Avoiding Exponential Parameter Growth in Fuzzy Systems

Mustafa K. Güven and Kevin M. Passino

***Abstract*—For standard fuzzy systems where the input membership functions are defined on a grid on the input space, and all possible combinations of rules are used, there is an exponential growth in the number of parameters of the fuzzy system as the number of input dimensions increases. This "curse of dimensionality" effect leads to problems with design of fuzzy controllers (e.g., how to tune all these parameters), training of fuzzy estimators (e.g., complexity of a gradient algorithm for training, and problems with "over parameterization" that lead to poor convergence properties), and with computational complexity in the implementation for practical problems. In this paper, we introduce a new fuzzy system whose number of parameters grows linearly depending upon the number of inputs, even though it is constructed by using all possible combinations of the membership functions in defining the rules. We prove that this new fuzzy system is equivalent to the standard fuzzy system as long as its parameters are specified in a certain way. Then, we show that it still holds the Universal Approximator Property by using the Stone–Weierstrass theorem. Finally, we illustrate the performance of the new fuzzy system via an application.**

***Index Terms*—Curse of dimensionality, number of parameters, standard fuzzy systems, Universal Approximation Property.**

I. INTRODUCTION

THE NUMBER of parameters is one of the main concerns for fuzzy systems, especially when it is desired to increase the number of inputs and rules, since for the standard fuzzy system the number of parameters increases exponentially when the number of inputs or rules are increased, and computational complexity increases accordingly. For instance, for a fuzzy system that has five inputs and five membership functions on each universe of discourse, the number of parameters in it will be 34 375 when it is assumed that all possible combinations of the membership functions are used for defining rules.

Different approaches have been proposed to solve the rule explosion problem. In the earlier works, rule reduction in fuzzy systems has been attempted via genetic algorithms, neural networks, and a variety of clustering techniques [1]–[3], in an effort to select only those rules that contribute the most to the inference outcome. Another approach is presented in [4]. In this paper, the authors attempt to eliminate the curse of dimensionality by providing a disjunctive form of the conjunctive rule in fuzzy IF–THEN rules. Another approach to eliminate the rule explosion problem is to introduce the hierarchical fuzzy system configuration in [5] and [6]. In this approach, instead of using a single fuzzy system with a high-dimensional input, a number of lower-dimensional fuzzy systems are linked in an hierarchical manner. While these methods are able to significantly reduce the number of rules, they do not address the exponential growth in parameters in fuzzy systems when the number of inputs or the number of rules are increased.

To avoid this problem, in Section II, we introduce a new fuzzy system whose number of parameters grows linearly depending upon the number of inputs. This fuzzy system is constructed by using all possible combinations of the membership functions and we prove that this new fuzzy system is in a certain sense equivalent to the standard fuzzy system. In Section III, we show that it still holds the Universal Approximator Property by using the Stone–Weierstrass theorem. In Section IV, we present the simulation results of a fuzzy estimator for a transformer's behavior during "inrush" by using the fuzzy system.

II. FUZZY SYSTEMS AND THE SIZE OF THEIR PARAMETERS

One way to define a fuzzy system is to let

$$y = f(x|\theta) = \frac{\sum_{i=1}^{R} b_i \mu_i(x)}{\sum_{i=1}^{R} \mu_i(x)} \tag{1}$$

where $x = [x_1, x_2, \ldots, x_n]^\top$ is the input, y is the output, b_i, $i = 1, 2, \ldots, R$ are the centers of the output membership functions and $\mu_i(x)$ is the certainty of the premise of the ith rule. Here, θ in (1) is an M_f-dimensional vector (M_f is the number of parameters in (1)) that holds the parameters of the fuzzy system (i.e., the b_i, c_j^i, and σ_j^i defined later). Suppose we use singleton fuzzification, product to represent the premises, and center-average defuzzification. Furthermore, we will use input membership functions that are Gaussian and of the form

$$\mu_j^i(x) = \exp\left(-\left(\frac{x_i - c_j^i}{\sigma_j^i}\right)^2\right)$$

so that

$$\mu_i(x) = \prod_{j=1}^{n} \mu_j^i(x) \tag{2}$$

where c_j^i (σ_j^i) is the center (relative width) of the membership function of the ith rule for jth universe of discourse.

