Journal of Intelligent & Fuzzy Systems 29 (2015) 1219–1227
DOI:10.3233/IFS-151730
IOS Press

A sufficient condition on a general class of interval type-2 Takagi-Sugeno fuzzy systems with linear rule consequent as universal approximators

Hao Ying*
Department of Electrical and Computer Engineering, Wayne State University, Detroit, Michigan, USA

Abstract. Whether a rule-based interval type-2 fuzzy system has the ability to approximate any continuous multivariate function arbitrarily well is a fundamentally important question for fuzzy control and modeling. The only approximation results available are the preliminary ones that we previously obtained. They state that two general classes of the interval T2 fuzzy systems, one for the Mamdani type and the other for the TS type, are universal approximators. We now further our investigation on the same T2 TS fuzzy systems to establish a quantitative sufficient approximation condition. These TS fuzzy systems use the linear rule consequent. Their input fuzzy sets are interval T2 and can be in any shape (e.g., Gaussian or trapezoidal) and the fuzzy AND operators in the rules can be any one type or mixed types. We first proved that the fuzzy systems could uniformly approximate any multivariate polynomials arbitrarily accurately and then utilized the Weierstrass approximation theorem to produce the sufficient approximation condition. The condition is a formula calculating the number of the input fuzzy sets needed to achieve any given approximation accuracy (the number of fuzzy rules needed can be easily computed afterward). A numerical example is provided for illustration.

Keywords: Function approximation, polynomials, interval type-2 fuzzy systems, TS fuzzy systems

1. Introduction

A type-2 (T2) fuzzy set employs upper and lower primary membership functions and a secondary membership function. This is in contrast to a type-1 (T1) fuzzy set that has only one primary membership function without a secondary membership function. As a result, a T2 fuzzy set is able to more naturally and possibly more effectively handle ranges of uncertainties and ambiguities. Any rule-based fuzzy system that uses a T2 fuzzy set can loosely be classified as a T2 fuzzy system. Although T2 fuzzy set was proposed and defined by Zadeh 40 years ago [38], not much research was reported in the literature until more recently. As the field of T1 fuzzy systems becomes mature after over 50 years of development, more and more attention is turned to T2 fuzzy systems partially due to their relative novelty and potential advantages over the T1 fuzzy systems. An important factor that makes current T2 fuzzy system study possible is the availability of a much more comprehensive foundation for T2 systems-related fuzzy mathematics and their algorithmic implementations, which were gradually accomplished only in the past decade or so [20, 22]. Taking advantage of the

*Corresponding author. Hao Ying, Department of Electrical and Computer Engineering, Wayne State University, Detroit, Michigan 48202, USA. Tel.: +313 577 3738; Fax: 313 577 1101; E-mail: hao.ying@wayne.edu.

improved infrastructure, interest in T2 fuzzy control and fuzzy modeling, arguably the most active areas of T2 fuzzy systems research so far, has surged [21].

The question whether a fuzzy system, T1 or T2, Mamdani type or Takagi-Sugeno (TS) type, is a universal approximator in that from input-output mapping standpoint, it can approximate any real multivariate continuous function to any degree of accuracy on a compact domain is of fundamental importance for the fields of fuzzy control, fuzzy modeling, and others (e.g., fuzzy signal processing). The answer for the various T1 fuzzy systems is affirmative and many results have been reported [2–4, 6–9, 13, 16–19, 23, 24–27, 29, 30, 34], including those that we originated as the first (quantitative) sufficient approximation conditions, as opposed to the existence result, covering a general classes of T1 TS fuzzy systems [32] and a general classes of T1 Mamdani fuzzy systems [30]. The topics cover existence theorems [11, 12, 27, 28], approximation accuracy [39, 40, 41], necessary conditions [5, 31, 37], and sufficient conditions [14, 30, 32, 33], to name a few. Together, these results provide a theoretical assurance that in most cases, if not all, a T1 fuzzy controller can be constructed to achieve desired system performance and a T1 fuzzy model can be made to closely represent the physical system's input-output relation. They also make the rigorous development of adaptive T1 fuzzy control possible because an adaptation scheme requires the universal approximation capability.

A comparable approximation theory for the T2 fuzzy systems is yet to establish. The only results currently available are ours concerning a general class of the T2 Mamdani fuzzy systems [35], and a general class of the T2 TS fuzzy systems [36]. Due to the nature of the T2 fuzzy sets, a T2 TS fuzzy system is more complicated than a T1 TS fuzzy system. Theoretically speaking, the functions in the rule consequent can be any type; but the vast majority of the fuzzy systems use the linear functions. Accordingly, we will deal with the linear rule consequent only in this paper. We will use the same two-step constructive proof process that we initially developed in [30]. We will answer the following questions: (1) Are the fuzzy systems function approximators? and (2) if so, how to determine the amounts of the input fuzzy sets and fuzzy rules needed to approximate any given function with a given approximation accuracy? The first question has been partially addressed in [36], but the second question has never been explored in the literature.

Because a T2 fuzzy system degrades to the T1 fuzzy system when the footprints of uncertainty of all the T2 fuzzy sets in the system become zero, some people might think that a separate T2 fuzzy approximation theory is not needed if the T1 fuzzy system has already been known to be a universal approximator. This view however is incomprehensive because the fundamental question remains open - whether or not the T2 fuzzy system with nonzero footprints of uncertainty (i.e., a genuine T2 system) is a universal approximator. It is this question that is the focus of our work in this paper.

In the next section, we will define various components of the T2 TS fuzzy systems. In Section 3, we will provide the mathematical proofs as well as the derivations of the sufficient condition. A numerical example is given in Section 4. Section 5 concludes the paper.

