An SVD-based Fuzzy Model Reduction Strategy

John Yen and Liang Wang
Center for Fuzzy Logic, Robotics, and Intelligent Systems
Department of Computer Science
Texas A&M University, College Station, TX 77843-3112

Abstract

This paper describes a novel fuzzy model reduction approach for overcoming the curse of dimensionality associated with high-dimensional data modeling problems. A numerically reliable orthogonal transformation technique, known as the singular value decomposition (SVD), is utilized to detect and select the dominant fuzzy rules from a rule base. The effectiveness of the proposed approach is illustrated using a nonlinear limit cycle modeling problem.

1. Introduction

A long-standing problem in fuzzy system modeling is the ***curse of dimensionality***, which occurs because the number of rules increases ***exponentially*** as the number of input variables increases. Several research efforts have been made to overcome the problem. For example, Takagi and Sugeno [24], Sugeno and Yasukawa [23] and Chiu [5] proposed several ***heuristic***-type algorithms to select the dominant input variables. Yen, Wang and Liao [33] suggested using ***principal component analysis*** in statistics to reduce the input dimension. These two efforts have focused on reducing the dimension of input space directly. Alternatively, Sugeno, Griffin and Bastian [21] constructed a ***hierarchical-structured*** fuzzy controller in which the number of fuzzy rules increases linearly rather than exponentially as the number of input variables increases.

Recently, Wang, Langari and Yen [29] proposed a fuzzy model reduction strategy (known as ***diagonal reduction***) based on ***Johansen's optimality theorem*** [13]. The basic philosophy of this approach is to select a subset of fuzzy rules from a complete rule base such that the resultant fuzzy rules satisfy the ***condition of completeness*** [31]. However, a possible drawback of this approach is that it may miss some dominant fuzzy rules in a rule base. This may sometimes lead to a large loss of accuracy.

In this paper we describe an alternative approach for fuzzy model reduction. A well-known orthogonal transformation technique, known as the ***singular value decomposition***, has been used to select the most dominant fuzzy rules from a rule base. We will illustrate the effectiveness of the proposed approach using a nonlinear system modeling and prediction problem.

2. Statement of the Problem

We will focus our discussion in this paper on the ***Sugeno-type model***[22][24], which has the following form

$$\begin{aligned} R_i: \quad & \textit{if } x_1 \textit{ is } A_{i1} \textit{ and } \dots \textit{ and } x_p \textit{ is } A_{ip} \\ & \textit{then } y = b_{i0} + b_{i1}x_1 + \dots + b_{ip}x_p \qquad (1) \\ & i = 1, 2, \dots, M \end{aligned}$$

where **p** and **M** are the number of input variables and rules respectively; y is the output variable; $\mathbf{A}_{ij}$(i=1,2,...,M, j=1,2,...,p) are the membership functions of input variables; b_{ij}(i=1,2,...,M, j=0,1,...,p) are the consequent parameters.

In our discussion we assume the model has a single output for notational simplicity; extension to the multiple-output case is straightforward, however.

Several methods have been suggested in the existing literature for identifying the parameters in Sugeno-type models. In their original paper, Takagi and Sugeno [24] proposed to use a combination of a nonlinear optimization algorithm with the well-known recursive least-square algorithm to determine the membership functions of input variables and the consequent parameters in the model. In a recent paper by Yager and Filev [32], a simplified identification approach was developed in which the concept of ***sample probability distribution*** was introduced. Also, in Jang [12], a back-propagation type algorithm was

0-7803-3645-3/96 $5.00©1996 IEEE

proposed to identify the model.

Various techniques based on statistics, clustering, and neural networks have been used to determine the membership functions of input variables. In this paper we apply ***B-splines*** to construct the membership functions. The motivation behind this is that B-splines are similar to fuzzy membership functions in both shape and mathematical properties. Moreover, they can be computed in a numerically reliable recursive procedure [6]. Once the membership functions of input variables are determined, the consequent parameters are then computed using the Kalman filtering algorithm.

