Pergamon

Automatica, Vol. 30, No. 3, pp. 521–525, 1994

0005-1098/94 $6.00 + 0.00

Brief Paper

Sufficient Conditions on General Fuzzy Systems as Function Approximators*

HAO YING†

Key Words—Approximation theory; fuzzy control; fuzzy systems; modeling; nonlinear systems; polynomials.

Abstract—In a constructive way, we have found sufficient conditions under which general fuzzy systems can uniformly approximate any real continuous function on a compact domain to any degree of accuracy. More importantly, we have revealed the underlying mechanism of such function approximation: the fuzzy systems can uniformly approximate any polynomials which, according to the Weierstrass approximation theorem, can uniformly approximate any continuous function on a compact domain. These findings lead to the following practical results: (1) explicit fuzzy rules of fuzzy systems can now be easily obtained according to functions to be approximated; and (2) formulas are derived which can calculate the number of input fuzzy sets, output fuzzy sets and fuzzy rules needed in order to satisfy any given approximation accuracy. The number increases as required approximation error decreases, and as approximation error approaches zero, the number approaches ∞. These formulas also reveal that the number is minimal when the maximum number of intersection between adjacent input fuzzy sets is one. Practical implications of these results will be discussed, especially in the context of fuzzy control and fuzzy modeling. Two examples are provided to demonstrate how to design fuzzy systems to approximate given functions with required approximation accuracy.

1. *Introduction*

FUZZY SYSTEMS, especially fuzzy controllers and fuzzy modellers, have been used to solve practical problems. From mathematics point of view, fuzzy systems are functions mapping inputs to outputs. Buckley (1992), Kosko (1992), Wang (1992), and Wang and Mendel (1992) proved that some particular configured fuzzy systems were 'universal approximators' (we call them function approximators instead in this paper) in the sense that these fuzzy systems could approximate any real continuous function to any degree of accuracy on a compact domain (closed and bounded in a finite dimensional space). Buckley used fuzzy expert system based fuzzy controllers with a linear defuzzifier. Kosko employed additive fuzzy systems while Wang and Mendel used fuzzy systems with the product inference, the centroid defuzzification and Gaussian membership functions. All these theorems are existence theorems proven by using the Stone-Weierstrass theorem (Rudin, 1976). Because of using the Stone–Weierstrass theorem in the proof, these theorems cannot provide answers to practical questions concerning why and how these particular fuzzy systems can uniformly approximate any continuous function with arbitrary accuracy.

In practice, a typical fuzzy control problem can be stated as follows: given a process, how can a fuzzy controller be designed to control the process with desired performance? A typical fuzzy modeling problem is this: given measured input and output data, how should a fuzzy system be constructed so that it can represent the true system with reasonable accuracy? These kinds of questions can be summarily represented by the following four fundamental questions which are crucial both theoretically and practically: (1) Are fuzzy systems with membership functions, fuzzy logic, fuzzy inference methods and defuzzification methods other than those used in the above-mentioned existence theorems function approximators? In other words, are general fuzzy systems function approximators? (2) If the answer to question (1) is yes, then, what are the (necessary or sufficient) conditions for general fuzzy systems as function approximators? (3) What are underlying mechanisms which enable general fuzzy systems to be function approximators? (4) Given a continuous function, how can a fuzzy system be designed to approximate the function with a required approximation accuracy? More specifically, how should input fuzzy sets, output fuzzy sets, fuzzy logic, fuzzy inference and fuzzy rules be selected?

The object of this research is to answer these questions. To be able to do so, we use a constructive proof approach. We will first define components of general fuzzy systems in the following section. In Section 3, we will constructively prove some sufficient conditions under which general fuzzy systems can uniformly approximate any continuous function on a compact domain. The Weierstrass approximation theorem will be utilized in the proof. Formulas will also be derived, which can calculate the number of input fuzzy sets, output fuzzy sets and fuzzy rules needed in order to satisfy required approximation accuracy. In Section 4, we will illustratively design two fuzzy systems for two given functions with required approximation accuracy. Conclusions are made in the last section.

2. *Components of general fuzzy systems*

The fuzzy systems studied in this paper are the general ones comprising of the following components.

An input vector with r scaled inputs of the fuzzy systems:

$$\underline{x(t)} := (x_1(t), x_2(t), \ldots, x_r(t)) = (a_1 z_1(t), a_2 z_2(t), \ldots, a_r z_r(t)),$$

where $z_i(t)$ and $x_i(t)$ $(i = 1, 2, \ldots, r)$ are unscaled and scaled continuous-time or discrete-time state variables, respectively. For simplicity, we denote $\underline{x(t)}$, $z_i(t)$ and $x_i(t)$ as $\underline{x}$, z_i and x_i, respectively, in the rest of the paper. a_i's are input scalars making

$$-1 \le x_i \le 1.$$

Without loss of generality, assume there are

$$N := 2n + 1, \quad (n \ge 1)$$

* Received 12 February 1993; received in final form 14 May 1993. This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Editor A. P. Sage. Corresponding author Hao Ying.

