Evolutionary ARMA Model Identification With Unknown Process Order*

G. N. BELIGIANNIS[1,2], E. N. DEMIRIS[1] and S. D. LIKOTHANASSIS[1,2,3]
[1]Department of Computer Engineering and Informatics
University of Patras
Rio, 26500 Patras
GREECE
[2]Computer Technology Institute (C.T.I.)
Kolokotroni 3, 26221 Patras
GREECE
[3]University of Patras Artificial Intelligence Research Center (U.P.A.I.R.C.)
University of Patras
Rio, 26500 Patras
GREECE

*The authors' names are in alphabetical order.

Abstract: - The fundamental problem of selecting the order and identifying the parameters of an AutoRegressive Moving Average model (ARMA) has been faced effectively, using the Multi Model Partitioning (MMP) theory. The disadvantage of this method is the use of a fixed population of candidate models. Depending on the a priori selection of the set of conditional models, this method gives (near) optimal solutions. This disadvantage can be alleviated using natural selection techniques, such as the Genetic Algorithms (GAs). Thus, a new Evolution Program (EP) is developed in this work. This method combines the effectiveness of the MMP theory with the robustness of the GAs. Although the parameters' coding is more complicated, simulation results show that the proposed algorithm succeeds better results compared to the conventional one, since it has the ability to search the whole parameter space.
Key-Words: - AutoRegressive Moving Average model, Multi Model Partitioning theory, Genetic Algorithms, Evolution Program.
CSCC'99 Proceedings: - Pages 1951-1958

1 Introduction

The adaptive filtering problem, with unknown time-invariant or time-varying parameters, has been a central issue in the field of signal processing. This is due to the fact that adaptive techniques have found applications in various fields and that the advances in VLSI technology together with the decreasing cost of hardware and digital signal processors have made it possible to implement complex algorithms at a reasonable cost. Selecting the correct order and estimating the parameters of a system model is a fundamental issue in linear prediction, system identification and spectral analysis. The problem of fitting an ARMA model to a given time series has attracted much attention because it arises in a large variety of applications, such as adaptive control, speech analysis and synthesis, radar, sonar, etc.

Various methods that represent information theoretic criteria exist for model order selection. The most well known of the proposed solutions for this problem include the Final Prediction Error (FPE), Akaike's Information Criterion (AIC) [1]-[3] and the Minimum Description Length (MDL) Criterion [4], [5]. Usually, these methods are based on the assumption that the data are Gaussian and they are two-pass methods. Therefore, they cannot be used in an on-line or adaptive fashion. A different adaptive approach, based on the Partitioning Theorem, is the Multi Model Adaptive Filter (MMAF) [6], that operates on general, not necessarily Gaussian data pdf's. The MMAF converges to the optimal solution, if the model supporting the data is included to the filter's bank. Otherwise, it converges to the closer model by mean of the Kullback information criterion minimization. This is due to the fact that the number of filters used in the MMAF bank is finite. Among the existing adaptive identification methods, we are particularly interested in the partitioned adaptive technique, since it is useful not

only for identifying the noise statistics but also for selecting the correct system order and estimating the unknown system parameters. Pioneer work in this area can be found in [7]-[9]. It is known that the linear filtering problem with unknown time-invariant or time-varying parameters reduces to a non-linear filtering problem, which has major difficulties in its realization. In particular, it is extremely difficult to access the effect of approximations made in the suboptimally realization of non-linear filters. However, partitioned adaptive filtering constitutes a partitioning of the original non-linear filters into a bank or set of much simpler linear elemental Kalman or Kalman-Busy filters. This realization is very simple to implement physically.

It is known that Genetic Algorithms (GAs) are one of the best methods for searching and optimization [12]-[16]. They apply genetic operators (reproduction, crossover and mutation), in a population of individuals (sets of unknown parameters properly coded), in order to achieve the optimum value of the fitness function. By evolving the best individuals, in each generation, they converge to the (near) optimal solution. The main advantage of the GAs is that they use the parameters' values instead of the parameters themselves. In this way they search the whole parameter space.

