

# Nonsingleton Fuzzy Logic Systems: Theory and Application

George C. Mouzouris and Jerry M. Mendel, *Fellow, IEEE*

**Abstract**—In this paper, we present a formal derivation of general nonsingleton fuzzy logic systems (NSFLS's) and show how they can be efficiently computed. We give examples for special cases of membership functions and inference and we show how an NSFLS can be expressed as a “nonsingleton fuzzy basis function” expansion and present an analytical comparison of the nonsingleton and singleton fuzzy logic systems formulations. We prove that an NSFLS can uniformly approximate any given continuous function on a compact set and show that our NSFLS does a much better job of predicting a noisy chaotic time series than does a singleton fuzzy logic system (FLS).

**Index Terms**—Fuzzy control, uniform approximation.

## I. INTRODUCTION

A fuzzy logic system processes crisp data at the input and produces crisp data at the output. Therefore, a *fuzzifier* is used at the front of the system to convert crisp data to fuzzy data, and a *defuzzifier* is used at the output of the system to convert fuzzy data into crisp data. The most widely used fuzzifier is the singleton fuzzifier [13], [35], [37] mainly because of its simplicity and lower computational requirements; however, this kind of fuzzifier may not always be adequate, especially in cases where noise is present in the training data or in the data which is later processed by the system. A different approach is necessary to account for uncertainty in the data, which is why we direct our attention in this paper at the *nonsingleton* fuzzifier and nonsingleton fuzzy logic systems (NSFLS's).

Nonsingleton fuzzifiers have been used successfully in a variety of applications [1], [10], [25], [28], [31], [33]. In neural fuzzy systems [1], [10], vectors of fuzzy sets are used both to train a fuzzy neural network and as inputs during processing. In [25], an optimizing control method for optimizing the fuel consumption rate of a marine diesel engine utilizes empirical rules that are expressed by fuzzy numbers. Nonsingleton input has also been used in turning process automation [31] to represent a human operator's actions in the fuzzy rule base, in the design of fuzzy control algorithms [28], and fuzzy information and decision-making [33]. These methods are largely heuristic and provide no closed-form expressions for fuzzy logic systems; hence, their generalizations are very difficult.

Manuscript received July 8, 1994; revised February 26, 1996. This work was supported by the University of Southern California under Grant MIP-9122018 of the National Science Foundation.

The authors are with the Signal and Image Processing Institute, Department of Electrical Engineering-Systems, University of Southern California, Los Angeles, CA 90089 USA.

Publisher Item Identifier S 1063-6706(97)00038-6.

In this paper, we develop a quantitative formulation of an NSFLS and its efficient computation. This formulation provides a tool for accounting for uncertainty in either the training data or the input to the system. In Section II, we give a brief discussion on nonsingleton fuzzification. In Section III, we derive from first principles the continuous and discrete forms of NSFLS's, and show how and the conditions under which a NSFLS reduces to a singleton fuzzy logic system (FLS). We also quantify the difference between the output fuzzy sets for nonsingleton and singleton fuzzification, show how our NSFLS can be expressed as a nonsingleton fuzzy basis function (FBF) expansion, and, present examples for i) product and minimum inference and Gaussian input, antecedent, and consequent fuzzy sets, and ii) minimum inference and triangular input, with parabolic antecedents and consequents. In Section IV, we show that our NSFLS's uniformly approximate any given continuous function on a compact set. In Section V, we apply our NSFLS to the prediction of a noisy Mackey–Glass chaotic time series and show, by means of simulations, that the NSFLS is much more successful in producing accurate forecasts of that series than is a comparable singleton FLS. In Section VI, we draw our conclusions.

## II. THE NONSINGLETON FUZZIFIER

Fuzzy sets have been interpreted as membership functions  $\mu_X$  [41] that associate with each element  $x$  of the universe of discourse  $U$ , a number  $\mu_X(x)$  in the interval  $[0, 1]$

$$\mu_X: U \rightarrow [0, 1]. \quad (1)$$

A fuzzifier maps a crisp point  $x \in U$  into a fuzzy set  $X \in U$ .

- 1) In the case of a *singleton fuzzifier*, the crisp point  $x \in U$  is mapped into a fuzzy set  $X$  with support  $x_i$ , where  $\mu_X(x_i) = 1$  for  $x_i = x$  and  $\mu_X(x_i) = 0$  for  $x_i \neq x$ , i.e., the *single* point in the support of  $X$  with nonzero membership function value is  $x_i = x$ .
- 2) In the case of a *nonsingleton fuzzifier*, the point  $x \in U$  is mapped into a fuzzy set  $X$  with support  $x_i$ , where  $\mu_X$  achieves maximum value at  $x_i = x$  and decreases while moving away from  $x_i = x$ . We assume that fuzzy set  $X$  is normalized so that  $\mu_X(x) = 1$ .

Nonsingleton fuzzification is especially useful in cases where the available training data, or the input data to the fuzzy logic system, are corrupted by noise. Conceptually, the nonsingleton fuzzifier implies that the given input value  $x$  is the most likely value to be the correct one from all the values

in its immediate neighborhood; however, because the input is corrupted by noise, neighboring points are also likely to be the correct values, but to a lesser degree.

It is up to the system designer to determine the shape of the membership function  $\mu_X$  based on an estimate of the kind and quantity of noise present. It would be the logical choice, though, for the membership function to be symmetric about  $x$  since the effect of noise is most likely to be equivalent on all points. Examples of such membership functions are: 1) the Gaussian  $\mu_X(x_i) = \exp[-(x - x_i)^2/2\sigma^2]$  where the variance  $\sigma^2$  reflects the width (spread) of  $\mu_X(x_i)$ , 2) triangular  $\mu_X(x_i) = \max(0, 1 - |(x - x_i)/c|)$ , and 3)  $\mu_X(x_i) = 1/(1 + |(x - x_i)/c|^p)$  where  $x$  and  $c$  are, respectively, the mean and spread of the fuzzy sets. Note that larger values of the spread of the above membership functions imply that more noise is anticipated to exist in the given data.

### III. NONSINGLETON FUZZY LOGIC SYSTEM FORMULATION

#### A. General Results

Consider a fuzzy logic system with a rule base of  $M$  rules, and let the  $l$ th rule be denoted by  $R^l$ . Let each rule have  $p$  antecedents and one consequent (as is well known, a rule with  $q$  consequents can be decomposed into  $q$  rules, each having the same antecedents and one different consequent), i.e., it is of the general form

$$\begin{aligned} R^l: & \text{IF } u_1 \text{ is } F_1^l \text{ and } u_2 \text{ is } F_2^l \text{ and } \dots \text{ and } u_p \text{ is } F_p^l \\ & \text{THEN } v \text{ is } G^l \end{aligned}$$

where  $u_k, k = 1, \dots, p$ , and  $v$  are the input and output linguistic variables, respectively. Each  $F_k^l$  and  $G^l$  are subsets of possibly different universes of discourse. Let  $F_k^l \subset U_k$  and  $G^l \subset V$ . Each rule can be viewed as a fuzzy relation  $R^l$  [42] from a set  $U$  to a set  $V$  where  $U$  is the Cartesian product  $U = U_1 \times \dots \times U_p$ .  $R^l$  itself is a subset of the Cartesian product  $U \times V = \{(\mathbf{x}, y) : \mathbf{x} \in U, y \in V\}$ , where  $\mathbf{x} \equiv (x_1, x_2, \dots, x_p)$ , and  $x_k$  and  $y$  are the points in the universes of discourse  $U_k$  and  $V$  of  $u_k$  and  $v$ .

First, we present a system formulation for the continuous case, and then for the discrete case. We also demonstrate that in the case of the *height defuzzifier*, the continuous and discrete cases produce identical systems.  $R^l$  is characterized by a continuous multivariate membership function  $\mu_{R^l}(\mathbf{x}, y)$ , and can be described by the following:

$$\begin{aligned} R^l & \equiv \int_{U \times V} \mu_{R^l}(\mathbf{x}, y) / (\mathbf{x}, y) \\ & = \int_{U_1} \dots \int_{U_p} \int_V \mu_{F_1^l}(x_1) * \dots * \mu_{F_p^l}(x_p) \\ & \quad * \mu_{G^l}(y) / (\mathbf{x}, y) \end{aligned} \quad (2)$$

where  $*$  denotes a  $t$ -norm.  $\int$  denotes the union of individual points of each set in the continuum [44].

Let the input to  $R^l$  be denoted by  $A$ , where  $A$  is a subset of a  $p$ -dimensional Cartesian product space and is given by

$$A = \int_{U_1} \dots \int_{U_p} \mu_{X_1}(x_1) * \dots * \mu_{X_p}(x_p) / (\mathbf{x}) \quad (3)$$

where  $X_k \subset U_k$  ( $k = 1, \dots, p$ ) are the fuzzy sets describing the inputs.

Up to this point, the formulation is identical to that of the singleton case. In the singleton case, though, it is assumed that each input fuzzy set  $X_k$  has nonzero membership value only at a single point, which reduces  $A$  to a set with a single point  $\mathbf{x}_r \in U$ . In our treatment here we do not make this assumption. Each input fuzzy set is represented by the more general nonsingleton form in (3), thereby allowing any uncertainty in the input to be represented in the system.

According to the *compositional rule of inference*, the fuzzy subset  $Y^l$  of  $V$  induced by  $A \in U$  is given by the composition of  $A$  and  $R^l$ :

$$Y^l = A \circ R^l = \sup_{\mathbf{x} \in U} (A * R^l). \quad (4)$$

Note that all unions (denoted by  $\int$ ) in  $A$  and  $R^l$ , over  $U_k, k = 1, \dots, p$ , are over the same spaces; therefore, we can write  $Y^l$  as

$$\begin{aligned} Y^l & = \sup_{\mathbf{x} \in U} \left( \int_{U_1} \dots \int_{U_p} \int_V \mu_{X_1}(x_1) * \dots * \mu_{X_p}(x_p) \right. \\ & \quad \left. * \mu_{F_1^l}(x_1) * \dots * \mu_{F_p^l}(x_p) * \mu_{G^l}(y) / (\mathbf{x}) \right) / (y). \end{aligned} \quad (5)$$

Since the supremum is only over  $\mathbf{x} \in U$ , then, by the commutativity and monotonicity properties of a  $t$ -norm we can rewrite  $Y^l$  as

$$\begin{aligned} Y^l & = \int_V \mu_{G^l}(y) * \sup_{\mathbf{x} \in U} \left( \int_{U_1} \dots \int_{U_p} \mu_{X_1}(x_1) * \dots \right. \\ & \quad \left. * \mu_{X_p}(x_p) * \mu_{F_1^l}(x_1) * \dots * \mu_{F_p^l}(x_p) / (\mathbf{x}) \right) / (y). \end{aligned} \quad (6)$$

Next, we recall that by definition, a  $t$ -norm is a two-place function from  $[0, 1] \times [0, 1]$  [44]; thus, we can consider every  $t$ -norm in (6) to be acting on a pair of membership functions. The calculation of the  $t$ -norm over all the points in the corresponding spaces of the two membership functions is easier to visualize if the membership functions are in the same space. Therefore, we rewrite (6) in the following manner:

$$\begin{aligned} Y^l & = \int_V \mu_{G^l}(y) * \sup_{\mathbf{x} \in U} \\ & \quad \cdot \left( \int_{U_1} \dots \int_{U_p} [\mu_{X_1}(x_1) * \mu_{F_1^l}(x_1)] * \dots \right. \\ & \quad \left. * [\mu_{X_p}(x_p) * \mu_{F_p^l}(x_p)] / (\mathbf{x}) \right) / (y). \end{aligned} \quad (7)$$

Note that the supremum in (7) is over all points  $\mathbf{x}$  in  $U$  (in a  $p$ -dimensional Cartesian product space). By the monotonicity property of a  $t$ -norm [44], that supremum is attained when each term in brackets attains its supremum. Let

$$\mu_{Q_k^l}(x_k) \equiv \mu_{X_k}(x_k) * \mu_{F_k^l}(x_k) \quad (8)$$

where  $\mu_{Q_k^l}, \mu_{X_k}, \mu_{F_k^l} \in U_k$ . Assuming that  $\mu_{Q_k^l}$  produced by (8) is a function whose supremum can be evaluated, let  $x_{k,\text{sup}}^l$  denote the point in  $U_k$  where that supremum is attained; then (7) becomes

$$Y^l = \int_V \mu_{G^l}(y) * \mathcal{T}_{k=1}^p \mu_{Q_k^l}(x_{k,\text{sup}}^l) / (y) \quad (9)$$

where  $\mathcal{T}_{k=1}^p$  denotes a sequence of  $p-1$   $t$ -norm operations.