We now can state the problem to be addressed in this paper using a two-input fuzzy system modeling example. The two input variables x_1 and x_2 have been assumed to have the same number of membership functions, say 6. Such a fuzzy model has a complete rule base for 36 rules. ***The goal is to select a subset of fuzzy rules from the complete rule base so that we can reduce the complexity of the fuzzy model without sacrificing the accuracy of the model significantly.***

Before we introduce our approach to fuzzy model reduction, we first briefly review a technique used in our approach - the singular value decomposition.

3. The Singular Value Decomposition

The singular value decomposition of an m-by-n matrix A is a factorization of A into a product of three matrices. That is

$$A = U \Sigma V^T \tag{2}$$

with

$$U^T U = V^T V = I_r \quad and \qquad \Sigma = diag(\sigma_1, \sigma_2, ..., \sigma_r) \tag{3}$$

where

$$r = \min(m, n) \tag{4}$$

The matrices U and V consist of the orthonormalized eigenvectors associated with the **r** largest eigenvalues of AA^T and A^TA, respectively. The diagonal elements of Σ are the nonnegative square roots of the **r** largest eigenvalues of AA^T and are called the ***singular values***.

Remark 1: The SVD here has been stated for the real case. For complex matrices the result is virtually identical with complex-conjugate transposes replacing transposes and unitary matrices replacing orthogonal matrices(see, e.g., [9]). For fuzzy model reduction, only the SVD for real matrices is needed.

SVD is one of the most useful and powerful tools of modern numerical analysis (particularly numerical linear algebra) and has found successful applications in many areas such as statistical analysis [10], image and signal processing [2], system identification [26], and control [15]. For a survey of the SVD, its theory, numerical details, and some newer applications, we refer the reader to [27].

4. Application of SVD in Fuzzy Model Reduction

4.1. Principle

The basic principle of using SVD for fuzzy model reduction is to eliminate the insignificant rules from the complete rule base. This is done by exploring the singular values of an N-by-M matrix Ω, where N denotes the number of available modeling data for each input variable and M denotes the number of fuzzy rules in the model. The **(i,j)***th* entry in Ω is the output of **j***th* rule for the **i***th* input data. If a fuzzy model contains insignificant rules, the outputs among the rules in the model will be ***correlated***, which will lead to a ***singular*** matrix Ω. The singularity of Ω, which is usually called ***rank-degeneracy*** in statistical and numerical literature (see, e.g., [20]), is indicated by the presence of "small" singular values. The number of these small singular values implies the number of insignificant fuzzy rules in the model.

4.2. Algorithm

Nearly all implemented SVD codes used today are based on the algorithm of Golub and Reinsch [8]. This algorithm first uses ***Householder transformation*** to bidiagonalize the given matrix and then an iterative algorithm to compute the singular values of the resultant bidiagonal form.

This algorithm has been implemented in ***Matlab***. However, the singular values produced appear in a descending order. These ordered singular values do not indicate the positions where they appear in the matrix Ω. This poses a difficulty for fuzzy rule selection since we need to relate the singular values with their corresponding rules.

Here we use a different method, known as the ***Jacobi method*** (see, e.g., [9]), to compute the SVD of Ω. We first apply a sequence of ***one-sided Jacobi transformation*** [16] to Ω and get

$$B = U^T \Omega \tag{5}$$

where B is row orthogonal and U is both row and column orthogonal, then computing the ***QR decomposition*** of B^T

and its transpose, that is

$$B = (B^T)^T = (Q_B R_B)^T = R_B^T Q_B^T \qquad (6)$$

where R_B is actually a ***diagonal*** matrix whose entries consist of the singular values of Ω (Assuming that a ***nonnegative handing step*** has been done [27]). Let

$$\Sigma = R_B^T \quad and \quad V^T = Q_B^T \qquad (7)$$

then from (5) and (6) we can get a complete SVD

$$\Omega = U \Sigma V^T \qquad (8)$$

Using this algorithm we can obtain the singular values of the N-by-M matrix Ω which appear in their original order.