In this work, a new Evolution Program which combines the effectiveness of adaptive multi model partitioning filters and GAs' robustness has been developed. This method was first introduced in [17]. Specifically, the a posteriori probability that a specific model, of a bank of the conditional models, is the true model, can be used as fitness function for the GA. In this way, the algorithm identifies the true model even in the case where it is not included in the filters' bank. It is clear that the evolution of the population of the filter's bank improves the filter's performance, since the algorithm can search the whole parameter space. The method is not restricted to the Gaussian case, it is applicable to on-line/adaptive operation and is computationally efficient. Furthermore, it can be realized in a parallel processing fashion, a fact which make it amenable to VLSI implementation.

The paper is organized as follows. In section 2 the ARMA model order selection and identification problem is stated. In sections 2.1 and 2.2 the proposed method and the EP structure are presented respectively. Section 3 contains the simulation results, while section 4 summarizes the conclusions.

2 ARMA Model Order Selection

A general model for ARMA can be represented as follows:

$$A(q)y(t)=B(q)e(t) \qquad (1)$$

or

$$y(t)=\sum_{i=1}^{n_a} a_i\, y(t-i)+\sum_{j=1}^{n_b} b_j e(t-j)+e(t) \qquad (2)$$

where y(t) is the observed data, e(t) is a zero-mean white noise process, with variance R, not necessarily Gaussian, $n=(n_a,n_b)$ is the order of the predictor and a_i $(i=1,...,n_a)$, b_j $(j=1,...,n_b)$ are the predictor coefficients. Clearly the problem is two-fold: one has both to select the order of the predictor and then to compute the predictor coefficients. Of course the most crucial part of the problem is the former.

Let us define the vector of coefficients θ(t) as follows:

$$\theta(t)=[a_1(t) \ldots a_{n_a}(t)\ b_1(t) \ldots b_{n_b}(t)]^T \qquad (3)$$

where $0\le t\le N$ (N denotes the number of samples). Notice that the coefficients a_i and b_j have been replaced by $a_i(t)$ and $b_j(t)$ respectively to reflect the possibility that the coefficients are subject to random perturbations. This can be modeled by assuming that:

$$\theta(t+1)=\theta(t)+w(t),\ t=1,...,N \qquad (4)$$

where

$$w(t)=[w_1(t)\ w_2(t) \ldots w_{n_a+n_b}(t)]$$

is a zero-mean white process, with variance Q , not necessarily Gaussian (e(t) and w(t) are assumed to be independent). Equation (2) can be written in the following form:

$$y(t)=h^T(t)\cdot\theta(t)+e(t) \qquad (5)$$

where

$$h^T=[y(t-1) \ldots y(t-n_a)\ e(t-1) \ldots e(t-n_b)] \qquad (6)$$

The formulation of equations (4) and (5) are in the standard space form, so that results of state space estimation techniques can be readily used to estimate the values of the θ(t)'s and can be found in [10]. First of all, we have to assign values to the variances of the processes w(k) and e(k), that can be denoted by Q and R respectively. Assessing the values of Q and R is not always an easy task. If R is not readily obtainable, it can be estimated using a technique described in [18]. The effect of estimating R via this technique is investigated in [19]. Q can

again be estimated using a technique described in [18]. As far as the vector θ(t) is concerned, we assume that an a-priori mean of the vector θ(0) can be set to zero in the case that no knowledge about their values is available before any measurements are taken (the most likely case). On the other hand, the usual choice of the initial variance of the vector θ(t), denoted by P_0 is $P_0=m\cdot I$, where m is a large integer. Finally, we assume that the measurements of y(t) and e(t) are set to zero for $t<0$ (prewindowing).

The ARMA model identification problem is now stated. Given a set of observations y(t), where $0\le t\le N$, from an unknown ARMA (n_a,n_b) process we have to determine the unknown parameter vector:

$$v=[n_a\ n_b\ \theta(t)\ Q\ R] \quad (7)$$

Let us now assume that the order n, or the parameters (n_a,n_b), is unknown and the only available knowledge about the true order is that it satisfies the condition $n_0\le n\le n_{MAX}$. It is clear then that the true model is one of a set of models described by equations (4) and (5) and is specified by the true value of parameters (n_a,n_b). The problem is then to select the correct model among various candidate models. In other words, we have to design an optimal
estimator (in the minimum variance sense), when uncertainty is incorporated in the signal model. The solution to this problem has been given by the multi model partitioning theory [7]-[9], which is described in the following.