† Department of Physiology and Biophysics; Biomedical Engineering Center; and Office of Academic Computing, University of Texas Medical Branch, Galveston, TX 77555, U.S.A.

input fuzzy sets for each input, X_i. Each fuzzy set is denoted by $A_{i,j}$ ($j = 0, \pm 1, \ldots, \pm n$). The positive, negative and zero subscripts j correspond to linguistic description of 'positive,' 'negative' and 'zero' (e.g. $A_{i,1}$ stands for x_i is 'positive small' and $A_{i,-4}$ means x_i is 'negative large'). $A_{i,j}$ has a continuous convex membership function $\mu_{A_{i,j}}$. It is supposed that membership functions of $A_{i,j}$ are the same for a specific i. However, membership functions may be different between $A_{a,j}$ and $A_{b,j}$, where $a \neq b$. The membership functions are defined over $[-1, 1]$ that is equally partitioned into $2n$ intervals, each of which is $[j/n, (j+1)/n]$. The value of $\mu_{A_{i,j}}$ is assumed to be one at j/n, and is nonincreasing for $x_i > j/n$ and $x_i < j/n$. We assume that, in $[j/n, (j+1)/n]$, the maximum number of intersection between adjacent $\mu_{A_{i,j}}$ is $c_i^{\max}$ ($c_i^{\max} > 0$). For typical fuzzy systems in the literature, $c_i^{\max}$ is one. That is, $\mu_{A_{i,j}}$ and $\mu_{A_{i,j+1}}$ intersect once in $[j/n, (j+1)/n]$.

N^r fuzzy rules are needed to cover all the possible combinations of $A_{i,j}$. The fuzzy rules in this study are described as

$$\text{IF } x_1 \text{ is } A_{1,p_1} \text{ AND } x_2 \text{ is } A_{2,p_2} \text{ AND} \cdots \text{AND } x_r \text{ is } A_{r,p_r} \text{ THEN } U(t) \text{ is } U_m,$$

where $u(t)$ is the crisp output of the fuzzy systems, which is constrained by

$$-H \leq u(t) \leq H, \quad H > 0.$$

U_m is an output fuzzy set. The positive, negative and zero subscript of U_m represents linguistic description of 'positive,' 'negative' and 'zero,' and is determined by an arbitrary real continuous function f whose output is integer over the integer inputs, p_i's:

$$m := f(\underline{p}), \quad \underline{p} := (p_1, p_2, \ldots, p_r).$$

m is subject to

$$-M(f, n) \leq m \leq M(f, n),$$

$$M(f, n) := \max \{|f(\underline{p})|, -n \leq p_i \leq n\}.$$

$M(f, n)$, determined by both f and n, is the 'maximum' positive output fuzzy set. Obviously, there are $2M(f, n) + 1$ fuzzy sets defined over $[-H, H]$ for $u(t)$. $[-H, H]$ is equally partitioned into $2M(f, n)$ intervals, each of which is $[m \cdot H/M(f, n), (m+1)H/M(f, n)]$. Membership function of U_m ($m = 0, \pm 1, \ldots, \pm 2m(f, n)$) is singleton. That is, membership is one at $m \cdot H/M(f, n)$ and zero anywhere else.

Fuzzy logic used to evaluate ANDs in each fuzzy rule can be any T-norm (e.g. Zadeh AND) and fuzzy logic employed to evaluate ORs between the fuzzy rules with the same output fuzzy sets can be any T-conorm (e.g. Zadeh OR) (Gupta and Qi, 1991a, b). Since the singleton membership functions are employed in the output fuzzy sets, any sensible inference method can infer U_m from $A_{1,p_1}, A_{2,p_2}, \ldots, A_{r,p_r}$. The most common inference methods include Mamdani's minimum inference method and Larsen's product inference method (Mizumoto, 1988). For the fuzzy systems in this paper, different fuzzy logic and inference methods may be used in different fuzzy rules.

The generalized defuzzification algorithm (Filev and Yager, 1991) is used to calculate $u(t)$ which actually is a mapping $F_n: C^r[-1, 1] \rightarrow [-H, H]$:*

$$F_n(\underline{x}) := u(t) = \frac{H}{M(f, n)} \frac{\sum_{h=1}^{P(t)} (\mu_h)^\alpha \cdot m_h}{\sum_{h=1}^{P(t)} (\mu_h)^\alpha}$$

$$= \frac{H}{M(f, n)} \frac{\sum_{h=1}^{P(t)} (\mu_h)^\alpha \cdot f(\underline{p} + \underline{c}_h)}{\sum_{h=1}^{P(t)} (\mu_h)^\alpha}, \tag{1}$$

* $C^r[-1, 1]$ represents r dimensional product of $[-1, 1]$.

where μ_h is the membership for U_{m_h} after fuzzy logic AND and OR, and fuzzy inference are executed. Also,

$$\underline{c}_h := (c_{h,1}, c_{h,2}, \ldots, c_{h,r}), \quad 0 \leq c_{h,i} \leq c_i^{\max}$$

and

$$-n \leq p_i + c_{h,i} \leq n.$$

$f(\underline{p} + \underline{c}_h)$ represents different fuzzy rules executed at time t. μ_h are resulting memberships on U_{m_h}. $P(t)$ represents number of (different) output fuzzy sets involved in the defuzzification at time t. Different defuzzification results can be obtained by using different α, where $0 \leq \alpha < +\infty$. Filev and Yager proved that the centroid defuzzification method was a special case of this generalized defuzzification method when $\alpha = 1$, and the mean of maximum defuzzification method was also a special case when $\alpha = 0$. It should be pointed out that the centroid and mean defuzzification methods are the most popular methods in practice (Lee, 1990).

