Algorithms for Minimal Model Structure Detection in Nonlinear Dynamic System Identification

K.Z.Mao S.A.Billings

Department of Automatic Control and Systems Engineering

University of Sheffield

Mappin Street, Sheffield S1 3JD

United Kingdom

Research Report No. 635

August 16, 1996

University of Sheffield

Algorithms for Minimal Model Structure Detection in Nonlinear Dynamic System Identification

K.Z.Mao S.A.Billings*
Department of Automatic Control and Systems Engineering
University of Sheffield
Sheffield S1 3JD, UK

Abstract

The minimal model structure detection (MMSD) problem in nonlinear dynamic system identification is formulated as a search for the optimal orthogonalization path. While an exhaustive search for a model with 20 candidate terms would involve 2.43×10^{18} possible paths, it is shown that this can typically be reduced to 2×10^3 by augmenting the orthogonal estimation algorithm with genetic search procedures. The MMSD algorithm provides the first practical solution for optimal structure detection in NARMAX modeling, training neural networks and fuzzy systems modeling. Based on the MMSD algorithm, a refined forward regression orthogonal (RFRO) algorithm is developed. The RFRO algorithm initially detects a parsimonious model structure using the forward regression orthogonal algorithm and then refines the model structure by applying the MMSD algorithm to the reduced model term set. The RFRO algorithm cannot guarantee to find the minimal model structure, but it is computationally more efficient than the MMSD algorithm and can find a smaller model than the forward regression orthogonal algorithm.

1 Introduction

Detecting the model structure or determining which terms to include in a model is vitally important in nonlinear system identification. Various approaches have been proposed to address this problem (Haber and Unbehauen 1990). One of the most efficient and most frequently used model structure detection techniques is the orthogonal algorithm (Korenberg *et al* 1988). The advantage of using the orthogonal algorithm is that the contributions of candidate terms are decoupled and consequently the significance of model terms can be measured based on the corresponding error reduction ratios. It is found, however, that the size of the error reduction ratios depends on the order in which candidate terms are orthogonalized into the regression equation. As a consequence simply orthogonalizing candidate model terms in an arbitrary order into the regression equation may produce

*corresponding author

incorrect information regarding the significance of terms. This problem can however be avoided by using the forward regression orthogonal algorithm which has been widely used in dynamic nonlinear systems identification (Billings *et al* 1988a, 1988b, Chen *et al* 1989, Billings and Zhu 1994a), radial basis function (RBF) neural networks training (Chen *et al* 1991, Chen and Billings 1992, Brouwn *et al* 1994, Zhang 1994, Arciniegus *et al* 1994) and fuzzy systems modeling (Wang and Mendel 1992, Jang and Sun 1993, Hohensohn and Mendel 1994, Wang and Langari 1995). However, this algorithm cannot guarantee that the model has minimal structure for a given accuracy because terms are selected on the basis of a local optimisation. This suboptimal property was discussed in Billings *et al* (1988b), and was recently re-analysed in Sherstinsky and Picard (1996). But neither of these studies suggested a solution to this problem.

When using the orthogonal algorithm to detect model structure previously selected terms can influence the selection of later terms. Therefore the selection of every term should be based on a global consideration. The minimal model structure detection problem can be considered as a search for the optimal *orthogonalization path* which is defined as the order in which candidate terms are orthogonalized into the regression equation. An intuitive method to address this optimal sequential decision problem is to detect the model structure using all the possible orthogonalization paths. This exhaustive search approach would guarantee finding the model with the minimal structure for a given accuracy. However, directly testing all possible paths one by one is totally impractical because the number of all possible orthogonalization paths for a model with n candidate terms is equal to $n!$ which becomes astronomical if n is large. Even for a model with only 20 candidate terms there are 2.4329×10^{18} possible paths. But detecting which terms to include in a model is important in many areas including NARMAX modeling, training neural networks and fuzzy systems model building and it is therefore important to investigate solutions to this problem.

In the present study a new minimal model structure detection (MMSD) algorithm is introduced by augmenting the orthogonal estimation procedure with a genetic algorithm. This maintains the simplicity of structure detection and parameter estimation based on the orthogonal estimator but uses the power of genetic algorithm to search for the optimal orthogonalization path without exhaustively testing every permutational posibility. For example a model with 20 candidate terms requires an total evaluation of 2×10^3 orthogonalization paths compared to the full set of 2.4329×10^{18} possible orthogonalization paths. Despite this massive reduction the amount of necessary computations is still large, but searching for the minimal model structure becomes practical for the first time.

Based on the MMSD algorithm, a refined forward regression orthogonal (RFRO) algorithm is developed. The RFRO algorithm consists of two steps. A parsimonious model structure is initially detected using the forward regression orthogonal algorithm and this

parsimonious model structure is then refined by applying the MMSD algorithm to the reduced model term set. The RFRO algorithm cannot guarantee to find the minimal model structure, but it is computationally more efficient than the MMSD algorithm and should find a smaller model than the forward regression orthogonal algorithm.

The paper is organized as follows. The forward regression orthogonal algorithm is briefly reviewed and the suboptimal property is further analysed in §2. In §3 the minimal model structure detection (MMSD) algorithm, which incorporates an optimal orthogonalization path search routine and an orthogonal parameter estimation and structure determination procedure, is developed for a wide class of nonlinear systems including polynomial NARMAX models, rational NARMAX models, radial basis function (RBF) neural networks and fuzzy inference systems. A refined forward regression orthogonal (RFRO) algorithm is developed in §4. Simulation examples are presented in §5 to demonstrate the performance of the new algorithms.

2 Forward regression orthogonal algorithm and the nonminimal model structure problem

2.1 A brief review of the forward regression orthogonal algorithm

Consider a dynamic nonlinear polynomial NARMAX (nonlinear autorgressive moving average model with exogenous input) model

$$\begin{aligned} y(k) &= F^l[y(k-1),\ldots,y(k-n_y),u(k-1),\ldots,u(k-n_u),e(k-1),\ldots,e(k-n_e)] + e(k) \\ &= \sum_{i=1}^{n} \varphi_i(k)\theta_i + e(k) \end{aligned} \tag{1}$$

where $F^l(\bullet)$ denotes a nonlinear polynomial of degree l, $\varphi_i(k)$ and θ_i denote nonlinear regressors and unknown parameters respectively, $\{e(k)\}$ is a white noise sequence with zero mean and finite variance. Notice that equation (1) can be used to represent a wide class of model types including radial basis function (RBF) neural network architectures (Chen and Billings 1992, and others) and fuzzy systems (Wang and Mendel 1992, Jang and Sun 1993, Wang and Langari 1995). Because the number of possible model terms in a NARMAX model is very large (typically hundreds), model structure detection is a vitally important problem in nonlinear systems identification. The orthogonal algorithm is one of the most efficient techniques which address this problem.