Using the same procedure and rationale as in the continuous case, we can derive corresponding expressions for discrete NSFLS's. Each  $U_i, (i = 1, \dots, p)$  and  $V$  are finite (and, thus, countable), and the  $i$ th antecedent membership function,  $\mu_{F_i^l}$ , and consequent membership function  $\mu_{G^l}$  have nonzero membership values at discrete points  $x_{i,1}, \dots, x_{i,n_i}$ , and  $y_1, \dots, y_m$ , respectively. The  $l$ th rule  $R^l$  is, therefore, given by

$$R^l = \sum_{i_1=1}^{n_1} \cdots \sum_{i_p=1}^{n_p} \sum_{j=1}^m \mu_{F_1^l}(x_{1,i_1}) * \cdots * \mu_{F_p^l}(x_{p,i_p}) * \mu_{G^l}(y_j) / (x_{1,i_1}, x_{2,i_2}, \dots, x_{p,i_p}, y_j) \quad (10)$$

where  $\Sigma$  denotes the union of the individual points of each set. The  $p$ -dimensional input to  $R^l$  is given by

$$A = \sum_{i_1=1}^{n_1} \cdots \sum_{i_p=1}^{n_p} \mu_{X_1}(x_{1,i_1}) * \cdots * \mu_{X_p}(x_{p,i_p}) / (\mathbf{x})$$

where now  $\mathbf{x} \equiv (x_{1,i_1}, x_{2,i_2}, \dots, x_{p,i_p})$ . Using the compositional rule of inference and the commutativity and monotonicity properties of a  $t$ -norm, we can write the discrete output fuzzy set  $Y^l$  as

$$Y^l = \sum_{j=1}^m \mu_{G^l}(y_j) * \sup_{\mathbf{x} \in U} \left( \sum_{i_1=1}^{n_1} \cdots \sum_{i_p=1}^{n_p} \mu_{X_1}(x_{1,i_1}) * \cdots * \mu_{X_p}(x_{p,i_p}) * \mu_{F_1^l}(x_{1,i_1}) * \cdots * \mu_{F_p^l}(x_{p,i_p}) / (\mathbf{x}) \right) / (y_j). \quad (11)$$

From the finiteness of each universe of discourse  $U_k$  the supremum is the same as the maximum; therefore, we need to calculate the maximum over all  $\mathbf{x} \in U$  of the parenthetical term in (11). That term will be maximized when every term

$$\mu_{Q_k^l}(x_{k,i_k}) \equiv \mu_{X_k}(x_{k,i_k}) * \mu_{F_k^l}(x_{k,i_k}) \quad (12)$$

is maximized. If the global maximum  $x_{k,\text{max}}^l$  of each  $\mu_{Q_k^l}(x_{k,i_k})$  can be evaluated, then

$$Y^l = \sum_{j=1}^m \mu_{G^l}(y_j) * \mathcal{T}_{k=1}^p \mu_{Q_k^l}(x_{k,\text{max}}^l) / (y_j) \quad (13)$$

which is the discrete counterpart to (9).

Because our FLS has  $M$  rules, the final output fuzzy set  $Y$  is obtained by  $t$ -conorm aggregation of the individual outputs of each rule, i.e.,

$$Y = Y^1 \dot{+} Y^2 \dot{+} \cdots \dot{+} Y^M \quad (14)$$

where  $\dot{+}$  denotes  $t$ -conorm ( $s$ -norm) [18], [44].

Kosko [13] has utilized arithmetic addition as a means of aggregation instead of a  $t$ -conorm. In most cases, this produces a central-limit-theorem effect that approximates a unimodal fuzzy set. The smaller the spread of this set, the higher the probability that the defuzzified value is the correct one. This additive structure is encountered in many  $t$ -conorms, which impose certain conditions (i.e., bounded sum) or normalization to keep the result in  $[0, 1]$  (e.g., Einstein sum, algebraic sum, Hamacher sum [44]).

We will use a special case of an additive fuzzy logic system, one that uses a weighted additive combination of the points where each output fuzzy set becomes a maximum; it is the *modified height defuzzifier* of [11] which was also proposed by Wang [35], [36]. Let  $\bar{y}^l$  denote the point in each individual output fuzzy set such that  $\mu_{G^l}(\bar{y}^l) = 1$ , and let  $\delta^l$  be proportional to the spread of  $\mu_{G^l}$ . Then, the modified height defuzzifier leads to the following output for our nonsingleton system

$$y = f_{ns}(\mathbf{x}) = \frac{\sum_{l=1}^M \bar{y}^l [\mathcal{T}_{k=1}^p \mu_{Q_k^l}(x_{k,\text{sup}}^l)] / (\delta^l)^2}{\sum_{l=1}^M [\mathcal{T}_{k=1}^p \mu_{Q_k^l}(x_{k,\text{sup}}^l)] / (\delta^l)^2}. \quad (15)$$

The spread  $\delta^l$  of the output fuzzy set is proportional to the amount of uncertainty that  $\bar{y}^l$  is close to the point where  $\mu_{G^l}(y^l) = 1$  (e.g., in the case of Gaussian consequent fuzzy sets  $\delta^l$  could be chosen to be the standard deviation); thus, if  $\delta^l$  is very small (high degree of confidence), then the corresponding point  $\bar{y}^l$  is more heavily weighted than if  $\delta^l$  is very large. For the case of the modified height defuzzifier, the discrete nonsingleton fuzzy logic system (with  $M$  rules in its rulebase), has the form of (15) with  $x_{k,\text{sup}}^l$  replaced by  $x_{k,\text{max}}^l$  and, assuming fine enough discretization, will give identical results as the continuous system.

By defining the following nonsingleton FBF's

$$p_l(\mathbf{x}) = \frac{[\mathcal{T}_{k=1}^p \mu_{Q_k^l}(x_{k,\text{sup}}^l)] / (\delta^l)^2}{\sum_{l=1}^M [\mathcal{T}_{k=1}^p \mu_{Q_k^l}(x_{k,\text{sup}}^l)] / (\delta^l)^2} \quad (16)$$

we can express our nonsingleton system in (15) as the following *fuzzy basis function expansion*

$$f_{ns}(\mathbf{x}) = \sum_{l=1}^M \bar{y}^l p_l(\mathbf{x}). \quad (17)$$

The modified height defuzzifier is a simple and fast method to implement a general FLS; however, other defuzzification methods that have different properties can also be used, e.g., *center-of-area*, *center-of-sums*, and *first-of-maxima* (see [11] for an overview and comparison of different defuzzification methods).

To contrast the nonsingleton and singleton discrete cases, and also to show that the singleton case is a special case of

(11), we assume (as mentioned before) that the input fuzzy sets are normalized. Consequently, for each input fuzzy set  $X_k$  there exists at least one point  $x_{k,q_k} \in X_k$  such that  $\mu_{X_k}(x_{k,q_k}) = 1$ . In addition, we assume that  $\bar{x}_k \equiv x_{k,q_k}$  is unique in  $X_k$  (extension to multiple such points is easy); then,  $\mu_{X_1}(\bar{x}_1) = \dots = \mu_{X_p}(\bar{x}_p) = 1$ , and  $\mu_{X_1}(\bar{x}_1) * \dots * \mu_{X_p}(\bar{x}_p) = 1$ , which is true for all  $t$ -norms. Consequently, the general nonsingleton case in (11) can be decomposed into contributions of the singleton components and the remaining components with membership values less than unity

$$\begin{aligned}
 Y^l = & \sum_{j=1}^m \mu_{G^l}(y_j) * \sup_{\mathbf{x} \in U} \left[ \mu_{F_1^l}(\bar{x}_1) * \dots * \mu_{F_p^l}(\bar{x}_p) \right. \\
 & + \sum_{i_1 \neq q_1} \sum_{i_2 \neq q_2} \dots \sum_{i_p \neq q_p} \mu_{X_1}(x_{1,i_1}) * \mu_{X_2}(x_{2,i_2}) * \dots \\
 & * \mu_{X_p}(x_{p,i_p}) * \mu_{F_1^l}(x_{1,i_1}) * \mu_{F_2^l}(x_{2,i_2}) * \dots \\
 & \left. * \mu_{F_p^l}(x_{p,i_p}) \right] / (y_j) \equiv \sum_{j=1}^m \mu_{G^l}(y_j) * S_{ns}/(y_j)
 \end{aligned} \quad (18)$$

where “+” denotes union. If the inputs are indeed singletons, then the term on the right-hand side of “+” becomes zero and the supremum is no longer necessary, in which case (18) reduces to the familiar singleton case, namely

$$\begin{aligned}
 Y_s^l = & \sum_{j=1}^m \mu_{G^l}(y_j) * \mu_{F_1^l}(\bar{x}_1) * \dots * \mu_{F_p^l}(\bar{x}_p) / (y_j) \\
 \equiv & \sum_{j=1}^m \mu_{G^l}(y_j) * S_s / (y_j).
 \end{aligned} \quad (19)$$

Equations (18) and (19) show that in the nonsingleton system the scaling of the output fuzzy set for each rule will, in general, be different than in the case of a singleton system.

A general expression for quantifying the scale factor difference  $\delta s = S_{ns} - S_s$  can be found as a direct consequence of (18) and (19), i.e.,

$$\begin{aligned}
 \delta s = & S_{ns} - S_s \\
 = & \sup_{\mathbf{x} \in U} \left[ \mu_{F_1^l}(\bar{x}_1) * \dots * \mu_{F_p^l}(\bar{x}_p) + \sum_{i_1 \neq q_1} \dots \right. \\
 & \cdot \sum_{i_p \neq q_p} \mu_{X_1}(x_{1,i_1}) * \dots * \mu_{X_p}(x_{p,i_p}) * \mu_{F_1^l}(x_{1,i_1}) * \dots \\
 & \left. * \mu_{F_p^l}(x_{p,i_p}) \right] - \mu_{F_1^l}(\bar{x}_1) * \dots * \mu_{F_p^l}(\bar{x}_p)
 \end{aligned} \quad (20)$$

where, again “+” denotes union, but “−” denotes arithmetic subtraction.

The most widely used subclasses of FLS’s are those with product or minimum inference as a  $t$ -norm and with Gaussian or triangular membership functions. These subclasses are considered next.

## B. Product Inference

The  $k$ th input fuzzy set and the corresponding  $k$ th antecedent are assumed to have the following forms:

$$\begin{aligned}
 \mu_{X_k}(x_{k,i_k}) &= \exp \left( -\frac{(x_{k,i_k} - m_{X_k})^2}{2\sigma_{X_k}^2} \right), \\
 \mu_{F_k^l}(x_{k,i_k}) &= \exp \left( -\frac{(x_{k,i_k} - m_{F_k^l})^2}{2\sigma_{F_k^l}^2} \right)
 \end{aligned} \quad (21)$$

and “\*” in (11) denotes algebraic product, i.e.,

$$\begin{aligned}
 Y^l = & \sum_{j=1}^m \mu_{G^l}(y_j) \cdot \max_{\mathbf{x} \in U} \left( \sum_{i_1=1}^{n_1} \dots \sum_{i_p=1}^{n_p} [\mu_{X_1}(x_{1,i_1}) \right. \\
 & \left. \cdot \mu_{F_1^l}(x_{1,i_1})] \dots [\mu_{X_p}(x_{p,i_p}) \cdot \mu_{F_p^l}(x_{p,i_p})] \right) / (y_j).
 \end{aligned} \quad (22)$$

By maximizing the function

$$\mu_{Q_k^l}(x_{k,i_k}) = \mu_{X_k}(x_{k,i_k}) \cdot \mu_{F_k^l}(x_{k,i_k}) \quad (23)$$

we find that it is maximum at

$$x_{k,\max}^l = \frac{\sigma_{X_k}^2 m_{F_k^l} + \sigma_{F_k^l}^2 m_{X_k}}{\sigma_{X_k}^2 + \sigma_{F_k^l}^2}. \quad (24)$$

Substituting (24) into (22), the latter becomes

$$Y^l = \sum_{j=1}^m \mu_{G^l}(y_j) \prod_{k=1}^p \mu_{Q_k^l}(x_{k,\max}^l) / (y_j). \quad (25)$$

An important special case occurs when all input points for each input variable have the same level of uncertainty (after they are fuzzified using the nonsingleton fuzzifier) so  $\sigma_{X_k}^2 = \sigma_X^2$ . In this case, we also set  $m_{X_k} = x'_k$  and, therefore

$$x_{k,\max}^l = \frac{\sigma_X^2 m_{F_k^l} + \sigma_{F_k^l}^2 x'_k}{\sigma_X^2 + \sigma_{F_k^l}^2}. \quad (26)$$