The multi model partitioning algorithm operates on the following discrete model:

$$\theta(t+1)=F(t+1,1/n)\cdot\theta(t)+w(t) \quad (8)$$

$$z(t)=h^T(t/n)\cdot\theta(t)+e(t) \quad (9)$$

where $n=(n_a,n_b)$ is the unknown model order, assumed to be a random variable with known a-priori pdf $p(n/0)=p(n)$, and $F(\cdot)$ is the state transition matrix. The optimal MMSE estimate of θ(t) is given by:

$$\hat{\theta}(t)=\int_n \hat{\theta}(t/t;n)p(n/t)dn \quad (10)$$

where $\hat{\theta}(t/t;n)$ is the conditional MMSE state vector estimate that is obtained by the corresponding Kalman filter matched to the model with parameter value n and initialized with initial conditions $\hat{\theta}(0/0;n)$ and P(0/0;n). The model-conditional pdf p(n/t) is given by:

$$p(n/t)=\frac{L(t/t;n)}{\int_n L(t/t;n)p(n/t-1)dn}p(n/t-1) \quad (11)$$

where

$$L(t/t;n)=\left|P_z(t/t-1;n)\right|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\left\|\tilde{z}(t/t-1;n)\right\|P_z^{-1}(t/t-1;n)\right\} \quad (12)$$

and

$$P_z(t/t-1;n)=\hat{h}^T(t/n)P(t/0;n)\hat{h}(t/n)+R(t) \quad (13)$$

$$\tilde{z}(t/t-1;n)=z(t)-\hat{h}^T(t/n)F(t+1,t/n)\hat{\theta}(t/t-1;n) \quad (14)$$

$$\hat{h}^T=\left[y(t-1)\ ...\ y(t-n_a)\ e(t-1/n)\ ...\ e(t-n_b/n)\right] \quad (15)$$

$$e(t/n)=z(t)-\hat{h}^T(t/n)\hat{\theta}(t/t-1;n) \quad (16)$$

Equations (8) and (9) pertain to the case where n's pdf is continuous in n. When this is the case, one is faced with the need for a nondenumerable infinity of Kalman filters for the exact realization of the optimal estimator. The usual approximation performed to overcome this difficulty is to somehow approximate n 's pdf by a finite sum [8]. Fortunately, in our case the sample space is naturally discrete, so that no approximations are necessary and the estimator is indeed optimal. Since this is the case, the integrals in (10) and (11) must be replaced by summations running over all possible values of $n=(n_a,n_b)$. If we rewrite equations (4) and (5) and define the state vector as a vector containing all model coefficients, then the model is easily extended to the multivariate
case [11].

2.1 ARMA Model Order Selection Using Evolutionary MMAF

In the ARMA model described at the previous paragraphs there are two quantities which must be estimated; the model order n and the unknown parameter vector θ(t). The estimation method is the MMAF as it is obtained from Lainiotis partition theorem. Our goal is to achieve the optimal estimation for both estimated quantities and particularly for the model order $n=(n_a,n_b)$. The only information we have for these parameters is that they belong to a set or a space (finite or infinite). It is obvious that if the unknown parameters belongs to a finite-discrete set with small cardinality, the MMAF is the only appropriate and most effective method to estimate them. It is also widely known that GAs perform better when the space which will be searched has a large number of elements. So, GAs can be used when the unknown parameters

belong to a space with large cardinality or belong to an infinite space or follow a probability distribution. Then, we should optimize with GAs the model-conditional pdf. That means that we have to optimize the probability described from relation (11), for discrete sample space, i.e. the following probability function:

$$p(n/t) = \frac{L(t/t;n)}{\sum_{n} L(t/t;n)p(n/t-1)} p(n/t-1) \qquad (17)$$

which will be the fitness function for the GA, for the several values of the unknown parameters $n=(n_a,n_b)$ and $\theta(t)$ underlying to the above constraints. The structure of the system of the proposed solution is shown in Fig. 1.

