MULTIOBJECTIVE GENETIC PROGRAMMING: A NONLINEAR SYSTEM IDENTIFICATION APPLICATION

Katya Rodríguez-Vázquez **Carlos M. Fonseca** **Peter J. Fleming**

Department of Automatic Control and System Engineering,

University of Sheffield, Mappin Street, Sheffield S1 3JD, UK.

email: katya@acse.shef.ac.uk, C.Fonseca@shef.ac.uk, P.Fleming@shef.ac.uk

Abstract

Genetic programming (GP) is applied to a multobjective optimisation problem and the advantages of its hierarchical tree encoding scheme are compared with an earlier use of a subset representation approach which used string-encoded genetic algorithms. The GP approach is applied to the identification of non-linear system polynomial models and provides a trade-off between the complexity and the performance of the models. Multiobjective Genetic Programming is not only presented as a tool for optimising multiple objectives of the problem under investigation, but also provides a means of controlling the size of tree structures used in the search process.

1. INTRODUCTION

Conventional genetic programming (GP), and evolutionary algorithms (EA), in general, assign a single performance measure to each individual based on the evaluation of a scalar fitness function. However, these population-based methods possess the characteristic of simultaneously searching for multiple solutions and, further, can evaluate several aspects of the problem simultaneously.

In the real world, most problems may be described as multiple objectives which must be satisfied. These multiobjective problems tend to be characterised by a set of alternatives that must be considered equivalent in the absence of information concerning the importance of each objective relative to the others. Thus, multiobjective or multicriteria optimisation is the problem of finding the vector of decision variables, $\bar{x} = [x_1, \ldots, x_m]$, which optimises the n components vector function

$$\bar{f}(\bar{x}) = \left(f_1(\bar{x}), \ldots, f_n(\bar{x})\right) \qquad (1)$$

The problem has a set of equally efficient alternative solutions that is known as the Pareto-Optimal set (Ben-Tal, 1980). This concept of Pareto-optimality or nondominance is explained thus: given two n components decision vectors, $\bar{u} = (u_1, \ldots, u_n)$ and $\bar{v} = (v_1, \ldots, v_n)$, one can say that $\bar{u}$ dominates $\bar{v}$ (is Pareto-optimal) if

$$\forall i \in \{1, \ldots, n\},\ u_i \geq v_i \ \wedge\ \exists i \in \{1, \ldots, n\},\ u_i > v_i$$

Then, the set of all Pareto-optimal decision vectors is called the Pareto-optimal or admissible set of the problem.

A number of research works have used evolutionary algorithms with the aim of simultaneously optimising multiple functions. Goldberg (1989) pointed out that the first attempt of multicriteria optimisation using evolutionary algorithms was given by Rosenberg (1967). He suggested a multiple properties function for the simulation of a population of single-celled organisms. However, he only considered a single property in his simulation, but it was the beginning of further multicriteria evolutionary approaches.

Since Rosenberg's work, a variety of approaches have appeared (Fonseca and Fleming, 1995). One class of these is the aggregating methods which combine multiple objectives into one fitness measure in order to provide scalar fitness information on which evolutionary algorithms can work. Here, the use of genetic algorithms has been widely studied. However, multiperforming genetic programming seems to offer significant potential benefits.

Previous aggregating approaches using genetic programming have had the aim of generating parsimonious computer programs by incorporating the number of nodes included in their associated tree representation into the fitness measure (Koza, 1992). Iba et al. (1994) and Zhang and Mühlenbein (1996) have proposed a Minimum Description Length (MDL) based fitness function which is a trade-off of the model code length and the error code length. Siegel and Chaffe (1996) have introduced a different approach that incorporates a time constraint in order to penalise evolved programs that take excessive execution time.

Other multiobjective evolutionary approaches do not use a combined fitness function. Instead, they assign fitness measure on the basis of a separate evaluation of each of the multiple objectives. An example is the MultiObjective Genetic Algorithm (MOGA) approach proposed by Fonseca and Fleming (1993). This is a Pareto-based technique which is based on the concept of nondominance or Pareto-optimality. Here fitness of an individual is assigned on the basis of relative non-dominance, i.e. all nondominated individuals in the population are assigned rank 1, those individuals dominated by one or more points are ranked 2 or higher. Fonseca and Fleming's approach also includes schemes to combat the formation of lethals (mating restriction) and genetic drift (fitness sharing) - special problems arising from the evolution of a set of Pareto-optimal solutions.