It is important to realize that $F_n(\underline{x})$ is a function sequence with respect to n. In the rest of the paper, $F_n(\underline{x})$ will be used to represent the general fuzzy systems.

3. *Sufficient conditions for the general fuzzy system as function approximators*

In this section, we will prove some sufficient conditions under which the general fuzzy systems can uniformly approximate any continuous function on a compact domain. In the meantime, we will also expose how the fuzzy systems approximate functions.

Without losing generality, from now on we always assume that at time t,

$$\frac{p_i}{n} \leq x_i \leq \frac{p_i + 1}{n}, \quad -n \leq p_i \leq n - 1. \tag{2}$$

Lemma 3.1.

$$\lim_{n\to\infty} \frac{p_i}{n} = \lim_{n\to\infty} \frac{p_i + c_i^{\max}}{n} = x_i.$$

Proof.

$$\frac{p_i}{n} \leq x_i \leq \frac{p_i + 1}{n} \Rightarrow \frac{p_i}{n} \leq x_i \leq \frac{p_i + c_i^{\max}}{n}.$$

Note $c_i^{\max}$ are finite positive integers. Therefore,

$$\lim_{n\to\infty} \frac{p_i}{n} \leq x_i \leq \lim_{n\to\infty} \frac{p_i + c_i^{\max}}{n}$$

$$\Rightarrow \lim_{n\to\infty} \frac{p_i}{n} = \lim_{n\to\infty} \frac{p_i + c_i^{\max}}{n} = x_i. \quad \blacksquare$$

Theorem 3.2. $F_n(\underline{x})$ can uniformly approximate any polynomial $P_d(\underline{x})$ defined in $C^r[-1, 1]$:

$$P_d(\underline{x}) = \sum_{d_i \geq 0} \left(\beta_{d_1,\ldots,d_r} \prod_{i=1}^{r} x_i^{d_i}\right), \quad \sum_{i=1}^{r} d_i \leq d,$$

where d is the order of the polynomial. That is, $\forall \varepsilon > 0$, there exists a sufficiently large positive integer n^* such that $\forall n > n^*$,

$$\|F_n - P_d\|_{C^r[-1,1]} = \max_{\underline{x} \in C^r[-1,1]} |F_n(\underline{x}) - P_d(\underline{x})| < \varepsilon.$$

Proof. Construct $f(\underline{p})$ as a dth order polynomial (the same order of polynomial as $P_d(\underline{x})$) with respect to p_i:

$$f(\underline{p}) = \sum_{d_i \geq 0} \left(L_{d_1,\ldots,d_r} \cdot n^d \cdot \prod_{i=1}^{r} \left(\frac{p_i}{n}\right)^{d_i}\right), \tag{3}$$

where $L_{d_1,\ldots,d_r}$ are integers calculated from $\beta_{d_1,\ldots,d_r}$ of $P_d(\underline{x})$:

$$L_{d_1,\ldots,d_r} = 10^s \times \beta_{d_1,\ldots,d_r}.$$

s is the smallest positive integer which makes all $10^s \times \beta_{d_1,\ldots,d_r}$ integers. For instance, if $P_d(\underline{x}) = 1.2 + 0.23x_1 + 1.542x_2 + 0.07823x_1x_2$, then $\beta_{0,0} = 1.2$, $\beta_{1,0} = 0.23$, $\beta_{0,1} = 1.542$, $\beta_{1,1} = 0.07823$. In this example, $s = 5$. Consequently, $L_{0,0} = 120{,}000$, $L_{1,0} = 23{,}000$, $L_{0,1} = 154{,}200$ and $L_{1,1} = 7{,}823$. Choosing such $L_{d_1,\ldots,d_r}$ is necessary because the value of $f(\underline{p})$ has to be integer with respect to integer inputs p_i. It is obvious that

$$M(f, n) = n^d \cdot \sum_{d_i \ge 0} |L_{d_1,\ldots,d_r}|.$$

If we choose

$$H = 10^{-s} \cdot \sum_{d_i \ge 0} |L_{d_1,\ldots,d_r}|, \tag{4}$$

then,

$$\frac{H}{M(f, n)} f(\underline{p}) = \sum_{d_i \ge 0} \left(\beta_{d_1,\ldots,d_r} \prod_{i=1}^{r} \left(\frac{p_i}{n} \right)^{d_i} \right) = P_d\left(\frac{\underline{p}}{n} \right), \tag{5}$$

where

$$\frac{\underline{p}}{n} := \left(\frac{p_1}{n}, \frac{p_2}{n}, \ldots, \frac{p_r}{n} \right).$$