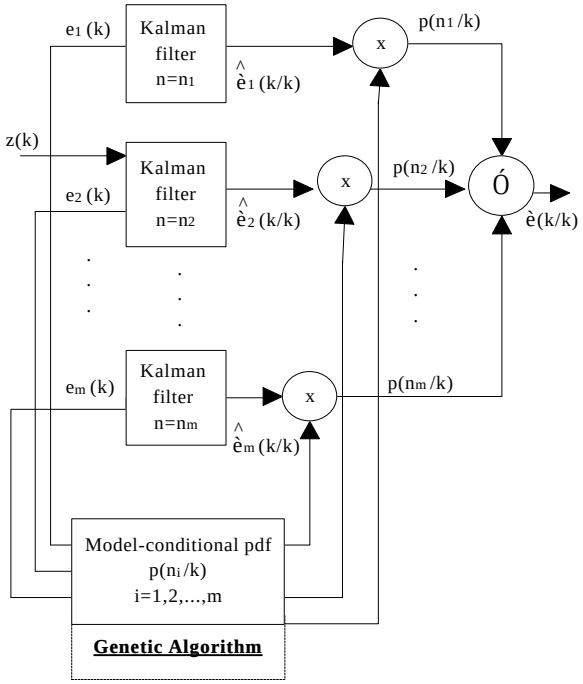
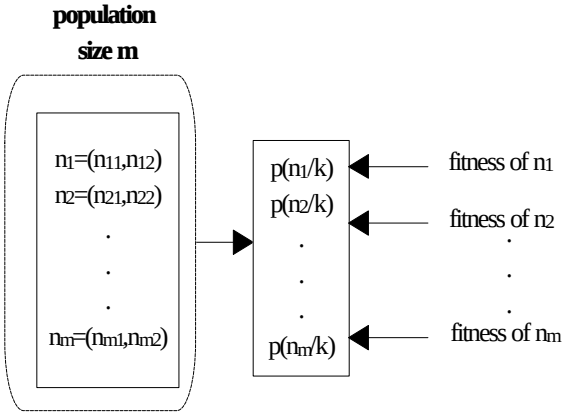


Fig.1: The Structure of the Evolutionary MMAF System



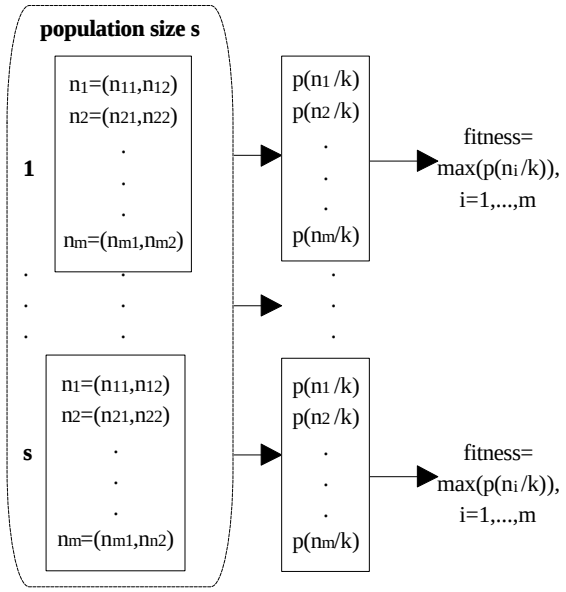
m = the number of Kalman filters in the Multi Model Adaptive Filter (MMAF).

Fig. 2: The first version of the EP

2.2 The Structure of the Evolution Program

The structure of the EP that has been developed has two versions. First of all, we must state that we use two different representations, binary and integer, for the possible solutions. According to the first version (Fig. 2), we first make an initial population of m pairs of integers each of them representing a possible value of the ARMA model order. For each population of possible solutions we apply an MMAF and have as result the model-conditional pdf. This is the fitness of each possible solution. According to the second version (Fig. 3), we first make an initial population of s vectors of m pairs of integers (each pair representing a possible value of the ARMA model order). For each such vector we apply a MMAF and have as result the model-conditional pdf of each value. The biggest pdf is the fitness of each vector.

