 7\kappa_4(y_i)^4 - 6\kappa_3(y_i)^2 \cdot \kappa_4(y_i) \right] \right) \qquad (7)$$

where $k_3(y_i) = m_3(y_i) = E\{y_i^3\}$, $k_4(y_i) = m_4(y_i) - 3 = E\{y_i^4\} - 3$ and C is a constant.

The proposed evaluation function splits the estimated signal y in a set of equal-size chunks. Subsequently, it approximates mutual information between the pieces of the signal according to equation (7). As mutual information must be minimal for the estimated source under the assumption of statistical independence, and the evaluation function is attempted to be maximized by the GA, the evaluation function for a given chromosome in this case will be:

$$\text{eval}_{IM}(W) = \frac{1}{48} \sum_{i=1}^{n} \left[4\kappa_3(y_i)^2 + \kappa_4(y_i)^2 + 7\kappa_4(y_i)^4 - 6\kappa_3(y_i)^2 \cdot \kappa_4(y_i) \right] \qquad (8)$$

where y is the signal obtained after applying the filter W to the observation x and y_i is the i-th chunk from estimation y.

- Measuring negentropy through approximation by cumulants. Negentropy is a non-egative measure of nongaussianity. Finally, using again higher-order cumulants and the Gram-Charlier polynomial expansion, gives the approximation:

$$J(y) \approx \frac{1}{12} E\{y^3\}^2 + \frac{1}{48} \left[E\{y^4\} - 3 \right]^2 \qquad (9)$$

The evaluation function is just equivalent to the approximation of negentropy, as the maximum values should give good estimations:

$$\text{eval}_{NEG}(W) = J(y) \qquad (10)$$

where y is the signal obtained after applying the filter W to the observation x.

4 Experimental Results

Finally, in order to verify the effectiveness of the proposed algorithm, some experimental results using uniform random sources are presented. In all the experiments, the source signal is an uniform random source with zero mean and unit variance. As the performance criterion, we have used the output Signal to the Noise Ratio (SNR) measured in decibels.

In the first experiment, the (unknown) filter is the low-pass filter $H(Z) = 1 + 0.5z^{-1}$. Then, the proposed algorithm is used to obtain the inverse system. The parameters of the algorithm are: $T = 5000$ (number of observed samples), $p = 7$ (order of inverse

filter), crossover probability was set to 0.8, mutation probability 0.075, population size is 80, and the stopping criterion is 100 generations.

In the second experiment, we diminish the number of available samples. Here, the problem is more difficult due to the lack of information. Parameters of the algorithm are set to: $T = 1000$ (number of observed samples), $p = 5$ (order of inverse filter), and we used the same parameters for the genetic part of the algorithm as in the first experiment. Figure 2 (center) shows the coefficients of the filters $W(Z)$ and $WH(Z)$ respectively when applying eval_{IM}.

W (Z)

WH (Z)
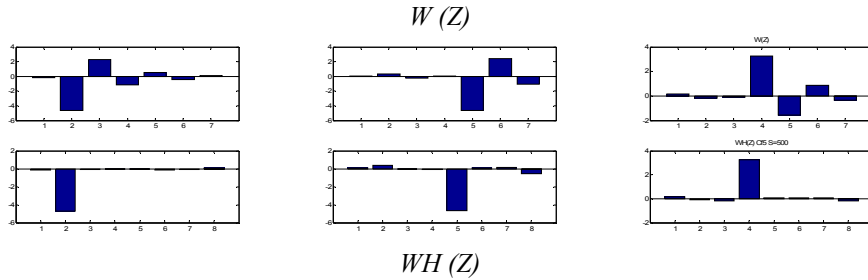
Fig. 2. On the top line, the inverse filter coefficients. Down figures represent coefficients of the convolution between filters $W(Z)$ and $H(Z)$. From left to right: filter coefficients gicen by the GA using $\text{eval}_{\text{Kurt}}$ with an observed signal of 2000 samples, eval_{IM} with 1000 samples, and eval_{Neg} with 500 samples. Note that in some of the estimated filters, a delay due to the indeterminacies may appear.