The MOGA approach is extended here to its use in conjunction with genetic programming to provide a multicriteria tool in which we can include an alternative way of controlling the growth of genetic programming parameters in the growth tree. The next section presents a general description of MOGA.

2 MULTOBJECTIVE GENETIC ALGORITHM

The multiobjective genetic algorithm approach proposed by Fonseca and Fleming (1993) uses on a rank-based fitness assignment, where the rank of a certain individual x_i at generation t is related to the number of individuals $p_i(t)$ in the current population by which it is dominated. It can be given by

$$rank(x_i, t) = 1 + p_i(t) \qquad (2)$$

All non-dominated individuals are assigned rank 1 and remaining individuals are penalised according to equation (2).

2.1. Fitness Assignment

Fitness is assigned by interpolating from the best individual (rank=1) to the worst, and then the fitness assigned to individuals with the same rank is averaged where the global population fitness is kept constant. However, this fitness assignment tends to produce premature convergence (Goldberg, 1989) due to the fact that all non-dominated (best rank) points are considered equally fit.

In order to overcome this deficiency, Fonseca and Fleming have used a niche induction method to assist the distribution of the population over the Pareto-optimal front in order to maintain diversity. This is done by the method of fitness sharing which promotes the reproduction of isolated individuals and favours diversification.

2.2. Preference Information.

Preference information is introduced in the form of a goal vector, which provides a means of evolving only a specific region of the search space. This allows the decision maker to focus on a region of the Pareto front by providing external information to the selection algorithm.

The ranking procedure described previously is modified to introduce the goal information by altering the way in which individuals are compared with one another. Degradation in vector components that meet their goals is acceptable here, provided it results in the improvement of other components that do not satisfy their goals and provided that goal boundaries are not violated. This allows discrimination between individuals (solutions) even though they are non-dominated. This concept is formalised in terms of a transitive relational operator (preferability), instead of the simple partially less than operator, based on Pareto-dominance. The preferability operator additionally takes into account whether or not the objectives meet their goals.

The combination of the notion of preferability coupled with the concept of nondominance introduces a preference articulation framework for multiobjective and constraint optimisation. Because the rank-based fitness assignment and fitness sharing of MOGA takes place in the objective function domain instead of the parameter domain, they can be directly applied to genetic programming. Thus, the structure of MOGA can be mapped onto genetic programming by introducing a hierarchical tree representation with its associated genetic operators.

3. GENETIC PROGRAMMING AND SYSTEM IDENTIFICATION

System identification is defined as the process of constructing a mathematical model from observations and prior knowledge. However, the problem of identifying non-linear model structures cannot be formulated using only a single criterion. Identification involves diverse characteristics as linearity, degree of non-linearity, model structure, performance and model validation, which have to be considered.

In earlier work (Fonseca and Fleming, 1996), genetic algorithms have been applied to identify non-linear model structures where system identification has been formulated as a subset selection problem. More recently, Rodríguez et al. (1997), have used a tree-structured representation (Koza, 1992) to this problem. This approach is based on an input-output model that describes the input-output relationship of a system. To represent these systems, Leontaris and Billings (1985) have introduced a non-linear difference equation model known as the NARMAX (Non-linear Autoregressive Moving Average with eXogenous inputs) model. This model represents a wide class of non-linear systems. Leontaris and Billings also claim that the polynomial representation is one of the most common and has been demonstrated to work well in practical applications. A polynomial NARMAX model of degree l can be expressed by

$$y(t) = \theta_0 + \sum_{i_1=1}^{n} \theta_{i_1} x_{i_1}(t) + \sum_{i_1=1}^{n} \sum_{i_2=i_1}^{n} \theta_{i_1}\theta_{i_2} x_{i_1}(t) x_{i_2}(t)$$
$$+ \ldots + \sum_{i_1=1}^{n} \cdots \sum_{i_l=i_{l-1}}^{n} \theta_{i_1} \ldots \theta_{i_l} x_{i_1}(t) \ldots x_{i_l}(t) + e(t) \qquad (3)$$

where $n=n_y+n_u+n_e$ (the sum of the corresponding output, input and noise maximum lags), θ_i's are scalar coefficients and $x_i(t)$ represents lagged terms in y, u and e. $e(t)$ is a zero mean independent sequence. This model is classified as non-linear in the input and output variables but linear in the parameters; the coefficients can be estimated by a least squares algorithm. Therefore, here the solution approach is formulated such that genetic programming is used for identifying the model structure of the non-linear system and least squares regression is applied to estimate the associated coefficients.