2. Configuration of the general interval type-2 TS fuzzy systems

The general TS fuzzy systems use the following components. The first component is an input vector, $\mathbf{x}(t)$, with r scaled input variables of the fuzzy systems:

$$\mathbf{x}(t) = [x_1(t)\; x_2(t) \cdots\cdots x_r(t)]$$
$$= [a_1 z_1(t)\; a_2 z_2(t) \cdots\cdots a_r z_r(t)]$$

where $z_i(t)$ and $x_i(t)$, $i = 1,\ 2,\ \cdots\cdots,\ r$, are the original and scaled continuous-time or discrete-time input variables, respectively. For simplicity, we use $\mathbf{x}$, z_i and x_i to represent $\mathbf{x}(t)$, $z_i(t)$ and $x_i(t)$, respectively. a_i is a scaling factor that makes $-1 \leq x_i \leq 1$. Without loss of generality, assume that there are

$$N = 2n + 1,\ n \geq 1$$

continuous interval T2 fuzzy sets for x_i, each of which is denoted $A_{i,j}$, $j = 0,\ \pm 1,\ \ldots\ldots,\ \pm n$. The positive, negative, and zero subscripts j can be regarded as linguistic labels of "positive," "negative" and "zero," respectively (e.g., $A_{i,1}$ means that x_i is "positive small" and $A_{i,-4}$ indicates that x_i is "negative large"). The upper and lower membership functions of $A_{i,j}$, designated as $\bar{\mu}_{A_{i,j}}$ and $\underline{\mu}_{A_{i,j}}$ respectively, are defined over [−1, 1] that is equally partitioned into $N - 1$ intervals. Hence, each interval is $[j/n,\ (j+1)/n]$. $\bar{\mu}_{A_{i,j}}$ and $\underline{\mu}_{A_{i,j}}$ are defined over $\left[\frac{j-k_{i,j}^L}{n},\ \frac{j+k_{i,j}^R}{n}\right]$, where $k_{i,j}^L$ and $k_{i,j}^R$ are positive integers and are greater than or equal to 1, and $j - k_{i,j}^L \geq -n$ and $j + k_{i,j}^R \leq n$. The membership value of $\bar{\mu}_{A_{i,j}}$ and $\underline{\mu}_{A_{i,j}}$ is 0 at the interval terminal points and beyond and is nonzero everywhere inside the interval. We introduce the notations:

$$c_i^{\max} := \max_j \left\{ k_{i,j}^L,\ k_{i,j}^R \right\}$$
$$c^{\max} := \max_i \left\{ c_i^{\max} \right\}.$$

In addition to indicating the interval size of $\left[\frac{j-k_{i,j}^L}{n},\ \frac{j+k_{i,j}^R}{n}\right]$, $c_i^{\max}$ also implies the largest possible number of intersections between $A_{i,j}$ and its adjacent right and left fuzzy sets. For instance, if $c_i^{\max} = 2$, it would mean $A_{i,j}$ possibly interests with $A_{i,j-2}$, $A_{i,j-1}$, $A_{i,j+1}$ and $A_{i,j+2}$ (the actual number of intersections depends on the shapes of these five fuzzy sets, and can be either 1 or 2). In most cases, $c_i^{\max}$ should be 1 for all i (i.e., $c^{\max} = 1$), meaning that $A_{i,j}$ should only intersect with $A_{i,j-1}$ and $A_{i,j+1}$ and only once. The secondary membership function of $A_{i,j}$ is constant 1 over its entire universe of discourse [0, 1] for any value of x_i, which is the definition of $A_{i,j}$ being an interval T2 fuzzy set.

Use of N^r fuzzy rules is necessary in order to cover all the possible combinations of the input fuzzy sets. Assume that $P(t)$, where $1 \le P(t) \le N^r$, represents the number of fuzzy rules with linear consequent executed at time t. The m-th rule is:

$$\text{IF } x_1 \textit{ is } A_{1,p_{1,m}} \text{ AND } x_2 \textit{ is } A_{2,p_{2,m}} \text{ AND } \ldots \text{ AND } x_r \textit{ is } A_{r,p_{r,m}} \text{ THEN } u(t) = a_{0,m} + a_{1,m}x_1 + \ldots\ldots + a_{r,m}x_r \quad (1)$$

where $u(t)$ is the output variable of the T2 fuzzy systems and $a_{0,m}, \ \ldots, \ a_{r,m}$ are design constants. We suppose that at time t, the computed value of $u(t)$ is U_m for the m-th rule. That is,

$$U_m = \sum_{i=0}^{r} a_{i,m} x_i$$

where x_0 is a dummy variable and $x_0 \equiv 1$. To evaluate the fuzzy ANDs in the fuzzy rules, any fuzzy T-norm may be used (e.g., the Zadeh AND operator or the product AND operator or both) [10]. To be general, we permit different fuzzy AND operators to be used in a rule or in different rules. The firing interval for the m-th rule is assumed to be $\left[\underline{\lambda}_m, \bar{\lambda}_m\right]$ where

$$\underline{\lambda}_m = \min\left(\underline{\mu}_{A1,_{p1,m}}, \underline{\mu}_{A2,_{p2,m}}, \cdots, \underline{\mu}_{Ar,_{pr,m}}\right)$$
$$\bar{\lambda}_m = \min\left(\bar{\mu}_{A1,_{p1,m}}, \bar{\mu}_{A2,_{p2,m}}, \cdots, \bar{\mu}_{Ar,_{pr,m}}\right).$$

One can view $\left[\underline{\lambda}_m, \bar{\lambda}_m\right]$ and U_m together as a singleton T2 fuzz set on $u(t)$ that is non-zero only at $u(t) = U_m$ and 0 elsewhere and whose upper and lower primary membership functions degenerate to points $\bar{\lambda}_m$ and $\underline{\lambda}_m$, respectively. Note that the number of the singleton T2 fuzz sets is $P(t)$ and they are formed at $u(t) = U_m$, $m = 0,\ 1, \ldots, P(t)$. Collectively, they can be viewed as one discrete interval T2 fuzzy set whose upper and lower membership functions are formed by the $P(t)$ points of $\bar{\lambda}_m$ and $\underline{\lambda}_m$, respectively.