When defining a fuzzy system, one can either use some of the input membership functions or all possible combinations of the input membership functions to construct the rule base. For the former case, the number of parameters may not increase exponentially by increasing the number of inputs or the number of rules. However, for latter case, the number of parameters M_f grows exponentially by the growth in the number of inputs or number of rules.

Manuscript received October 29, 1999; revised May 19, 2000.
The authors are with the Department of Electrical Engineering, The Ohio State University, Columbus, OH 43210 USA.
Publisher Item Identifier S 1063-6706(01)01358-3.

For (1), the number of parameters which can be tuned is

$$M_f = (2n+1)R \tag{3}$$

where n is the number of inputs, R is the number of rules, nR of M_f are the c_j^i, nR of M_f are the σ_j^i, and R of M_f are the b_i, where $i = 1, 2, \ldots, R$ and $j = 1, 2, \ldots, n$. Assume that the jth universe of discourse has N_j membership functions, where $j = 1, 2, \ldots, n$, and the rules result from all possible combinations of these membership functions. This is usual case in fuzzy control applications [7]. In this case the number of rules can be represented as

$$R = \prod_{j=1}^{n} N_j$$

so that using (3), the total number becomes

$$M_f = (2n+1) \prod_{j=1}^{n} N_j.$$

Clearly, for either large n or N_j, M_f can be very large and there is an exponential increase in the number of parameters for additional inputs leading to the "curse of dimensionality." Here we will focus on how to reduce the number of parameters needed to define a fuzzy system.

First, note that in many practical applications (e.g., most control applications [7]) we *first* define the membership functions on the input and the output universes of discourse and then define the rules based on these membership functions. Define $\overline{c}_j^i$ to be the center of the ith input membership function on the jth universe of discourse where $j = 1, 2, \ldots, n$, $i = 1, 2, \ldots, N_j$. Note that this notation is not to be confused with the c_j^i above. In particular, for $\overline{c}_j^i$, $i = 1, 2, \ldots, N_j$ there are N_j centers on the jth universe of discourse, but for c_j^i, there is a center on the jth universe of discourse for *each* rule, $i = 1, 2, \ldots, R$. Similarly, we define $\overline{\sigma}_j^i$ to be the spread of the ith membership function on the jth universe of discourse, where $i = 1, 2, \ldots, N_j$, $j = 1, 2, \ldots, n$. Using these definitions we introduce a function (fuzzy system)

$$\overline{f}(x|\overline{\theta}) = \frac{\prod_{j=1}^{n}\left(\sum_{i=1}^{N_j} \overline{b}_j^i \overline{\mu}_j^i(x)\right)}{\prod_{j=1}^{n}\left(\sum_{i=1}^{N_j} \overline{\mu}_j^i(x)\right)} \tag{4}$$

where

$$\overline{\theta} = \left[\overline{b}_1, \overline{b}_2, \ldots, \overline{b}_n, \overline{c}_1, \overline{c}_2, \ldots, \overline{c}_n, \overline{\sigma}_1, \overline{\sigma}_2, \ldots, \overline{\sigma}_n\right]$$

such that

$$\overline{b}_j = \left[\overline{b}_j^1, \overline{b}_j^2, \ldots, \overline{b}_j^{N_j}\right]$$

$$\overline{c}_j = \left[\overline{c}_j^1, \overline{c}_j^2, \ldots, \overline{c}_j^{N_j}\right]$$

and

$$\overline{\sigma}_j = \left[\overline{\sigma}_j^1, \overline{\sigma}_j^2, \ldots, \overline{\sigma}_j^{N_j}\right]$$

where $j = 1, 2, \cdots, n$, and

$$\overline{\mu}_j^i(x) = \exp\left(-\left(\frac{x_j - \overline{c}_j^i}{\overline{\sigma}_j^i}\right)^2\right).$$

Here, $\overline{\mu}_j^i$ is the ith membership function on the jth universe of discourse. The parameters $\overline{b}_j^i$ will be used later to specify the output membership function centers b_j^i, which are defined the same as the output membership function centers in standard fuzzy systems. Note that with this approach output membership function centers will be able to be specified after the identification process through the $\overline{b}_j^i$ as in (6). Obviously, the consequent part of the fuzzy IF–THEN rules cannot be defined before the identification process, and this might make the use of this structure limited in the sense of standard fuzzy controller design.