Remark 2: For fuzzy model reduction, only the singular values of the N-by-M matrix Ω are needed and it is not necessary to explicitly store and compute the orthogonal matrices U and V.

4.3. Computational considerations

In building a Sugeno-type model, the number of available modeling data is usually much larger than the number of fuzzy rules in the model being constructed. This implies that the row dimension of the matrix Ω is much larger than its column dimension, that is, $N \gg M$. A faster method for the SVD of Ω may result if we first transform Ω into an M-by-M upper triangular form R by applying Householder transformation on its left, and then proceed to process R. The important difference is that this enables us to work with a much smaller matrix R, than Ω, and so it is conceivable that the work required is much less than that originally done. The idea of computing the bidiagonalization in this manner is more fully analyzed in [4].

4.4. Fuzzy model reduction procedure

The proposed fuzzy model reduction approach consists of the following steps.

Step 1: Identify input variables. If the number of input variables is too large, principal component analysis can be used to reduce the dimension of the input space [29] [33]. However, this is not the focus of the paper.

Step 2: Construct membership functions of input variables using B-splines.

Step 3: Estimate consequent parameters using the Kalman filter to minimize the objective function

$$J = \sum_{k=1}^{N} (d(k) - y(k))^2 \qquad (9)$$

where d(k) and y(k) denote the target output and the total output of the Sugeno-type model for ***kth*** training data respectively. The total output y(k) is computed by

$$y(k) = \sum_{i=1}^{M} y_i(k) \qquad (10)$$

where $y_i(k)$ denotes the output of ***ith*** rule in the Sugeno-type model and is computed by

$$y_i(k) = \frac{w_i\,(b_{i0} + b_{i1}x_1(k) + \dots + b_{ip}x_p(k))}{\sum_{i=1}^{M} w_i} \qquad (11)$$

where w_i denotes the truth value of ***ith*** rule in the Sugeno-type model.

Step 4: For each training data, compute the individual output of each rule using (11). As a result, an N-by-M matrix, denoted by Ω, is formed, where $\Omega(k,i) = y_i(k)$.

Step 5: Compute the SVD of Ω using the Jacobi method. The number of small singular values in Ω implies the number of insignificant fuzzy rules and their positions indicate which rules should be discarded in the model.

Step 6: Estimate the consequent parameters of the reduced fuzzy model using the Kalman filter.

Step 7: Evaluate the model. To see whether the reduced fuzzy model adequately represents the given set of modeling data, a diagnostic test should be performed. This can be done based on some ***statistical information criteria***(see, e.g., [1]) or the ***correlation analysis of residuals*** [25]. If any inadequacy is detected, the above iterative cycle is repeated until a satisfactory reduced fuzzy model is found.

5. Example

In this section we provide a limit cycle modeling example to illustrate the proposed fuzzy model reduction approach. We consider the two-input, single-output nonlinear time series model

$$\begin{aligned} y(k) = (0.8 - 0.5\exp(-y^2(k-1)))y(k-1) - (0.3 + \\ 0.9\exp(-y^2(k-1)))y(k-2) + 0.1\sin(\pi y(k-1)) + e(k) \end{aligned} \qquad (12)$$

where e(k) is a zero mean Gaussian white noise process with variance 0.1. The origin of this time series is an unstable equilibria which is enclosed by a stable attracting limit cycle. Harris, Moore and Brown [11] have modelled this system using a neural network. Here we model this system using the Sugeno-type model. We generate 2000 noisy observations from (12) and select y(k-1) and y(k-2) as the input variables and y(k) as the output variable of the model. We use the first 1000 data to identify the model and the performance of the resulting model is tested using the remaining 1000 data. We use the ***cubic B-splines*** with the ***interior knots*** {-2, -2/3, 2/3, 2} to construct the membership functions of y(k-1) and y(k-2). The number of interior knots, together with the type of B-splines determines the order of the B-splines. Since the cubic B-splines have been used, the corresponding order is equal to 6 [19]. The order of B-splines also determines the number of membership functions.