The discrete interval T2 fuzzy set needs to be converted to a discrete T1 fuzzy set using a fuzzy set type-reducer. We use the center-of-sets reducer for this purpose [15]. The T1 fuzzy set contains a total of $2^{P(t)}$ points on $u(t)$, the membership value of all of which is all 1. We only need to calculate the smallest (μ_L) and largest (μ_R) elements of these $2^{P(t)}$ points:

$$u_L = \min_{\lambda_m \in [\underline{\lambda}_m, \bar{\lambda}_m]} \left\{ \frac{\sum_{m=1}^{P(t)} \lambda_m \cdot U_m}{\sum_{m=1}^{P(t)} \lambda_m} \right\}$$

$$u_R = \max_{\lambda_m \in [\underline{\lambda}_m, \bar{\lambda}_m]} \left\{ \frac{\sum_{m=1}^{P(t)} \lambda_m \cdot U_m}{\sum_{m=1}^{P(t)} \lambda_m} \right\}$$

The rest of the $2^{P(t)}$-2 points can be ignored here as they will not be needed. The average of μ_L and μ_R represents the final output of the fuzzy systems [22]. From the input-output standpoint, we treat the systems as a mapping F_n: $C^r[-1,\ 1] \to (-\infty,\ \infty)$, where $C^r[-1,\ 1]$ represents a r-dimensional product space for $\mathbf{x}$ and $(-\infty, \infty)$ is the universe of discourse for $u(t)$. That is,

$$F_n(\mathbf{x}) = \frac{u_L + u_R}{2} = \frac{1}{2} \min_{\lambda_m \in [\underline{\lambda}_m, \bar{\lambda}_m]} \left\{ \frac{\sum_{m=1}^{P(t)} \lambda_m \cdot U_m}{\sum_{m=1}^{P(t)} \lambda_m} \right\} + \frac{1}{2} \max_{\lambda_m \in [\underline{\lambda}_m, \bar{\lambda}_m]} \left\{ \frac{\sum_{m=1}^{P(t)} \lambda_m \cdot U_m}{\sum_{m=1}^{P(t)} \lambda_m} \right\} \quad (2)$$

Denote θ_m the values of such λ_m's that make the first part of (2) (i.e., the min part) achieve the minimization and ζ_m the values of such λ_m's that make the second part of (2) (i.e., the max portion) attain the maximization. We can rewrite (2) as

$$F_n(\mathbf{x}) = \frac{1}{2}\frac{\sum_{m=1}^{P(t)}\theta_m \cdot U_m}{\sum_{m=1}^{P(t)}\theta_m} + \frac{1}{2}\frac{\sum_{m=1}^{P(t)}\zeta_m \cdot U_m}{\sum_{m=1}^{P(t)}\zeta_m}$$

$$= \frac{1}{2}\frac{\sum_{m=1}^{P(t)}\theta_m \cdot \sum_{i=0}^{r} a_{i,m}x_i}{\sum_{m=1}^{P(t)}\theta_m} + \frac{1}{2}\frac{\sum_{m=1}^{P(t)}\zeta_m \cdot \sum_{i=0}^{r} a_{i,m}x_i}{\sum_{m=1}^{P(t)}\zeta_m} \tag{3}$$

The detail on how θ_m and ζ_m (and hence $\bar{\lambda}_m$ and $\underline{\lambda}_m$) are obtained and what their values are is not important. Indeed, they will not needed in the mathematical proofs in the next section. Note that $F_n(\mathbf{x})$ is a function sequence with respect to $n \left(= \frac{N-1}{2}\right)$. In the rest of the paper, $F_n(\mathbf{x})$ represents the general T2 TS fuzzy systems.

3. Results

We now establish a sufficient condition under which the general interval T2 TS fuzzy systems defined in the previous section can uniformly approximate any multivariate continuous function with any degree of accuracy on a compact domain. The proof is a two-step constructive procedure. The first step is to establish Theorem 1 below.

Without loss of generality, we suppose that $P_d(\mathbf{x})$ is a multivariate polynomial function of degree d defined in $C^r[-1,\ 1]$:

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

Theorem 1. *A T2 TS fuzzy system of the general class can uniformly approximate $P_d(\mathbf{x})$ with an arbitrarily small approximation error bound. In other words, $\forall \varepsilon > 0$, there exists a positive integer n^* such that $\forall n > n^*$,*

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

Proof. Note that $\|\cdot\|$ means the norm. We have

$$\|F_n - P_d\| = \max_{\mathbf{x}\in C^r[-1,1]}\left|\frac{1}{2}\frac{\sum_{m=1}^{P(t)}\theta_m \cdot \sum_{i=0}^{r} a_{i,m}x_i}{\sum_{m=1}^{P(t)}\theta_m} + \frac{1}{2}\frac{\sum_{m=1}^{P(t)}\zeta_m \cdot \sum_{i=0}^{r} a_{i,m}x_i}{\sum_{m=1}^{P(t)}\zeta_m} - P_d(\mathbf{x})\right|$$
$$\leq \frac{1}{2}\max_{\mathbf{x}\in C^r[-1,1]}\left|\frac{\sum_{m=1}^{P(t)}\theta_m \cdot \sum_{i=0}^{r} a_{i,m}x_i}{\sum_{m=1}^{P(t)}\theta_m} - P_d(\mathbf{x})\right| + \frac{1}{2}\max_{\mathbf{x}\in C^r[-1,1]}\left|\frac{\sum_{m=1}^{P(t)}\zeta_m \cdot \sum_{i=0}^{r} a_{i,m}x_i}{\sum_{m=1}^{P(t)}\zeta_m} - P_d(\mathbf{x})\right| \tag{4}$$

We construct the N^r fuzzy rules with linear consequent in such a way that the parameter values in the consequent of the m-th rule are chosen as follows:

$$\begin{aligned} a_{0,m} &= P_d\left(\frac{p_{1,m}}{n}, \ldots, \frac{p_{r,m}}{n}\right) - \left(\beta_{1,0\ldots,0}\frac{p_{1,m}}{n} + \beta_{0,1\ldots,0}\frac{p_{2,m}}{n} + \ldots + \beta_{0,\ldots,1}\frac{p_{r,m}}{n}\right) \\ a_{1,m} &= \beta_{1,0,\ldots,0} \\ a_{2,m} &= \beta_{0,1,\ldots,0} \\ &\vdots \\ a_{r,m} &= \beta_{0,\ldots,1} \end{aligned} \tag{5}$$