For both versions since we have the fitness of each possible solution (or vector of possible solutions) we are able to perform the other genetic operators, i.e. reproduction, crossover and mutation. The reproduction operator is the classic biased roulette wheel selection according to the fitness function value of each possible solution (or vector of possible solutions). As far as crossover is concerned, we use four crossover operators for the binary representation (Uniform Crossover, Even-Odd Crossover, One-Point Crossover, Two-Point Crossover) and four crossover operators for the integer representation (Uniform Crossover, Blend Crossover, Arithmetic Crossover, Neighbor Crossover-a new crossover operator implemented, giving new integer values near the old ones).



m = the number of Kalman filters in the Multi Model Adaptive Filter (MMAF).

Fig. 3: The second version of the EP

Finally, we use one mutation operator for binary representation (Flip Mutator) and two mutation operators for integer representation (Flip Mutator, Gaussian Mutator). Every new generation of possible solutions (or vectors of possible solutions) iterates the same process as the old ones and all this process may be repeated as many generations as we desire or till the fitness function has value 1 (one) which is the maximum value it is able to have as a probability[17].

The first version results to a final population of pairs of integers always containing the true order of the ARMA model. The second version results to a vector of pairs of integers having the best model-conditional pdf always containing the true order of the ARMA model. After that, the estimation of $\theta(t)$, for both the EP's versions, is straightforward using the conventional MMAF.

3 Experimental Results

The presented algorithm has been run extensively on several simulation experiments. All experiments were carried out 100 times (100 Monte Carlo runs). In this section, two examples are discussed. On both examples, the true order $n_i=(n_{a_i}, n_{b_i})$ of each ARMA model satisfies the conditions $1\leq n_{a_i}\leq 64$ and $1\leq n_{b_i}\leq 64$. We also assume, R=1, Q=0.01, P(0/0;n)=100 and F(t+1,t/n)=1.

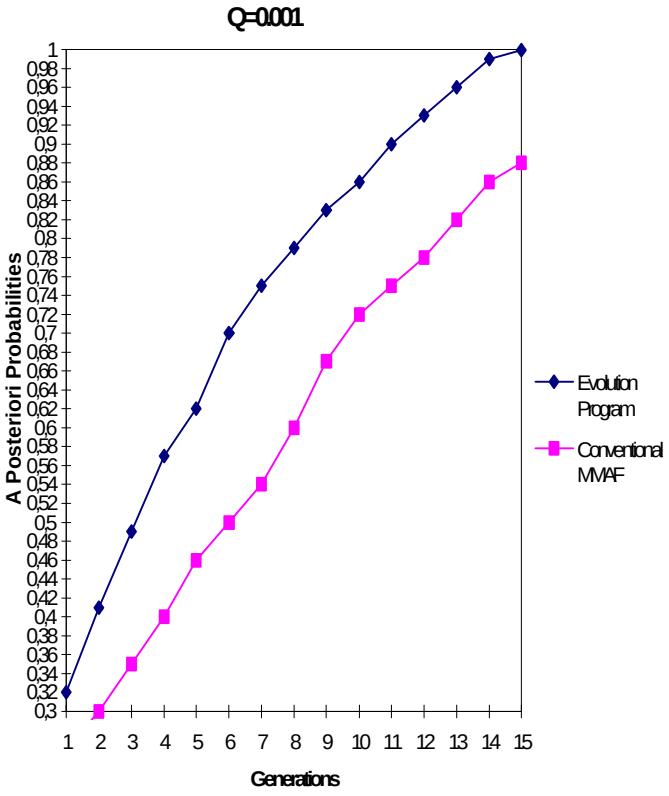


Fig. 4: The evolution of the A Posteriori Probabilities of the best genome (1st version of the EP, 1st example) compared to the conventional MMAF.