Potential model structures of the form of equation (3) can then be represented in a hierarchical tree structure as shown in Figure 1. This model is expressed in Polish notation as

(ADD (ADD X1 X4) (MULT (ADD X2 X3) (ADD X1 X2)))

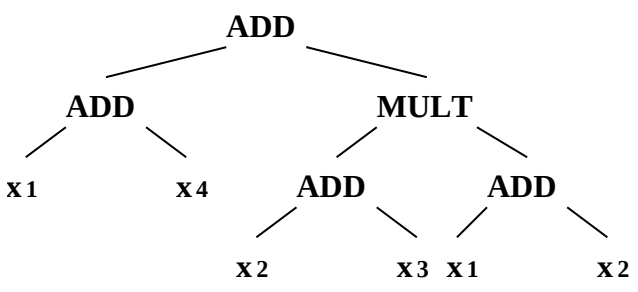

Figure 1. Tree-structured encoding for a non-linear polynomial model.

This expression represents the non-linear model defined as

$$y(t) = \theta_0 + \theta_1 y(t-1) + \theta_2 y(t-2) + \theta_3 u(t-1)$$
$$+\theta_4 y(t-1)^2 + \theta_5 y(t-1) y(t-2),$$

where X1, X2, X3 and X4 are equal to 1.0 (constant term), y(t-1), y(t-2) and u(t-1), respectively. Then, a least squares algorithm is applied to compute the coefficient vector θ. In order to calculate the associated coefficients, Iba et al. (1994) have also proposed a least squares method and Sharman et al. (1995) have proposed a simulated annealing-based tool in order to estimate these constant values.

3.1. Fitness Evaluation

In Rodríguez-Vázquez et al (1997) fitness is based on the evaluation of an information criterion that is a compromise function between the non-linear model structure and the performance of the model (minimisation of the residual variance). This is expressed as

$$IC = ln(\hat{\sigma}^2_{\hat{\theta}_p}) + \frac{4p}{N} \qquad (4)$$

where $\hat{\sigma}^2_{\hat{\theta}_p}$ is the variance of the residuals associated with the p model terms, $\hat{\theta}_p$ is the p estimated coeffiecnt vector, and N is the number of observed data points.

However, this approach does not consider the validation process. Thus, with the aim of the aforementioned multiobjective and constraint optimisation framework, the GP/NARMAX approach is extended to a multicriteria form.

3.2. Multiobjective Genetic Programming (MOGP)

MOGP is applied to 1000 input-output pairs of data generated by a pilot scale level-system (Voon, 1984), excited by a zero mean Gaussian input-signal. Data is divided into two groups. The first 500 points are used in the identification process, while the last 500 points are used for validation. Here, a polynomial NARX (Non-linear AutoRegressive with eXogenous inputs) model was considered. This is a special case of the general NARMAX model where $n_e=0$.

Applying MOGP to the problem of identifying polynomial NARX model structures, the terminal set consists of all the linear terms with maximum input and output lags defined by n_u and n_y, respectively, and the real value 1.0 representing the constant term. Thus, the terminal set was

$$\mathcal{T} = \{Xo, \ldots, Xn_y+n_u)\}$$

where $n_u = n_y = 10$, and Xo, ..., Xn_y+n_u are linear terms as expressed in Table 1.

Table 1. Terminal Set.

Element of $\mathcal{T}$	Linear Term
X_0	1.0
X_1	y(t-1)
...	...
Xn_y	$y(t-n_y)$
Xn_{y+1}	u(t-1)
...	...
Xn_y+n_u	$u(t-n_u)$

In the case of polynomial models, the function set was composed of just two functions. Thus, the function or non-terminal set is defined by

$$\mathcal{F} = \{ \text{ADD, MULT} \}$$

With these two functions, one is able to build any polynomial model of the form of equation (3).