Here, $p_{i,m}$ means j used in Section 2 and the subscripts i and m are needed to indicate which input variable and fuzzy rule are involved. By the same token, we employ $k_{i,m}^L$ and $k_{i,m}^R$ in place of $k_{i,j}^L$ and $k_{i,j}^R$. Because at any time, it is always true that

$$\frac{p_{i,m} - k_{i,m}^L}{n} \leq x_i \leq \frac{p_{i,m} + k_{i,m}^R}{n}$$

and hence

$$\left|x_i - \frac{p_{i,m}}{n}\right| \leq \frac{c_i^{\max}}{n} \leq \frac{c^{\max}}{n} \tag{6}$$

Noting that when $n \to \infty$, $p_{i,m} \to \infty$, we obtain

$$\lim_{n\to\infty}\left|x_i - \frac{p_{i,m}}{n}\right| \leq \lim_{n\to\infty}\frac{c^{\max}}{n} = 0$$

which means

$$\lim_{n\to\infty} \frac{p_{i,m}}{n} = x_i$$

Therefore, based on (5)

$$\lim_{n\to\infty} a_{0,m} = P_d(\mathbf{x}) - \left(\beta_{1,0\ldots,0}x_1 + \beta_{0,1\ldots,0}x_2 + \ldots + \beta_{0,\ldots,1}x_r\right)$$

which results in

$$\lim_{n\to\infty} \sum_{i=0}^{r} a_{i,m}x_i = P_d(\mathbf{x})$$

That is to say when $n \to \infty$, $\|F_n - P_d\| \to 0$ for (4) because in (4)

$$\left| \frac{\sum_{m=1}^{P(t)} \theta_m \cdot \sum_{i=0}^{r} a_{i,m}x_i}{\sum_{m=1}^{P(t)} \theta_m} - P_d(\mathbf{x}) \right| \to 0$$

and

$$\left| \frac{\sum_{m=1}^{P(t)} \zeta_m \cdot \sum_{i=0}^{r} a_{i,m}x_i}{\sum_{m=1}^{P(t)} \zeta_m} - P_d(\mathbf{x}) \right| \to 0$$

Now we prove that the approximations are uniform. To accomplish this task, we will derive a formula that can be used to calculate a positive integer n^*, based on given approximation error ε, such that $\forall n > n^*$, $\|F_n - P_d\| < \varepsilon$. Note that

$$\max_{\mathbf{x}\in C^r[-1,1]} \left| \frac{\sum_{m=1}^{P(t)} \theta_m \cdot \sum_{i=0}^{r} a_{i,m}x_i}{\sum_{m=1}^{P(t)} \theta_m} - P_d(\mathbf{x}) \right|$$

$$= \max_{\mathbf{x}\in C^r[-1,1]} \left| \frac{\sum_{m=1}^{P(t)} \theta_m \cdot a_{0,m} + \sum_{m=1}^{P(t)} \theta_m \cdot \sum_{i=1}^{r} a_{i,m}x_i}{\sum_{m=1}^{P(t)} \theta_m} - P_d(\mathbf{x}) \right|$$

$$= \max_{\mathbf{x}\in C^r[-1,1]} \left| \frac{\sum_{m=1}^{P(t)} \theta_m \cdot a_{0,m}}{\sum_{m=1}^{P(t)} \theta_m} - P_d(\mathbf{x}) + \beta_{1,0\ldots,0}x_1 + \beta_{0,1\ldots,0}x_2 + \ldots + \beta_{0\ldots,r}x_r \right|$$

$$= \max_{\mathbf{x}\in C^r[-1,1]} \left| \frac{\sum_{m=1}^{P(t)} \theta_m P_d\left(\frac{p_{1,m}}{n}, \ldots, \frac{p_{r,m}}{n}\right)}{\sum_{m=1}^{P(t)} \theta_m} - P_d(\mathbf{x}) - \frac{\sum_{m=1}^{P(t)} \theta_m \left(\beta_{1,0\ldots,0}\frac{p_{1,m}}{n} + \beta_{0,1\ldots,0}\frac{p_{2,m}}{n} + \ldots + \beta_{0,\ldots,1}\frac{p_{r,m}}{n}\right)}{\sum_{m=1}^{P(t)} \theta_m} + \beta_{1,0\ldots,0}x_1 + \beta_{0,1\ldots,0}x_2 + \cdots + \beta_{0\ldots,r}x_r \right|$$

$$\le \max_{\mathbf{x}\in C^r[-1,1]} \left| \frac{\sum_{m=1}^{P(t)} \theta_m P_d\left(\frac{p_{1,m}}{n}, \cdots, \frac{p_{r,m}}{n}\right)}{\sum_{m=1}^{P(t)} \theta_m} - P_d(\mathbf{x}) \right| + \max_{\mathbf{x}\in C^r[-1,1]} \left| \left(\sum_{m=1}^{P(t)} \theta_m \left(\beta_{1,0\ldots,0}\frac{p_{1,m}}{n} + \beta_{0,1\ldots,0}\frac{p_{2,m}}{n} + \cdots + \beta_{0,\ldots,1} \times \frac{p_{r,m}}{n}\right)\right) / \sum_{m=1}^{P(t)} \theta_m - (\beta_{1,0\ldots,0}x_1 + \beta_{0,1\ldots,0}x_2 + \cdots + \beta_{0\ldots,r}x_r) \right|$$

Using (6), the part following + sign after $\le$ in (7) can be written as

$$\max_{\mathbf{x}\in C^r[-1,1]} \left| \left(\sum_{m=1}^{P(t)} \theta_m \left[\beta_{1,0\ldots,0}\left(\frac{p_{1,m}}{n} - x_1\right) + \beta_{0,1\ldots,0} \times \left(\frac{p_{2,m}}{n} - x_2\right) + \cdots + \beta_{0,\ldots,1}\left(\frac{p_{r,m}}{n} - x_r\right) \right] \right) / \sum_{m=1}^{P(t)} \theta_m \right|$$

$$\le \frac{c^{\max}}{n} \max_{\mathbf{x}\in C^r[-1,1]} \left| \frac{\sum_{m=1}^{P(t)} \theta_m \left(\beta_{1,0\ldots,0} + \beta_{0,1\ldots,0} + \ldots + \beta_{0,\ldots,1}\right)}{\sum_{m=1}^{P(t)} \theta_m} \right|$$

$$\le \frac{c^{\max}\left|\beta_{1,0\ldots,0} + \beta_{0,1\ldots,0} + \ldots + \beta_{0,\ldots,1}\right|}{n}$$