Let $a_i(i=1,2,...,6)$ and $b_i(i=1,2,...,6)$ denote the membership functions of y(k-1) and y(k-2), respectively. The complete combinations of the two sets are shown in Table 1. Using the model reduction strategy suggested in [29], we can obtain a model consisting of 6 fuzzy rules. This is done by taking the combinations appeared in the main diagonal of the Table 1, i.e., a_1b_1, a_2b_2, a_3b_3, a_4b_4, a_5b_5, a_6b_6. Whereas the model comprises a small number of fuzzy rules, it may not be optimal. Here we use the proposed SVD-based model reduction strategy to construct reduced fuzzy models. Based on the complete combinations shown in Table 1, we first build a fuzzy model consisting of 36 rules. By computing the individual outputs of the 36 rules with the given 1000 modeling data, we can construct a 1000-by-36 output matrix. We then compute the SVD of the matrix, and the resulting singular values are shown in Fig. 1.

Table 1. Combinations of membership functions of input variables x_1 and x_2.

a_1b_1	a_1b_2	a_1b_3	a_1b_4	a_1b_5	a_1b_6
a_2b_1	a_2b_2	a_2b_3	a_2b_4	a_2b_5	a_2b_6
a_3b_1	a_3b_2	a_3b_3	a_3b_4	a_3b_5	a_3b_6
a_4b_1	a_4b_2	a_4b_3	a_4b_4	a_4b_5	a_4b_6
a_5b_1	a_5b_2	a_5b_3	a_5b_4	a_5b_5	a_5b_6
a_6b_1	a_6b_2	a_6b_3	a_6b_4	a_6b_5	a_6b_6

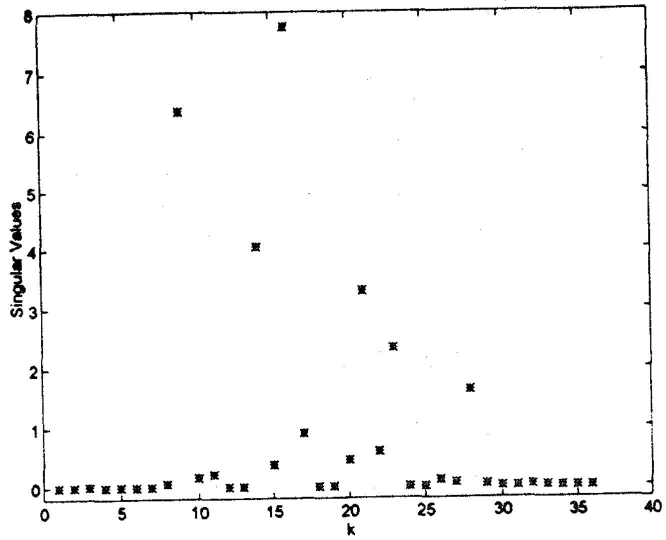

Fig. 1. Singular values distribution of the 1000-by-36 output matrix.

Based on the distribution of singular values we select the rules corresponding to the first 12 largest singular values and obtain a reduced fuzzy model consisting of 12 rules. The positions of the 12 largest singular values indicate the positions of the dominant combinations of the membership functions in Table 1 which correspond to a_3b_4, a_2b_3, a_3b_2, a_4b_3, a_4b_5, a_5b_4, a_3b_5, a_4b_4, a_4b_2, a_3b_3, a_2b_5, and a_2b_4,

For comparison with the diagonal reduction approach [29], we also select the rules corresponding to the first 6 largest singular values and obtain a reduced fuzzy model consisting of 6 rules. The 6 largest singular values correspond to a_3b_4, a_2b_3, a_3b_2, a_4b_3, a_4b_5, and a_5b_4 in Table 1.

Table 2 compares the ***mean-squared errors*** (***MSE***) of the SVD-based reduced fuzzy models and that of the reduced fuzzy model obtained using diagonal reduction. It can be seen that given the same number of fuzzy rules, 6, the reduced fuzzy model using SVD gives less MSE than the reduced fuzzy model using diagonal reduction. As a comparison, Table 2 also shows the MSE of the fuzzy model consisting of 36 rules (we name it ***complete fuzzy model***). An important result in our experiment is that the SVD-based reduced fuzzy model with 12 rules gives nearly the same MSE as the complete fuzzy model. Table 3 shows the number of consequent parameters in these fuzzy models and corresponding computational time on a PC Pentium-90. Fig. 2 shows the dynamics of the real model, the complete fuzzy model, the reduced fuzzy model using SVD with 6 rules, and the reduced fuzzy model using the diagonal reduction.