Next, we will provide conditions under which (1) and (4) are equivalent.

Theorem 1: Assume that N_j is the number of membership functions on the jth universe of discourse where $j = 1, 2, \ldots, n$ and for $f(x|\theta)$ rules are constructed by all possible combinations of membership functions on the universes of discourses. In this case,

$$f(x|\theta) = \overline{f}\left(x|\overline{\theta}\right)$$

if we specify $\overline{\mu}_j^i$ and $\overline{b}_j^i$ properly.

Proof: We will first show that if $\overline{f}(x|\overline{\theta})$ is given, an equivalent $f(x|\theta)$ can be defined, and second, if $f(x|\theta)$ is given, $\overline{f}(x|\overline{\theta})$ can be defined. Assume that we have

$$\overline{f}(x|\overline{\theta}) = \frac{\prod_{j=1}^{n}\left(\sum_{i=1}^{N_j} \overline{b}_j^i \overline{\mu}_j^i\right)}{\prod_{j=1}^{n}\left(\sum_{i=1}^{N_j} \overline{\mu}_j^i(x)\right)}.$$

Note that as shown in the equation at the bottom of the next page. Assume that μ_i is the certainty of the premise of ith rule for $f(x|\theta)$, so

$$\mu_i = \prod_{j=1}^{n} \mu_j^i$$

and b_i is the center of output membership function of the ith rule. Since we assume that the rules are constructed such that the rule base contains all possible combinations of the membership functions in the universes of discourse, we can define the certainty of the premise of the rules as

$$\begin{aligned}
\mu_1 &= \overline{\mu}_1^1 \ldots \overline{\mu}_n^1 \\
\mu_2 &= \overline{\mu}_1^1 \ldots \overline{\mu}_{n-1}^1 \overline{\mu}_n^2 \\
\mu_3 &= \overline{\mu}_1^1 \ldots \overline{\mu}_{n-1}^2 \overline{\mu}_n^1 \\
&\vdots \quad \vdots \quad \vdots \\
\mu_R &= \overline{\mu}_1^{N_1} \ldots \overline{\mu}_{n-1}^{N_{n-1}} \overline{\mu}_n^{N_n}
\end{aligned} \tag{5}$$

and

$$\begin{aligned} b_1 &= \overline{b}_1^1 \dots \overline{b}_n^1 \\ b_2 &= \overline{b}_1^1 \dots \overline{b}_{n-1}^1 \overline{b}_n^2 \\ b_3 &= \overline{b}_1^1 \dots \overline{b}_{n-1}^2 \overline{b}_n^1 \\ &\vdots \; \vdots \; \vdots \\ b_R &= \overline{b}_1^{N_1} \dots \overline{b}_{n-1}^{N_{n-1}} \overline{b}_n^{N_n} \end{aligned} \tag{6}$$

where

$$R = \prod_{j=1}^{n} N_j.$$

So with these definitions $\overline{f}$ becomes

$$\overline{f}(x|\overline{\theta}) = \frac{b_1\mu_1 + b_2\mu_2 + \cdots + b_R\mu_R}{\mu_1 + \mu_2 + \cdots + \mu_R} = \frac{\sum_{i=1}^{R} b_i\mu_i}{\sum_{i=1}^{R} \mu_i}.$$

Hence,

$$\overline{f}(x|\overline{\theta}) = \frac{\sum_{i=1}^{R} b_i \prod_{j=1}^{n} \mu_j^i}{\sum_{i=1}^{R} \prod_{j=1}^{n} \mu_j^i} = f(x|\theta).$$

To show that given $f(x|\theta)$ we can define $\overline{f}(x|\overline{\theta})$, we simply use (5) and (6) in a similar manner. This process is actually involves finding the equivalent polynomial in the form of multiplications of summations for the one in the summations of multiplications. Since the standard fuzzy system is in the form of the latter, the transition from the standard fuzzy system back to the new one merely involves some algebra and taking into the consideration the new definitions for the input membership function centers $\overline{c}_j^i$ and spreads $\overline{\sigma}_j^i$. ■