For the special case of product inference, Gaussian membership functions, and equal uncertainty on all inputs, the nonsingleton FBF’s given by (16) become

$$p_l(\mathbf{x}) = \frac{\prod_{k=1}^p \mu_{Q_k^l} \left( \frac{\sigma_X^2 m_{F_k^l} + \sigma_{F_k^l}^2 x'_k}{\sigma_X^2 + \sigma_{F_k^l}^2} \right) / (\delta^l)^2}{\sum_{l=1}^M \prod_{k=1}^p \mu_{Q_k^l} \left( \frac{\sigma_X^2 m_{F_k^l} + \sigma_{F_k^l}^2 x'_k}{\sigma_X^2 + \sigma_{F_k^l}^2} \right) / (\delta^l)^2}. \quad (27)$$

When the uncertainty of the input is zero,  $\sigma_X^2 = 0$ , and (26) reduces to the singleton case, i.e.,  $x_{k,\max}^l = x'_k$ . In this case, each  $\mu_{X_k}(x_{k,\max}^l) = \mu_{F_k^l}(x'_k)$ ,  $k = 1, \dots, p$  is unity [see (21)] and, therefore,  $\mu_{Q_k^l}(x_{k,\max}^l) = \mu_{F_k^l}(x'_k)$  so (25) reduces to

$$Y^l = \sum_{j=1}^m \mu_{G^l}(y_j) \prod_{k=1}^p \mu_{F_k^l}(x'_k) / (y_j) \quad (28)$$

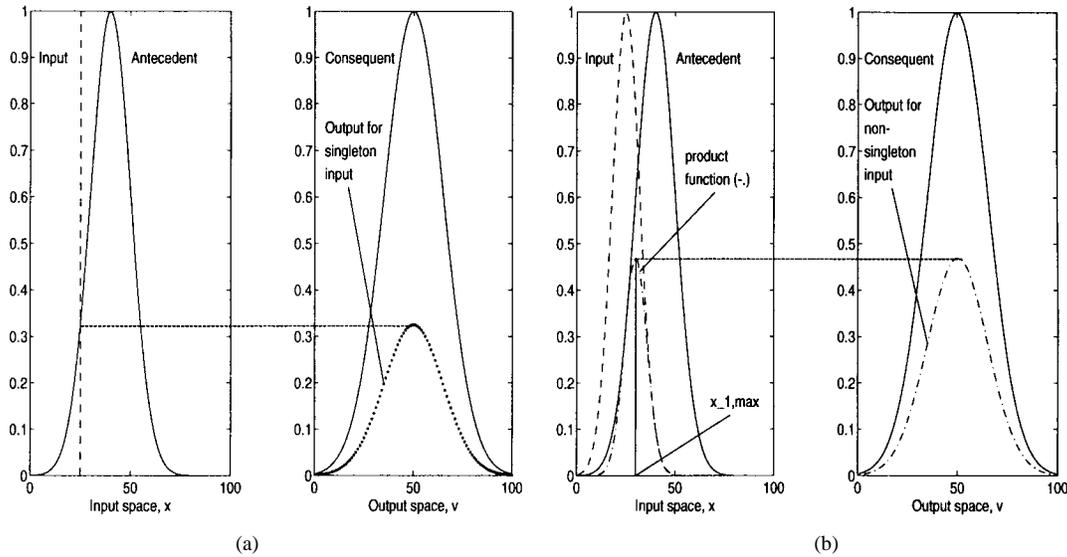


Fig. 1. (a) SISO FLS: composition of rule with singleton input and corresponding output. (b) SISO NSFLS: composition of rule with Gaussian input and corresponding output. This figure also illustrates (25) and (26) for a one-rule system ( $p = 1$ ,  $m_{F_1^l} = 40$ ,  $\sigma_{F_1^l} = 10$ ,  $x_1^l = 25$ ,  $\sigma_X = 7$ ,  $m_{G_1} = 50$ , and  $\sigma_{G_1} = 15$ ) for Gaussian membership functions. Equation (25) reduces to (19) if the input is singleton with “\*” replaced by product.

and the NSFLS in (15) reduces to the following singleton FLS:

$$f_s(\mathbf{x}') = \sum_{l=1}^M \bar{y}^l p_l^s(\mathbf{x}') = \frac{\sum_{l=1}^M \bar{y}^l \prod_{k=1}^p \mu_{F_k^l}(x'_k) / (\delta^l)^2}{\sum_{l=1}^M \prod_{k=1}^p \mu_{F_k^l}(x'_k) / (\delta^l)^2}. \quad (29)$$

Observe that we could have relabeled  $f_s(\mathbf{x}')$  to  $f_s(\mathbf{x})$ . Comparing (15) and (29), we see that nonsingleton and singleton FLS's are structurally the same.

To illustrate the differences between a NSFLS and a singleton FLS, we present the case of a one-rule ( $M = 1$ ) single-input single-output (SISO) FLS with Gaussian membership functions and product inference. For the nonsingleton system, the input is also fuzzified using a Gaussian membership function. Fig. 1 shows the output fuzzy sets corresponding to each system. Observe the Gaussian nature in Fig. 1(b) of the product function  $\mu_{Q_1^1}(x)$  and the location of its maximum value at  $x = x_{1,\max}^1$ . In this case, the output membership function for the NSFLS is of greater maximum height than that of the singleton FLS.

Fig. 2(a) demonstrates (20) for product inference and Gaussian membership functions. Fig. 2(b) demonstrates (20) for parabolic antecedent ( $\mu_{F_1^l} = \max(0, -((x - m_{F_1^l})/a)^2 + 1)$ ) and triangular input membership functions, for both minimum (“-”, “- -”) and product (“×”, “+”) inference. The bimodal curves in Fig. 2 indicate the scale factor difference between the two systems as the input sweeps over the entire universe of discourse of the antecedent membership function. The larger amplitude scale factor difference curves correspond to a nonsingleton input with higher level of uncertainty ( $\sigma_X \approx 0.7\sigma_F$ ), whereas the smaller amplitude difference curves correspond to nonsingleton input with lower level of uncertainty ( $\sigma_X \approx 0.2\sigma_F$ ). When the input singleton and the mean of the corresponding nonsingleton input are at the mean of the

antecedent [i.e.,  $m_{F_k^l} = m_{X_k}$  so that  $x_{k,\max}^l = m_{F_k^l}$  (see (24)], the difference in the scale factor is zero.

Figs. 1 and 2 clearly imply that (especially in the cases of more complex systems with more than one rule and multiple inputs) the defuzzified nonsingleton system output may be significantly different than the corresponding singleton system output.

Fig. 3(a), which is a direct implementation of (27) (with  $p = 1$ ,  $M = 4$ ,  $x_1 \in [1, 100]$ ,  $\delta^l = 1$ ,  $m_{F_1^l} = 20 \cdot l$ , and  $\sigma_{F_1^l} = 10$  for  $l = 1, \dots, M$ ) compares the fuzzy basis functions in a singleton system versus the fuzzy basis functions in a nonsingleton system. When the degree of uncertainty in the Gaussian input is large, the basis functions of (27) (shown in Fig. 3(a) by the dash-dotted line) are significantly different than in the singleton case (solid line). Note that, when the uncertainty at the input is small, then the nonsingleton case is essentially identical to the singleton case, and the nonsingleton (NS) FBF's plotted in Fig. 3 are indistinguishable from the singleton FBF's. For the example presented here, the higher uncertainty input had variance of the order of the variance of the antecedents, whereas the lower uncertainty input had variance about one twentieth of the variance of the antecedents. Fig. 3(b) depicts the fuzzy basis functions for a singleton (solid line) and a nonsingleton FLS with four rules, product inference, and triangular membership functions. Observe that the two exterior FBF's saturate at unity.

It should be noted that the FBF's given in these examples are symmetric because the centers of the antecedent membership functions were equispaced. In general, FBF's are not symmetric.

Greater uncertainty in the input not only “fires” rules at a higher level than a singleton system [(20) and Figs. 1 and 2], but it also usually fires more rules than a singleton system would. This is due to the fact that a nonsingleton input may have membership (or nonzero subsethood [12]) in more antecedent fuzzy regions than a singleton input. For example,

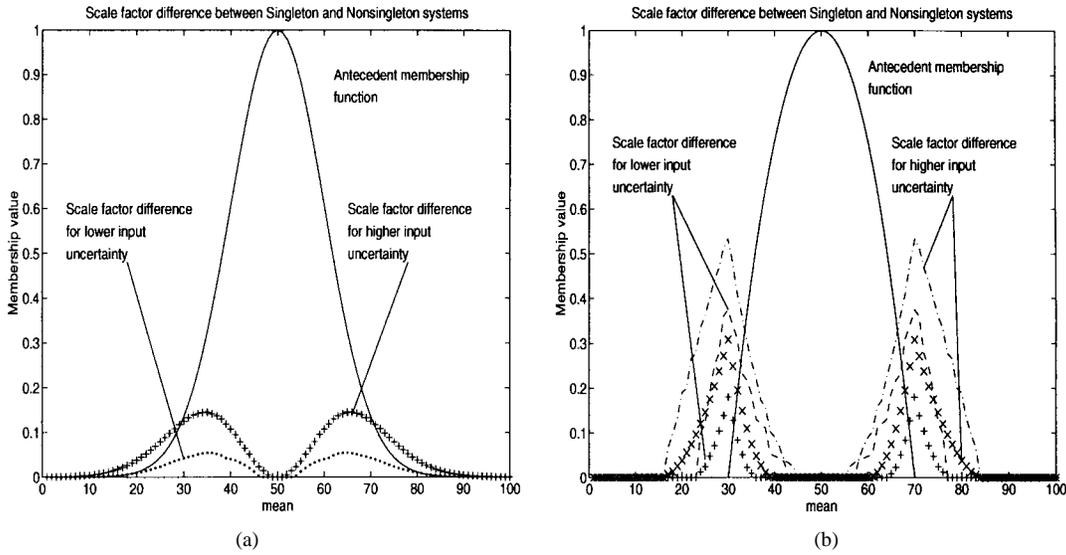


Fig. 2. (a) Scale-factor difference between SISO singleton and nonsingleton systems over the entire input range for one antecedent. For higher levels of uncertainty in the input (represented by larger variance in the Gaussian nonsingleton input), the scale factor difference is larger. This difference in scale factors is also evident in Fig. 1 for one particular case of singleton and nonsingleton inputs. The horizontal axis represents the means of the Gaussian input in the nonsingleton case and the location of the singleton input in the singleton case. (b) Scale-factor difference between SISO singleton and nonsingleton systems over the entire input range for one parabolic antecedent and triangular input (“-.” and “- -” represent minimum inference and “×” and “+” represent product inference).

a vertical line at  $x = 40$  in Fig. 3(a) intersects three singleton FBF’s and four nonsingleton FBF’s.

During the review process, one of the reviewers observed that in the special case of Gaussian membership functions and product inference, it is sometimes possible to interpret the NSFLS as a SFLS. This can be easily seen by substituting (26) into (23) and carrying out some simple algebraic calculations. Of course, even then that system contains the same information about the nonsingleton input set. This observation, however, is not true in general, e.g., the case of parabolic antecedent and triangular membership functions with any of the commonly used  $t$ -norms. One of the strengths of fuzzy logic is that fuzzy system formulas can be described by meaningful linguistic rules, and the rules can be described by an underlying logic. Shifting fuzziness from the input into rule antecedents changes the meaning of rules and in many applications this does not make a lot of sense (see, for example, [44, pp. 185–189]). In many other applications, the input can only be defined as a set or an interval [21], [23], [43].

### C. Minimum Inference

Now, we let “\*” in (11) be the  $t$ -norm ‘min’, in which case  $\mu_{Q_k^l}(x_k, i_k)$  in (12) is

$$\mu_{Q_k^l}(x_k, i_k) \equiv \min(\mu_{X_k}(x_k, i_k), \mu_{F_k^l}(x_k, i_k)) \quad (30)$$

for finite universes of discourse. As in the case of the algebraic product  $t$ -norm, the term in the parenthesis in (11) is maximized when every grouped pair is a maximum, i.e.,  $\mu_{Q_k^l}(x_k, i_k)$  must be maximized over all  $k$  and all  $x$ .

