Using Genetic Algorithms in System Identification

Ecaterina Vladu

Department of Electrical Engineering and Information Technology, University of Oradea, Universitatii, 410087 Oradea, România
Phone: +40259408435, Fax: +40259408408, e-mail: evladu@rdslink.ro

Abstract: A system identification problem can be formulated as an optimization task where the objective is to find a model and a set of parameters that minimize the prediction error between the plant outputs i.e., the measured data, and the model output. The most existing system identification approaches are highly analytical and based on mathematical derivation of the system's model. As an alternative to these methods, evolutionary computation seems to be a very promising approach, because it needs only few knowledge about the problem and id can be easily combined with a number of other techniques from control engineering, machine learning, artificial intelligence and so on. This paper considers an evolutionary approach for system identification and attempts to show how GAs can be applied in system identification tasks. Some study cases confirm that good performance can be achieved by this method.

Keywords: system identification, evolutionary techniques, genetic algorithms

1 Introduction

System identification consists of two tasks. The first task is structural identification of the equations and the second one is an estimation of the model's parameters. In control engineering, system identification is used to find a model of the plant to control. In this context, system models describe the behavior of the plant over time.

In the case the structure of the model is known in advance, the needed knowledge relies to the numerical values of a number of parameters.

In the following paragraphs, the experimental estimation of parameters will be referred. These methods use the measurements carried out on input and output signals, having the goal to find the mathematical models, which better describe, very close to reality, the behavior of the plant.

In order to apply GAs in systems identification, each individual in the population must represent a model of the plant and the objective becomes a quality measure of the model, by evaluating its capacity of predicting the evolution of the measured outputs. The measured output predictions, inherent to each individual *i*, is compared with the measurements made on the real plant. The obtained error is a function of the individual's quality. As less is this error, as more performing the individual is.There are many ways in which the GAs can be used to solve system identification tasks. The main tendencies are described in [1], [2], [3], [5], [7], [8].

2 Experimental Methods for Parameters Estimation

The phases to be passed for experimental parameters estimation are presented in Figure 1 [4].

The described method uses as starting point an approximate plant model. The model's outputs are compared with the experimental results and an error criteria related to the plant outputs and the mathematical model outputs is used. The mathematical model's those parameters are determined which lead to an output that fits the best to the plant outputs carried out by experimental measurements. These stages are then continued until the error criterion is met. [4]

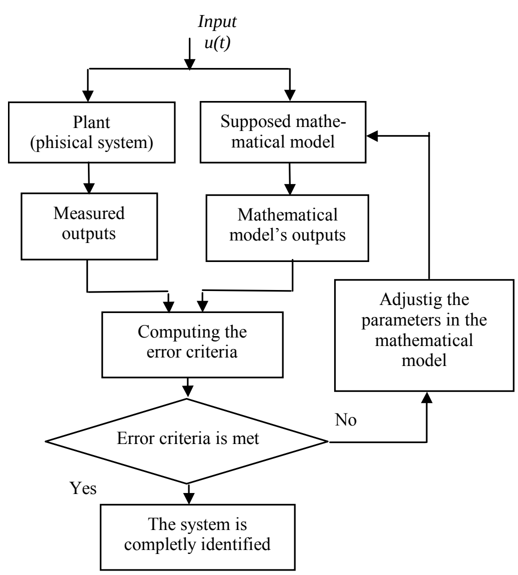

Figure 1

The experimental parameters estimation phases

The identification can be carried out on-line or off-line. In the on-line case, the input and output signals used are those, which appear in the usual operation of the plant and the model of the system is obtained in real time. In the off-line case, also the signals that appear in the usual operation of the plant are employed, but these signals are previously collected, by employing laboratory measurements. [4].

In the case the plant structure is not known in advance, the following phases are to be performed:

- Experimentally investigate if the plant has constant or adjustable parameters;
- Experimentally identify the linear domain of the plant;
- Adopt a mathematical describing principle.

Related to the input *u(t)* in Figure 1, the following assignation is necessary: this signal must have a suitable structure in order to punctuate in the reaction of the system all its characteristics. Consequently, it can be or a simple signal – as step or ramp – or a complex signal established from successive typical signals. Also, in order to obtain accurate resuts, an apropriate data management is necessary, as presented in [7].

Regarding the measured plant output, related to the type of the model to find – discrete time or continuous time – the samples must be oriented in the ascending order of time, they must correspond to the same equidistant moments in the case of discrete time systems and arbitrary in case of continuous time systems.

If the error criterion is of integral type, the practical usage consists in replacing the integral by an equivalent sum.