The first step in the orthogonal algorithm is to orthogonalize the terms in eqn (1) into an

auxilary orthogonal model

$$y(k) = \sum_{i=1}^{n} w_i(k) g_i + e(k) \tag{2}$$

where $w_i(k)$ and g_i denote orthogonal regressors and unknown parameters respectively. Several methods are available for the orthogonal transformation. Details of these methods can be found in Chen *et al* (1989).

Once the auxilary orthogonal model is obtained, parameter estimation and error reduction ratios can be computed using

$$\hat{g}_i = \frac{\sum_{k=1}^{N} w_i(k) y(k)}{\sum_{k=1}^{N} w_i^2(k)} \tag{3}$$

$$err_i = \frac{\hat{g}_i^2 \sum_{k=1}^{N} w_i^2(k)}{\sum_{k=1}^{N} y^2(k)} \tag{4}$$

The error reduction ratio values are used as a measure of the significance of each candidate model term. However, the size of error reduction ratios depends on the order in which candidate terms are orthogonalized into equation (2). Therefore simply orthogonalizing candidate terms into the orthogonal equation (2) in the order in which they happen to be written down in equation (1) may produce an incorrect evaluation of the terms significance. The forward regression orthogonal estimation algorithm was developed to solve this problem (Billings *et al* 1988a and 1988b). At the first step, all the terms $\varphi_i(k)$, $i = 1, 2, \ldots, n$ in eqn (1) are considered as the possible candidates for the first term in the orthogonal model eqn (2)

$$w_1^{(i)}(k) = \varphi_i(k) \qquad 1 \leq i \leq n$$

the parameters are estimated

$$\hat{g}_1^{(i)} = \frac{\sum_{k=1}^{N} w_1^{(i)}(k) y(k)}{\sum_{k=1}^{N} [w_1^{(i)}(k)]^2} \qquad 1 \leq i \leq n$$

and the corresponding error reduction ratios are computed

$$err_1^{(i)} = \frac{\sum_{k=1}^{N} [\hat{g}_1^{(i)} w_1^{(i)}(k)]^2}{\sum_{k=1}^{N} y^2(k)} \qquad 1 \leq i \leq n$$

The term corresponding to the maximum error reduction ratio, assume $\varphi_j(k)$, is selected as the first term $w_1(k)$ in the auxilary orthogonal model eqn (2).

At the second step in the forward regression procedure, all other terms, except $\varphi_j(k)$, are

considered as candidates to be orthogonalized into eqn (2). Compute

$$w_2^{(i)}(k) = \varphi_i(k) - \alpha_{12}^{(i)} w_1(k) \qquad 1 \le i \le n, \quad i \ne j$$

where

$$\alpha_{12}^{(i)} = \frac{\sum_{k=1}^{N} w_1(k)\varphi_i(k)}{\sum_{k=1}^{N} w_1^2(k)}$$

estimate the parameters

$$\hat{g}_2^{(i)} = \frac{\sum_{k=1}^{N} w_2^{(i)}(k) y(k)}{\sum_{k=1}^{N} [w_2^{(i)}(k)]^2} \qquad 1 \le i \le n, \quad i \ne j$$

and compute the corresponding error reduction ratios

$$err_2^{(i)} = \frac{\sum_{k=1}^{N} [\hat{g}_2^{(i)} w_2^{(i)}(k)]^2}{\sum_{k=1}^{N} y^2(k)} \qquad 1 \le i \le n, \quad i \ne j$$

The term with the maximum error reduction ratio is then selected to produce the second term $w_2(k)$. The above procedure is continued until the accuracy requirement is met.

Because the most significant term is selected at every step, the forward regression orthogonal algorithm provides a parsimonous model structure.

2.2 Analysis of the forward regression orthogonal algorithm

Definition 1: Orthogonalization Path

An orthogonalization path is defined as a vector $O = [o_1, o_2, \ldots, o_n]$ which represents the order in which candidate terms are orthogonalized into the regression equation (2), where $o_i \in \{1, 2, \ldots, n\}$, $i = 1, 2, \ldots, n$. The physical interpretation of the orthogonalization path O is that the i^{th} term that is orthogonalized into the orthogonal regression equation (2) is $\varphi_{o_i}(k)$.

Exchanging values between o_i yields different orthogonalization paths. For a model which has n candidate terms, the number of all possible orthogonalization paths amounts to $n \times (n-1) \times \ldots \times 2 \times 1$.

Definition 2: Minimal Model

The minimal model is defined as a model that meets the accuracy requirement and has the least necessary terms.

The error reduction ratios vary with the order in which the associated candidate terms are orthogonalized into the orthogonal equation. This suggests that different orthogonalization paths may give different results when applying the orthogonal algorithm to detect the model structure. Assume that the model size or the number of necessary terms corresponding to path O_j is n_j, the minimal model is the one that satisfies the two conditions

$$\sum_{i=1}^{n_j} err_i > h \tag{5}$$

$$n_{min} = \min(n_j) \qquad j = 1, 2, \ldots, m. \tag{6}$$

where h is the required approximation accuracy, m is the number of all possible orthogonalization paths.

Eqn (5) shows that the number of necessary model terms depends on the size of error reduction ratios of selected terms. The forward regression orthogonal algorithm therefore solves the minimisation problem eqn (6) by selecting the term with maximum error reduction ratio at each step. This appears to be equivalent to minimising the objective function equation (6), and consequently the model obtained has the least terms. But this is not quite true because former selected terms influence the selection of later terms so that a local optimal selection may not be optimal from a global sense. In fact the minimisation of eqn (6) is a very complex nonlinear dynamic programming problem where the selection of every term should be based on a global consideration.