For the first example, the data generating process at time instant t=0, is given by:

y(t)=1.8y(t-1)-0.9y(t-2)+0.4e(t-1)-0.8e(t-2)+e(t)

For the second example, the process model changes during the operation and is given by:

a) for the first 400 samples at time instant t=0:

y(t)=1.8y(t-1)-0.9y(t-2)+0.4e(t-1)-0.8e(t-2)+e(t)

b) for the last 600 samples at time instant t=400:

y(t)=1.5y(t-1)-1.4y(t-2)+0.8y(t-3)+0.7e(t-1)
-1.1e(t-2)-2.3e(t-3)+1.5e(t-4)+e(t)

The size of the population we used in both our examples was 10, the crossover probability was 0.95 and the mutation probability was 0.15. Also, the number of Kalman filters in the MMAF was 10 and for every generation of the GA the MMAF was applied for 50 runs.

The presented results in Fig. 4 show the evolution of the model-conditional pdf of the best genome (first version of the EP) for the first example, compared to the conventional MMAF.

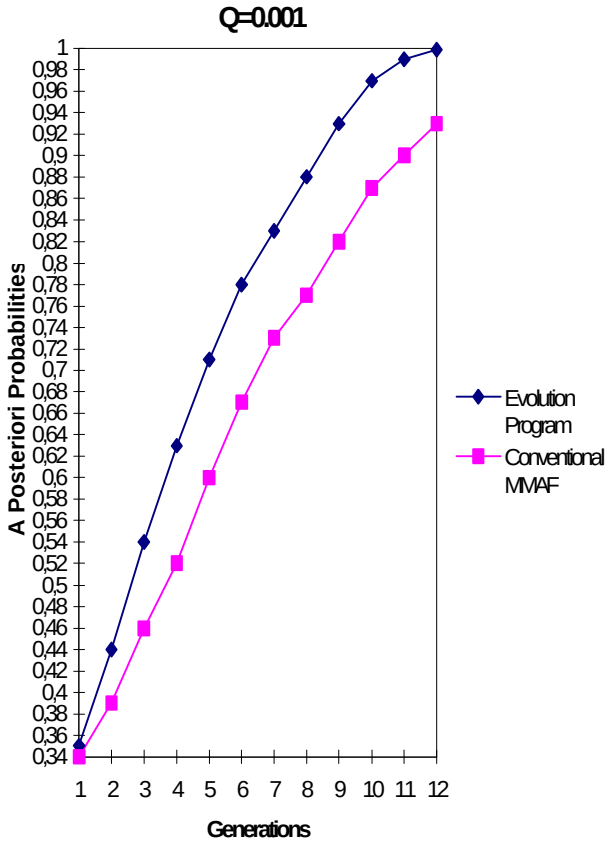


Fig. 5: The evolution of the A Posteriori Probabilities of the best genome (2nd version of the EP, 1st example) compared to the conventional MMAF.

The presented results in Fig. 5 show the evolution of the model-conditional pdf of the best genome (second version of the EP) for the first example for Q=0.001, compared to the conventional MMAF.

The presented results in Fig. 6 show the evolution of the model-conditional pdf of the best genome (second version of the EP) for the first example for Q=0.01, compared to the conventional MMAF.

The presented results in Fig. 7 show the evolution of the model-conditional pdf of the best genome (second version of the EP) for the second example for Q=0.001, compared to the conventional MMAF.

Furthermore, according to all experiments' results the following conclusions were come to:

- As the population size grows the first version of the EP converges faster giving a model-conditional pdf higher than 0.99.
- As the population size grows the second version of the EP converges faster giving a model-conditional pdf higher than 0.999.
- The crossover probability must be higher than 0.9 for both versions in order to identify the true order of the system model.
- The mutation probability must be higher than 0.05 and smaller than 0.2 for the first example and higher than 0.1 and smaller than 0.25 for the second example for both versions in order to identify the true order of the system model.