For the input fuzzy set  $X_k$  to activate the fuzzy rule  $R^l$  at  $F_k^l$ , some of the points in the support of  $X_k$  must coincide with some points in the support of  $F_k^l$ . Let this set of points define the closed interval  $[a_k, c_k]$  such that  $b_k \in (a_k, c_k)$  and  $a_k < b_k < c_k$  (Fig. 4). Without loss of

generality, let  $\mu_{F_k^l}(x_k, i_k)$  be strictly increasing and  $\mu_{X_k}(x_k, i_k)$  be strictly decreasing on the interval  $[a_k, c_k]$  and, consequently, on any closed subintervals  $[a_k, b_k], [b_k, c_k]$ . [In the case of  $\mu_{X_k}(x_k, i_k)$  being a proper subset of  $\mu_{F_k^l}(x_k, i_k)$ , i.e.,  $\mu_{X_k}(x_k, i_k) \subset \mu_{F_k^l}(x_k, i_k)$ , and  $\mu_{X_k}(x_k, i_k) \neq \mu_{F_k^l}(x_k, i_k)$  for at least one  $x_k, i_k$ , or if  $\mu_{X_k}(x_k, i_k) = \mu_{F_k^l}(x_k, i_k)$ , then obviously  $\max_{\forall x} \mu_{Q_k^l}(x_k, i_k) \equiv \max_{\forall x} \mu_{X_k}(x_k, i_k)$ .] Let  $b_k$  be the point of intersection of  $\mu_{F_k^l}$  and  $\mu_{X_k}$ . From the monotonicity conditions on the membership functions in these intervals we have

$$\begin{aligned} \forall x \in [a_k, b_k], \quad \min(\mu_{X_k}(x_k, i_k), \mu_{F_k^l}(x_k, i_k)) &= \mu_{F_k^l}(x_k, i_k) \\ \text{where } \mu_{F_k^l}(x_k, 1) < \mu_{F_k^l}(x_k, 2) &\text{ if } x_k, 1 < x_k, 2. \end{aligned} \quad (31)$$

Thus

$$\max_{x \in [a_k, b_k]} \mu_{F_k^l}(x_k, i_k) = \mu_{F_k^l}(b_k). \quad (32)$$

Similarly

$$\begin{aligned} \forall x \in [b_k, c_k], \quad \min(\mu_{X_k}(x_k, i_k), \mu_{F_k^l}(x_k, i_k)) &= \mu_{X_k}(x_k, i_k) \\ \text{where } \mu_{X_k}(x_k, 3) > \mu_{X_k}(x_k, 4) &\text{ if } x_k, 3 < x_k, 4 \end{aligned} \quad (33)$$

so that

$$\max_{x \in [b_k, c_k]} \mu_{X_k}(x_k, i_k) = \mu_{X_k}(b_k). \quad (34)$$

Consequently,  $\mu_{Q_k^l}(x_k, i_k)$  becomes a maximum at the point of intersection of the two membership functions, which is also intuitively satisfying. For example, for the previously

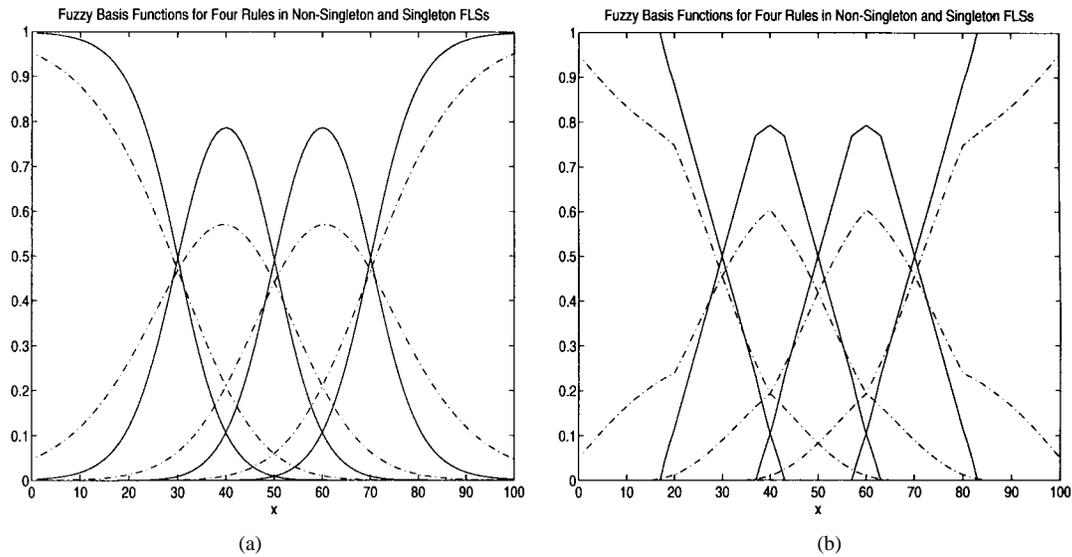


Fig. 3. (a) Fuzzy basis functions for nonsingleton (dash-dotted line) and singleton fuzzy logic systems with four rules, for Gaussian input and antecedent membership functions, and for product inference. (b) Fuzzy basis functions for nonsingleton (dash-dotted line) and singleton fuzzy logic systems with four rules, for triangular input and antecedent membership functions, and for product inference. The antecedent membership functions were chosen to be the same in the NS FBF's and singleton FBF's.

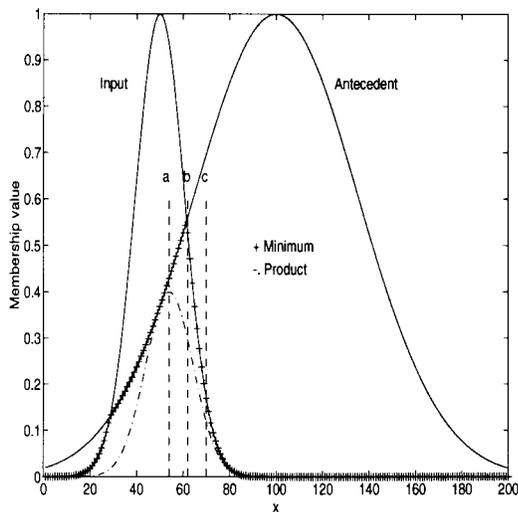


Fig. 4. Membership functions of nonsingleton fuzzy input  $\mu_{X_k}(x_{k,i_k})$  and corresponding antecedent  $\mu_{F_k^l}(x_{k,i_k})$ , their product and minimum.

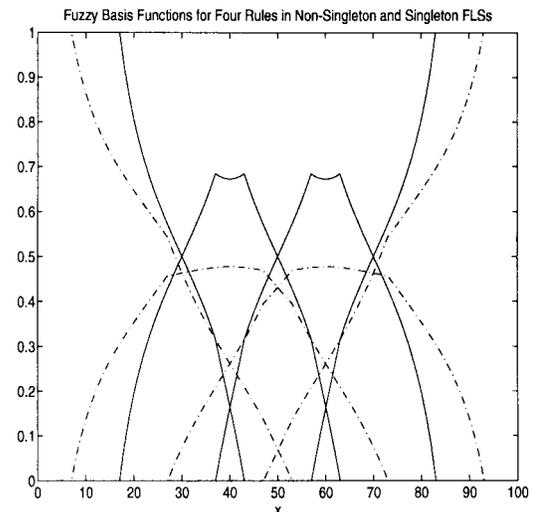


Fig. 5. Fuzzy basis functions for nonsingleton (dash-dotted line) and singleton fuzzy logic systems with four rules, for parabolic antecedent and triangular input membership functions, and minimum inference.

described Gaussian membership functions (21), by equating their exponents and solving a second-order equation, we find the point of intersection to be given by

$$x_{k,\max}^l = \frac{\sigma_X m_{F_k^l} + \sigma_{F_k^l} x_k}{\sigma_X + \sigma_{F_k^l}}. \quad (35)$$

The output fuzzy set  $Y^l$  is computed exactly as in (25), after replacing product with  $\min$ , and  $x_{k,\max}^l$  with the expression in (35). Fig. 5 depicts the fuzzy basis functions for a singleton (solid line) and a nonsingleton FLS with four rules, minimum inference, parabolic antecedent, and triangular input membership functions.

The complete solutions for the maximization of  $\mu_{Q_k^l}$  for algebraic product and minimum  $t$ -norms are summarized in

Table I, along with the corresponding solutions for the case of triangular membership functions. The centers of the antecedent and input triangular membership functions are denoted by  $m_{F_k^l}$  and  $m_{X_k}$ , respectively, and their corresponding left and right spreads by  $l_{F_k^l}$ ,  $l_{X_k}$ ,  $r_{F_k^l}$ , and  $r_{X_k}$ . Observe, from Table I, that  $x_{k,\max}^l$  is easiest to calculate in the cases of product and minimum inference with Gaussian membership functions, and minimum inference with triangular membership functions.

#### IV. UNIFORM APPROXIMATION PROPERTY OF NONSINGLETON FUZZY LOGIC SYSTEMS

Wang and Mendel [38] and Kosko [13] have shown that specific subclasses of singleton fuzzy logic systems are universal approximators. In [38], the Stone–Weierstrass theorem

TABLE I  
 $x_{k,\max}^l$  FOR GAUSSIAN AND TRIANGULAR MEMBERSHIP FUNCTIONS IN THE CASES OF ALGEBRAIC PRODUCT  
 AND MINIMUM  $t$ -NORMS (DUE TO SPACE LIMITATIONS, DETAILS OF DERIVATIONS ARE OMITTED)

t-norms	$x_{k,\max}^l$ for Gaussian membership functions	$x_{k,\max}^l$ for triangular membership functions
Product	$\frac{\sigma_{X_k}^2 m_{F_k^l} + \sigma_{F_k^l}^2 m_{X_k}}{\sigma_{X_k}^2 + \sigma_{F_k^l}^2}$	if $m_{X_k} < m_{F_k^l}$ and $\gamma_1 \in [m_{X_k}, m_{F_k^l}]$ , then $x_{k,\max}^l = \gamma_1$ $\text{where } \gamma_1 \equiv \frac{m_{X_k} + m_{F_k^l} - l_{F_k^l} + r_{X_k}}{2}$ if $l_{F_k^l} > r_{X_k}$ and $\gamma_1 \notin [m_{X_k}, m_{F_k^l}]$ , then $x_{k,\max}^l = m_{X_k}$ if $l_{F_k^l} < r_{X_k}$ and $\gamma_1 \notin [m_{X_k}, m_{F_k^l}]$ , then $x_{k,\max}^l = m_{F_k^l}$ , if $m_{X_k} > m_{F_k^l}$ and $\gamma_2 \in [m_{X_k}, m_{F_k^l}]$ , then $x_{k,\max}^l = \gamma_2$ $\text{where } \gamma_2 \equiv \frac{m_{F_k^l} + m_{X_k} - l_{X_k} + r_{F_k^l}}{2}$ if $l_{X_k} > r_{F_k^l}$ and $\gamma_2 \notin [m_{F_k^l}, m_{X_k}]$ , then $x_{k,\max}^l = m_{F_k^l}$ if $l_{X_k} < r_{F_k^l}$ and $\gamma_2 \notin [m_{F_k^l}, m_{X_k}]$ , then $x_{k,\max}^l = m_{X_k}$ if $m_{X_k} = m_{F_k^l} = m$ , then $x_{k,\max}^l = m$
Minimum	$\frac{\sigma_{X_k} m_{F_k^l} + \sigma_{F_k^l} m_{X_k}}{\sigma_{X_k} + \sigma_{F_k^l}}$	$\frac{l_{F_k^l} m_{X_k} + r_{X_k} m_{F_k^l}}{l_{F_k^l} + r_{X_k}} \text{ if } m_{X_k} \leq m_{F_k^l}$ $\frac{l_{F_k^l} m_{X_k} + r_{X_k} m_{F_k^l}}{l_{F_k^l} + r_{X_k}} \text{ if } m_{X_k} \leq m_{F_k^l}$

(involving algebras of functions) was used for approximations in the uniform topology on a compact set. This result was proven only for singleton FLS's with product inference and Gaussian membership functions. In [13], continuity was used to prove uniform approximation capability of singleton additive systems. Buckley has also used the Stone–Weierstrass theorem to show that Sugeno-type controllers are universal controllers [3]. Additional results on the approximation capabilities of different structures of FLS's are presented in [9]. Other approaches parallel results from the neural networks literature (such as in [14]).

Although (15) bears structural resemblance to Kolmogorov's representation theorem [19], the two are quite different. Kolmogorov's theorem (which resolved Hilbert's 13th problem) provides a representation of a continuous function of several variables defined on an  $n$ -dimensional cube by sums and superpositions of continuous functions of one variable. In contrast to fuzzy basis functions (16), the nonlinear functions employed in Kolmogorov's theorem, as well as in the subsequent improvements of that theorem by Lorentz [19] and Sprecher [32], are highly nonsmooth and strictly increasing. Even though these nonsmooth functions can be viewed as limits of uniformly converging series of smooth *sigmoidal*-type functions [15], [16] the monotonicity requirements are too restrictive for our basis functions.

Utilizing concepts from real analysis [30], we will show that a NSFLS can uniformly approximate any continuous function  $g: U \rightarrow V$  on a compact set. This approach includes the systems described in [38] as a special case.