The model of a dynamic system is not necessarily unique and this is another factor that affects minimal model structure detection in dynamic system identification. Consider the following NARX model (nonlinear autoregressive model with exogenous input)

$$y(k) = F^l[y(k-1), \ldots, y(k-n_y), u(k-1), \ldots, u(k-n_u)] \tag{7}$$

From model (7), the multi-step ahead predictions are given by

$$\begin{aligned} y(k+1) &= F^l[y(k), \ldots, y(k-n_y+1), u(k), \ldots, u(k-n_u+1)] \\ &= F^l[F^l[y(k-1), \ldots, y(k-n_y), u(k-1), \ldots, u(k-n_u)], \ldots, \\ &\quad y(k-n_y+1), u(k), \ldots, u(k-n_u+1)] \end{aligned} \tag{8}$$

$$\begin{aligned} y(k+2) &= F^l[y(k+1), y(k), \ldots, y(k-n_y+2), u(k+1), u(k), \ldots, u(k-n_u+2)] \\ &= F^l[F^l[F^l[y(k-1), \ldots, y(k-n_y), u(k-1), \ldots, u(k-n_u)], \ldots, y(k-n_y+1), \\ &\quad u(k), \ldots, u(k-n_u+1)], F^l[y(k-1), \ldots, y(k-n_y), u(k-1), \ldots, \\ &\quad u(k-n_u)], \ldots, y(k-n_y+2), u(k+1), u(k), \ldots, u(k-n_u+2)] \end{aligned} \tag{9}$$

By repeated application of $F^l[\bullet]$, a series of multistep ahead prediction equations can be obtained. These prediction equations can be written as

$$\begin{aligned} y(k) &= F^l[F^l[y(k-2),\ldots,y(k-n_y-1),u(k-2),\ldots,u(k-n_u-1)],\ldots, \\ &\quad y(k-n_y),u(k-1),\ldots,u(k-n_u)] \end{aligned} \tag{10}$$

$$\begin{aligned} y(k) &= F^l[F^l[F^l[y(k-3),\ldots,y(k-n_y-2),u(k-3),\ldots,u(k-n_u-2)],\ldots, \\ &\quad y(k-n_y-1),u(k-2),\ldots,u(k-n_u-1)],F^l[y(k-3),\ldots,y(k-n_y-2), \\ &\quad u(k-3),\ldots,u(k-n_u-2)],\ldots,y(k-n_y),u(k-1),\ldots,u(k-n_u)] \end{aligned} \tag{11}$$

Eqn (10)-(11) show that many other terms which are not included in model (7) are also relevant to the output. Thus, when applying the forward regression orthogonal algorithm to detect the model structure, the original model (7) will be obtained only if the error reduction ratios of the terms in model (7) are larger than that of other relevant terms. This condition is, however, system dependent and is genarally not true. To illustrate this, consider a manufactured example based on the second order linear system

$$y(k) = a_1 y(k-1) + a_2 y(k-2) + b_1 u(k-1) + b_2 u(k-2) + e(k) \tag{12}$$

where $y(k)$, $u(k)$ and $e(k)$ denote the output, input and noise respectively. The input is a uniformly distributed random sequence with amplitute ± 1, the noise is a normally distributed white random sequence with zero mean and 0.01 variance. Three pairs of parameters are considered

(I) $a_1 = -1.7$, $a_2 = -0.8$, $b_1 = 1$, $b_2 = 0.3$

(II) $a_1 = -1.7$, $a_2 = -0.9$, $b_1 = 1$, $b_2 = 1.5$

(III) $a_1 = -1.7$, $a_2 = -0.8$, $b_1 = 1$, $b_2 = 0.8$

Assume that the maximum lags of the input and output are both 5 so that the algorithm searches over the set

$$\begin{aligned} \mathcal{S}_- &= \{u(k-1),u(k-2),u(k-3),u(k-4),u(k-5)\} \\ &\quad \cup \{y(k-1),y(k-2),y(k-3),y(k-4),y(k-5)\} \end{aligned}$$

Applying the forward regression orthogonal algorithm to these systems, gives the model structures shown in Tables 1, 2 and 3 respectively. The minimal model structure is obtained in case (I). In case (II) and (III), besides the terms in the original model eqn (12) some others terms which are not in eqn (12) are also selected. Although the spurious terms tend to have much smaller coefficients, this condition is worthy of further investigation

and can briefly be explained as follows. Eqn (12) can be written as (noise term is omitted

Term	True Value	Estimate	Error Reduction Ratio (%)
$y(k-1)$	-1.7	-1.70	86.75
$u(k-1)$	1	0.99	7.33
$y(k-2)$	-0.8	-0.804	5.40
$u(k-2)$	0.3	0.3079	0.32

Table 1: Model structure selection for Case (I) (maximum lags=5)

Term	True value	Estimate	Error Reduction Ratio (%)
$y(k-1)$	-1.7	-1.701	69.30
$y(k-3)$	0	-0.003	10.86
$u(k-1)$	1	1.002	7.82
$u(k-2)$	1.5	1.51	6.65
$y(k-2)$	-0.9	-0.903	5.10

Table 2: Model structure selection for Case (II) (maximum lags=5)

Term	True value	Estimate	Error Reduction Ratio (%)
$y(k-1)$	-1.7	-1.69	67.35
$u(k-1)$	1	0.995	26.85
$y(k-4)$	0	-0.0014	2.64
$u(k-4)$	0	0.017	0.65
$u(k-3)$	0	-0.0012	0.58
$u(k-2)$	0.8	0.785	0.62
$y(k-2)$	-0.8	-0.785	0.43

Table 3: Model structure for Case (III) (maximum lags=5)

for simplicity of exposition)

$$\begin{aligned} y(k) &= a_1[a_1y(k-2)+a_2y(k-3)+b_1u(k-2)+b_2u(k-3)]+a_2y(k-2) \\ &\quad +b_1u(k-1)+b_2u(k-2) \\ &= (a_1^2+a_2)y(k-2)+a_1a_2y(k-3)+b_1u(k-1) \\ &\quad +(a_1b_1+b_2)u(k-2)+a_1b_2u(k-3) \\ &= (a_1^2+a_2)[a_1y(k-3)+a_2y(k-4)+b_1u(k-3)+b_2u(k-4)] \end{aligned} \tag{13}$$

$$
\begin{aligned}
& +a_1a_2[a_1y(k-4)+a_2y(k-5)+b_1u(k-4)+b_2u(k-5)] \\
& +b_1u(k-1)+(a_1b_1+b_2)u(k-2)+a_1b_2u(k-3) \\
= \; & (a_1^3+a_1a_2)y(k-3)+(a_2^2+2a_1^2a_2)y(k-4)+a_1a_2^2y(k-5) \\
& +b_1u(k-1)+(a_1b_1+b_2)u(k-2)+(a_1^2b_1+a_2b_1+a_1b_2)u(k-3) \\
& +(a_1^2b_2+a_2b_2+a_1a_2b_1)u(k-4)+a_1a_2b_2u(k-5) \qquad (14)
\end{aligned}
$$

Obviously many other terms $y(k-3)$, $y(k-4)$, $y(k-5)$, $u(k-3)$, $u(k-4)$ and $u(k-5)$ which are not included in model (9) may also be relevant to the output. Which terms are selected will depend on the system parameters, a_1, a_2, b_1 and b_2 in the above example.