So,

$$F_n(\underline{x}) = \frac{\sum_{h=1}^{P(t)} \left((\mu_h)^\alpha P_d\left(\frac{\underline{p} + \underline{c}_h}{n} \right) \right)}{\sum_{h=1}^{P(t)} (\mu_h)^\alpha} \tag{6}$$

and

$$\|F_n - P_d\| = \max_{\underline{x} \in C^r[-1,1]} \left| \frac{\sum_{h=1}^{P(t)} \left((\mu_h)^\alpha P_d\left(\frac{\underline{p} + \underline{c}_h}{n} \right) \right)}{\sum_{h=1}^{P(t)} (\mu_h)^\alpha} - P_d(\underline{x}) \right|.$$

According to Lemma 3.1, as $n \to \infty$,

$$P_d\left(\frac{\underline{p} + \underline{c}_h}{n} \right) \to P_d(\underline{x}) \Rightarrow F_n \to P_d.$$

In fact, $F_n(\underline{x})$ approximates $P_d(\underline{x})$ uniformly. For simplicity, we will only prove $r = 1$ case ($r > 1$ cases can be treated similary). When $r = 1$,

$$\begin{aligned}
\|F_n - P_d\| &= \max_{x_1 \in [-1,1]} \left| \frac{\sum_{h=1}^{P(t)} \left((\mu_h)^\alpha \cdot P_d\left(\frac{p_1 + c_{h,1}}{n} \right) \right)}{\sum_{h=1}^{P(t)} (\mu_h)^\alpha} - P_d(x_1) \right| \\
&\le \max_{x_1 \in [-1,1]} \left\{ \frac{\sum_{h=1}^{P(t)} \left((\mu_h)^\alpha \left| P_d\left(\frac{p_1 + c_{h,1}}{n} \right) - P_d(x_1) \right| \right)}{\sum_{h=1}^{P(t)} (\mu_h)^\alpha} \right\} \\
&\le \max_{x_1 \in [-1,1]} \left\{ \frac{\sum_{h=1}^{P(t)} \left((\mu_h)^\alpha \sum_{d_1=0}^{d} |\beta_{d_1}| \left| \left(\frac{p_1 + c_{h,1}}{n} \right)^{d_1} - x_1^{d_1} \right| \right)}{\sum_{h=1}^{P(t)} (\mu_h)^\alpha} \right\}.
\end{aligned} \tag{7}$$

According to (2),

$$\left| \frac{p_1 + c_{h,1}}{n} - x_1 \right| \le \frac{c_{h,1}}{n} \le \frac{c_1^{\max}}{n}.$$

Also, note that $|x_1| \le 1$ and $|p_1 + c_{h,1}|/n \le 1$. Therefore, for $1 \le d_1 \le d$,

$$\begin{aligned}
&\left| \left(\frac{p_1 + c_{h,1}}{n} \right)^{d_1} - x_1^{d_1} \right| \\
&\le \left| \frac{p_1 + c_{h,1}}{n} - x_1 \right| \left| \sum_{v=1}^{d_1} \left(\frac{p_1 + c_{h,1}}{n} \right)^{d_1 - v} x_1^{v-1} \right| \\
&\le \frac{c_1^{\max}}{n} \left(\sum_{v=1}^{d_1} \left| \frac{p_1 + c_{h,1}}{n} \right|^{d_1 - v} |x_1|^{v-1} \right) \\
&< \frac{c_1^{\max}}{n} \sum_{v=1}^{d_1} 1 < \frac{c_1^{\max} \cdot d_1}{n}.
\end{aligned} \tag{8}$$

Substituting (8) into (7), we get

$$\begin{aligned}
\|F_n - P_d\| &\le \frac{c_1^{\max}}{n} \sum_{d_1=0}^{d} (|\beta_{d_1}| \cdot d_1) \\
&= \frac{c_1^{\max}}{n} \sum_{d_1=1}^{d} (|\beta_{d_1}| \cdot d_1).
\end{aligned} \tag{9}$$

$\forall \varepsilon > 0$, if we choose

$$n^* > \frac{c_1^{\max}}{\varepsilon} \sum_{d_1=1}^{d} (|\beta_{d_1}| \cdot d_1),$$

then $\|F_n - P_d\| < \varepsilon$, $\forall n > n^*$, indicating $F_n(\underline{x})$ approximates $P_d(\underline{x})$ uniformly. ■

We call $P_d(\underline{p}/n)$ in (5) 'transformed fuzzy rules,' which can be directly derived from $P_d(\underline{x})$ by replacing $\underline{x}$ with $\underline{p}/n$. It should be noted that the value of $P_d(\underline{p}/n)$ at $\underline{p}$ may not be integer. Using the transformed fuzzy rules, $F_n(\underline{x})$ in (1) can be neatly expressed as $F_n(\underline{x})$ in (6), and calculation of $F_n(\underline{x})$ in (6) is also simpler.