On the other hand, the part immediately after $\le$ in (7) equals to

$$\max_{\mathbf{x}\in C^r[-1,1]} \left| \frac{\sum_{m=1}^{P(t)} \theta_m \left[P_d\left(\frac{p_{1,m}}{n}, \ldots, \frac{p_{r,m}}{n}\right) - P_d(\mathbf{x})\right]}{\sum_{m=1}^{P(t)} \theta_m} \right|$$

$$\le \max_{\mathbf{x}\in C^r[-1,1]} \left\{ \frac{\sum_{m=1}^{P(t)} \theta_m \left|P_d\left(\frac{p_{1,m}}{n}, \ldots, \frac{p_{r,m}}{n}\right) - P_d(\mathbf{x})\right|}{\sum_{m=1}^{P(t)} \theta_m} \right\}$$

$$= \max_{\mathbf{x}\in C^r[-1,1]} \left\{ \left(\sum_{m=1}^{P(t)} \theta_m \left| \sum_{d_i\ge 0} \left(\beta_{d_1,\ldots,d_r} \prod_{i=1}^{r} \left(\frac{p_{i,m}}{n}\right)^{d_i} \right) - \sum_{d_i\ge 0} \left(\beta_{d_1,\ldots,d_r} \prod_{i=1}^{r} x_i^{d_i} \right) \right| \right) \Big/ \sum_{m=1}^{P(t)} \theta_m \right\}$$

$$\le \max_{\mathbf{x}\in C^r[-1,1]} \left\{ \frac{\sum_{m=1}^{P(t)} \theta_m \sum_{d_i\ge 0} \left|\beta_{d_1,\cdots,d_r}\right| \left| \prod_{i=1}^{r} \left(\frac{p_{i,m}}{n}\right)^{d_i} - \prod_{i=1}^{r} x_i^{d_i} \right|}{\sum_{m=1}^{P(t)} \theta_m} \right\} \tag{7}$$

According to (6), one can always find such $\delta_{i,m}$, where $-k_{i,m}^L \le \delta_{i,m} \le k_{i,m}^R$, that makes $\frac{p_{i,m}+\delta_{i,m}}{n} = x_i$. Also, keep in mind that $|x_i| \le 1$. Therefore,

$$\sum_{d_i\ge 0} \left|\beta_{d_1,\cdots,d_r}\right| \left| \prod_{i=1}^{r} \left(\frac{p_{i,m}}{n}\right)^{d_i} - \prod_{i=1}^{r} x_i^{d_i} \right|$$

$$= \sum_{d_i\ge 0} \left|\beta_{d_1,\cdots,d_r}\right| \left| \prod_{i=1}^{r} \left(x_i - \frac{\delta_{i,m}}{n}\right)^{d_i} - \prod_{i=1}^{r} x_i^{d_i} \right|$$

Note that

$$\prod_{i=1}^{r} \left(x_i - \frac{\delta_{i,m}}{n}\right)^{d_i} = \prod_{i=1}^{r} \left(\sum_{k_i=0}^{d_i} C_{d_i}^{k_i} x_i^{d_i-k_i} \left(-\frac{\delta_{i,m}}{n}\right)^{k_i} \right)$$

$$= \prod_{i=1}^{r} \left(x_i^{d_i} + \sum_{k_i=1}^{d_i} C_{d_i}^{k_i} x_i^{d_i-k_i} \left(-\frac{\delta_{i,m}}{n}\right)^{k_i} \right)$$

$$\le \prod_{i=1}^{r} \left(x_i^{d_i} + \sum_{k_i=1}^{d_i} C_{d_i}^{k_i} x_i^{d_i-k_i} \left(\frac{\left|\delta_{i,m}\right|}{n}\right)^{k_i} \right)$$

$$\le \prod_{i=1}^{r} \left(x_i^{d_i} + \sum_{k_i=1}^{d_i} C_{d_i}^{k_i} \left(\frac{c^{\max}}{n}\right)^{k_i} \right)$$

$$\le \prod_{i=1}^{r} \left(x_i^{d_i} + \sum_{k=1}^{d} C_d^w \frac{c^{\max}}{n} \right)$$

$$= \prod_{i=1}^{r} \left(x_i^{d_i} + \frac{d \cdot C_d^w \cdot c^{\max}}{n} \right) \tag{8}$$

where we use the standard notation for the combinations

$$C_{d_i}^{k_i} = \frac{d_i!}{(d_i-k_i)!k_i!}$$

and utilize the facts $|x_i| \le 1$, $x_i^{d_i-k_i} \le 1$, and $\frac{|\delta_{i,m}|}{n} \le \frac{c^{\max}}{n}$. To reach the last inequality in (8), note that for $k_i = 1, 2, \cdots\cdots, d_i$, the maximum of $C_{d_i}^{k_i}$ is reached when $d_i = d$ and k_i is the integer that is the middle term between 1 and d. We denote such a value w. We also use the fact $\left(\frac{c^{\max}}{n}\right)^{k_i} \le \frac{c^{\max}}{n}$. Image $x_i^{d_i} = 1$. One attains from (8) (for better presentation, we still use $x_i^{d_i}$ instead of its value 1 for the $\prod_{i=1}^{r} x_i^{d_i}$ term)

$$\prod_{i=1}^{r} \left(x_i^{d_i} + \frac{d \cdot C_d^w \cdot c^{\max}}{n} \right)$$

$$= \prod_{i=1}^{r} x_i^{d_i} + \left(1 + \frac{d \cdot C_d^k \cdot c^{\max}}{n}\right)^r - 1$$

$$= \prod_{i=1}^{r} x_i^{d_i} + \sum_{q=0}^{r} C_r^q \left(\frac{d \cdot C_d^k \cdot c^{\max}}{n}\right)^{r-q} - 1$$

$$= \prod_{i=1}^{r} x_i^{d_i} + \sum_{q=0}^{r-1} C_r^q \left(\frac{d \cdot C_d^k \cdot c^{\max}}{n}\right)^{r-q}$$

$$\le \prod_{i=1}^{r} x_i^{d_i} + \frac{c^{\max}}{n} \sum_{q=0}^{r-1} C_r^q \left(d \cdot C_d^w\right)^{r-q}$$

$$= \prod_{i=1}^{r} x_i^{d_i} + \frac{c^{\max}}{n} \left(\sum_{q=0}^{r} C_r^q \left(d \cdot C_d^w\right)^{r-q} - C_r^r \right)$$

$$= \prod_{i=1}^{r} x_i^{d_i} + \frac{c^{\max}}{n} \left[\left(d \cdot C_d^w + 1\right)^r - 1 \right].$$