The size of the allowed maximum lags in the input and output terms (and degree of nonlinearity for nonlinear systems) can also affect term selection. If the maximum allowed lags in the input and output are assumed to be 3, applying the forward regression orthogonal algorithm to Case (III) yields the model structure shown in Table 4. Tables 3

Term	True value	Estimate	Error Reduction Ratio (%)
$y(k-1)$	-1.7	-1.67	67.35
$u(k-1)$	1	1.00	26.85
$y(k-3)$	0	0.044	2.41
$u(k-3)$	0	-0.04	0.78
$u(k-2)$	0.8	0.77	1.52
$y(k-2)$	-0.8	-0.73	0.2

Table 4: Model structure for case (III) (maximum lags=3)

and 4 show that different model structures are obtained for the same system because of the different lags employed when forming the set of candidate terms. For any given system, the possibility of producing the nonminimal model is proportional to the number of candidate terms which is determined by the maximum allowed lags in the input, output, noise and the degree of nonlinearity. Equations (10) and (11) show that other relevent terms are usually associated with higher order nonlinear degree and larger lags. These relevant terms can be excluded from the candidate term set if a relatively low nonlinear degree and small number of maximum lags are used and as a consequence a smaller model can be obtained. Smaller maximum lags and a lower nonlinear degree should therefore be tried first from the point of view of building a smaller model.

3 The Minimal Model Structure Detection (MMSD) algorithm

3.1 Outline of the MMSD algorithm

The minimal model detection problem can be considered as a constrained nonlinear programming problem

$$n_{min} = \min(n_j) \qquad j = 1, 2, \ldots \tag{15}$$

subject to

$$\sum_{i=1}^{n_j} err_i > h \tag{16}$$

It was shown above that sorting regression terms according to the size of the error reduction ratios cannot minimise objective function (15). One way to solve the above problem is to detect the model structure using all the possible orthogonalization paths. Assume that a model has 3 candidate terms, $\varphi_1(k)$, $\varphi_2(k)$ and $\varphi_3(k)$

$$y(k) = \sum_{i=1}^{3} \varphi_i(k)\theta_i \tag{17}$$

Changing the order in which these 3 terms enter the regression equation (17) yields 6 orthogonalization paths

$$[1\ 2\ 3],\ [1\ 3\ 2],\ [2\ 1\ 3],\ [2\ 3\ 1],\ [3\ 1\ 2],\ [3\ 2\ 1]$$

Applying the standard orthogonal estimation algorithm (Korenberg *et al* 1988) to eqn (17) using these 6 orthogonalization paths yields 6 model structures, some of which may be the same. The minimal model is the one that has the least number of terms and meets the accuracy requirement. This exhaustive search approach guarantees finding the minimal model structure. However, this approach is difficult to implement in practice because the search space quickly becomes astronomically large. Even for a model with only 20 candidate terms, the all possible orthogonalization paths amount to 2.4329×10^{18}. Directly searching for the optimal path from all the possible paths is therefore impractical and this is why the forward regression orthogonal algorithm was introduced as a suboptimal solution to this problem.

An algorithm which addresses the optimal path search problem for minimal model structure detection (MMSD) is formulated by combining ideas of exhaustive search with the orthogonal estimation algorithm and genetic algorithms (GAs) (Goldberg 1989, Fonseca and Fleming 1995, and others). Parameter estimation and structure determination are done by the orthogonal estimation algorithm. The advantage of employing GAs to seach

for the optimal path is that GAs preserve the good paths in the search process. Consequently, the optimal path can be found before all possible paths have been tested. For example, for a model with 20 candidate terms, the GA routine typically evaluates 2×10^3 orthogonalization paths which is much less than the full set of 2.4329×10^{18} paths. Despite this massive reduction the amount of necessary computations is still large, but searching for the minimal model structure becomes practical for the first time.

The MMSD is a totally different algorithm from the forward regression orthogonal algorithm. In this new algorithm, the optimal orthogonalization path is determined by directly minimising the final objective (the model size) and the selection of every term is based on a global consideration. As a consequence, the MMSD algorithm should find the minimal model structure.

3.2 Details of the MMSD algorithm

The minimal model structure detection algorithm is an improvement of the exhaustive search idea which uses both the orthogonal estimation algorithm and genetic algorithms. Details of the minimal model structure detection algorithm are developed below.

3.2.1 Encoding

The individuals represent the orthogonalization paths, therefore each bit of an individual should represent the order of a corresponding term entering the regression equation (1). For example the code of an individual of a model with 8 candidate terms is of the following form

$$\overbrace{1}^{P_1}\ \overbrace{6}^{P_2}\ \overbrace{3}^{P_3}\ \overbrace{7}^{P_4}\ \overbrace{2}^{P_5}\ \overbrace{4}^{P_6}\ \overbrace{8}^{P_7}\ \overbrace{5}^{P_8}$$

where P_i denotes the i^{th} element in eqn (1). The physical interpretation of the above string is that the 1^{st}-8^{th} term that are orthogonalized into the regression equation (2) are respectively the 1^{st}, 6^{th}, 3^{rd}, 7^{th}, 2^{nd}, 4^{th}, 8^{th} and 5^{th} term in eqn (1).

3.2.2 Fitness function

The purpose of employing genetic algorithms is to find the minimal model. Under the condition of satisfying the required accuracy, the less the number of model terms, the better the model is. Therefore the fitness value should be inversely proportional to the

number of necessary model terms n_j

$$f_j \propto \frac{1}{n_j}$$

The windowing mapping scheme (Davis 1989) was employed in the present study to map the inverse relation. A zero or constant minimum fitness value is initially assigned to the worst individual, then each individual of the population is assigned a fitness value proportional to the number of selected terms.

$$f_j = f_{max} - \frac{f_{max} - f_{min}}{n_{most} - n_{least}}[n_j - n_{least}] \tag{18}$$

where n_{least}, n_{most}, and f_{min}, f_{max} denote the least number and most number of selected terms, and the minimum and maximum fitness values respectively.