For simplicity and considerates of good approximation, the time moments when the comparison is carried out are considered equidistant, disregarding if the models are of discrete or continuous type.

3 Applying GAs in Systems Identification

This paragraph presents a method intended to the estimation of the plant parameters by using GAs. The method uses the principle scheme depicted in Figure 2.

In Figure 2, the bloc named Plant has the unknown parameters, which are to be found in the genetic search. The bloc Model has adjustable parameters, which are transmitted from GAs in the evaluation step. By comparing the $y(t)$ and $y_{mi}(t)$ outputs, a measure of the performance J_i is obtained, on base of which the individual i has assigned the $Fitness_i$ function.

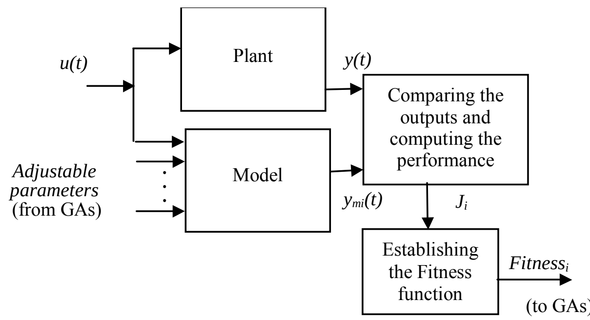

Figure 2
The principle scheme for parameters estimation

3.1 Implementing the Proposed Method by Simulation

The evaluation step for the proposed method is performed by simulation. For this purpose, the bloc diagram depicted in Figure 2 is implemented in a Simulink model. By performing a simulation for the individual i, the plant output *y(t)* and the model output $y_{mi}(t)$ are obtained.

GAs use individuals encoded as real numbers vector, that are the parameters searched in the estimation process. The individual encoding, in the case of estimating a number of *n* parameters is showed in Figure 3.

Figure 4 represents the output of a given plant to a step input signal. Such a curve is compared with the output of the model having adjustable parameters, at equidistant dt time moments, belonging to the interval [0, t_{max}], where t_{max} is the maximum simulation time.

If the model's parameters are identical to the plant parameters, that is, if they are correctly estimated, the response curve $y_{mi}(t)$ of the model completely overlaps the plant output curve *y(t)*.

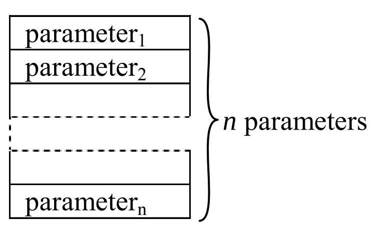

Figure 3
An individual encoding

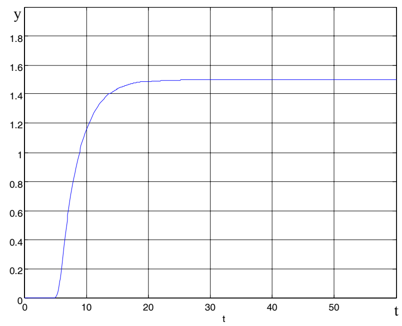

Figure 4
The output of a given plant to a step input signal

Contrary, the two response curves differ and the estimation error for the individual i can be defined by relation (1).

$$\varepsilon_i(t) = abs[y(t) - y_{mi}(t)] \quad (1)$$

Adding the both members of the relation (1) for all the time moments and denoting:

$\sum_{t=0}^{t_{\max}} \varepsilon(t)dt$ with $Ji(param_i)$, the relation (2) is obtained, where $Ji(param_i)$ becomes a performance criteria for the individual i:

$$Ji(param_i) = \sum_{t=0}^{t_{\max}} abs[y\ (t) - y_{mi}(t)] \quad (2)$$

The sum given by relation (2) is considered to be built up with such a small step, as it can be approximated with an integral.

The geometrical interpretation of the relation (2) is presented in Figure 5, where the plant output is presented relative to the model's output.

Denoting with S_y and S_{ymi} the domains delimited by *y(t)* curve, respectively $y_{mi}(t)$ and the time axis, showed hachured in the figure, due to the non-overlapping of the curves, between them a series of "non-overlapping" surfaces are delimited, which are those hachured in the same direction in the Figure 5.

These surfaces can be considered as performance criteria for the estimation quality: as smaller they are, as better the parameters estimation is. Their sum is obtained with relation (2).

In the problems of parameters estimation, it is not sufficient to use the system's response relative to a unique input signal. In order to obtain estimation as good as possible, a number *Ne* of test vectors are to be used, corresponding to a number *Ne* of input-output experiments [6], [7], [8].

