

# Crossover of genetic algorithm for linear and nonlinear system identification

F.A. Zainuddin<sup>1</sup>, M.F. Abd Samad<sup>1,\*</sup>

<sup>1</sup>Faculty of Mechanical Engineering, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia

\*Corresponding e-mail: mdifahmi@utem.edu.my

**Keywords:** System identification; genetic algorithm; crossover

**ABSTRACT** –System identification is the field of study involving the derivation of a mathematical model to explain dynamical behaviour of a system. Model structure selection is a problem in system identification which addresses selecting an adequate model i.e. a model that has a good balance between parsimony and accuracy in approximating a dynamic system. Genetic algorithm (GA) as a popular search method, is used for selecting a model structure. The advantage of using GA is in reduction of computational burden. This paper investigates the effect of different types of crossover in producing an optimum model structure. This shall be tested using computational software on simulated data, generated from linear and nonlinear model. As a conclusion, single point crossover produces the model with the best balance according to Akaike Information Criterion as objective function in both model types.

## 1. INTRODUCTION

System identification (SI) is a method of determining a mathematical relation between variables and terms of a process based on observed input-output data with the aim to enable better control of a system [1]. SI is done through 4 main steps involving data acquisition, selection of the model structure, parameter estimation and model validation [2]. An optimal model is normally described as having adequate predictive accuracy in the response to the system, yet parsimonious in structure. GA can work efficiently and deliver better results in the area of optimisation and search problems [3]-[4]. Samad & Nasir [5] used GA with single-point crossover to identify a discrete-time system based on a novel information criterion.

There are many varieties of crossover type e.g. single-point crossover, double-point crossover, multi-point crossover and uniform crossover [6]. Some researchers introduced new crossover types to solve their target problems with good quality solutions. Some researchers still use traditional crossover to deal with such problems. The purpose of this paper is to clarify the performances of traditional crossovers known in GA, specifically for use in binary representation of model structure selection problem in discrete-time system identification. Traditional crossover such as one-point, two-point, multi-point and uniform crossover has long been the benchmark to many other new developments in crossover mechanism introduction. Unsuitable crossover method may cause inefficiency to explore the space of possible solutions thoroughly and

effectively. It is expected that with various types of crossovers, better insight in exploring new search spaces may be gained, and thus further varying the offsprings, especially in the application of system identification.

## 2. METHODOLOGY

The methodology is to develop further the approach of using GA, in particular, the one used in [5]. GA can be considered as a multi-directional search method to solve problems as inheritance includes three evolutionary operations: reproduction, crossover, and mutation. Traditional gradient method however searches for solution in the search space only from a single direction. In GA, all the variables of interest must be encoded as binary digits (genes) in the traditional binary-coded GA, and a collection of binary digits further form a string (chromosome). The final binary digits, after an evolution, are decoded based on original problem representation after conversion of the chromosome to give the final solution.

To obtain a set of single-input-single-output data, a linear and a nonlinear model were simulated. Throughout the evolution, the criteria used for evaluation is Akaike information criterion (AIC) [7]. AIC is composed as:

$$AIC = n \ln \frac{RSS}{n} + 2k \quad (1)$$

where  $n$  is the number of observations,  $RSS$  (the residual sum of square) is the maximised value of the likelihood function for the estimated model and  $k$  is the number of parameters in the model. The identification simulation run were made 15 times for each crossover in order to come up with an average performance of crossover.

In this simulation, a linear and a Nonlinear AutoRegressive with eXogenous input (ARX and NARX, respectively) models were simulated using computer simulation software MATLAB. The models are denoted as model 1 and model 2. With,  $l$  = nonlinearity,  $n_u$  = maximum order of lag for input,  $n_y$  = maximum order of lag for output and  $n_k$  = time lag, the following are the models written as linear regression models, number of correct regressors, the search specification and number of possible models:

Simulated model 1 (ARX):

$$y(t) = 0.5y(t-1) - 0.2y(t-4) + 0.5y(t-8) + 0.6u(t-2) - 0.2u(t-9) + e(t) \quad (2)$$

Number of correct regressors = 5 out of 20

Search space specification:  $l = 1, n_y = 8, n_u = 8, n_k = 2$   
 Number of possible models = 1048575

Simulated model 2 (NARX):

$$y(t) = 0.4y(t-1) + 0.4u(t-1) - 0.6u(t-3) - 0.7y(t-1)u(t-1) - 0.2y^2(t-2) + 0.2u^2(t-3) + e(t) \quad (3)$$

Number of correct regressors = 6 out of 20  
 Search space specification:  $l = 2, n_y = 2, n_u = 3, n_k = 1$   
 Number of possible models = 1048575

### 3. RESULT and DISCUSSION

In this section, comparison of performance is made between different crossover types. Figures 1 and 2 show the results on model 1 and model 2 in terms of the attainment of objective function.