Table 2. Comparison of MSE's of different fuzzy models in the modeling stage.

Fuzzy Model	MSE (modeling)
Complete fuzzy models (36 rules)	1.0705e-4
Reduced fuzzy model using SVD (12 rules)	1.0924e-4
Reduced fuzzy model using SVD (6 rules)	1.6164e-4
Reduced fuzzy model using diagonal (6 rules)	5.6848e-4

Table 3. Number of consequent parameters and corresponding computational time.

Fuzzy Model	Number of Consequent Parameters	Computational Time (Second)
Complete fuzzy model (36 rules)	108	986
Reduced fuzzy model using SVD (12 rules)	36	67
Reduced fuzzy model using SVD (6 rules)	18	12
Reduced fuzzy model using diagonal(6 rules)	18	12

Table 4. Comparison of MSE's of different fuzzy models in the prediction stage.

Fuzzy Model	MSE (prediction)
Complete fuzzy models (36 rules)	1.0985e-4
Reduced fuzzy model using SVD (12 rules)	1.0974e-4
Reduced fuzzy model using SVD (6 rules)	1.5285e-4
Reduced fuzzy model using diagonal (6 rules)	5.5532e-4

The prediction performance of these fuzzy models is tested using the remaining 1000 data. Table 4 lists the MSE's of these fuzzy models. As before, given the same number of fuzzy rules, 6, the reduced model using SVD shows superior performance over the reduced model using the diagonal reduction. A surprising result here is that the reduced fuzzy model using SVD with 12 rules gives slightly better prediction accuracy compared to the complete fuzzy model with 36 rules. This implies that the latter might have been ***overfitted***. An overfitted model may show good performance in fitting, but it cannot always assure good performance in prediction. We will discuss this issue in great detail in a future paper.

6. Conclusion

This paper proposed a novel approach for fuzzy model reduction based on the singular value decomposition method. The approach is capable of detecting and selecting the important fuzzy rule from a rule base. Our