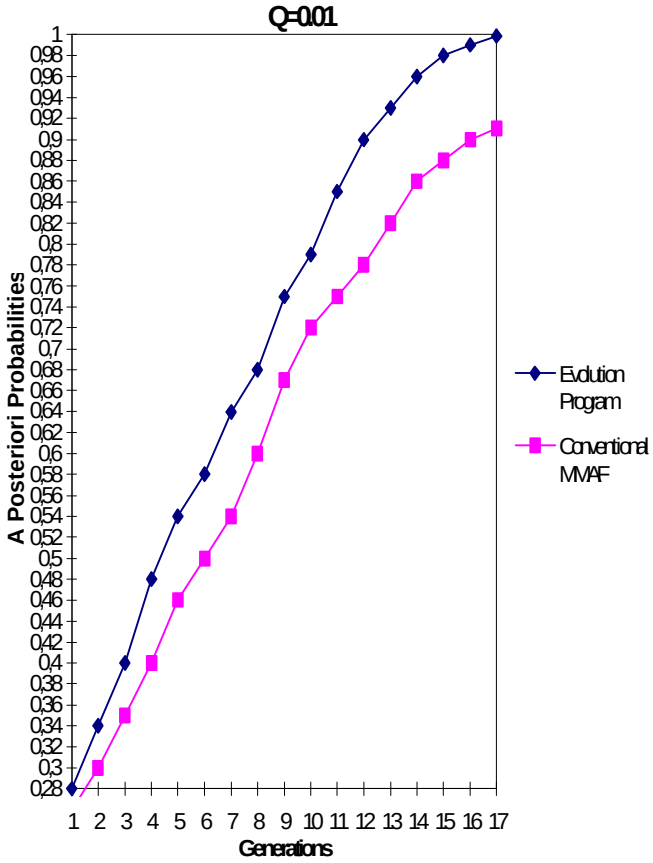


Fig. 6: The evolution of the A Posteriori Probabilities of the best genome (2nd version of the EP, 1st example) compared to the conventional MMAF.

- The ratio between the population size and the size of the domain set, for the first version, must be at least 1/16 in order to identify the true order of the system model.
- The ratio between the size of the genome of pairs of integers and the size of the domain set, for the second version, must be at least 1/16 in order to identify the true order of the system model.
- As far as crossover operators are concerned, faster convergence is reached by using the Uniform Crossover, for binary representation,

and Neighbor Crossover, for integer representation.

- As far as mutation operators are concerned, faster convergence is reached by using the Flip Mutator, for binary representation, and Gaussian Mutator, for integer representation.

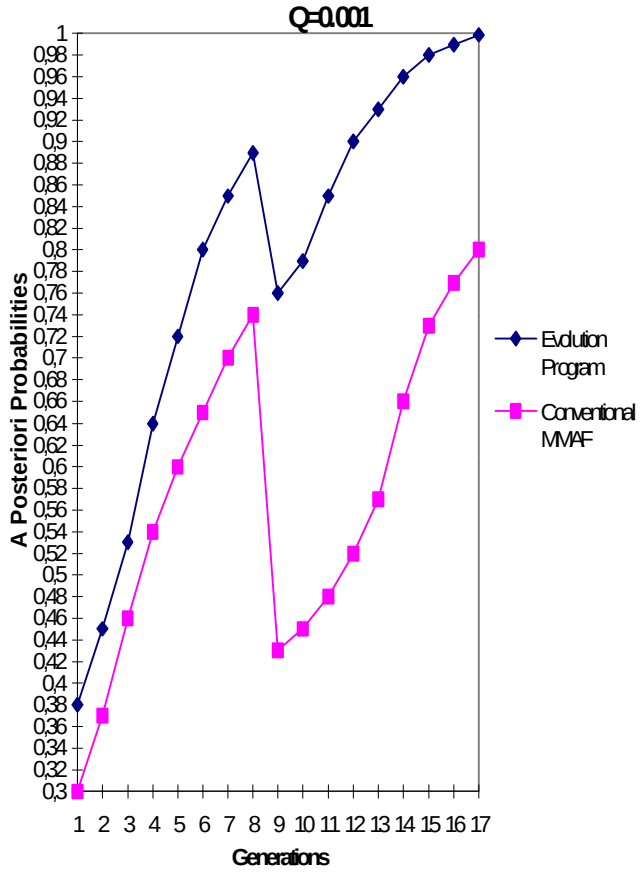


Fig. 7: The evolution of the A Posteriori Probabilities of the best genome (2nd version of the EP, 2nd example) compared to the conventional MMAF.