We are now ready to prove that the general fuzzy systems are function approximators based on the Weierstrass approximation theorem (Bronshtein and Semendyayev, 1985) and its extended theorems for multiple input variables (e.g. Picard, 1891). The Weierstrass approximation theorem states: to any function $G(x)$ that is continuous in $[a, b]$ and to any real number $\varepsilon > 0$, there corresponds a polynomial $P_d(x)$ such that $\|P_d - G\|_{C[a,b]} < \varepsilon$. It sould be noted that the smaller the ε, the larger the d.

Theorem 3.3 (*general fuzzy systems approximation theorem*). The general fuzzy systems, $F_n(\underline{x})$, can uniformly approximate any function $G(\underline{x})$ which is continuous in $C^r[-1, 1]$ to any degree of accuracy. In other words, $\forall \varepsilon > 0$, there exists a sufficiently large positive integer n^* such that $\forall n > n^*$,

$$\|F_n - G\| = \max_{\underline{x} \in C^r[-1,1]} |F_n(\underline{x}) - G(\underline{x})| < \varepsilon.$$

Proof. Acccoding to the Weierstrass approximation theorem, for any given function $G(\underline{x})$ which is continuous in $C^r[-1, 1]$, there always exists a polynomial $P_d(\underline{x})$ which can uniformly approximate $G(\underline{x})$ to arbitrary accuracy. That is, $\forall \varepsilon_1 > 0$, $\|P_d - G\| < \varepsilon_1$. Further, according to Theorem 3.2, $\forall \varepsilon_2 > 0$, we can always find a sufficiently large positive integer n^* such that $\forall n > n^*$, $\|F_n - P_d\| < \varepsilon_2$. Therefore,

$$\|F_n - G\| \le \|F_n - P_d\| + \|P_d - G\| < \varepsilon, \quad \varepsilon = \varepsilon_1 + \varepsilon_2,$$

meaning $F_n(\underline{x})$ can uniformly approximate $G(\underline{x})$. ■

Based on above two theorems, one can see that the general fuzzy systems can uniformly approximate any continuous function because they have the ability to uniformly approximate any polynomial. Combining the theorems, we get the following theorem concerning sufficient conditions for the general fuzzy systems as function approximators.

Theorem 3.4 (*sufficient conditions*). Assume $\|P_d - G\| < \varepsilon_1$, $\|F_n - P_d\| < \varepsilon_2$ and $\varepsilon = \varepsilon_1 + \varepsilon_2 (\forall \varepsilon > 0)$. $\|F_n - G\| < \varepsilon$, if (i) $f(\underline{p})$ and H are chosen according to (3) and (4), respectively, to form the transformed fuzzy rules $P_d(\underline{p}/n)$ in (5); and (ii) n is such chosen that $n > n^*$ where

(a) $$n^* = \frac{c_1^{\max}}{\varepsilon_2} \sum_{d_1=1}^{d} (|\beta_{d_1}| \cdot d_1), \quad \text{when } r = 1; \tag{10}$$

(b) $$n^* \approx \frac{1}{\varepsilon_2} \sum_{d_i \ge 1} \left(|\beta_{d_1,\ldots,d_r}| \sum_{i=1}^{r} (d_i \cdot c_i^{\max}) \right), \quad \text{when } r > 1. \tag{11}$$

For $r > 1$ cases, the calculated n^* should be checked against the inequality

$$\sum_{d_i \ge 0} |\beta_{d_1,\ldots,d_r}| \left[\prod_{i=1}^{r} \left(1 + \frac{c_i^{\max}}{n^*} \right)^{d_i} - 1 \right] < \varepsilon,$$

to see if the inequality holds. If not, n^* needs a slight increase to make the inequality hold.

Proof. (1) This condition is evident in the context of the proof of Theorem 3.2.

(2) When $G(x)$ is a one variable function ($r=1$). As a by-product of proving Theorem 3.2 (see (9)), we immediately get

$$n^* = \frac{c_1^{\max}}{\varepsilon_2} \sum_{d_1=1}^{d} (|\beta_{d_1}| \cdot d_1).$$

For multiple input variable cases ($r>1$), we want

$$\begin{aligned}
&\|F_n - P_d\| \\
&= \max_{x \in C^r[-1,1]} \left| \frac{\sum_{h=1}^{P(t)} \left((\mu_h)^\alpha \cdot P_d\left(\frac{p + c_h}{n^*}\right)\right)}{\sum_{h=1}^{P(t)} (\mu_h)^\alpha} - P_d(x) \right| \\
&\le \max_{x \in C^r[-1,1]} \\
&\quad \left\{ \frac{\sum_{h=1}^{P(t)} (\mu_h)^\alpha \left(\sum_{d_i \ge 0} |\beta_{d_1,\ldots,d_r}| \left| \prod_{i=1}^{r} \left(\frac{p_i + c_{h,i}}{n^*}\right)^{d_i} - \prod_{i=1}^{r} x_i^{d_i} \right| \right)}{\sum_{h=1}^{P(t)} (\mu_h)^\alpha} \right\} \\
&< \varepsilon_2. \qquad (12)
\end{aligned}$$