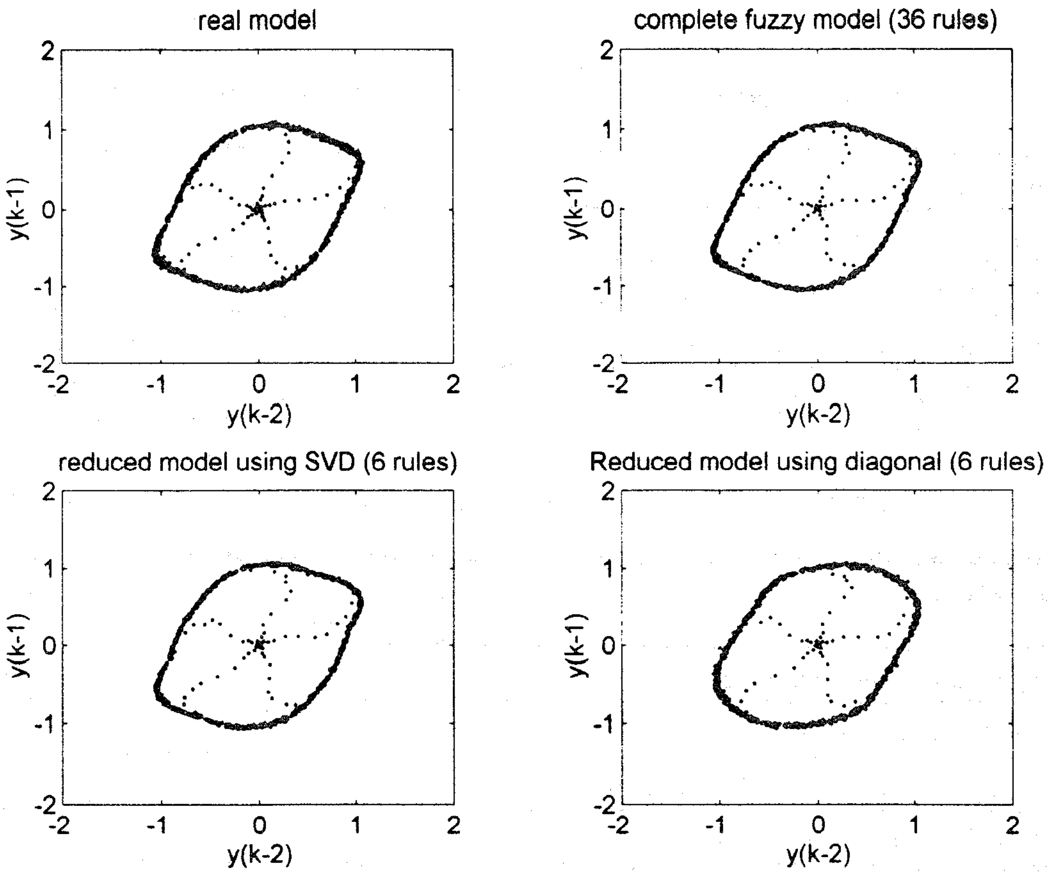

Fig. 2. Limit cycle modeling using Sugeno-type models.

experiments showed the reduced fuzzy models can perform well in nonlinear systems modeling and prediction.

The proposed approach is not necessarily confined to the construction of Sugeno-type models. It can also be extended to other types of fuzzy models such as the ***Mamdani model*** [17] and the fuzzy basis functions model [30] as well as the ***radial basis function networks*** introduced in [18].It may also be applied to ***splines modeling***(see, e.g.,[7]) where the curse of dimensionality may pose particular difficulties because of the introduction of high-dimensional splines.

We are currently developing objective criteria to assist users in determining the range of small singular values. We are also investigating theoretical and computational issues regarding the proposed SVD-based fuzzy model reduction technique.

Acknowledgements

This research is supported by National Science Foundation Young Investigator Award IRI-9257293.

References

[1] H. Akaike, A new look at statistical model identification, *IEEE Trans. Automat. Contr.*,Vol.19, 1974, pp.716-723.

[2] H.C.Andrews and C.L.Patterson, Singular value decomposition and digital image processing, *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. 24, 1976, pp. 26-53.

[3] J.C.Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.

[4] T.F. Chan, An improved algorithm for computing the singular value decomposition, *ACM Trans. Math. Softw.*, Vol. 8, 1982, pp. 72-83.

[5] S.L.Chiu, An efficient method for input variable selection in fuzzy modeling, in *Proc. 34th IEEE Conf. Decision Control*, New Orleans, Louisiana, December, 1995.

[6] C. De Boor, On calculating with B-splines, *Journal of Approximation Theory*, Vol. 6, 1972, pp. 50-62.

[7] P. Dierckx, *Curve and Surface Fitting with Splines*, Clarendon Press, Oxford, 1993.

[8] G.H. Golub and C. Reinsch, Singular value decomposition and least squares solutions, *Numer. Math.*, Vol. 14, 1970, pp. 403-420.