The number of necessary terms is determined by using the standard orthogonal algorithm and the orthogonalization path represented by the corresponding individual. Consider an individual as follows

$$\overbrace{1}^{P_1}\ \overbrace{6}^{P_2}\ \overbrace{3}^{P_3}\ \overbrace{7}^{P_4}\ \overbrace{2}^{P_5}\ \overbrace{4}^{P_6}\ \overbrace{8}^{P_7}\ \overbrace{5}^{P_8}$$

The corresponding regression equation would be given by

$$y(k) = \theta_1\varphi_1(k)+\theta_2\varphi_6(k)+\theta_3\varphi_3(k)+\theta_4\varphi_7(k)+\theta_5\varphi_2(k)+\theta_6\varphi_4(k)+\theta_7\varphi_8(k)+\theta_8\varphi_5(k) \tag{19}$$

At the first step, let the first term in equation (19) be the first term of the auxilary orthogonal model

$$w_1(k) = \varphi_1(k) \tag{20}$$

estimate the orthogonal parameter and compute the error reduction ratio using

$$\hat{g}_1 = \frac{\sum_{k=1}^{N} w_1(k)y(k)}{\sum_{k=1}^{N} w_1^2(k)} \tag{21}$$

$$err_1 = \frac{\hat{g}_1^2 \sum_{k=1}^{N} w_1^2(k)}{\sum_{k=1}^{N} y^2(k)} \tag{22}$$

At the second step, orthogonalize the second term $\varphi_6(k)$ into the orthogonal equation

$$w_2(k) = \varphi_6(k) - \sum_{j=1}^{2-1} \alpha_{j2} w_j(k) \tag{23}$$

where

$$\alpha_{j2} = \frac{\sum_{k=1}^{N} w_j(k)\varphi_6(k)}{\sum_{k=1}^{N} w_j^2(k)}$$

estimate the parameter

$$\hat{g}_2 = \frac{\sum_{k=1}^{N} w_2(k)y(k)}{\sum_{k=1}^{N} w_2^2(k)} \tag{24}$$

and compute the error reduction ratio

$$err_2 = \frac{\hat{g}_2^2 \sum_{k=1}^{N} w_2^2(k)}{\sum_{k=1}^{N} y^2(k)} \tag{25}$$

The above procedure is continued until

$$\sum_{i=1}^{n_j} err_i > h \tag{26}$$

Remark

Notice that if the system is disturbed by colored noise, a procedure for noise modeling should be included in the orthognal algorithm to detect the model size corresponding to one specific individual. This can be done by the following iterative routine

(i) Initially set $e(k) = 0$.

(ii) Apply the standard orthogonal algorithm to eqn (1) using the orthogonalization path represented by the corresponding individual, estimate the parameters, compute the error reduction ratios, select the significant terms and calculate the number of necessary terms.

(iii) Compute one-step ahead preditions of the output using the selected terms at step (ii)

$$\hat{y}(k) = \sum \phi_i(k)\hat{\theta}_i \tag{27}$$

estimate the noise sequence and compute the variance of noise

$$\hat{e}(k) = y(k) - \hat{y}(k) \tag{28}$$

$$\sigma_e^2 = \frac{\sum_{k=1}^{N} \hat{e}^2(k)}{N} \tag{29}$$

(iv) Set $e(k) = \hat{e}(k)$ in eqn (1), repeat steps (ii)-(iii) until the variance of the noise

converges to a constant.

(v) The number of selected model terms at the last iteration is considered as the model size corresponding to the considered individual.

3.2.3 Reproduction

The *roulette wheel* approach is employed to implement the reproduction procedure in this study. Each string is allocated a slot of the roulette wheel subtending an angle proportional to its fitness at the center of the wheel. A random number in the range of 0 to 2π is generated. A copy of a string goes to the mating pool if the random number falls in the slot corresponding to the string. For a population with size l, the reproduction process is repeated l times and l strings go into the mating pool.

3.2.4 Crossover

Crossover is the most important genetic operator. This operator first randomly selects two strings from the mating pool, then exchanges the right parts of the two strings from a randomly selected crossover point. However, this crossover style can not be applied to path selection. Consider for example, two randomly selected parent strings

P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	
1	6	3	7	2	4	8	5	(I)
1	3	8	5	4	7	2	6	(II)

The parent strings

Exchanging the right parts of the two parent strings from a randomly selected crossover point, for example P_5, produces two off-springs

P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	
1	6	3	7	4	7	2	6	(A)
1	3	8	5	2	4	8	5	(B)

The off-springs

Obviously the fifth and the eighth term are missing in offspring (A), and the sixth and the seventh term are missing in offspring (B) due to the crossover operation. To deal with this problem, a new crossover operation was developed. Assume that two randomly

selected parent strings are given by (I) and (II) above. First randomly select the term to change the order in the regression equation, for example the seventh term. Then detect the position of this in the parent strings, *i.e.* P_4 and P_6. Respectively exchanging the terms at position P_4 and P_6 in each string, yields

$$\overbrace{1}^{P_1}\ \overbrace{6}^{P_2}\ \overbrace{3}^{P_3}\ \overbrace{4}^{P_4}\ \overbrace{2}^{P_5}\ \overbrace{7}^{P_6}\ \overbrace{8}^{P_7}\ \overbrace{5}^{P_8} \quad \text{(I)}$$

$$\overbrace{1}^{P_1}\ \overbrace{3}^{P_2}\ \overbrace{8}^{P_3}\ \overbrace{7}^{P_4}\ \overbrace{4}^{P_5}\ \overbrace{5}^{P_6}\ \overbrace{2}^{P_7}\ \overbrace{6}^{P_8} \quad \text{(II)}$$

The off-springs

3.2.5 Mutation

Mutation is a local operator so that if the string is binary encoded, a 1 is replaced by a 0, and a 0 is replaced by a 1. Obviously, this approach cannot be applied to the optimal path search problem. In this study mutation is achieved by exchanging the selected bit with one of the other bits of the same string. Consider for example, the following string

$$\overbrace{1}^{P_1}\ \overbrace{6}^{P_2}\ \overbrace{3}^{P_3}\ \overbrace{7}^{P_4}\ \overbrace{2}^{P_5}\ \overbrace{4}^{P_6}\ \overbrace{8}^{P_7}\ \overbrace{5}^{P_8}$$

If bit P_1 is supposed to mutate, exchanging this bit with a randomly selected bit from $P_2 - P_8$, for example P_6, yields the mutated string

$$\overbrace{4}^{P_1}\ \overbrace{6}^{P_2}\ \overbrace{3}^{P_3}\ \overbrace{7}^{P_4}\ \overbrace{2}^{P_5}\ \overbrace{1}^{P_6}\ \overbrace{8}^{P_7}\ \overbrace{5}^{P_8}$$

The mutated string

3.2.6 Summary of the minimal model structure detection (MMSD) algorithm

The MMSD algorithm can be summarized as follows

(I) Generate an initial population set $\mathcal{P}$ consisting of l individuals, each individual represents an orthogonalization path. Set the current generation number $i = 1$.

(II) Apply the standard orthogonal algorithm to eqn (1) using the orthogonalization path represented by each individual, compute the error reduction ratios and calculate the corresponding model size and fitness value. Form a mating pool $\mathcal{M}$ using all individuals in the population set $\mathcal{P}$ at the probabilties assigned to each individual according to the corresponding fitness value.

(III) Randomly select a pair of parent strings from the mating pool $\mathcal{M}$. Choose a random crossover point and exchange the parent string bits to produce two offsprings and put the offsprings in the offspring set $\mathcal{O}$. Repeat this procedure $l/2$ times.