In the first run, seven objectives were evaluated. These were: the number of terms in the model, model degree, model lag, the residual variance (VAR) and the long-term prediction error (LTPE). The autocorrelation function (ACF) of the residuals and the crosscorrelation function (CCF) between the input and the residuals, were also included. These latter two functions were treated as hard objectives which had to be satisfied when optimising the remaining objectives. In terms of system identification, the model had to be validated before it was considered a 'good' representation of the dynamic system. With respect to the goal information, setting all the goals to $\pm\infty$ forces the method to attempt to evolve a discretised description of the whole Pareto set. Thus, the first five goal values were set to the maximum value of their range. ACF and CCF were assigned a $1.95/\sqrt{N}$ goal value which corresponds to the 95% confidence interval. Note, however, that the multiobjective optimisation framework permits changes to the goals of the objectives.

In a second experiment, the information criterion (equation 4) was included instead of residual variance. This criterion is an aggregating function that implicitly involves the optimisation of the number of model terms. The objective regarding the *number of model terms* was, for this reason, 'ignored', but it remains in Fig. 2 for monitoring purposes. Both runs 1 and 2 may be considered similar. However the second approach provides a better determination of the number of model terms by reaching the minimum of the IC curve, as shown in Rodríguez et al. (1997). Results of this run are shown in Figure 2.

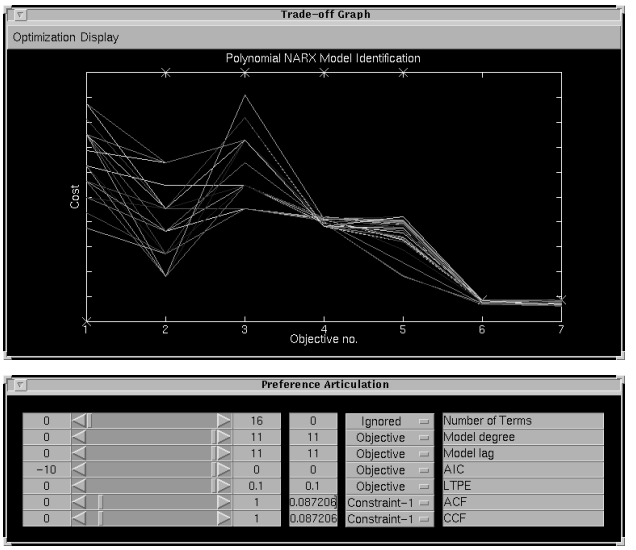

Figure 2. Trade-off graph of run 2.

Since tree size plays an important role in genetic programming, a third run was carried out. Here, the size of the tree was controlled by introducing the GP parameter of maximum number of nodes as an objective. Figure 3 presents the resulting trade-off graph.

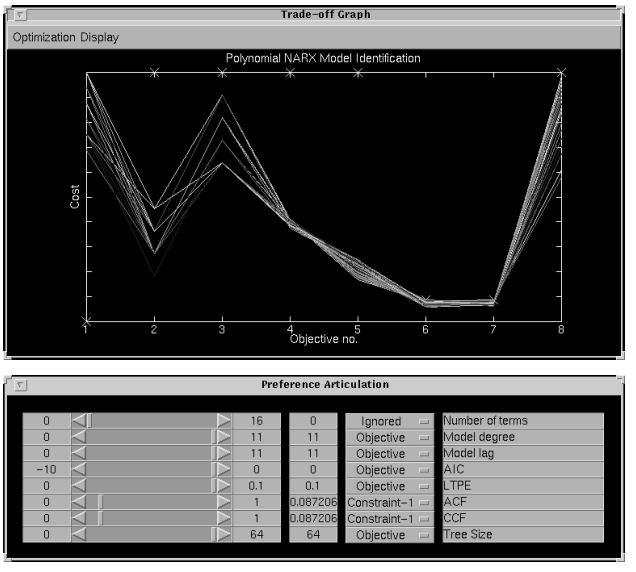

Figure 3. Trade-off graph of run 3.