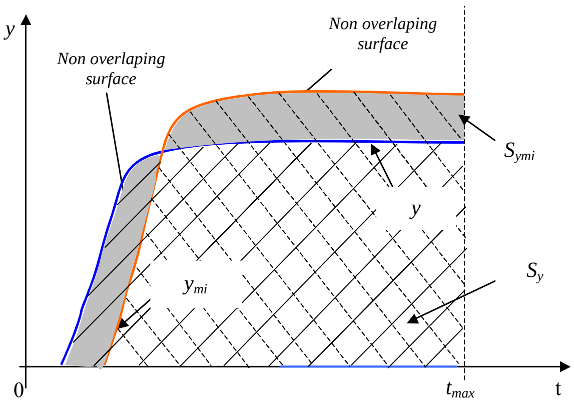

Figure 5

The plant output relative to the model's output

A test vector associates a plant input signal to an output signal, having the form (u_j, y_j), where u_j, is a given input signal and y_j is the corresponding output signal, where $j = 1 \ldots Ne$.

By this way, to the input of both plant and model, successively are applied de input signals u_j and the outputs y_j and y_{mij} are compared. This comparison supposes that a simulation having the inputs u_j is performed and the value J_{ij} is computed with relation (2) on base of the obtained output, for each individual i.

The performance indicator becomes the sum of the computed values of J_{ij} for each simulation. As smaller this sum is, as better the parameters estimation is. The evaluation step becomes more complex, as showed in Figure 6. This step is repeated for each individual i, in each generation of the algorithm.

In mathematical form, the performance criteria is given by relation (3):

$$Ji(param_i) = \sum_{\substack{j=1 \\ u(t)=uj(t)}}^{Ne} \int_0^{t_{\max}} \{abs[y_j(t) - y_{mij}(t)]\} dt \tag{3}$$

It has to be noted that, in the expression (3) the performance criteria uses equal weights for all of the input signals only in the cases these signals have a comparable amplitude. In the case the input signals amplitude are very different, weighting coefficients can be used in order to compensate these amplitude differences, as presented in relation (4).

$$Ji(param_i) = \sum_{\substack{j=1 \\ u(t)=uj(t)}}^{Ne} \int_0^{t_{\max}} k_j \{abs[y_j(t) - y_{mij}(t)]\} dt \tag{4}$$

where the coefficients k_j are choosed by using apriori knowledges, based in the previous analyse of the measurements performed on the real plant (for example contrariwise with the signal amplitude).

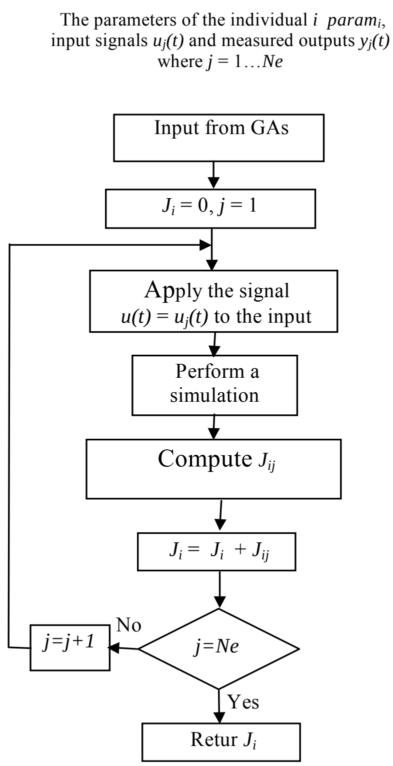

Figure 6

Evaluation step for an individual

The Fitness function is defined by relation (5).

$$Fitness_i = \frac{1}{1 + J_i(param_i)} \tag{5}$$

where the maximum value of $Fitness_i$ is 1, this situation corresponding to the complete achievement of performance requirements.

This method was described with the goal of testing it by simulation, supposing the system's structure is known in advance, but the parameters are unknown, the goal of appling GAs being exactly the estimation of these parameters.

3.2 Implementing the Proposed Method in the Real Plants

In real situations, the evaluation step is also implemented by simulation, unlike the parameters estimation is carried out off-line, by using previously performed measurements on the real plant.

In this case, the following elements are supposed to be known:

- the plant structure, for example having the form of a transfer function with unknown parameters;
- a number *Ne* of sample vectors, as pears of input signals *u(t)* and output signals *y(t)* obtained by measurements;
- the maximum recording time t_{max} of the signals collected from the real plant.

By using this input data, the evaluation of an individual is carried out in a number of *Ne* phases, using a Simulink model for simulation.