*Definition 1:* Let  $c_1^i < c_2^i < \dots < c_M^i \in U_i, i = 1, 2, \dots, p$  form a partition of  $U_i$ . Then, the largest partition interval in  $U_i$  is given by  $\delta^i = \max_j (c_j^i - c_{j-1}^i)$ .

*Definition 2:* Let  $\mathcal{S}$  denote the set of points of support and  $m \in \mathcal{S}$  the mean of a one-dimensional convex fuzzy set with membership function  $\mu$  and  $|\sup(\mathcal{S}) - \inf(\mathcal{S})| < \infty$ . The *left spread* of the fuzzy set is defined as  $s_L = m - \inf(\mathcal{S})$ , and the *right spread* is defined as  $s_R = \sup(\mathcal{S}) - m$ . For fuzzy sets with infinite support and strictly decreasing membership values while moving away from  $m$  (such as the Gaussian), the left and right spreads are defined as  $s_L = m - x_L, s_R = x_R - m$ , where  $x_L \leq m \leq x_R, \mu(x_L) = \mu(x_R) = \eta$  and  $\eta$  is a small, nonzero membership value. Alternatively,  $s_L = m - \inf(\mathcal{S}^\eta), s_R = \sup(\mathcal{S}^\eta) - m$ , where  $\mathcal{S}^\eta$  is the (finite)  $\eta$ -level set of the fuzzy set with infinite support. Clearly, for symmetric fuzzy sets,  $s_L = s_R$ . As has been illustrated in (2) and (3), the  $p$ -dimensional antecedent and input membership functions can be composed from their corresponding  $p$  one-dimensional fuzzy sets.

The following definition imposes certain restrictions on the type of membership functions that can be used and some of their parameters.

*Definition 3:* The class of *admissible* antecedent and input membership functions defined on  $U$  should have the following properties:

- [P1] membership functions should be normalized convex fuzzy sets (fuzzy numbers) or fuzzy intervals;
- [P2] membership function spread should be no larger than  $\delta = [\sum_{i=1}^p (\delta^i)^2]^{1/2}$ , but large enough so that the union of the supports, or the  $\eta$ -level sets of all antecedents cover the entire input space; input membership function spread should not be larger than the antecedent membership function spread.

Fig. 6 shows examples of admissible and inadmissible membership functions.

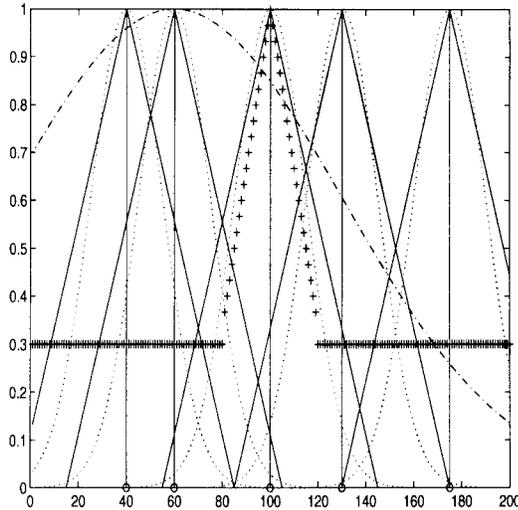


Fig. 6. Examples of admissible (solid and dotted lines) and inadmissible (cross and dash-dotted lines) membership functions.

*Theorem:* Let nonsingleton fuzzy logic system  $f_{ns}$  in (15) be defined for admissible membership functions and any  $t$ -norm such that  $center(Y^t = \sum_{j=1}^m \mu_{G^t}(y_j) * S_{ns}/(y_j)) = center(G^t)$ . Then, any given continuous function  $g: U \rightarrow V$  on a compact set  $U \in R^p$  can be uniformly approximated by  $f_{ns}$ , as  $M$  becomes sufficiently large.

Note that in the case of minimum  $t$ -norm the consequent fuzzy sets must also be symmetric. Note, also, that the condition  $center(Y^t = \sum_{j=1}^m \mu_{G^t}(y_j) * S_{ns}/(y_j)) = center(G^t)$  is satisfied by  $t$ -norms such as algebraic product, Einstein product, bounded difference, Hamacher product, and for consequent membership functions of arbitrary shape.

*Proof:* (See Appendix A.)

This is an existence theorem which gives us assurance that there exists an NSFLS capable of uniformly approximating any continuous function on compact a. Although the construction of the proof provides some insight, it does not tell us how to choose some of the parameters of the NSFLS—nor does it tell us how many basis functions will be needed to achieve such performance.

## V. MODELING AND PREDICTION OF COMPLEX SYSTEMS: CHAOTIC TIME SERIES

Several well-known algorithms exist [2], [5], [40] for the calculation of dynamical and geometric invariants such as fractal dimension (capacity, correlation, information) and Lyapunov exponents of an underlying strange attractor of time series. The largest Lyapunov exponent is essentially a measure of how predictable a system can be from observations of the past, whereas the fractal dimension is an indication of the complexity of a system. Unfortunately, data requirements and extreme sensitivity to noisy measurements make these algorithms prohibitively expensive in practical applications. Even if these invariants could be calculated accurately, they still do not provide enough information for the construction of a predictive model.

Other approaches attempt to construct a predictive model based solely on the given time-series data. Several such approaches (e.g., Gabor polynomials, linear prediction) have

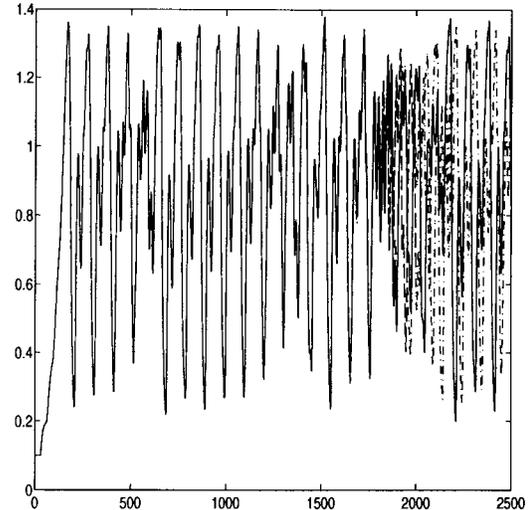


Fig. 7. Sample waveforms of the Mackey–Glass chaotic time series ( $\tau = 30$ ) for two different, but very close, initial conditions (all values in the initial value interval are set to 0.1 for the first waveform and to 0.100001 for the second waveform). Observe that after a certain point, the two realizations evolve completely differently.

been compared to prediction using feedforward neural networks with the best results occurring for the neural networks [17]. In [35] it was shown that a numerical-fuzzy approach based on a singleton FLS can give results equivalent to those or even better than that of a neural network. Here, we show that our nonsingleton FLS can significantly outperform an equivalent singleton FLS in the predictive modeling of a chaotic time series, especially in cases where either or both the training data and input data are corrupted by noise.

Chaos is having an impact on many different fields including physics, biology, chemistry, economics, and medicine [4], [6], [29]. The chaotic time series that we examine is a model for blood production due to Mackey and Glass [20]. This chaotic time series is described by the following delay differential equation that is now known as the “Mackey–Glass equation”:

$$\frac{dx(t)}{dt} = \frac{0.2x(t-\tau)}{1+x^{10}(t-\tau)} - 0.1x(t). \quad (36)$$

For  $\tau$  greater than 17, (36) exhibits chaotic behavior.

Chaotic behavior can be described as bounded fluctuations of the output of a system with high degree of sensitivity to initial conditions [4], i.e., trajectories with nearly identical initial conditions diverge exponentially (note that exponential divergence of two distinct trajectories does not imply unboundedness. Fig. 7 illustrates this kind of behavior). A system such as the one in (36), exhibiting chaotic dynamics, evolves in a deterministic manner. However, the correlation of observations appears to be limited; thus, prediction of the state of the system is particularly difficult [29].

To demonstrate the qualitative nature of the dynamical system given by (36), we display representative portions of the time series for different values of  $\tau$  in Fig. 8(a) and (b). We also depict the amplitude spectrum estimates of these representative sections in Fig. 8(c) and (d), and their corresponding two-dimensional phase plots in Fig. 8(e) and (f). From these plots, we are able to distinguish periodic behavior for small  $\tau$ 's and chaotic behavior for larger  $\tau$ 's. Most commonly, the

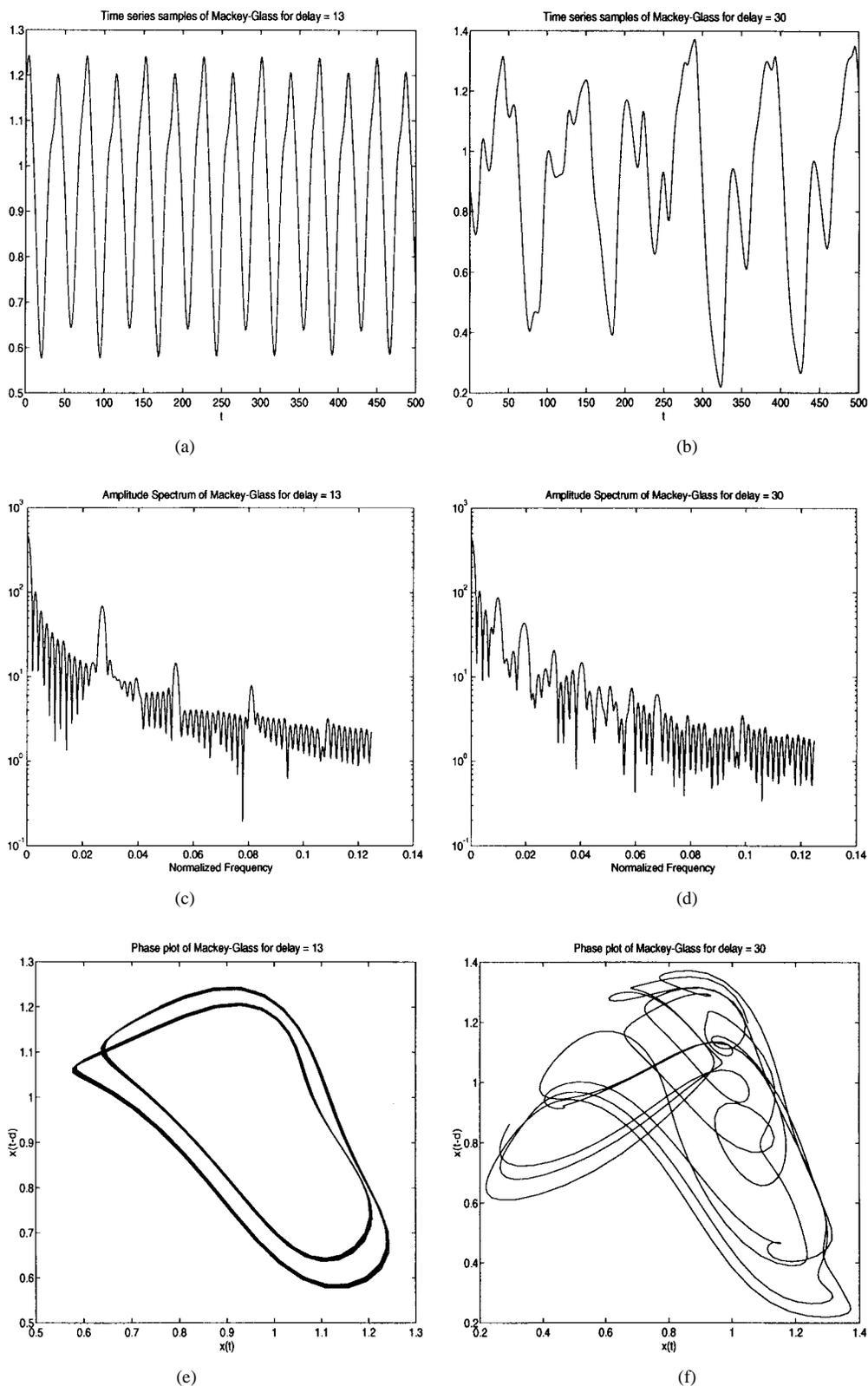


Fig. 8. (a) and (b) Representative samples of the Mackey–Glass time series after letting transients relax. (c) and (d) Amplitude spectra plotted on a semilog scale for the examples illustrated in (a) and (b). (e) and (f) Corresponding phase plots of the time series segments depicted in (a) and (b). (“ $d$ ” in  $x(t - d)$  on the vertical axis denotes the delay used in the Mackey–Glass equation; it is either 13 or 30.)

transition of a system from predictability to chaos occurs via a cascade of bifurcations which are qualitative changes in the limit sets of a system, where a limit set is the set of points whose every neighborhood  $\mathcal{B}$  is repeatedly entered by the

solution  $\phi_t(x)$  of the differential equation as  $t \rightarrow \infty$ . For  $\tau$  close to 17, which is the onset of chaos for the Mackey–Glass equation, the bifurcations get closer and closer together until the system breaks down into chaotic behavior.