Based on (2), $x_i = (p_i + \theta_i)/n$, where $0 \le \theta_i \le 1$. Hence,

$$\begin{aligned}
&\sum_{d_i \ge 0} |\beta_{d_1,\ldots,d_r}| \left| \prod_{i=1}^{r} \left(\frac{p_i + c_{h,i}}{n^*}\right)^{d_i} - \prod_{i=1}^{r} x_i^{d_i} \right| \\
&= \sum_{d_i \ge 0} |\beta_{d_1,\ldots,d_r}| \left| \prod_{i=1}^{r} \left(x_i + \frac{c_{h,i} - \theta_i}{n^*}\right)^{d_i} - \prod_{i=1}^{r} x_i^{d_i} \right| \qquad (13) \\
&\le \sum_{d_i \ge 0} |\beta_{d_1,\ldots,d_r}| \left[\prod_{i=1}^{r} \left(1 + \frac{|c_{h,i} - \theta_i|}{n^*}\right)^{d_i} - 1 \right] \\
&\approx \sum_{d_i \ge 1} |\beta_{d_1,\ldots,d_r}| \left[\prod_{i=1}^{r} \left(1 + d_i \frac{|c_{h,i} - \theta_i|}{n^*}\right) - 1 \right] \\
&\approx \sum_{d_i \ge 1} \left(|\beta_{d_1,\ldots,d_r}| \sum_{i=1}^{r} \frac{d_i\, |c_{h,i} - \theta_i|}{n^*} \right) \\
&\approx \sum_{d_i \ge 1} \left(|\beta_{d_1,\ldots,d_r}| \sum_{i=1}^{r} \frac{d_i \cdot c_i^{\max}}{n^*} \right). \qquad (14)
\end{aligned}$$

Substituting (14) into (12), we get

$$\begin{aligned}
&\varepsilon_2 \approx \sum_{d_i \ge 1} \left(|\beta_{d_1,\ldots,d_r}| \sum_{i=1}^{r} \frac{d_i \cdot c_i^{\max}}{n^*} \right) \Rightarrow \\
&n^* \approx \frac{1}{\varepsilon_2} \sum_{d_i \ge 1} \left(|\beta_{d_1,\ldots,d_r}| \sum_{i=1}^{r} (d_i \cdot c_i^{\max}) \right). \qquad (15)
\end{aligned}$$

Due to the minor approximation in the above derivation, n^* calculated based on (15) needs to be verified to make sure it is large enough. Note from the expression (13)

$$\begin{aligned}
&\sum_{d_i \ge 0} |\beta_{d_1,\ldots,d_r}| \left| \prod_{i=1}^{r} \left(x_i + \frac{c_{h,i} - \theta_i}{n^*}\right)^{d_i} - \prod_{i=1}^{r} x_i^{d_i} \right| \\
&\le \sum_{d_i \ge 0} |\beta_{d_1,\ldots,d_r}| \left[\prod_{i=1}^{r} \left(1 + \frac{c_i^{\max}}{n^*}\right)^{d_i} - 1 \right].
\end{aligned}$$

Hence, we can use

$$\sum_{d_i \ge 0} |\beta_{d_1,\ldots,d_r}| \left[\prod_{i=1}^{r} \left(1 + \frac{c_i^{\max}}{n^*}\right)^{d_i} - 1 \right] < \varepsilon_2 \qquad (16)$$

to check the adequacy of the calculated n^*. ■

If a calculated n^* is not an integer, we choose the smallest integer larger than the calculated n^* as n^*. If a n^* calculated according to (11) does not satisfy (16), we need to slightly increase it, usually by one or two, as an example below shows. It is interesting to note that if we let $r=1$ and replace the sign '$\approx$' by the sign '$=$' in (11), (11) coincides with (10). This coincidence implies that the approximation adopted in the above derivation is tight and that n^* calculated by (11) is close to what actual n^* should be.

In the spirit of deriving (11), we also derived another sufficient condition which can directly calculate n^* for $r>1$ cases without any verification:

$$n^* > \frac{1}{\varepsilon_2} \sum_{d_i \ge 1} \left(|\beta_{d_1,\ldots,d_r}| \prod_{i=1}^{r} (2^{d_i} \cdot c_i^{\max}) \right) \qquad (17)$$

However, as a numerical example in this paper will show, the calculated n^* is much more conservative (too large) than n^* computed by (11).

For easy implementation of fuzzy systems in practice, a small n^* is always desirable. In (10) and (11), $\beta_{d_1,\ldots,d_r}$ are determined by $G(x)$ and we have no control on them. Consequently, one way to obtain a smaller n^* is to use the smallest $c_i^{\max}$, which are one. This finding can be stated formally:

Proposition 3.5. To reduce the number of fuzzy rules needed in a fuzzy system, the maximum number of intersection between adjacent input fuzzy sets should be one.