[9] G.H.Golub. and C.F.Van Loan, *Matrix Computations*, 2nd

edition, John Hopkins University Press, Baltimore, MD, 1989.
[10] S.J.Hammarling, The singular value decomposition in multivariate statistics, *ACM Signum Letter*, Vol. 20, 1985, pp. 2-25.
[11] C.J. Harris, C.G. Moore, and M. Brown, *Intelligent Control: Aspects of Fuzzy Logic and Neural Networks*, World Scientific, Singapore, 1993.
[12] J.-S.R. Jang, ANFIS: Adaptive-network-based fuzzy inference system, *IEEE Trans. Syst., Man,Cybern.*, Vol. 23, 1993, pp. 665-685.
[13] T.A. Johansen, On the Optimality of the Takagi-Sugeno-Kang fuzzy inference mechanism, in *Proc. FUZZ-IEEE/IFES'95*, Yokohama, Japan, March, 1995, pp. 97-102.
[14] M.J. Karson, *Multivariate Statistical Methods*, The Iowa State University Press, Ames, Iowa, 1982.
[15] A.J. Laub, Numerical linear algebra aspects of control design computations, *IEEE Trans. Automat. Contr.*, Vol. 30, 1985, pp. 97-108.
[16] F.T. Luk, Computing the singular value decomposition on the ILLIAC IV, *ACM Trans. Math. Softw.*, Vol. 6, 1980, pp. 524-539.
[17] E.H.Mamdani, Application of fuzzy algorithms for simple dynamic plant,*Proc.IEE*,Vol.121,1974, pp.1585-1588.
[18] J. Moody and C.J. Darken, Fast learning in networks of locally-tuned processing units, *Neural Computation*, Vol. 1, 1989, pp. 281-294.
[19] M.J.D. Powell, *Approximation Theory and Methods*, Cambridge University Press, London, 1981.
[20] G.W. Stewart, Rank degeneracy, *SIAM J. Sci. Stat. Comput.*, Vol. 5, 1984, pp. 403-413.
[21] M. Sugeno, M.F. Griffin and A. Bastian, Fuzzy hierarchical control of an unmanned helicopter, in *Proc. 5th Int. Fuzzy Systems Association World Congress*, Korea, 1993, pp. 179-182.
[22] M.Sugeno and G. Kang, Structure identification of fuzzy model, *Fuzzy Sets Syst.*, Vol. 28, 1988, pp. 15-33.
[23] M.Sugeno and T.Yasukawa, A fuzzy logic based approach to qualitative modeling, *IEEE Trans. Fuzzy Syst.*, Vol. 1, 1993, pp. 7-31.
[24] T. Takagi and M. Sugeno, Fuzzy identification of systems and its applications to modeling and control, *IEEE Trans. Syst., Man, Cybern.*, Vol. 15, 1985, pp. 116-132.
[25] G.C. Tiao and G.E.P. Box, Modeling multiple time series with applications, *J. Amer. Stat. Assoc.*, Vol. 76, pp. 802-816, 1981.
[26] J. Vandewalle and B. De Moor, A variety of applications of singular value decomposition in identification and signal processing, in *SVD and Signal Processing: Algorithms, Applications and Architectures*, E. Deprettere, edit, North-Holland, Amsterdam, 1988, pp. 43-91.
[27] L. Wang, *SVD-based State Space Modeling and Kalman Filtering Techniques*, Ph.D Thesis, Faculté Polytechnique de Mons, Mons, Belgium, 1993.
[28] L. Wang and R. Langari, Complex systems modeling via fuzzy logic, *IEEE Trans. Syst., Man, Cybern.*, 1995(accepted for publication).
[29] L. Wang, R. Langari and J. Yen, Principal components, B-splines, and fuzzy systems reduction, in *Proc. CFSA/IFIS/SOFT'95*, Taiwan, December, 1995.
[30] L.X. Wang, *Adaptive Fuzzy Systems and Control: Design and Stability Analysis*, Prentice-Hall, Englewood Cliffs, NJ, 1994.
[31] D.Willaeys and N.Malvache, Contribution of the fuzzy sets to man-machine system, in *Advances in Fuzzy Set Theory and Applications*, M.M.Gupta, R.Ragade and R.Yager, edits, North-Holland, Amsterdam, 1979,pp.481-499.
[32] R.R.Yager and D.P. Filev, Unified structure and parameter identification of fuzzy models, *IEEE Trans. Syst., Man, Cybern.*, Vol. 23, 1993, pp. 1198-1205.
[33] J. Yen, H. Wang and J. Liao,. A method for automatic generation of fuzzy model, in *Proc. 3rd Int. Conf. Industrial Fuzzy Control Intelligent Systems*, Houston, Texas, December, 1993, pp. 88-92.