Hence

$$\sum_{d_i\ge 0} \left|\beta_{d_1,\cdots,d_r}\right| \left| \prod_{i=1}^{r} \left(\frac{p_{i,m}}{n}\right)^{d_i} - \prod_{i=1}^{r} x_i^{d_i} \right| \le \frac{c^{\max} \left[\left(d \cdot C_d^w + 1\right)^r - 1 \right] \sum_{d_i\ge 0} \left|\beta_{d_1,\cdots,d_r}\right|}{n}$$

and consequently the last expression in (7) is less than or equal to

$$\frac{c^{\max} \left[\left(d \cdot C_d^w + 1\right)^r - 1 \right] \sum_{d_i\ge 0} \left|\beta_{d_1,\cdots,d_r}\right|}{n}.$$

As a result, the portion after the last $\leq$ sign in (7) is less than or equal to

$$\frac{c^{\max}}{n}\Big(\left|\beta_{1,0\ldots,0}+\beta_{0,1\ldots,0}+\ldots+\beta_{0,\ldots,1}\right| + \left[\left(d\cdot C_d^w+1\right)^r-1\right]\sum_{d_i\geq 0}\left|\beta_{d_1,\cdots,d_r}\right|\Big).$$

Clearly,

$$\max_{\mathbf{x}\in C^r[-1,1]}\left|\frac{\sum_{m=1}^{P(t)}\theta_m\cdot\sum_{i=0}^{r}a_{i,m}x_i}{\sum_{m=1}^{P(t)}\theta_m}-P_d(\mathbf{x})\right| \leq \frac{c^{\max}}{n}\Big(\left|\beta_{1,0\ldots,0}+\beta_{0,1\ldots,0}+\ldots+\beta_{0,\ldots,1}\right| + \left[\left(d\cdot C_d^w+1\right)^r-1\right]\sum_{d_i\geq 0}\left|\beta_{d_1,\cdots,d_r}\right|\Big)$$

Obviously, this inequality also holds for $\max_{\mathbf{x}\in C^r[-1,1]}\left|\frac{\sum_{m=1}^{P(t)}\zeta_m\cdot\sum_{i=0}^{r}a_{i,m}x_i}{\sum_{m=1}^{P(t)}\zeta_m}-P_d(\mathbf{x})\right|$ above. So, from (4), one attains

$$\|F_n-P_d\|\leq\frac{c^{\max}}{n}\Big(\left|\beta_{1,0\ldots,0}+\beta_{0,1\ldots,0}+\ldots+\beta_{0,\ldots,1}\right| + \left[\left(d\cdot C_d^w+1\right)^r-1\right]\sum_{d_i\geq 0}\left|\beta_{d_1,\cdots,d_r}\right|\Big).$$

This is to say that for any given approximation error bound ε, n^* can be computed by the following formula:

$$n^*=\frac{c^{\max}}{\varepsilon}\left\{\left|\beta_{1,0\ldots,0}+\beta_{0,1\ldots,0}+\cdots+\beta_{0,\ldots,1}\right| + \left[\left(d\cdot C_d^w+1\right)^r-1\right]\sum_{d_i\geq 0}\left|\beta_{d_1,\cdots,d_r}\right|\right\} \tag{9}$$

If the calculated n^* is not an integer, we choose the smallest integer that is larger than the calculated n^* as n^*. □

We are now ready to prove that the general T2 TS fuzzy systems are function approximators based on Weierstrass approximation theorem [1]. In essence, the theorem states that for any multivariate function $G(\mathbf{x})$ that is continuous on a compact domain, there corresponds a polynomial $P_d(\mathbf{x})$ such that $\|\mathrm{P}_d-G\|<\varepsilon$ on the domain.

Theorem 2 (general interval T2 TS fuzzy system approximation theorem). *A T2 TS fuzzy system of the general class can uniformly approximate any multivariate function $G(\mathbf{x})$ that is continuous in $C^r[-1,\ 1]$ to any degree of accuracy. In other words, $\forall\varepsilon>0$, there exists a positive integer n^* computed by (9) such that $\forall n>n^*$,*

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

Proof. According to Weierstrass approximation theorem, for any given function $G(\mathbf{x})$ which is continuous in $C^r[-1,\ 1]$, there always exits a polynomial $P_d(\mathbf{x})$ capable of uniformly approximating $G(\mathbf{x})$ to arbitrary accuracy. That is, $\forall\varepsilon_1>0$, $\|P_d-G\|<\varepsilon_1$. Further, according to Theorem 1, $\forall\varepsilon_2>0$, we can always use (9) to find a positive integer n^* such that $\forall n>n^*$, $\|F_n-P_d\|<\varepsilon_2$. Therefore,

$$\|F_n-G\|\leq\|P_d-G\|+\|F_n-P_d\|<\varepsilon_1+\varepsilon_2<\varepsilon$$

meaning $F_n(\mathbf{x})$ can uniformly approximate $G(\mathbf{x})$. □

We would like to stress that n^* computed using (9) represents a sufficient condition, not a necessary one. As such, it can be conservative. In other words, n^* can potentially be a large number. Very importantly though, the reader needs to understand that such a seemingly large number guarantees to achieve the required multivariate function approximation uniformly regardless of the shape of the T2 input fuzzy sets, and the type of the fuzzy T-norm (recall that the process of formulating the interval T2 TS fuzzy systems whose final input-output expression is (3) never requires specifying these component choices. We only need to assume their outcomes). Another important factor is that (9) is obtained after derivations involving a series of inequalities, which may not be very "tight".

From practical implementation standpoint, the smallest n^* is always desirable. In (9), $\beta_{d_1,\ldots,d_r}$ are determined by $G(\mathbf{x})$ and we have no control over their magnitudes. Consequently, one way to obtain a smaller n^* is to use the smallest possible $c^{\max}$, which is 1 (i.e., an input fuzzy set only intersects with its immediately adjacent right and left neighbors).