4 Summary

In this work, a new evolutionary method for adaptive estimation of ARMA discrete time systems, with unknown order and parameters, has been proposed. The method combines the well known Adaptive Multi Model Partitioning theory with the effectiveness of the GAs. Simulation results show that the method performs significally better than the conventional MMAF. Although the parameters' coding is more complicated a variety of defined crossover and mutation operators was investigated, resulting in accelerations of the algorithms convergence. Furthermore, the evolution of the initial population results to the identification of the true model, even in the case where it is not included in the initial population of the filter's bank. Finally, the method can be implemented in a parallel environment, since the MMAF as well as the GA are naturally parallel structured, thus increasing the computational speed.

References:

[1] H. Akaike Fitting Autoregressive Models for Prediction, *Ann. Inst. Of Stat. Math.*, vol. 21, 1969, pp. 243-347.

[2] H. Akaike, Information Theory and an Extension of the Maximum Likelihood Principle, *in Proc. 2nd Int. Symp. Inform. Theory*, Budapest, Hungary: Akademia Kiado, 1973, pp. 267-281.

[3] H. Akaike, A New Look at the Statistical Model Identification, *IEEE Trans. Automat. Contr.*, vol. 26, 1977, pp. 1-18.

[4] G. Schwartz, Estimation of the Dimension of a Model, *Ann. Stat.*, vol. 6, 1978, pp. 461-464.

[5] J. Rissanen, Modeling by Shortest Data Description, *Automatica*, vol. 14, 1978, pp. 465-471.

[6] K. Watanabe, *Adaptive Estimation and Control. Partitioning Approach*, Prentice-Hall, U.K., 1992.

[7] D. G. Lainiotis, Optimal Adaptive Estimation: Structure and Parameter Adaptation, *IEEE Trans. Automat. Contr.*, vol. AC-16, 1971, pp. 160-170.

[8] D. G. Lainiotis, Partitioning: A Unifying Framework for Adaptive Systems I: Estimation, *Proc. IEEE*, vol. 64, 1976, pp. 1126-1143.

[9] D. G. Lainiotis, Partitioning: A Unifying Framework for Adaptive Systems II: Estimation, *Proc. IEEE*, vol. 64, 1976, pp. 1182-1198.

[10] B. O. Anderson and J. B. Moore, *Optimal Filtering*, Englewood Cliffs, N.J., Prentice-Hall, 1979.

[11] S. D. Likothanassis and S. K. Katsikas, Multivariable AR Model Identification and Control, *Proc. IASTED Int. Symp. Modeling, Identification and Control*, Grindelwald, Switzerland, 1987, pp. 248-252.

[12] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Reading, Mass., 1989.

[13] Z.~Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer-Verlag, N.Y., 1992.

[14] M. Mitchell, *An Introduction to Genetic Algorithms*, A Bradford Book, The MIT Press,

Cambridge, Massachusetts, London, England, 1995.
[15] L. D. Whitley, *Foundations of Genetic Algorithms 2*, Morgan Kaufmann, 1993.
[16] L. D. Whitley and M. D. Vose, *Foundations of Genetic Algorithms 3*, Morgan Kaufmann, 1995.
[17] K. G. Berketis, S. K. Katsikas and S. D. Likothanassis, Multimodel Partitioning Filters and Genetic Algorithms, *Journal of Non-Linear Analysis*, vol. 30-4, Dec. 1997, pp. 2421-2428.
[18] A. Sage and G. Husa, *Algorithms for Sequential Adaptive Estimation of Prior Statistics*, 1969.
[19] A. K. Mahalanabis and S. Prasad, On the Application of the Fast Kalman Algorithm to Adaptive Deconvolution of Seismic Data, *IEEE Trans. Geosci. Remote Sensing*, Vol. GE-21, No. 4, 1983.
[20] S. D. Likothanassis, E. N. Demiris and D. G. Karelis, A Real Time ARMA Model Structure Identification Method, *Int. Conference on Digital Signal Processing, DSP '97*, 2-4 July 1997, Greece.