As in previous runs, the goals were set up to take their maximum values. However, introducing a new goal, in this case, the tree size objective, the search process is encouraged to generate parsimonious models (Figure 4).

A population size of 100 individuals and probabilities of 0.9 and 0.1 for crossover and mutation, respectively, were used in all of these experiments.

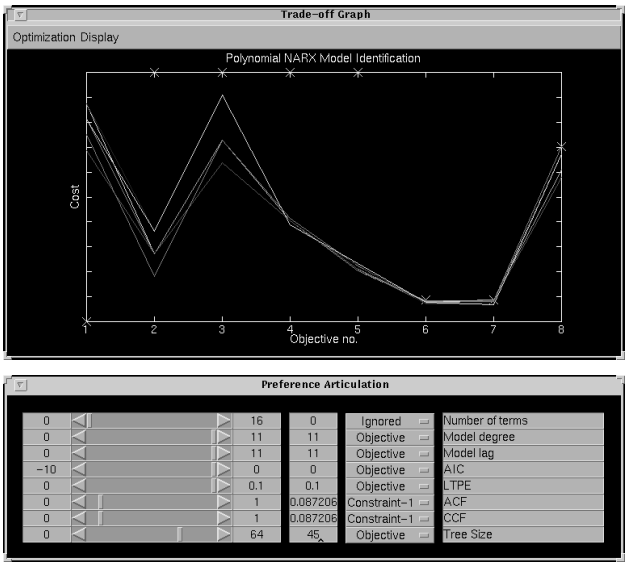

Figure 4. Trade-off graph of run 3 with a modified tree size goal.

4. DISCUSSION

The previous section introduced a tree-structured representation into the multicriteria optimisation framework. Two different aspects have been emphasised and are discussed below.

4.1. Improved Performance of MOGP

As can be seen from runs 1 and 2, the multiperforming genetic programming tool provides a set of potential models which can address separate objectives concerning the complexity, flexibility and validity of a model. It also provides the opportunity of manipulating the family of solutions by changing goals values of the objective functions depending on the purpose of the identification procedure.

While the framework has previously been applied to identification problems using the subset representation approach (Fonseca and Fleming, 1996), the hierarchical tree encoding used here appears to be more powerful. Because, in subset selection, the total set of all possible linear and non-linear terms and the maximum length of the chromosome, corresponding to the maximum number of terms, have to be set up a priori, the richness of the exploration of the search space is restricted. This situation could be ameliorated if the total set contains a wide range of terms but, it has been found that, if the number of terms is increased, the identification process can become unnecessarily time consuming. Even though the GP approach is also restricted by the number of nodes permissible in a tree, the search space is still extremely large and its variable size and dynamic representation provides diversification in the population. Thus, it is important to note that MOGP was able to evolve a tree structure of variable size which results in models that have better prediction capabilities than models resulting from the subset selection approach. Setting the goals for model degree and model lag to 3 and 8, respectively, MOGP approach provides better performing models in terms of residual variance (implicitly computed by means of the information criterion) and long term prediction error, than models produced by the previous multiobjective subset selection genetic algorithm. These comparative results are summarised in Table 2. As can be seen an increase in the performance has been realised by models with a slightly higher degree and number of terms. Note that, in all cases, the hard objectives (ACF and CCF functions) are satisfied.

4.2. Restricting tree size

Optimising the information criterion IC, is by itself a form of controlling the tree size. However, a particular model may have more than one tree structure representation, one of them embodying duplicated terms or redundant nodes. This is reflected in an increase of the tree's length and the time taken to evaluate the associated program.

In the third run, an additional objective was considered which had the aim of penalising tree structures that included redundant or non-functional nodes. In the trade-off graph shown in Figure 4, the goal for *tree size* objective was set to a lower value than the maximum range. This run tended to discourage redundancy which meant smaller tree structures. This, however, did not lead to decrease in performance.

5. CONCLUSIONS AND FUTURE WORK

This paper has presented an exploration in the area of multiobjective genetic programming providing an alternative to control the search process by restricting the search space.

Further work will consider the control of genetic programming parameters in general by means of this approach. It is also considered the exploration of parallel genetic programming schemes in this multiobjective algorithm.