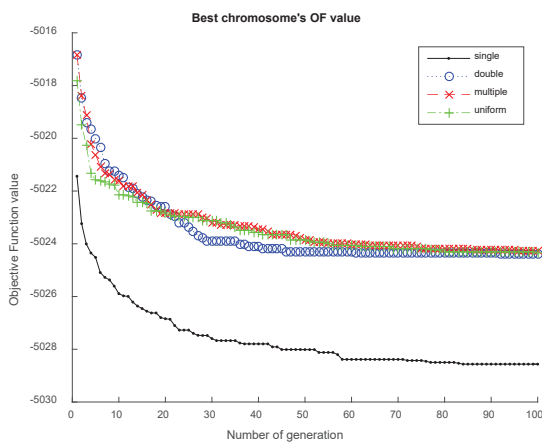


Figure 1 Best chromosomes's OF value for simulated model 1

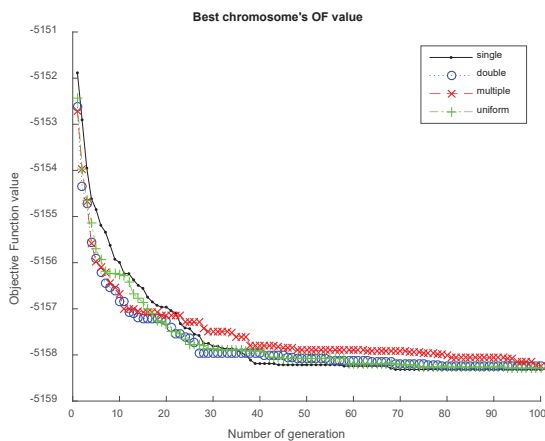


Figure 2 Best chromosomes's OF value for simulated model 2

Figure 1 shows that single-point produces model with the lowest OF compared to other crossover types. The lowest OF model has eight regressors. Other than all other correct regressors, the model included  $y(t-5)$ ,  $u(t-3)$  and  $u(t-7)$ . Figure 2 also indicates that single-point crossover produced the model with the lowest OF compared to other crossover types, even though seen slightly. Like for model 1, it is still unable to select the same model as the original model as, other than the original regressors, it included  $y^2(t-1)$  and  $y(t-1)y(t-2)$ , totaling to 8 regressors.

### 4. CONCLUSION

GA is an efficient approach to computing which is effectively used in various issues. Their output depends on the encoding scheme and the choice of genetic operators, in particular, the operators for selection, crossover and mutation. Before applying the operators to solve a new problem, it is essential to overview the search space and understand its modality. Throughout this test, the single-point crossover has been promising in producing models with needed balance between accuracy and parsimony based on Akaike information criterion for linear and nonlinear discrete-time system identification

### ACKNOWLEDGEMENT

The authors would like to thank Ministry of Higher Education Malaysia and Universiti Teknikal Malaysia Melaka (UTeM) for their financial support through FRGS/1/2018/TK03/UTEM/02/13 grant, technical advice and suggestion

### REFERENCES

- [1] V. Duong, A.R. Stubberud, Proc. IEEE Aerospace Conference, paper 270, March 2002, pp. 2331–2337.
- [2] M.F. Abd Samad, H. Jamaluddin, R. Ahmad, M. Shafiek, *International Journal of Intelligent Control and Systems*, Vol. 16, n. 3, pp. 182-190, 2011.
- [3] A. Mazahery, M.O. Shabani, *Metallurgical and Materials Transactions A*, Vol. 43, n. 13, pp. 5279-5285, 2012.
- [4] B. Koohestani, *Expert Systems with Applications*, Vol. 151, 113381, 2020.
- [5] M.F. Abd Samad, A.R.M. Nasir, *Journal of Fundamental and Applied Sciences*, Vol. 9, n. 7S, pp. 584-599, 2017.
- [6] M. F. Abd Samad, *International Review of Automatic Control*, Vol. 7, n. 2, pp. 208-216, 2014.
- [7] H. Akaike, *IEEE Transactions on Automatic Control*, Vol. 19, n. 6, pp. 716-723, 1974.