(IV) Mutate each bit of each offspring in the set $\mathcal{O}$ with a pre-specified mutation rate and calculate the fitness value of each mutated offspring using the procedure summarised in step (II).

(V) Select the l fitest individuals from sets $\mathcal{P}$ and $\mathcal{O}$ by comparing fitness values.

(VI) Reset the set $\mathcal{P}$ with the newly selected l individuals, reset the number of generations $i = i + 1$, and nullify the offspring set $\mathcal{O}$.

(VII) Steps (II)-(VI) are repeated until a pre-specified number of generations arrives.

3.3 MMSD algorithm for fuzzy system modelling

A fuzzy inference system is composed of a set of fuzzy if-then rules and a database containing membership functions with linguistic labels. Suppose that the fuzzy if-then rules are as follows

Rule 1 If x_1 is $\mathcal{X}_1^1$, x_2 is $\mathcal{X}_2^1$, ...,x_n is $\mathcal{X}_n^1$, then $y = f_1(x_1, x_2, \ldots, x_n)$

Rule 2 If x_1 is $\mathcal{X}_1^2$, x_2 is $\mathcal{X}_2^2$, ...,x_n is $\mathcal{X}_n^2$, then $y = f_2(x_1, x_2, \ldots, x_n)$

$\vdots$

Rule m If x_1 is $\mathcal{X}_1^m$, x_2 is $\mathcal{X}_2^m$, ...,x_n is $\mathcal{X}_n^m$, then $y = f_m(x_1, x_2, \ldots, x_n)$

The overall output is the weighted average of all the rules

$$y = \frac{\sum_{i=1}^{m} w_i f_i(x_1, x_2, \ldots, x_n)}{\sum_{i=1}^{m} w_i} = \sum_{i=1}^{m} f_i(x_1, x_2, \ldots, x_n)\theta_i \tag{30}$$

where

$$\theta_i = \frac{w_i}{\sum_{i=1}^{m} w_i}$$

and w_i denotes the membership value.

Eqn (30) can be viewed as a special case of a polynomial NARMAX model eqn (1), and the MMSD algorithm developed in §3.2 is therefore equally applicable to fuzzy systems modeling.

3.4 MMSD algorithm for radial basis function (RBF) network training

The input-ouput mapping of a radial basis function neural network can be described by

$$y = \sum_{i=1}^{n_c} \theta_i \varphi(||x - c_i||, \rho) \tag{31}$$

where θ_i, $i = 1, 2, \ldots, n_c$ are the weights, c_i, $i = 1, 2, \ldots, n_c$ are the center vectors, $\varphi(\bullet)$ is a radial basis function and ρ is its width, and $||x - c_i||$ is the Euclidean distance from x to c_i. $\varphi(||x - c_i||, \rho)$ is in fact the output of i^{th} node.

Aagin this can be interpreted as a form of model eqn (1), the MMSD algorithm can therefore be applied directly to RBF neural network traning.

3.5 MMSD algorithm for rational models

The nonlinear rational model formulation, defined as the ratio of two nonlinear polynomial expansions, was recently introduced as an alternative to the polynomial model (Billings and Chen 1988)

$$\begin{aligned} y(k) &= \frac{F_n[y(k-1)\ldots y(k-n_{ny}), u(k-1)\ldots u(k-n_{nu}), e(k-1)\ldots e(k-n_{ne})]}{F_d[y(k-1)\ldots y(k-n_{dy}), u(k-1)\ldots u(k-n_{du}), e(k-1)\ldots e(k-n_{de})]} + e(k) \\ &= \frac{\sum_{j=1}^{n_{num}} \varphi_{nj}(k)\theta_{nj}}{\sum_{j=1}^{n_{den}} \varphi_{dj}(k)\theta_{dj}} + e(k) \end{aligned} \tag{32}$$

where $F_n(\bullet)$ and $F_d(\bullet)$ are nonlinear polynomials functions, $n_\bullet$ denotes the order, $\varphi_\bullet(k)$ and $\theta_\bullet$ denote the regressor and the parameters respectively. The definitions of $y(k)$, $u(k)$ and $e(k)$ are same to that in the polynomial model eqn (1).

Genetic algorithms can be applied to a wide range of problems. But the MMSD algorithm cannot directly be applied to search for the minimal rational model because a linear-in-the-parameters regression equation is required when using the orthogonal algorithm to detect the model size in step (II). Eqn (32) can be multipled out to form a linear-in-the-parameters form (Billings and Zhu 1991, 1994a)

$$Y(k) = \sum_{j=1}^{n_{num}} \varphi_{nj}(k)\theta_{nj} - \sum_{j=2}^{n_{den}} \varphi_{dj}(k)y(k)\theta_{dj} + \sum_{j=1}^{n_{den}} \varphi_{dj}(k)\theta_{dj}e(k) \tag{33}$$

where

$$Y(k) = y(k)\varphi_{d1}(k)\theta_{d1}|_{\theta_{d1}=1}$$

But notice that now the residue $\sum_{j=1}^{n_{den}} \varphi_{dj}(k)\theta_{dj}e(k)$ becomes highly correlated with the regressor $\varphi_{dj}(k)y(k)\theta_{dj}$. This occurs even when the noise in eqn (32) is purely additive and white. As a consequence estimates based on eqn (33) will be biased if ordinary least squares algorithms are applied directly. Eqn (33) can be written as

$$\begin{aligned} Y(k) &= \sum_{j=1}^{n_{num}} \varphi_{nj}(k)\theta_{nj} - \sum_{j=2}^{n_{den}} \varphi_{dj}(k)[\overline{y}(k) + e(k)]\theta_{dj} + \sum_{j=1}^{n_{den}} \varphi_{dj}(k)\theta_{dj}e(k) \\ &= \sum_{j=1}^{n_{num}} \varphi_{nj}(k)\theta_{nj} - \sum_{j=2}^{n_{den}} \varphi_{dj}(k)\overline{y}(k)\theta_{dj} + \varphi_{d1}(k)e(k) \\ &= \phi_{free}(k)\Theta + \xi_1(k) \end{aligned} \tag{34}$$

where

$$\varphi_{free}(k) = [\varphi_{n1}(k) \ldots \varphi_{nn_{num}}(k), \ -\varphi_{d2}(k)\overline{y}(k) \ldots - \varphi_{dn_{den}}(k)\overline{y}(k)]$$

$$\xi_1(k) = \varphi_{d1}(k)e(k)$$

$$\overline{y}(k) = y(k) - e(k) = \frac{F_n(\bullet)}{F_d(\bullet)}$$

and $\overline{y}(k)$ is the current noise free part of the noisy output $y(k)$, which is uncorrelated with the residue $\xi_1(k)$. Billings and Mao (1996) showed that if a NARMAX smoothing algorithm is initially used to preprocess the raw data to yield the signal $\overline{y}(k)$ and the estimation is then performed based on eqn (34), unbiased parameter estimates can be obtained for the nonlinear stochastic rational model.