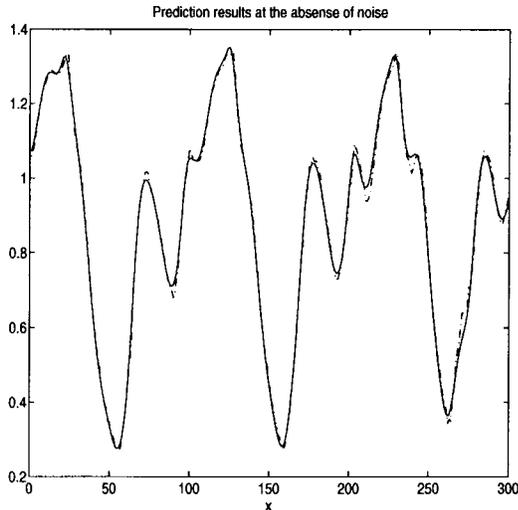


Fig. 9. Prediction results in the absence of noise, for 300 points of (36) ( $x(1005)$  to  $x(1304)$ ), after allowing 1000 points for the transients to die out). The solid line is the actual time series, whereas the dash-dotted line shows the predicted estimates. In this case, the NSFLS and the singleton FLS produce identical results. The same parameters of the singleton FLS used here are also used in the subsequent cases when noise is added. All the common parameters of the NSFLS and the singleton FLS are identical.

The largest Lyapunov exponent [6]  $\lambda$  which quantifies the rate of divergence of the time series and, therefore, the degree to which the system is predictable in the long term, can be estimated using nonlinear regression [26]. If  $\lambda > 0$ , then the system will exhibit chaotic behavior (this property makes it possible to detect chaotic dynamics in a noisy environment just by examining the sign of  $\lambda$ ). We estimated  $\lambda$  to be 0.007 for the Mackey–Glass equation with  $\tau = 30$ , thus confirming that the system is indeed chaotic.

We designed the NSFLS in (17) to be a single-step predictor of the Mackey–Glass chaotic time series. Our design is based on  $N$ -training pairs that we obtained by simulating (36) for  $\tau = 30$ , using a finite-dimensional iterated mapping. A higher order Adams–Bashforth integration algorithm [27] with sufficiently small step size would be a good choice for the simulation of (36) (note that for the higher order algorithms to yield better results than the lower order ones, the integration step should be chosen to be sufficiently small). In our simulations, we assumed that only noisy measured values of  $x(t)$  are available, namely  $\zeta(t) = x(t) + n(t)$ ,  $t = 1, \dots, N$ . The input vector to our fuzzy logic system is  $\underline{u}(t-1) = [\zeta(t-1), \zeta(t-2), \dots, \zeta(t-p)]^T$  so that the estimate of the time series based on our NSFLS is given by

$$\hat{x}(t) = f_{ns}(\underline{u}(t-1)). \quad (37)$$

Fig. 9 depicts the prediction results in the absence of noise for 300 points of (36) [ $x(1005)$  to  $x(1304)$ ] after allowing 1000 points for the transients to die out), produced by our singleton FLS using a numerical fuzzy algorithm (all fuzzy rules were generated from the training data). Five hundred input-output training pairs of five inputs ( $p = 5$ ) and a single output were used to construct the system; these were drawn from  $x(1)$  to  $x(1004)$  where the  $i$ th input of the  $l$ th training pair was selected as  $x(2(l-1) + 6 - i)$ ,  $i = 1, \dots, p$ , and

the  $l$ th output was selected as  $x(2(l-1) + 6)$ ,  $l = 1, \dots, 500$ . The same numerical-fuzzy algorithm and number of training pairs were used to construct a nonsingleton system of the form of (15). All the common parameters between the singleton and nonsingleton FLS's were chosen to be the same. For all simulations,  $\sigma_X$  in the NSFLS case was chosen equal to the standard deviation of the additive noise. In a noise-free environment for which  $\sigma_X = 0$ , the performance of the nonsingleton system was identical to that of the singleton system. As shown in Fig. 9, both systems can closely predict (36) in the given range. Both systems were trained using the one-pass method described in [37], with  $M = 500$  (basis functions) and  $\delta = 1$ .

Using the same common parameters for both systems (as in the noise free case) we added zero-mean uniform noise to the training and test sets and we retrained both systems. Again, 500 input-output training pairs of five inputs and a single output were used to construct the system drawn from  $\zeta(1)$  to  $\zeta(1004)$ , as described above. Fig. 10 shows typical prediction results for both the singleton and the nonsingleton systems at signal-to-noise ratios (SNR's) (defined as  $20 \cdot \log \sigma_{\text{signal}}/\sigma_{\text{noise}}$ , where  $\sigma$  denotes standard deviation) of  $-5$ ,  $0$ , and  $10$  dB. Fig. 11 summarizes the performance of the nonsingleton and singleton systems for these SNR's averaged over 100 realizations. The average mean-squared error and average standard deviation for each case are given in Table II. Fig. 12(a) shows the prediction results based on our NSFLS for 20 different realizations and SNR = 0 dB, overlapped on the same figure to give a visual representation of the standard deviation of the predicted estimates. Fig. 12(b) shows the prediction results based on the corresponding singleton FLS for 20 realizations. It is evident from this figure that the singleton FLS does not appear to be able to deal effectively with the uncertainty during the prediction process due to the noise in the training data and the input. Our NSFLS is significantly more successful in producing better predicted estimates of the chaotic time series.

Table III summarizes the average mean-squared errors and average standard deviations for the same experiments as above, but with parabolic antecedent and triangular input membership functions and minimum inference. Observe that the results in Tables II and III are quite similar, hence, the improved forecasting ability of our NSFLS is *not* limited to NSFLS's with Gaussian membership functions.

Note that no attempt was made to optimize the FLS's used here. Algorithms aimed at obtaining optimum or near-optimum NSFLS's are described in [24].

## VI. CONCLUSIONS

We have presented the continuous and discrete cases of a general NSFLS, and its efficient computation for the special subclasses of Gaussian and triangular membership functions with product or minimum inference. We have also developed a simple equation that describes the scale-factor difference between singleton and nonsingleton systems; it lets us examine and compare the behavior of singleton and nonsingleton systems for different types of

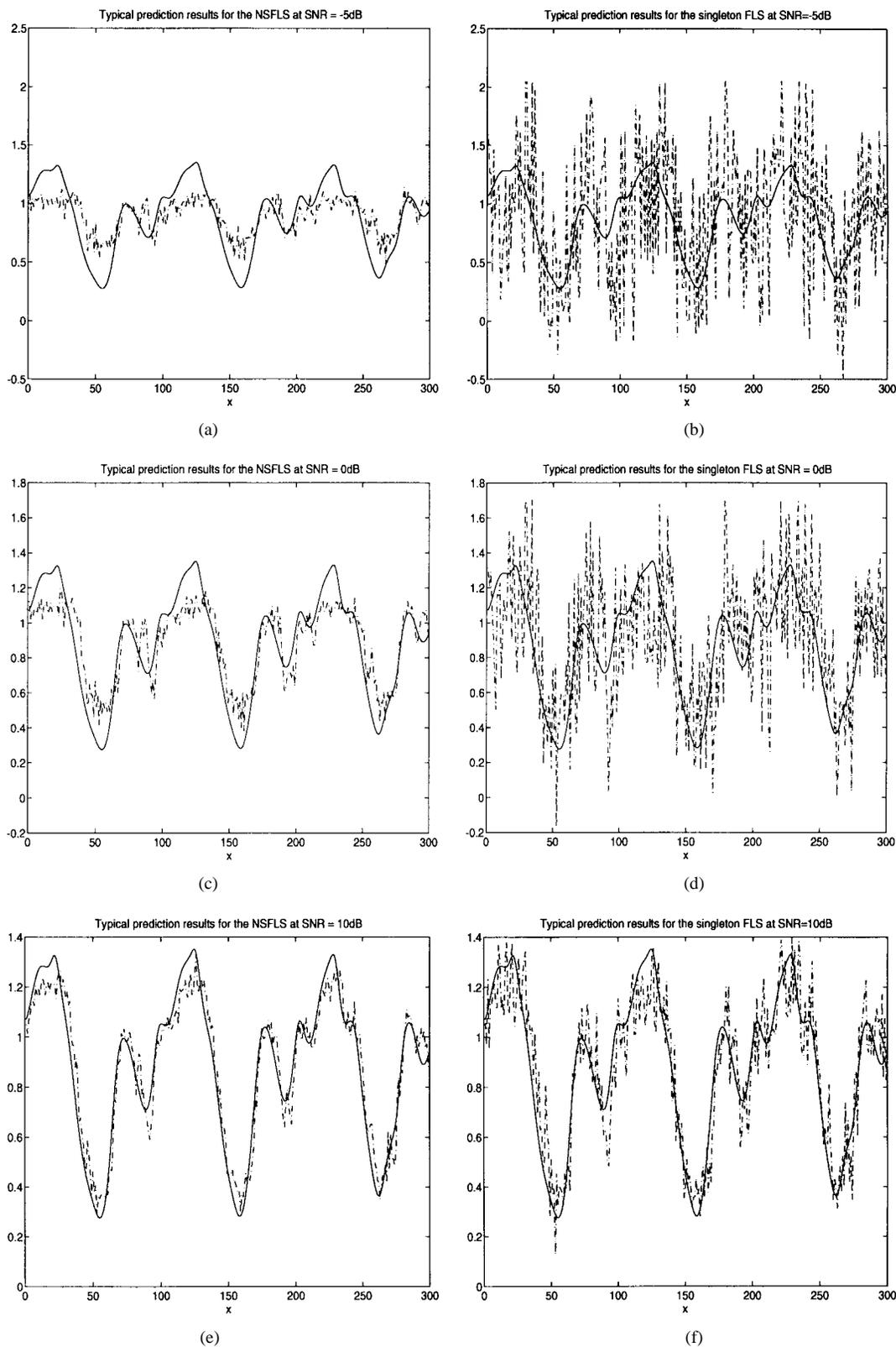


Fig. 10. Typical prediction results (single realization) for nonsingleton and singleton fuzzy logic systems for different signal-to-noise ratios (solid line: actual time series, dash-dotted line: predicted estimates), for 300 points of (36) ( $x(1005)$  to  $x(1304)$ ) after allowing 1000 points for the transients to die out).

inputs and uncertainty levels. We have also shown that a NSFLS can be expressed as a linear combination of nonsingleton fuzzy basis functions. Although, conceptually, singleton and nonsingleton FLS's are very different, the latter does have essential structural similarities with the

former, and they become identical in the absence of noise or uncertainty at the input. We have also shown that the NSFLS (15) can uniformly approximate a continuous function on a compact set, i.e., it is a universal approximator.