It should be pointed out that $G(\mathbf{x})$ is unavailable in real-world problems. In a typical control problem, $G(\mathbf{x})$ represents a control solution, while in a modeling problem $G(\mathbf{x})$ represents the best mathematical representation of the system to be modeled. In both situations, one's task is to realize $G(\mathbf{x})$ or approximate it with enough accuracy through heuristically manipulating different components of the T2 fuzzy systems.

4. Illustrative numerical example

Example. What are the sufficiently large numbers of the input fuzzy sets and fuzzy rules for the general T2 TS fuzzy systems to uniformly approximate $G(x_1, x_2) = e^{x_1+x_2}$, where $x_1, x_2 \in [-1,\ 1]$, with approximation error less than 0.6?

Solution. Because $G(x_1, x_2)$ is defined in $[-1,\ 1] \times [-1,\ 1]$ already, no variable scaling is needed (that is, $a_1 = a_2 = 1$). $G(x_1, x_2)$ can be approximated uniformly by the following third-order polynomials:

$$P_3(x_1, x_2) = \frac{191}{192} + x_1 + x_2 + \frac{13}{24}x_1^2 + \frac{13}{24}x_2^2 + \frac{13}{12}x_1x_2 + \frac{1}{6}x_1^3 + \frac{1}{6}x_2^3 + \frac{1}{2}x_1^2x_2 + \frac{1}{2}x_1x_2^2$$

with a truncation error slightly less than 0.071 (i.e., $\varepsilon_1 = 0.071$). Note that $r = 2$, $d = 3$, $w = 1$ (or 2), $\beta_{0,0} = \frac{191}{192}$, $\beta_{1,0} = 1$, $\beta_{0,1} = 1$, $\beta_{2,0} = \frac{13}{24}$, $\beta_{0,2} = \frac{13}{24}$, $\beta_{1,1} = \frac{13}{12}$, $\beta_{3,0} = \frac{1}{6}$, $\beta_{0,3} = \frac{1}{6}$, $\beta_{2,1} = \frac{1}{2}$, and $\beta_{1,2} = \frac{1}{2}$. Obviously, $\varepsilon_2 = 1 - \varepsilon_1 = 0.529$. To minimize the numbers of fuzzy sets and rules, we choose $c^{\max} = 1$. Based on (9), the calculated $n^* = 1,219.3$. Thus n^* should be 1,220. That translates to 2,441 (i.e., $2n^* + 1$) input fuzzy sets and 5,958,481 (i.e., $(2n^* + 1)^2$) fuzzy rules.

These requirements appear to be large. However, the reader should keep in mind that they represent a sufficient condition, not a necessary one. The numbers are necessarily big because they hold regardless of the shapes of the input fuzzy sets and types of fuzzy logic AND operators (also different types of AND operators can be used in the same rule or different rules). The numbers actually needed for the approximation should be dramatically less than the numbers computed here for any given fuzzy system when its components are concretely specified. A compounding reason is that formula (10) is derived using a series of inequalities, leading to a conservative estimation of n^*.

5. Conclusion

We have constructively proved that a general class of the interval T2 TS fuzzy systems can uniformly approximate any multivariate continuous function on a compact domain to arbitrary accuracy. We have established a sufficient condition (i.e., (9)) to calculate the numbers of fuzzy sets and fuzzy rules required for any given approximation problem. This is the first and only quantitative result in the area of T2 fuzzy systems as universal approximators. The theoretical significance of our new results should be obvious (e.g., a much-needed theoretical foundation for developing mathematically rigorous T2 adaptive fuzzy control).

A less conservative sufficient condition can possibly be achieved if the fuzzy sets are limited to certain type (e.g., the trapezoidal type only) and only one particular kind of AND operator (e.g., the product AND operator) is allowed for the rules. Such an approximation condition would be less general but more practical. Another interesting open question is the theoretical comparison of a T2 TS fuzzy system with a comparable T1 TS fuzzy system (or a T2 system class vs. a T1 system class). In theory, a T2 system has the potential to deal with input and/or output system uncertainties more intuitively and comprehensively via the footprint of uncertainty. How does this enhanced capability impact the numbers of fuzzy sets and rules needed? The general class of the T1 TS fuzzy systems with linear rule consequent that we studied in [32] are somewhat comparable to the T2 systems in the present paper as far as the configuration is concerned. But comparing the sufficient condition established in [32] with (9) does not appear meaningful partially due to the series of the inequalities used in the derivations of the both sufficient conditions in [32] and in (10). A better study design seems to be necessary for a fair and direct comparison. This would not be easy and is left as an interesting future research project.

References

[1] I.N. Bronshtein and K.A. Semendyayev, *Handbook of Mathematics*, Van Nostrand Reinhold Company, 1985, p. 693.

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

[3] J.J. Buckley, Sugeno type controllers are universal controllers, *Fuzzy Sets and Systems* **53**(2) (1993), 299–303.

[4] J.L. Castro, Fuzzy logic controllers are universal approximators, *IEEE Trans Syst, Man, Cybern* **25**(4) (1995), 629–635.

[5] Y.-S. Ding, H. Ying and S.-H. Shao, Necessary conditions on minimal system configuration for general MISO Mamdani fuzzy systems as universal approximators, *IEEE Trans Syst, Man, Cybern A* **30**(6) (2000), 857–864.

[6] R. Hassine, F. Karrayb, A.M. Alimic and M. Selmid, Approximation properties of fuzzy systems for smooth functions and their first-order derivative, *IEEE Trans Syst, Man, Cybern A, Syst Humans* **33**(2) (2003), 160–168.