The used Simulink model corresponds to the principle scheme in Figure 2, in which the Model block has the transfer function with the same form as the Plant block, but the Model block has adjustable parameters collected from the GAs and the Plant block is replaced by the measurements performed on the real Plant.

3.3 Possbilities to Extend the Proposed Method to Systems Having Unknown Structure

Although in the previous paragraph GAs were used only to estimate the parameters of systems having a known structure, this approach can be extended to identify some systems having an unknown structure.

The procedure described in 3.2 can be applied to systems having the transfer function given by relation (6).

$$H_P = \frac{K_p}{T_p s + 1} e^{-\tau s} , \qquad (6)$$

This transfer function has a minimum order for the part without delay time. The coefficients K_p, T_p are τ the unknown parameters, which are to be estimated.

If by applying GAs in these conditions the obtained results are not satisfying, the rational part of the transfer function is completed with additional zeros and/or poles, until satisfying results are obtained [4].

By this way, the transfer function has the form given by relation (7), where the values of m and k are gradually enlarged in each step, and the procedure described in the paragraph 3.2. is applied.

$$H_P = K_p \cdot \frac{a_k s^k + a_{k-1} s^{k-1} + ... + a_1 s + 1}{b_m s^m + b_{m-1} s^{m-1} + ... + b_1 s + 1} \cdot e^{-\tau s} \quad (7)$$

In this situation, GAs will use a population having a high number of individuals and they will evolve for a great number of generations. Additionally, GAs need to apply some techniques intended to maintain the population diversity.

In the real situations, the computing process can be simplified, since the coefficients K_p and τ can be find experimentally without difficulties [4], such as these coefficients can become constants having known values, or parameters defined in a small interval, which takes in consideration the experimental identification errors, as in relation (8).

$$K_p \in [K_{p\exp} \cdot (1 - \frac{p_{1\%}}{100}), K_{p\exp} \cdot (1 + \frac{p_{1\%}}{100})],$$

$$\tau \in [\tau_{\exp} \cdot (1 - \frac{p_{1\%}}{100}), \tau_{\exp} \cdot (1 + \frac{p_{2\%}}{100})] \quad (8)$$

where $K_{p\ \exp}$ and $\tau_{\exp}$ represents the experimentally determined coefficients, and the constants p_1 and p_2 are established depending on the accuracy of the experimental measurements performed.

3.4 Case Study

The following problem of parameters estimation is considered:

- The system has the transfer function with the form of the relation (6);
- The unknown coefficients K_p, T_p and τ have values in the domain [0.1 …20].

As input signals u_j a number of k = 10 sample signals are used, exprimed by relation (9), where the coefficients $a_{jk}, \omega_{jk}, \varphi_{jk}$ are randomly chosen.

$$u_k(t) = \sum_{j=1}^{4} a_{jk} \sin(\omega_{jk} t + \varphi_{jk}) \quad (9)$$

The problem is implemented by simulation, accordingly to paragraph 2. In the Simulink scheme, a plant having the transfer function (10) is used.

$$H_P = \frac{1,5}{5s+1} e^{-10s} \quad (10)$$

GAs have to identify the coefficients in the relation (10) in such a way that, at the end of the algorithm we expect to obtain the solutions:

$K_p = 1{,}5$, $T_p = 5$, $\tau = 10$.

The block diagram of the Simulink model is given in Figure 7.

GAs have the following parameters:

- Population size N = 50;
- Tournament selection with the tour size T=5;
- Arithmetic crossover;
- Uniform mutation;
- Crossover probability 0,6;
- Mutation probability 0,1;
- Stopping criteria 15 generations.

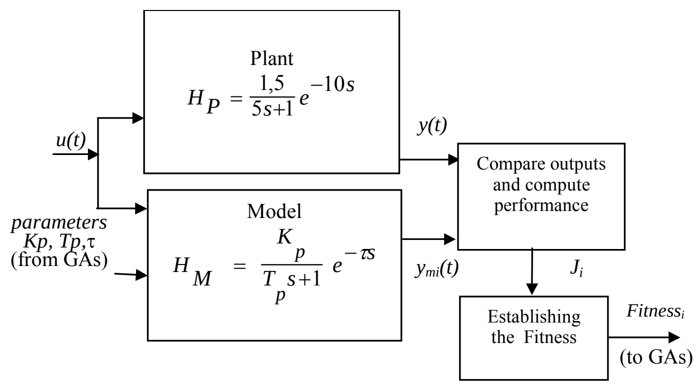

Figure 7
The block diagram of the Simulink model used in the case study

By applying GAs in the manner described in paragraph 3.2, the following solution was found:

$K_p = 1.4907$, $T_p = 4.4977$, $\tau = 9.7982$, *Fitness* = 0.96592

It can be senn that this solution approximates the searched parameters with an acuracy of 96%.