In the last experiment, we reduced the number of available to T=500. The rest of the parameters remain the same as in the former simulations.

Figure 3 summarizes the results of the experiments in terms of crosstalk (left) and computational time (right). These experiments show that although the number of samples are low, the algorithm has been capable of estimating the inverse system. Crosstalk between the estimation and the source is situated between 20-30dB, depending on the length of the signal and the contrast function applied. When compared to a typical deconvolution gradient descent algorithm, the GA presents a better performance as the number of samples in the observed signal decreases.

5   Conclusion

A GA algorithm is presented for blind inversion of channels. The use of this technique is justified here because of the small number of samples. In this situation, gradient-like algorithm fails because it is very difficult to obtain a good estimation of statistics (score function, pdf, etc.). Optimization using GA avoids these calculations and gives us good results for inverting the unknown filter. Future research should extend the idea to Wiener systems (linear filter plus nonlinear function), where a Hammerstein structure can be used and all the parameters should be found by these optimization techniques.
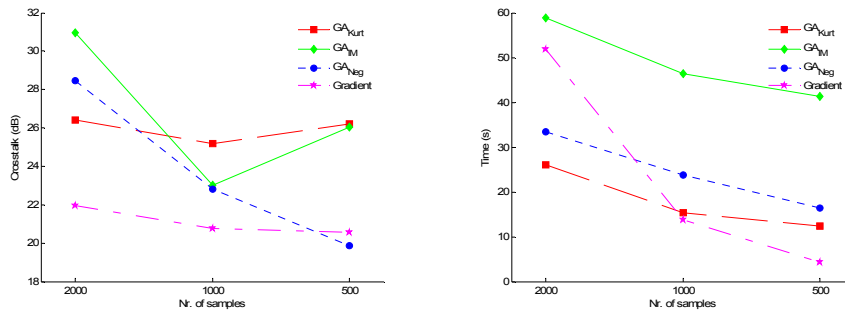
Fig. 3. Crosstalk and time comparison of the proposed GA using the three different evaluation functions and a gradient descent algorithm.

References

1. Jutten, C., Hérault, J., Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture, Signal Processing 24, 1991.
2. Comon, P., Independent component analysis, a new concept?, Signal Processing 36, 1994.
3. Amari, S., Cichocki A., Yang, H.H., A new learning algorithm for blind signal separation, NIPS 95, MIT Press, 8, 1996
4. Comon, P., Separation of sources using higher order cumulants Advanced Algorithms and Architectures for Signal Processing, 1989.
5. Cardoso, J.P., Source separation using higher order moments, in Proc. ICASSP, 1989.
6. Puntonet, C., Mansour A., Jutten C., A geometrical algorithm for blind separation of sources, GRETSI, 1995
7. Bell, A.J., Sejnowski T.J., An information-maximization approach to blind separation and blind deconvolution, Neural Computation 7, 1995.
8. Nguyen Thi, H.L., Jutten, C., Blind source separation for convolutive mixtures, IEEE Transactions on Signal Processing, vol. 45, 2, 1995.
9. Taleb, A., Solé-Casals, J., Jutten, C., Quasi-Nonparametric Blind Inversion of Wiener Systems, IEEE Transactions on Signal Processing, Vol. 49, n°5, pp.917-924 (2001).
10. Solé-Casals, J., Jutten, C., Taleb A., Parametric approach to blind deconvolution of nonlinear channels. Neurocomputing, Vol. 48, pp.339-355 (2002)
11. Michalewicz, Z.: Genetic Algorithms + Data Structures = Evolution Programs. 3rd edn. Springer-Verlag, Berlin Heidelberg New York (1996)
12. 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, Vol. 34 (4), pp. 407 - 416, November 2004.
13. Shalvi, O.; Weinstein, E. New criteria for blind deconvolution of nonminimum phase systems (channels). IEEE Transactions on Information Theory, Volume 36, Issue 2, March 1990 Page(s):312 – 321.
14. P. Comon, Independent Component Analysis, A new concept?. Signal Processing, vol.36, nº. 3, pp. 287-314. 1994.