[7] R. Hassine, F. Karrayb, A.M. Alimic and M. Selmid, Approximation properties of piece-wise parabolic functions fuzzy logic systems, *Fuzzy Sets and Systems* **157** (2006), 501–515.
[8] O. Huwendiek and W. Brockmann, Function approximation with decomposed fuzzy systems, *Fuzzy Sets and Systems* **101** (1999), 273–296.
[9] E.P. Klement, L.T. Koczy and B. Moser, Are fuzzy systems universal approximators, *International Journal of General Systems* **28**(2–3) (1999), 259–282.
[10] G. Klir and B. Yuan, Fuzzy sets and fuzzy logic: Theory and applications, Upper Saddle River, NJ: Prentice Hall; 1995.
[11] B. Kosko, Fuzzy systems as universal approximators, *Proceedings of IEEE Int Conf Fuzzy Systems*, San Diego, CA, 1992, pp. 1153–1162.
[12] B. Kosko, Fuzzy systems as universal approximators, *IEEE Trans Comput* **43** (1994), 1329–1333.
[13] M. Landajo, M.J. Rio and R. Perez, A note on smooth approximation capabilities of fuzzy systems, *IEEE Trans Fuzzy Syst* **9**(2) (2001), 229–236.
[14] D.-C. Li, Z.-K. Shi and Y.-M. Li, Sufficient and necessary conditions for Boolean fuzzy systems as universal approximators, *Information Sciences* **178** (2008), 414–424.
[15] Q. Liang and J.M. Mendel, Interval type-2 fuzzy logic systems: Theory and design, *IEEE Trans Fuzzy Syst* **8**(5) (2000), 535–550.
[16] P.-Y. Liu, Mamdani fuzzy system: Universal approximator to a class of random processes, *IEEE Transactions on Fuzzy Systems* **10**(6) (2002).
[17] P.-Y. Liu and H.-X. Li, Analyzes for Lp (mu)-norm approximation capability of generalized Mamdani fuzzy systems, *Inform Sci* **138**(3–4) (2001), 195–210.
[18] P.-Y. Liu and H.-X. Li, Approximation of stochastic processes by T–S fuzzy systems, *Fuzzy Sets and Systems* **155** (2005), 215–235.
[19] Q. Luo, W. Yang and D.-Y. Yi, Kernel shapes of fuzzy sets in fuzzy systems for function approximation, *Information Sciences* **178** (2008), 836–857.
[20] J.M. Mendel, Type-2 fuzzy sets and systems: An overview, *IEEE Computational Intelligence Magazine* **2**(2) (2007), 20–29.
[21] J. Mendel, H. Hagras, W.-W. Tan, W.W. Melek and H. Ying, Introduction to Type-2 fuzzy logic control: Theory and applications, John Wiley & Sons, 2014.
[22] J.M. Mendel, R.I. John and F. Liu, Interval Type-2 fuzzy logic systems made simple, *IEEE Transactions on Fuzzy Systems* **14**(6) (2006), 808–821.
[23] S. Mitaim and B. Kosko, The shape of fuzzy sets in adaptive function approximation, *IEEE Transactions on Fuzzy Systems* **9**(4) (2001), 637–656.
[24] I. Perfilieva and V. Kreinovich, A New Universal approximation result for fuzzy systems, which reflects CNF–DNF duality, *International Journal of Intelligent Systems* **17** (2002), 1121–1130.
[25] R. Rovatti, Fuzzy piecewise multilinear and piecewise linear systems as universal approximators in Sobolev norms, *IEEE Transactions on Fuzzy Systems* **6**(2) (1998).
[26] D. Tikk, L.T. Koczy and T.D. Gedeon, A survey on the universal approximation and its limits in soft computing techniques, *International Journal of Approximate Reasoning* **33**(2) (2003), 185–202.
[27] L.-X. Wang, Fuzzy systems are universal approximators, *Proceedings of IEEE Int Conf Fuzzy Systems*, San Diego, CA, 1992, pp. 1163–1170.
[28] L.-X. Wang and J.M. Mendel, Fuzzy basis functions, universal approximation, and orthogonal least-squares learning, *IEEE Transactions on Neural Networks* **3** (1992), 807–814.
[29] R.R. Yager and V. Kreinovich, Universal approximation theorem for uninorm-based fuzzy systems modeling, *Fuzzy Sets and Systems* **140** (2003), 331–339.
[30] H. Ying, Sufficient conditions on general fuzzy systems as function approximators, *Automatica* **30** (1994), 521–525.
[31] H. Ying and G.-R. Chen, Necessary conditions for some typical fuzzy systems as universal approximators, *Automatica* **33** (1997), 1333–1338.
[32] H. Ying, Sufficient condition on uniform approximation of multivariate functions by Takagi-Sugeno fuzzy systems with linear rule consequent, *IEEE Transactions on Systems, Man and Cybernetics – Part A: Systems and Human* **28** (1998), 515–520.
[33] H. Ying, General Takagi-Sugeno fuzzy systems with simplified linear rule consequent are universal controllers, models and filters, *Information Sciences* **108** (1998), 91–107.
[34] H. Ying, *Fuzzy Control and Modeling: Analytical Foundations and Applications*, IEEE Press, 2000.
[35] H. Ying, General interval type-2 Mamdani fuzzy systems are universal approximators, *Proceedings of North American Fuzzy Information Processing Society Conference*, New York, NY, 2008.
[36] H. Ying, Interval type-2 Takagi-Sugeno fuzzy systems with linear rule consequent are universal approximators, *Proceedings of the 28th North American Fuzzy Information Processing Society Conference*, Cincinnati, Ohio, USA, 2009.
[37] H. Ying, Y.-S. Ding, S.-K. Li and S.-H. Shao, Comparison of necessary condition for typical Takagi-Sugeno and Mamdani fuzzy systems as universal approximators, *IEEE Transactions on Systems, Man and Cybernetics – Part A: Systems and Human* **29**(5) (1999), 508–514.
[38] L.A. Zadeh, The concept of a linguistic variable and its application to approximate reasoning, *Information Science* **8** (1975), 199–249.
[39] K. Zeng, N.Y. Zhang and W.L. Xu, A comparative study on sufficient conditions for Takagi–Sugeno fuzzy systems as universal approximators, *IEEE Transactions on Fuzzy Systems* **8**(6) (2000), 773–780.
[40] X.-J. Zeng and M.G. Singh, Approximation theory of fuzzy systems—MIMO case, *IEEE Transactions on Fuzzy Systems* **3**(2) (1995), 219–235.
[41] X.-J. Zeng and M.G. Singh, Approximation accuracy analysis of fuzzy systems as function approximators, *IEEE Transactions on Fuzzy Systems* **4**(1) (1996), 44–63.