This application has confined itself to non-linear polynomial models. In new work, non-linear rational models are being considered using this genetic programming based identification technique.

Table 2. Comparative results of MOGP and MOGA applied in non-linear system identification.

Model	No.Terms	Degree	Lag	AIC	LTPE	ACF	CCF	VAR$_{x10^{-3}}$
MOGP_1	12	3	8	-6.2311	0.0201	✓	✓	1.7872.
MOGP_2	12	3	8	-6.2307	0.0190	✓	✓	1.7879
MOGP_3	10	3	8	-6.1754	0.0283	✓	✓	1.9200
MOGP_4	11	4	6	-6.0660	0.0291	✓	✓	2.1249
MOGP_5	12	3	8	-6.0419	0.0179	✓	✓	2.1590
MOGA_1	8	2	8	-5.86768	0.0262	✓	✓	2.6540
MOGA_2	9	2	7	-5.8555	0.0249	✓	✓	2.6653
MOGA_3	9	2	6	-5.80792	0.0652	✓	✓	2.7951
MOGA_4	8	2	7	-5.80560	0.0472	✓	✓	2.8240

ACKNOWLEDGEMENTS

Katya Rodríguez-Vázquez gratefully acknowledges the financial support of Consejo Nacional de Ciencia y Tecnología (CONACyT). We also wish to acknowledge the support of the Engineering and Physical Sciences Research Council (UK) through the grant GR/K31343.

REFERENCES

Ben-Tal, A.(1980) Characterization of Pareto and Lexicographic Optimal Solutions, Multiple Criteria Decision Making, Theory and Applications, Vol. 77 of Lecture Notes in Economics and Mathematical Systems, (Fandel and Gal, eds.), 1-11.

Fonseca, C.M. and P.J. Fleming. (1993) Genetic Algorithms for Multi-Objective Optimization: Formulation, Discussion and Generalization. Proc. of the Fifth Int. Conf. on Genetic Algorithms, (Forrest, Ed.), 416-423.

Fonseca, C.M. and P.J. Fleming. (1995) An Overview of Evolutionary Algorithms in Multiobjective Optimization, *Evolutionary Computation* **3(1)**, 1-16.

Fonseca, C.M. and P.J. Fleming. (1996) Nonlinear System Identification with Multiobjective Genetic Algorithms, *Proc. of the 13th World Congress of IFAC*, San. Fco., 187-192.

Goldberg, D.A. (1989) Genetic Algorithms in Search, Optimisation and Machine Learning, Addison-Wesley.

Iba, H., H. de Garis and T. Sato. (1994) Genetic Programming Using a Minimum Desciption Length Principle, Advances in Genetic Programming, (Kinnear, Ed.), 265-284.

Koza, J.R. (1992) *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, MIT Press.

Leontaris,I.J. and S.A. Billings. (1985) Input-Output Parametric Models for Non-Linear Systems. *Int. J. Control* **41(2)**, 311-341.

Rodríguez-Vázquez, K., C.M. Fonseca and P.J. Fleming. (1997) An Evolutionary Approach to Non-Linear System Identification, to appear in *11th IFAC Symposium on System Identification*.

Rosenberg, R.S. (1967) *Simulation of Genetic Populations with Biochemical Properties*. PhD Thesis, University of Michigan.

Sharman, K.C., A.I. Esparcia-Alcázar and Y.Li. (1995) Evolving Signal Processing Algorithms by Genetic Programming, *Proc. of the IEE/IEEE GALESIA*, 473-480.

Siegel, E.V. and A.D. Chaffe. (1996) Genetically Optimizing the Speed of Programs Evolved to Play Tetris, *Advances in Genetic Programming 2*, (Angeline and Kinnear, Eds.), 279-298.

Voon, W.S.F. (1984) Nonlinear Parameter Estimation Techniques. PhD Thesis. Dept. Automatic Control and Systems Eng., Univ. of Sheffield, Sheffield, U.K.

Zhang, B.T. and H. Mühlenbein. (1996) Adaptive Fitness Functions for Dynamic Growing/Pruning, *Advances in Genetic Programming 2*, (Angeline and Kinnear, Eds.), 241-256.