According to (10) and (11), it is desirable to use a large ε_2 and a small d in order to yield a smaller n^*. Unfortunately, they cannot be achieved simultaneously. This is because, for a fixed ε, a larger ε_2 means a smaller ε_1 which in turn results in a larger d. When ε is very small, both ε_1 and ε_2 are very small while d is very large, making n^* extremely large. As an extreme, if $\varepsilon = 0$, $n^* = \infty$. Since $n^* = \infty$ is practically unachievable, we conclude that (see also Ying, 1993):

Proposition 3.6. The general fuzzy systems cannot always be exactly equivalent to a given continuous function.

It should be pointed out that $G(x)$ is unavailable in real-world problems. In a control problem, $G(x)$ represents a control solution, while in a modeling problem $G(x)$ represents the true model of the system to be modeled. In both situations, one's task is to realize $G(x)$ or approach it with enough approximation accuracy through manipulating different components of the fuzzy systems.

4. *Examples*

In the following examples, we always use $c_i^{\max} = 1$ for all i to minimize the number of fuzzy rules needed.

Example 1. Design a fuzzy system to uniformly approximate the continuous function $G(z) = \sin z/z$ defined on $[-3, 3]$ with $\varepsilon = 0.2$.

Let $x_1 = z/3$ (i.e. $a_1 = \frac{1}{3}$) and consequently $G(x_1) = \sin 3x_1/3x_1$ is defined on $[-1, 1]$. According to Taylor expansion,

$$\begin{aligned}
G(x_1) = \frac{\sin 3x_1}{3x_1} &= 1 - 1.5x_1^2 + 0.675x_1^4 \\
&\quad - 0.14464x_1^6 + 0.01808x_1^8 - \cdots.
\end{aligned}$$

If we use the first three terms as a polynomial to approximate $G(x_1)$, then the absolute value of the truncation error is less than

$$\varepsilon_1 = 0.14464,$$

and consequently

$$\varepsilon_2 = 0.2 - \varepsilon_1 = 0.05536.$$

Obviously, $d = 4$ and

$$P_4(x_1) = 1 - 1.5x_1^2 + 0.675x_1^4.$$

$\beta_0 = 1$, $\beta_2 = -1.5$, $\beta_4 = 0.675$, and $s = 3$. According to (10), the number of fuzzy rules needed is

$$\begin{aligned}
n^* &= \frac{c_1^{\max}}{\varepsilon_2} \sum_{d_1=1}^{d} (|\beta_{d_1}| \cdot d_1) \\
&= \frac{1}{0.05536} (1.5 \times 2 + 0.675 \times 4) \\
&= 102.96,
\end{aligned}$$

meaning n^* should be 103. As a result, we should use at least $N = 2n^* + 1 = 207$ fuzzy rules. The transformed fuzzy rules are described by

$$P_4\left(\frac{p_1}{n}\right) = 1 - 1.5\left(\frac{p_1}{103}\right)^2 + 0.675\left(\frac{p_1}{103}\right)^4.$$

If $\varepsilon = 0.02$ is desired, then $d = 6$, $\varepsilon_1 = 0.01808$ and $\varepsilon_2 = 0.00192$, making $n^* = 3{,}421$. Consequently, at least 6,843 fuzzy rules are needed.

Example 2. Design a fuzzy system to uniformly approximate the polynomial defined on $C^2[-1, 1]$:

$$P_2(\underline{x}) = 0.52 + 0.1x_1 + 0.38x_2 - 0.06x_1x_2,$$

with $\varepsilon = 0.1$.

Obviously, $d = 2$, $\beta_{0,0} = 0.52$, $\beta_{1,0} = 0.1$, $\beta_{0,1} = 0.38$, $\beta_{1,1} = -0.06$, and $s = 2$. We shall estimate n^* based on (11) (note $\varepsilon_2 = \varepsilon$ in this example):

$$n \approx \frac{1}{\varepsilon_2} \sum_{d_i \ge 1} \left(|\beta_{d_1,\ldots,d_r}| \sum_{i=1}^{r} (d_i \cdot c_i^{\max}) \right)$$
$$= \frac{1}{0.1}[0.1(1 \times 1) + 0.38(1 \times 1) + 0.06(1 \times 1 + 1 \times 1)] = 6.$$

According to (16),

$$0.1\left[\left(1 + \frac{1}{6}\right) - 1\right] + 0.38\left[\left(1 + \frac{1}{6}\right) - 1\right] + 0.06\left[\left(1 + \frac{1}{6}\right)\left(1 + \frac{1}{6}\right) - 1\right] = 0.10167 > \varepsilon = 0.01.$$

So, we need to increase n^* to seven which makes $\varepsilon = 0.087 < 0.1$. As a result, $N^2 = (2n^* + 1)^2 = 225$ fuzzy rules are needed. The transformed fuzzy rules are

$$P_2\left(\frac{p}{n}\right) = 0.52 + 0.1\frac{p_1}{7} + 0.38\frac{p_2}{7} - 0.06\frac{p_1}{7} \cdot \frac{p_2}{7}.$$

If $\varepsilon = 0.01$ is wanted, then $n^* = 61$ and 15,129 fuzzy rules are needed.