The current noise free output $\overline{y}(k)$ can also be obtained using filtering. The advantage of using filtering is that the estimation of $\overline{y}(k)$ and the model identification can be carried out using the same iterative routine, with no need for an additional signal preprocessor.

For a specific individual, the model size determination routine for the rational model can be summarized as follows

(II-1) Initially set $e(k) = 0$, $\overline{y}(k) = y(k)$.

(II-2) Apply the standard orthogonal algorithm to eqn (34) using the orthogonalization path represented by the individual, estimate the parameters, compute the error reduction ratios, select the significant terms and calculate the number of selected model terms.

(II-3) Construct a rational model using the selected terms at step (II-2)

$$y(k) = \frac{\sum \phi_{ni}(k)\hat{\theta}_{ni}}{\sum \phi_{di}(k)\hat{\theta}_{di}} \tag{35}$$

estimate the current noise free part of $y(k)$ and the unknown noise

$$\overline{y}(k) = \frac{\sum \phi_{ni}(k)\hat{\theta}_{ni}}{\sum \phi_{di}(k)\hat{\theta}_{di}} \tag{36}$$

$$\hat{e}(k) = y(k) - \overline{y}(k) \tag{37}$$

and compute the noise variance

$$\sigma_e^2 = \frac{\sum_{k=1}^{N} \hat{e}^2(k)}{N} \tag{38}$$

(II-4) Set $e(k) = \hat{e}(k)$ in eqn (34), repeat steps (II-2)-(II-3) until the variance of the noise converges to a constant.

(II-5) The number of selected model terms at the last iteration is considered as the model size of the corresponding individual.

If step (II) of the MMSD algorithm in §3.2.6 is replaced with steps (II-1)-(II-5), the MMSD algorithm is applicable for rational model identification.

4 A Refined Forward Regression Orthogonal (RFRO) algorithm

In theory the MMSD algorithm in section §3.2 should find the minimal model structure no matter how large the full model size is. However, the search space of the MMSD algorithm increases dramatically with the number of candidate model terms. For example, a model with 50 candidate terms has 3.04×10^{64} possible orthogonalization paths, and a model with 100 candidate terms has 9.33×10^{157} possible orthogonalization paths. Searching for the optimal solution in such a large space involves at least two difficulties. First, a large search space means a large amount of computation. Second, the larger the search space is, the more local minima there are, and hence a greater possibility that the solution will converge to a local minimum although in theory the GA algorithm should find the global near-optimal solution.

The forward regression orthogonal algorithm cannot guarantee that the model has the minimal structure, but all the terms of the minimal model should be contained in this non-minimal model if a conservative cut-off value is used. This is possibly because when a bad basis is selected by the forward regression orthogonal algorithm, the energy reduction of the latter terms in this basis will be small, and as a consequence the approximation accuracy will not be met until all the important terms are selected. This suggests that

if the forward regression orthogonal algorithm is initially used as a presearch to build a parsimonious model, the MMSD algorithm can then be applied to this reduced model term set to refine the final model structure so that the minimal model structure can be readily obtained. The above consideration motivates the following two-step Refined Forward Regression Orthogonal (RFRO) algorithm

(A) Detect the model structure using the forward regression orthogonal algorithm with a conservative cut off value.

(B) Refine the model structure by applying the MMSD algorithm to the reduced model term set selected in step (A).

Even if some terms of the minimal model are not selected in step (A), the above procedure can still be used to refine the model structure. Provided model validity tests (Billings and Zhu 1994b) are used to check for missing model terms, it should be possible to detect when more terms need to be added in step (A). It is therefore possible to apply the MMSD algorithm to a reduced model term set and to refine the model and to produce a computationally efficient algorithm for model structure detection.

5 Simulation Examples

Example 1

Consider the following nonlinear dynamic system

$$\begin{aligned} y(k) &= 0.2y^3(k-1) + 0.7u(k-1)y(k-1) + 0.6u^2(k-2) \\ &\quad -0.5y(k-2) - 0.7y(k-2)u^2(k-2) + e(k) \end{aligned} \tag{39}$$

with a uniformly distributed random input with zero mean and amplititute ± 1, and a normally distributed white noise sequence with zero mean and variance 0.0004.

Initially a parsimonious model structure was determined using the forward regression orthogonal algorithm. The maximum lags of both input and output were assumed to be 4 and the maximum degree of nonlinearity to be 3. The model structure and parameter estimates are shown in Table 5. A comparison with eqn (39) shows that an incorrect term $y(k-4)u^2(k-2)$ has been selected.

Applying the MMSD algorithm summarized in §3.2.6 to refine the model structure in Table 5 with a population size of 50, mutation rate 0.06, two cross-over points, and where each string consisted of 6 decimal numbers which represented the order in which the 6 candidate terms were orthogonalized into the regression equation. At the 15^{th} generation, the refined model structure and parameter estimation were as shown in Table 6.

Term	Estimate	Error Reduction Ratio (%)
$y(k-4)u^2(k-2)$	-0.023	46.95
$y(k-1)u(k-1)$	0.6742	10.31
$y(k-2)$	-0.4906	11.15
$u^2(k-2)$	0.6134	24.77
$y^3(k-1)$	0.1942	2.60
$y(k-2)u^2(k-2)$	-0.7536	2.64

Table 5: Model structure for Example 1 using the forward regression orthogonal algorithm (maximum lags 4, maximum degree of nonlinearity 3)

Term	True Value	Estimate	Error Reduction Ratio (%)
$y(k-1)u(k-1)$	0.7	0.6741	28.44
$y(k-2)u^2(k-2)$	-0.7	-0.7371	9.57
$y(k-2)$	-0.5	-0.4910	5.82
$u^2(k-2)$	0.6	0.6091	44.89
$y^3(k-1)$	0.2	0.1939	9.69

Table 6: Model structure for Example 1 using the RFRO algorithm (maximum lags 4, maximum degree of nonlinearity 3)

If the maximum lags in the input and output are set to 2, and the nonlinearity degree to 3. Applying the forward regression orthogonal algorithm yields the final model structure shown in Table 7.

In this case the forward regression orthogonal algorithm provides the minimal model structure because the restriction on the maximum lags has excluded the spurious term which appeared in Table 5.