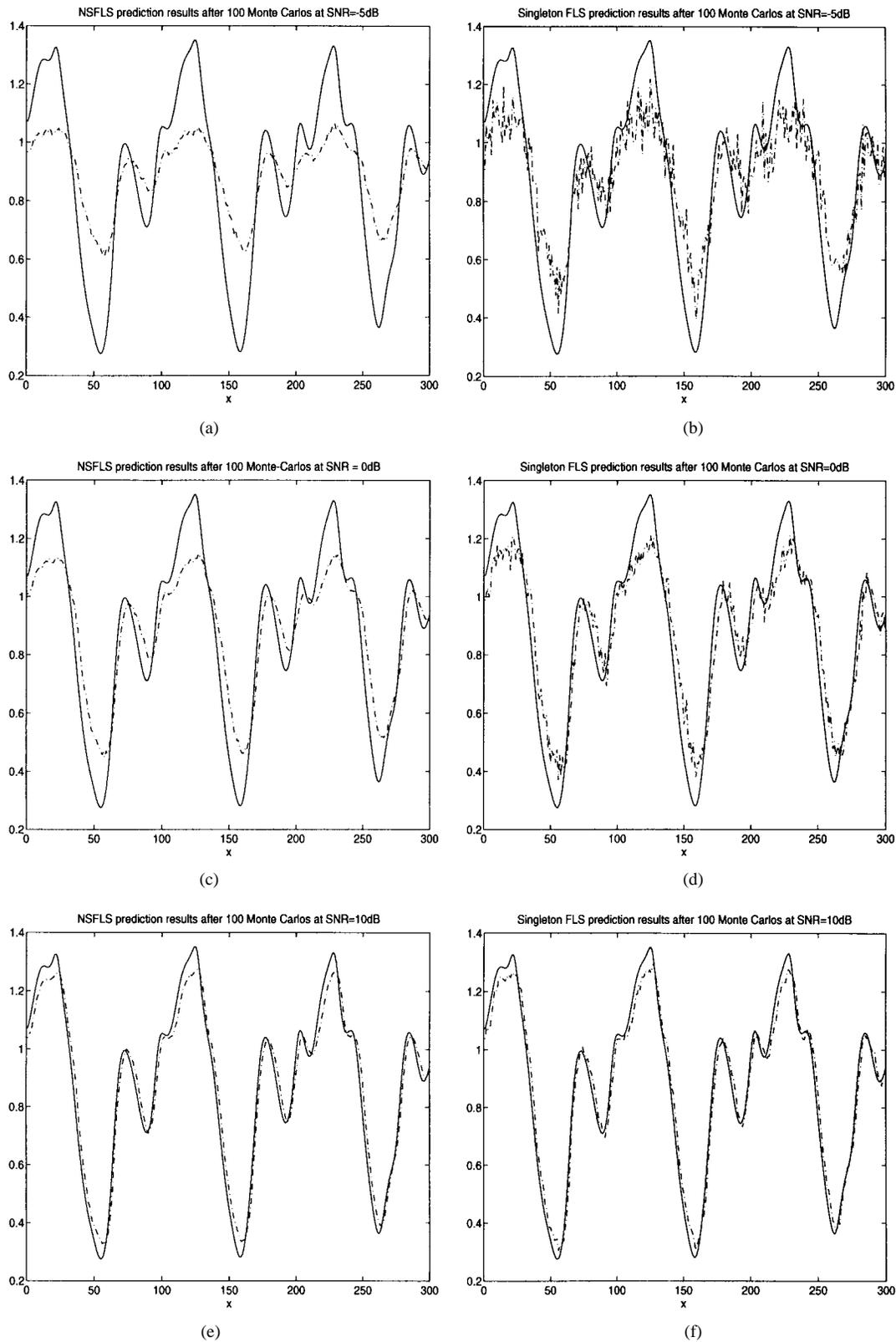


Fig. 11. Averaged prediction results for 100 Monte Carlo runs (solid line: actual time-series, dash-dotted line: predicted estimates) for 300 points of (36) ( $x(1005)$  to  $x(1304)$ ) after allowing 1000 points for the transients to die out). In all cases, the mean-squared errors and the corresponding standard deviations are significantly smaller in the nonsingleton case (see Table II).

The utility of our NSFLS and its superiority over a singleton FLS was demonstrated through our example on the prediction of a chaotic time series from noisy observations. In all cases, our NSFLS significantly outperformed a cor-

responding singleton FLS. Both were trained in the same manner with identical common parameters and were capable of closely approximating the time series in a noise-free environment.

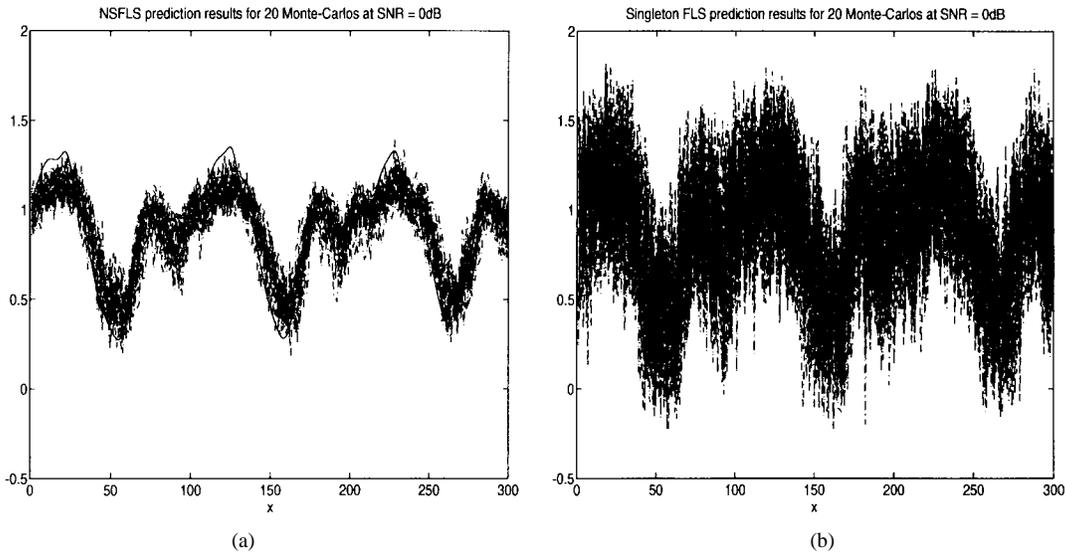


Fig. 12. (a) Prediction results for the NSFLS for 20 different noise realizations at 0 dB. (b) Prediction results for the singleton FLS for 20 different noise realizations at 0 dB. These figures show that the standard deviation of the prediction results is much higher in the singleton case than in the nonsingleton case. Predictions are given for 300 points of (36) ( $x(1005)$  to  $x(1304)$ ) after allowing 1000 points for the transients to die out).

TABLE II

AVERAGE STANDARD DEVIATIONS AND AVERAGE MEAN-SQUARED ERRORS FOR DIFFERENT SIGNAL-TO-NOISE RATIOS, FOR SINGLETON AND NONSINGLETON SYSTEMS WITH GAUSSIAN MEMBERSHIP FUNCTIONS. IN THE NOISE-FREE CASE, THE AVERAGE STANDARD DEVIATION AND MEAN-SQUARED ERROR WERE  $0.0168$  AND  $2.8329 \times 10^{-4}$ , RESPECTIVELY, FOR BOTH SYSTEMS

SNR	-5dB	0dB	10dB
Singleton FLS, avg. std	0.5517	0.3308	0.1281
Nonsingleton FLS, avg. std	0.2128	0.1496	0.0765
Singleton FLS avg. mse	0.3066	0.1101	0.0165
Nonsingleton FLS avg. mse	0.0462	0.0228	0.0059

TABLE III

AVERAGE STANDARD DEVIATIONS AND AVERAGE MEAN-SQUARED ERRORS FOR DIFFERENT SIGNAL-TO-NOISE RATIOS FOR SINGLETON AND NONSINGLETON SYSTEMS WITH PARABOLIC ANTECEDENT AND TRIANGULAR INPUT MEMBERSHIP FUNCTIONS. IN THE NOISE-FREE CASE, THE AVERAGE STANDARD DEVIATION AND MEAN-SQUARED ERROR WERE  $0.0159$  AND  $2.5233 \times 10^{-4}$ , RESPECTIVELY, FOR BOTH SYSTEMS

SNR	-5dB	0dB	10dB
Singleton FLS, avg. std	0.5418	0.3545	0.1123
Nonsingleton FLS, avg. std	0.2227	0.1345	0.0766
Singleton FLS avg. mse	0.3185	0.1345	0.0132
Nonsingleton FLS avg. mse	0.0504	0.0242	0.0057

The fuzzification subsystem within a FLS lets us handle *uncertainty* in a very natural way. It provides a mapping of noisy data  $\mathbf{x}$  into  $\mathbf{x}_{\max}^t$ , e.g., (26), that can be interpreted as a form of nonlinear data prefiltering [22]. To date, there does not seem to be a comparable way to handle uncertainty in a feedforward neural network.

In summary, we have found NSFLS's to be very promising and an effective tool for dealing with uncertainty and for designing systems with set-valued inputs totally within the framework of fuzzy logic.

#### APPENDIX A

##### PROOF OF UNIFORM APPROXIMATION THEOREM

*Proof:* Uniform approximation requires that

$$\forall \epsilon > 0 \exists |f_{ns}(\mathbf{x}) - g(\mathbf{x})| < \epsilon \forall \mathbf{x} \in U. \quad (\text{A.1})$$

$f_{ns}(\mathbf{x})$  is given by (15) and represents the crisp output produced by the modified height method defuzzifier of the nonsingleton system. Since  $U$  is a compact subset of  $R^p$ , continuity of  $g$  implies uniform continuity [30]. Thus

$$\forall \epsilon_1 > 0 \exists \delta > 0 \exists |g(\mathbf{x}_r) - g(\mathbf{x}_s)| < \epsilon_1 \forall \mathbf{x}_r, \mathbf{x}_s \in U \ni |\mathbf{x}_r - \mathbf{x}_s| < \delta. \quad (\text{A.2})$$

Construct a “grid” of points in  $U$  such that

- 1) given any  $\mathbf{x} \in U$ , there is a point  $\mathbf{c}_l$  in the grid so that  $|\mathbf{x} - \mathbf{c}_l| < \delta$ ;
- 2) the family of open balls  $\{B(\mathbf{c}_l, \delta)\}$  (each open ball is centered at  $\mathbf{c}_l$  and has radius  $\delta$ ) forms a finite open subcover of  $U$ , i.e.,  $U \subset \bigcup_{l=1}^M B(\mathbf{c}_l, \delta)$ , so that assuming  $U$  is connected and  $B(\mathbf{c}_l, \delta)$  and  $B(\mathbf{c}_k, \delta)$  are nearest neighbors

$$I_{lk} \equiv B(\mathbf{c}_l, \delta) \cap B(\mathbf{c}_k, \delta) \neq \emptyset \quad \text{for } l, k \leq M. \quad (\text{A.3})$$

$\delta$  is obtained as outlined in Definitions 1–3. For  $f_{ns}(\mathbf{x})$  in (15),  $\mathbf{c}_l$  corresponds to the mean of the  $p$ -dimensional antecedent membership function and  $\delta$  corresponds to an upper bound of the spread of each of its components. The image of the center of each open ball under  $g$  is given by  $g(\mathbf{c}_l)$ . Since  $g$  is a continuous function, the image of  $U$  under  $g$ ,  $g(U)$  in  $V$  will also be compact and connected and

$$g(U) \subset \bigcup_{l=1}^M g(B(\mathbf{c}_l, \delta)). \quad (\text{A.4})$$

Select fuzzy sets  $G^l \in V$  centered at  $\bar{\mathbf{y}}^l = g(\mathbf{c}_l)$  such that the set of points of support of  $G^l$ ,  $S_{G^l}$ , covers  $g(B(\mathbf{c}_l, \delta) \cap U)$ ,

and  $\mu_{G^l}(\bar{y}^l) = 1$ . Hence,  $\cup_{l=1}^M S_{G^l}$  covers  $g(U)$ . Let  $\mathbf{x}_1 \in B(\mathbf{c}_l, \delta)$ ,  $\mathbf{x}_2 \in B(\mathbf{c}_k, \delta)$  and  $\mathbf{x}_3 \in I_{lk}$ ; then,  $|\mathbf{x}_1 - \mathbf{x}_3| < \delta$  and  $|\mathbf{x}_3 - \mathbf{x}_2| < \delta$ . From (A.2) and the triangular inequality

$$\begin{aligned} |g(\mathbf{x}_1) - g(\mathbf{x}_2)| &\leq |g(\mathbf{x}_1) - g(\mathbf{x}_3)| + |g(\mathbf{x}_3) - g(\mathbf{x}_2)| \\ &< \epsilon_1 + \epsilon_1 = \epsilon_2. \end{aligned} \quad (\text{A.5})$$

The distances of  $\mathbf{x}_1$  from  $\mathbf{c}_l$  and  $\mathbf{x}_2$  from  $\mathbf{c}_k$  are both bounded by  $\delta$ ; thus,  $\mathbf{x}_1$  can come arbitrarily close to  $\mathbf{c}_l$ , and  $\mathbf{x}_2$  can come arbitrarily close to  $\mathbf{c}_k$ . By letting  $\mathbf{x}_1$  and  $\mathbf{x}_2$  approach  $\mathbf{c}_l$  and  $\mathbf{c}_k$ , respectively, then (A.5) bounds the difference of their corresponding images  $|g(\mathbf{c}_l) - g(\mathbf{c}_k)| = |\bar{y}^l - \bar{y}^k| < \epsilon_2$ .

Any point  $\mathbf{x} \in U$  that is not a center of a fuzzy set belongs to some  $I_{lk}$ ; therefore, the distance of its image from the image of the center  $\mathbf{c}_l$  is given by (A.5):  $|g(\mathbf{c}_l) - g(\mathbf{x})| < \epsilon_2$ . Let this point  $\mathbf{x}$  be the point  $\mathbf{x}_{k,\text{sup}}^l$  at which the functions  $\mu_{Q_k^l}$  in  $T_{k=1}^P \mu_{Q_k^l}(\mathbf{x}_{k,\text{sup}}^l)$  in (15) are maximized.