From this example, one can see that n^* can be quite accurately estimated using (11) even when n^* is as low as six. The estimation accuracy is better when n^* is larger.

If we use (17) to estimate n^*, then $n^* = 12$ ($N^2 = 625$) when $\varepsilon = 0.1$ and $n^* = 120$ ($N^2 = 58{,}081$) when $\varepsilon = 0.01$, which are much larger numbers (too conservative).

In these two examples, it is obvious that the fuzzy rules can be easily obtained using (3). The rest of the components can be easily determined according to the requirements stated in sections.

5. *Conclusions*

The components of the general fuzzy systems studied in this paper are almost arbitrary. Moreover, both the centroid defuzzification and the mean of maximum defuzzification are included. Utilizing the Weierstrass approximation theorem, we have not only constructively proven that the general fuzzy systems can uniformly approximate any real continuous function on a compact domain to arbitrary accuracy, but we also have disclosed the underlying mechanism of such function approximation. Further, we have found sufficient conditions for the general fuzzy systems as function approximators.

Generally speaking, accurately approximating (complex) functions requires partitioning universes of discourse of input variables into many intervals. The more accurate the approximation desired, the greater the number of the partitions must be, and consequently, the greater the number of fuzzy rules must be. This is intuitionally reasonable because if one wants to describe every aspect of a system in every detail, the number of fuzzy rules is bound to be overwhelmingly large. Fuzzy systems are based on fuzzy rules which describe experience and knowledge of human experts. In practice, it is difficult, if not impossible, for experts to state an arbitrarily large number of fuzzy rules. Therefore, it is logical to conclude that, in theory, fuzzy systems are not ideal function approximators when high approximation accuracy is required. Theoretically speaking, some nonfuzzy function approximators are better.

Practically speaking, however, the fuzzy system approach may be a better approach because human experience and knowledge can be utilized and built into systems. It is this distinct advantage that makes the fuzzy system approach more practical than many other theoretically favorable approaches. Taking control as an example, nonfuzzy control techniques need mathematically explicit and precise process models, which is often a difficult requirement to be satisfied in practice. On the other hand, even if a nonlinear process model is available, it is possible that the existing nonlinear control theory still cannot produce an appropriate controller. In contrast, such a nonlinear controller may be constructed using fuzzy control techniques without the process model. Fuzzy controllers with a relatively small number of fuzzy rules have reportedly solved some real-world complicated control problems effectively.

Finally, in light of the results achieved in this paper, we believe that it is important to develop novel fuzzy systems whose structures will enable the fuzzy systems as better function approximators in the sense that less fuzzy rules are needed in order to approximate any continuous functions to any degree of accuracy.

Acknowledgements—The author would like to thank Professor L. C. Sheppard for his support in the research. Thanks also go to Drs A. Chakrabarty and G. R. Chen for their helpful comments and suggestions.

References

Bronshtein, I. N. and K. A. Semendyayev (1985). *Handbook of Mathematics*, p. 693. Van Nostrand Reinhold, NY.

Buckley, J. J. (1992). Universal fuzzy controllers. *Automatica*, **28**, 1245–1248.

Filev, D. P. and R. R. Yager (1991). A generalized defuzzification method via BAD distributions. *Int. J. of Intelligent Systems*, **6**, 687–697.

Gupta, M. M. and J. Qi (1991a). Theory of T-norms and fuzzy inference methods. *Fuzzy Sets and Sysems*, **40**, 431–450.

Gupta, M. M. and J. Qi (1991b). Design of fuzzy logic controllers based on generalized T-operators. *Fuzzy Sets and Systems*, **40**, 473–490.

Kosko, B. (1992). Fuzzy systems as universal approximators. *Proc. of IEEE International Conference on Fuzzy Systems*. San Diego, CA.

Lee, C. C. (1990). Fuzzy logic in control systems: Fuzzy logic controller. *IEEE Trans. on Systems, Man and Cybernetics*, **20**, 404–435.

Mizumoto, M. (1988). Fuzzy controls under various fuzzy reasoning methods. *Information Sciences*, **45**, 129–151.

Picard, E. (1891). Sur la représentation approachée des fonctions. *CR*, **112**, 183–186.

Rudin, W. (1976). *Principles of Mathematical Analysis*, 3rd ed. McGraw-Hill, NY.

Wang, L. X. (1992) Fuzzy systems are universal approximators. *Proc. of IEEE International Conference on Fuzzy Systems*. San Diego, CA.

Wang, L. X. and J. M. Mendel (1992). Fuzzy basis functions, universal approximation, and orthogonal least-squares learning. *IEEE Trans. on Neural Networks*, **3**, 807–814.

Ying, H. (1993). General analytical structure of typical fuzzy controllers and their limiting structure theorems, *Automatica*, **29**, 1139–1143.