The genetic evolution is presented in Figure 8, where the the average Fitness is denoted M and the standard deviation with S. In this figure it can be seen that the average Fitness keeped its ascending charasteristic until the last generation. It can conclude that it is possible to obtain better solutions by increasing the number of the generations. For this reason, another run of GAs was performed, this time with 50 generations.

By applying GAs the following solution was found:

$K_p = 1{,}4999$

$T_p = 4{,}9916$

$\tau = 10{,}0049$

$Fitness = 0{,}9995$

It can be seen that this solution approximates the searched parameters with an acuracy of 99,9%.

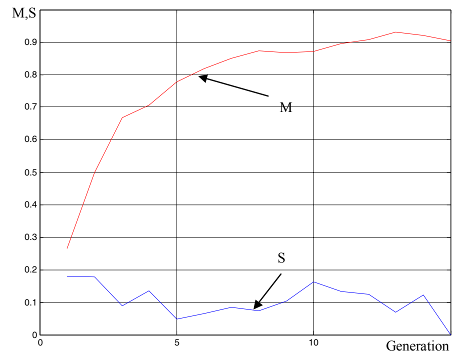

Figure 8

The GAs evolution in 15 generations

The genetic process is depicted in Figure 9. By analyzing this figure, it can conclude that the GAs parameters were well selected, since in the final generations a local search was performed, that lead to obtain an estimation with a higher acuracy.

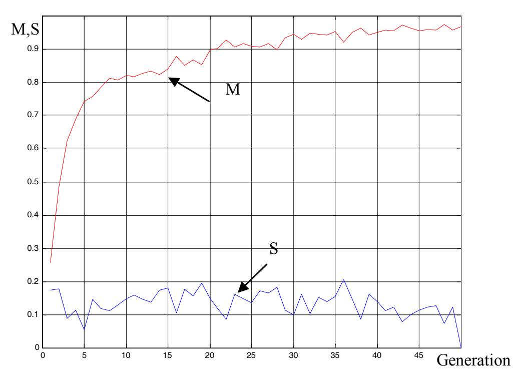

Figure 9

The GAs evolution in 50 generations

Conclusions

This paper presented a system identification method based on an evolutionary strategy. The genetic algorithm approach has shown to be versatile when applied to parameters estimation without requiring a detailed mathematical representation of the problem. The presented method is flexible and applicable in a wide range of problems. The obtained results show that GAs can estimate the parameters values with a high accuracy.

This work can be continued with other studies and experiments regarding the input data management or approaching other techniques, such as genetic programming, neural networks, fuzzy logic and combinations of the last ones and GAs [9], [10].

References

[1] J. Abonyi, J. Madar, L. Nagy, and F. Szeifert, "Interactive Evolutionary Computation in Identification of Dynamical Systems", *Comp.& Chem. Engineering, Carnegie Mellon University,* IF: 0.784, 2004

[2] R. Belea, "Contribution to the system identification using genetic algorithms", PhD Thesys, DUNĂREA DE JOS – University of Galați, 2004

[3] I. Dumitrache, and C. Buiu, Genetic Algorithms. *Fundamental principes and applications in control system*, Ed. Mediamira, Cluj-Napoca, 2000

[4] I. Dumitrache, *Techniques of control*, Ed. Didactic and Pedagogic, Bucureşti, 1980

[5] G. Franco, R. Betti, and H. Lus, "*Identification of Structural Systems Using an Evolution Strategy*", Journal of Engineering Mechanics, 130, 10, 1125-1139, 2004

[6] R. K. Ursem, "Models for Evolutionary Algorithms and Their Applications in System Identification and Control Optimization" PhD Thesys, Faculty of Science of the University of Aarhus, 2003

[7] E. E. Vladu, "Contributions in using genetic algorithms in engineering", PhD Thesis, "Politehnica" University of Timisoara, 2003

[8] E. E. Vladu, T. L. Dragomir, "On data management in elementary system identification approaches using Genetic Algorithms", Proceedings SOFA 2005

[9] A. Zakharov, and S. Halasz, "Parameter identification of a robot arm using Genetic Algorithms", *Periodica Politehnica Ser. Eng.* Vol. 45, No 3-4, pp 195-209, 2001

[10] Q. Zhao, B. H. Krogh, and P. Hubbard, "Generating test inputs for embedded control systems", *IEEE Control Systems Magazine*, 49-57, 2003