Example 2

Consider the following dynamic nonlinear rational model

$$y(k) = \frac{y^3(k-1) + u(k-1)u(k-2) + e(k-1)}{1 + y^2(k-1) + y^2(k-2)} + e(k) \tag{40}$$

where the input $u(k)$ is as in Example 1, $\{e(k)\}$ is a normally distributed white noise sequence with zero mean and variance 0.01. Multiplying out the rational model, yields

$$y(k) = y^3(k-1) + u(k-1)u(k-2) + e(k-1) - \overline{y}(k)y^2(k-1) - \overline{y}(k)y^2(k-2) + e(k) \tag{41}$$

Term	True Value	Estimate	Error Reduction Ratio (%)
$y(k-2)$	-0.5	-0.4910	31.23
$u^2(k-2)$	0.6	0.6091	45.93
$y(k-1)u(k-1)$	0.7	0.6741	14.33
$y(k-2)u^2(k-2)$	-0.7	-0.7371	4.13
$y^3(k-1)$	0.2	0.1939	2.79

Table 7: Model structure for Example 1 using the forward regression orthogonal algorithm (maximum lags 2, maximum degree of nonlinearity 3)

where

$$\overline{y}(k) = y(k) - e(k)$$

The maximum lag was set to 4, and the maximum degree of nonlinearity to 3. Applying the MMSD algorithm summarized in §3.2.6 and the procedure summarized in §3.5 with a population size, mutation rate, number of crossover point and string structure as in example 1, gave the model structure and parameters shown in Table 8.

Numerator polynomial	True Value	Estimate	Error Reduction Ratio(%)
$u(k-1)u(k-2)$	1	1.07	78.51
$e(k-1)$	1	0.96	8.22
$y^3(k-1)$	0.5	0.48	0.70
Denominator polynomial	True Value	Estimate	Error Reduction Ratio (%)
$y^2(k-1)$	1	1.03	1.62
$y^2(k-2)$	1	0.92	0.98

Table 8: Model structure for Example 2 using the MMSD algorithm (maximum lags 4, maximum degree of nonlinearity 3)

Applying the forward regression orthogonal algorithm and the refined forward regression orthogonal (RFRO) algorithm to this problem, produced the same result as in Table 8. Obviously all three algorithms provided the minimal model structure in this example.

6 Conclusions

A new algorithm for Minimal Model Structure Detection (MMSD) has been derived based on the standard orthogonal algorithm and genetic search procedures. This allows, for the first time, the practical search for the optimal orthogonalization path in nonlinear dy-

namic system identification and provides a solution to the combined problem of model structure detection and parameter estimation. Although the new approach results in a massive reduction in computations compared to an optimal exhaustive search, the necessary computation is still quite large. To overcome this problem, a Refined Forward Regression Orthogonal (RFRO) algorithm has been developed. Simulated results were used to demonstrate the performance of the two new algorithms which can be used in NARMAX modeling, the configuration and training of radial basis function (RBF) neural networks and fuzzy model building.

7 Acknowledgement

SAB gratefully acknowledges that part of this work was supported by EPSRC.

References

[1] Arciniegus, J.I., A.H.Eltimshy and K.J.Cios, 1994, Identification of flexible robotic manupulators using neural networks, *Integrated Computer-Aided Engineering*, 1(3), 195-208.

[2] Billings, S.A., M.J.Korenberg and S.Chen, 1988a, Identification of output-affine systems using an orthogonal least squares algorithm, *International Journal of Systems Science*, 19, 1559-1568.

[3] Billings, S.A., S.Chen and M.J.Korenberg, 1988b, Identification of MIMO non-linear systems using a forward-regression orthogonal estimator, *International Journal of Control*, 49, 2157-2189.

[4] Billings, S.A. and Q.M.Zhu, 1994a, A structure detection algorithm for nonlinear dynamic rational models, *International Journal of Control*, 59, 1439-1463.

[5] Billings, S.A. and Q.M.Zhu, 1994b, Nonlinear model validation using correlation tests, *International Journal of Control* , 60, 1107-1120.

[6] Brouwn, G.G., A.Krigsman, H.B.Verbruggen, and P.M.Bruijn, 1994, Single layer networks for nonlinear system identification, *Engineering Applications of Artificial Intelligence*, 7(3), 227-243.

[7] Chen, S., S.A.Billings and W.Luo, 1989, Orthogonal least squares methods and their application to nonlinear system identification, *International Journal of Control*, 50, 1873-1896.

[8] Chen, S., C.F.N.Cowan, and P.M.Grant, 1991, Orthogonal least squares learning algorithm for radial basis function networks, *IEEE Transaction on Neural Networks*, 2, 302-309.

[9] Chen, S. and S.A.Billings, 1992, Neural networks for nonlinear dynamic system modelling and identification, *International Journal of Control*, 56, 319-346.

[10] Davis, L., Adaptive operator probabilities in genetic algorithms, *Proc. 3rd International Conference on Genetic Algorithms*, Morgan Kauffman Publisher, Cambridge, Massachusetts, 1989, 61-69.

[11] Fonseca, C.M. and P.J.Fleming, 1995, An overview of evolutionary algorithms in multiobjective optimization, *Evolutionary Computation*, 3, 1-16.

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

[13] Haber, R. and H.Unbehauen, 1990, Structure identification of nonlinear dynamic systems–a survey on input/output approach, *Automatica*, 26, 651-677.

[14] Hohensohn, J. and J.M.Mendel, 1994, Two-pass orthogonal least-squares algorithm to train and reduce fuzzy logic systems, *Proc. 3rd IEEE Conference on Fuzzy Systems*, Orlando, USA, 696-700.

[15] Jang, J.S.R. and C.T.Sun, 1993, Functional equivalence between radial basis function networks and fuzzy inference systems, *IEEE Transaction on Neural Networks*, 4, 156-159.

[16] Korenberg, M.J., S.A.Billings, Y.P.Liu and P.J.McIlroy, 1988, Orthogonal paramter estimation algorithm for nonlinear stochastic systems, *International Journal of Control*, 48, 193-210.

[17] Sherstinsky, A. and R.W.Picard, 1996, On the efficiency of the orthogonal least squares training method for radial basis function networks, *IEEE Transaction on Neural Networks*, 7, 195-200.

[18] Wang, L.X. and J.M.Mendel, 1992, Fuzzy basis functions universal approximation, and orthogonal least squares learning, *IEEE Transaction on Neural Networks*, 3, 807-814.

[19] Wang, L. and R.Langari, 1995, Building sugeno-type models using fuzzy discretization and orthogonal parameter estimation techniques, *IEEE Transaction on Fuzzy System*, 3(4), 454-458.

[20] Zhang, Q., 1994, Regression selection and wavelet network construction, *Proc. 32nd IEEE Conference on Decision and Control*, San Antonio, USA, 3688-3693.