Now suppose that  $\mathbf{x}_{k,\text{sup}}^l$  activates  $N \leq M$  rules. Then, we have  $N$  nonzero output fuzzy sets (and, thus,  $N$  individual centers) which are combined according to the modified height method defuzzifier to produce the final system output. The system output lies somewhere inbetween or exactly on the centers of the output fuzzy sets  $Y^l$  and, consequently, on the centers of the sets  $G^l$  by construction. This is ensured by the fact that for any  $\mathbf{x} \in U$  the sum of the coefficients (16) of each  $\bar{y}^l$  in (15) equals one, i.e.,  $\sum_{l=1}^M p_l(\mathbf{x}) = 1$ . [Note that from the properties of the most commonly used  $t$ -norms (algebraic product, Einstein product, bounded difference, and Hamacher product) and (13) we know that the output fuzzy set  $Y^l$  and the consequent fuzzy set of the  $l$ th rule  $G^l$  achieve their maximum at the same point  $\bar{y}^l$ . In the case of *minimum* as a  $t$ -norm,  $G^l$  is required to be symmetric.] We have already shown that the distance between centers  $G^l$  and  $G^k$  is bounded by  $\epsilon_2$ ; thus, the distance of the output  $f_{ns}(\mathbf{x})$  produced by the nonsingleton fuzzy logic system (15) from the centroid of  $G^l$  is also bounded by  $\epsilon_2$ , i.e.,  $|f_{ns}(\mathbf{x}) - g(\mathbf{c}_l)| < \epsilon_2$ . Then, by the triangular inequality

$$\begin{aligned} |f_{ns}(\mathbf{x}) - g(\mathbf{x})| &\leq |f_{ns}(\mathbf{x}) - g(\mathbf{c}_l)| + |g(\mathbf{c}_l) - g(\mathbf{x})| \\ &< \epsilon_2 + \epsilon_2 = \epsilon. \end{aligned} \quad (\text{A.6})$$

Thus, the nonsingleton fuzzy logic system  $f_{ns}(\mathbf{x})$  in (15) uniformly approximates  $g(\mathbf{x})$ .

It should be noted that as the number of pairs  $(\mathbf{c}_l, \bar{y}^l)$  becomes arbitrarily large (i.e.,  $M \rightarrow \infty$ ), then neighboring antecedent centers  $\mathbf{c}_{l-1}$  and  $\mathbf{c}_l$  get arbitrarily close ( $\delta \rightarrow 0$ ) and so do consequent fuzzy set centers  $\bar{y}^{l-1}$  and  $\bar{y}^l$ , implying that  $\epsilon \rightarrow 0$ .  $\square$

The above proof includes the singleton FLS's of [38] as a special case. It can also be easily extended to nonsingleton additive FLS's [13].

#### ACKNOWLEDGMENT

The authors would like to thank the anonymous referee who suggested the possible interpretation of a NSFLS as a SFLS in the special case of Gaussian membership functions and product inference, and the referees who suggested improvements to the statement and proof of the universal approximation theorem.

#### REFERENCES

- [1] M. Balazinski, E. Czogala, and T. Sadowski, "Control of metal-cutting process using neural fuzzy controller," in *2nd IEEE Int. Conf. Fuzzy Syst.*, San Francisco, CA, Mar. 1993, vol. 1, pp. 161–166.
- [2] R. Brown, P. Bryant, and H. D. I. Abarbanel, "Computing the Lyapunov spectrum of a dynamical system from an observed time series," *Phys. Rev. A*, vol. 43, no. 6, pp. 2787–2806, Mar. 1991.
- [3] J. J. Buckley, "Sugeno-type controllers are universal controllers," *Fuzzy Sets Syst.*, vol. 53, no. 3, pp. 299–303, 1993.
- [4] M. Casdagli, "A dynamical systems approach to modeling input-output systems," *Nonlinear Modeling and Forecasting, SFI Studies in the Sciences of Complexity Process—Vol. XII*. New York: Addison-Wesley, 1992, pp. 265–281.
- [5] J. P. Eckmann, S. O. Kamphorst, D. Ruelle, and S. Ciliberto, "Lyapunov exponents from time series," *Phys. Rev. A*, vol. 34, no. 6, pp. 4971–4979, Dec. 1986.
- [6] J. D. Farmer, "Chaotic attractors of an infinite-dimensional dynamical system," *Physica*, vol. 4-D, pp. 366–393, 1982.
- [7] J. D. Farmer and J. J. Sidorowich, "Predicting chaotic time series," *Physical Rev. Lett.*, vol. 59, no. 8, pp. 845–848, Aug. 1987.
- [8] D. P. Filev and R. R. Yager, "A generalized defuzzification method via BAD distributions," *Int. J. Intell. Syst.*, vol. 6, pp. 687–697, 1991.
- [9] M. Grabisch, H. T. Nguyen, and E. A. Walker, *Fundamentals of Uncertainty Calculi with Applications to Fuzzy Inference*. Dordrecht, The Netherlands: Kluwer, 1995.
- [10] Y. Hayashi, J. J. Buckley, and E. Czogala, "Fuzzy neural network with fuzzy signals and weights," *Int. J. Intell. Syst.*, vol. 8, no. 4, pp. 527–537, 1993.
- [11] H. Hellendoorn and C. Thomas, "Defuzzification in fuzzy controllers," *J. Intell. Syst.*, vol. 1, no. 2, pp. 109–123, 1993.
- [12] B. Kosko, *Neural Networks and Fuzzy Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1992.
- [13] ———, "Fuzzy systems as universal approximators," in *IEEE Int. Conf. Fuzzy Syst.*, San Diego, CA, Mar. 1992, pp. 1153–1166; also in *IEEE Trans. Comput.*, vol. 43, pp. 1329–1333, Nov. 1994.
- [14] V. Y. Kreinovich, "Arbitrary nonlinearity is sufficient to represent all functions by neural networks: A theorem," *Neural Networks*, vol. 4, no. 3, pp. 381–383, 1991.
- [15] V. Kůrková, "Kolmogorov's theorem is relevant," *Neural Computat.*, vol. 3, pp. 617–622, 1991.
- [16] V. Kůrková, "Kolmogorov's theorem and multilayer neural networks," *Neural Networks*, vol. 5, no. 3, pp. 501–506, 1992.
- [17] A. Lapedes and R. Farber, "Nonlinear signal processing using neural networks: prediction and system modeling," Los Alamos, NM, Rep. LA-UR-87-2662, 1987.
- [18] C. C. Lee, "Fuzzy logic in control systems: fuzzy logic controller—Part II," *IEEE Trans. Syst., Man, Cybern.*, vol. 20, pp. 419–435, 1990.
- [19] G. G. Lorentz, *Approximation of Functions*. New York: Holt, Reinhart, Winston, 1966.
- [20] M. C. Mackey and L. Glass, "Oscillation and chaos in physiological control systems," *Sci.*, vol. 197, pp. 287–289, 1977.
- [21] J. Matthews, R. Broadwater, and L. Long, "The application of interval mathematics to utility economic analysis," *IEEE Trans. Power Syst.*, vol. 5, pp. 177–181, Feb. 1990.
- [22] J. M. Mendel, "Fuzzy logic systems for engineering: A tutorial," *Proc. IEEE*, vol. 83, pp. 345–377, Mar. 1995.
- [23] R. E. Moore, *Interval Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1966.
- [24] G. C. Mouzouris and J. M. Mendel, "A singular-value-qr decomposition based method for training fuzzy logic systems in uncertain environments," *J. Intell. Fuzzy Syst.*, to be published.
- [25] Y. Muyaram, T. Terano, S. Masui, and N. Akiyama, "Optimizing control of a diesel engine," in *Industrial Applications of Fuzzy Control*, M. Sugeno, Ed. Amsterdam, The Netherlands: North-Holland, 1992, pp. 63–71.
- [26] D. W. Nychka, S. Ellner, D. McCaffrey, and A. R. Gallant, "Finding chaos in noisy systems," *J. Royal Statistical Soc.*, ser. B54, pp. 399–426, 1992.
- [27] T. S. Parker and L. O. Chua, *Practical Numerical Algorithms for Chaotic Systems*. New York: Springer-Verlag, 1989.
- [28] W. Pedrycz, "Design of fuzzy control algorithms with the aid of fuzzy models," in *Industrial Applications of Fuzzy Control*, M. Sugeno, Ed. Amsterdam, The Netherlands: North-Holland, 1992, pp. 153–173.
- [29] S. N. Rasband, *Chaotic Dynamics of Nonlinear Systems*. New York: Wiley, 1990.
- [30] W. Rudin, *Principles of Mathematical Analysis, Third Edition*. New York: McGraw-Hill, 1976.

- [31] Y. Sakai and K. Ohkusa, "A fuzzy controller in turning process automation," *Industrial Applications of Fuzzy Control*, M. Sugeno, Ed. Amsterdam, The Netherlands: North-Holland, 1992, pp. 139–151.
- [32] D. A. Sprecher, "On the structure of continuous functions of several variables," *Trans. Amer. Mathematical Soc.*, vol. 115, pp. 340–355, Mar. 1965.
- [33] H. Tanaka, T. Okuda, and K. Asai, "Fuzzy information and decision in statistical model," in *Advances in Fuzzy Set Theory and Applications*, M. M. Gupta, R. K. Ragade, and R. R. Yager, Eds. Amsterdam, The Netherlands: North-Holland, 1979, pp. 303–320.
- [34] V. R. Vemuri and R. D. Roberts, Eds., *Artificial Neural Networks: Forecasting Time Series*. Los Angeles, CA: IEEE Comput. Soc. Press, 1994.
- [35] L. X. Wang, "Analysis and design of fuzzy systems," Ph.D. dissertation, Univ. Southern California, Los Angeles, CA, 1992.
- [36] L. X. Wang, *Adaptive Fuzzy Systems and Control*. Englewood-Cliffs, NJ: Prentice-Hall, 1994.
- [37] L. X. Wang and J. M. Mendel, "Generating fuzzy rules from numerical data, with applications," USC-SIPI Rep. 169, 1991; also in *IEEE Trans. Syst., Man, Cybern.*, vol. 22, pp. 1414–1427, Nov./Dec. 1992.
- [38] ———, "Fuzzy basis functions, universal approximation, and orthogonal least squares learning," *IEEE Trans. Neural Networks*, vol. 3, pp. 807–814, Sept. 1992.
- [39] ———, "Back-propagation fuzzy systems as non-linear dynamic system identifiers," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, San Diego, CA, Mar. 1992, pp. 1409–1418.
- [40] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, "Determining Lyapunov exponents from a time series," *Physica*, vol. 16-D, pp. 285–317, 1985.
- [41] L. A. Zadeh, "Fuzzy sets," *Inform. Contr.*, vol. 8, pp. 338–353, 1965.
- [42] ———, "Outline of a new approach to the analysis of complex systems and decision processes," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, pp. 28–44, Jan. 1973.
- [43] ———, "Fuzzy sets and information granularity," in *Advances in Fuzzy Set Theory and Applications*, M. M. Gupta, R. K. Ragade, and R. R. Yager, Eds. Amsterdam, The Netherlands: North-Holland, 1979, pp. 3–18.
- [44] H. J. Zimmermann, *Fuzzy Set Theory And Its Applications*, 2nd ed. Boston, MA: Kluwer, 1991.



**George C. Mouzouris** was born in Cyprus in 1966. He received the B.S. (honors) and M.S. degrees from Brown University, Providence, RI, both in 1990, and the Ph.D. degree from the University of Southern California, Los Angeles, in 1996.

From 1990 to 1992, he was with the Digital Signal Processing Group of Texas Instruments, Houston, TX, where he dealt with the design and implementation of signal and image processing algorithms on specialized processors. He is currently with the Signal and Image Processing Institute, University of Southern California, Los Angeles. His research interests lie in the areas of nonlinear dynamic modeling, fuzzy systems, radial basis functions, neural networks, and financial modeling.

Dr. Mouzouris is a member of Tau Beta Pi and Sigma Xi.



**Jerry M. Mendel** (S'59–M'61–SM'72–F'78) received the Ph.D. degree in electrical engineering from the Polytechnic Institute, Brooklyn, NY, in 1963.

Currently, he is a Professor of electrical engineering, Director of Special Education Projects for the School of Engineering, and Associate Director for Education of the Integrated Media Systems Center at the University of Southern California, Los Angeles, where he has been since 1974. He has published over 325 technical papers and is the author and/or editor of seven books. His present research interests include higher order statistics and neural networks applied to array processing and prediction of nonlinear time-series, and fuzzy logic applied to prediction of nonlinear time-series, classification problems, and social science problems.

Dr. Mendel is a Distinguished Member of the IEEE Control Systems Society. He was President of the IEEE Control Systems Society in 1986. Among his awards are the 1983 Best Transactions Paper Award of the IEEE Geoscience and Remote Sensing Society, the 1992 Signal Processing Society Paper award, and a 1984 IEEE Centennial Medal.