Note that the number of parameters, introduced by the new fuzzy system, $\overline{f}$, is

$$M_{\overline{f}} = 3\sum_{j=1}^{n} N_j$$

since the number of each $\overline{b}_j^i$, $\overline{c}_j^i$, and $\overline{\sigma}_j^i$ are $\sum_{j=1}^{n} N_j$, even though $\overline{f}$ is formed from a fuzzy system that uses all possible combinations of the membership functions as rules. Clearly, $M_{\overline{f}}$ and M_f satisfy

$$\begin{aligned} M_{\overline{f}} &= M_f \quad \text{if } n = 1 \\ M_{\overline{f}} &< M_f \quad \text{if } n > 1. \end{aligned}$$

Also, it is clear that $M_{\overline{f}}$ will only grow by $3N_j$ with each unit increase in n.

III. Meeting the Universal Approximation Property with Fewer Parameters

The question is, however, whether by reducing the number of parameters we have reduced the representational capability of $\overline{f}(x|\overline{\theta})$ to be less than $f(x|\theta)$. We will answer this question by investigating whether $\overline{f}$ has the universal approximation property using *Universal Approximation Theorem*. One proof of Universal Approximation Theorem is based on the Stone–Weierstrass Theorem.

Theorem 2: The fuzzy system $\overline{f}$ satisfies the Stone–Weierstrass theorem and hence it possesses the universal approximation property (as does f).

Proof: In [8], it is shown that the fuzzy system f satisfies the "universal approximation property"; hence it can approximate any continuous function on a compact set with an arbitrary degree of accuracy (if we allow for an arbitrary number of parameters in $\overline{\theta}$). Next, we show that $\overline{f}(x|\overline{\theta})$ also satisfies the universal approximation property so we then know that even though $\overline{f}(x|\overline{\theta})$ is described with fewer parameters, it has the same basic representation capabilities.

Let U be a compact set, and Y be the set of all fuzzy systems in the form of $\overline{f}$. We now show that Y is an algebra, Y separates points on U, and Y vanishes at no point of U.

Let $\overline{f}_1$, $\overline{f}_1 \in Y$ so that we can write them as

$$\overline{f}_1 = \frac{\prod_{j=1}^{n}\left(\sum_{i_1=1}^{N1_j} \overline{b1}_j^{i_1}\,\overline{\mu1}_j^{i_1}\right)}{\prod_{j=1}^{n}\left(\sum_{i_1=1}^{N1_j} \overline{\mu1}_j^{i_1}\right)}$$

$$\overline{f}_2 = \frac{\prod_{j=1}^{n}\left(\sum_{i_2=1}^{N2_j} \overline{b2}_j^{i_2}\,\overline{\mu2}_j^{i_2}\right)}{\prod_{j=1}^{n}\left(\sum_{i_2=1}^{N2_j} \overline{\mu2}_j^{i_2}\right)}$$

where n is the number of inputs, $N1_j$ and $N2_j$ are the number of membership functions for each universes of discourse, where

---

$$\begin{aligned} \overline{f}(x|\overline{\theta}) &= \frac{\left(\overline{b}_1^1\overline{\mu}_1^1 + \cdots + \overline{b}_1^{N_1}\overline{\mu}_1^{N_1}\right)\left(\overline{b}_2^1\overline{\mu}_2^1 + \cdots + \overline{b}_2^{N_2}\overline{\mu}_2^{N_2}\right)\dots\left(\overline{b}_n^1\overline{\mu}_n^1 + \cdots + \overline{b}_n^{N_n}\overline{\mu}_n^{N_n}\right)}{\left(\overline{\mu}_1^1 + \cdots + \overline{\mu}_1^{N_1}\right)\dots\left(\overline{\mu}_n^1 + \cdots + \overline{\mu}_n^{N_n}\right)} \\ &= \frac{\overline{b}_1^1\dots\overline{b}_n^1\overline{\mu}_1^1\dots\overline{\mu}_n^1 + \overline{b}_1^1\dots\overline{b}_{n-1}^1\overline{b}_n^2\overline{\mu}_1^1\dots\overline{\mu}_{n-1}^1\overline{\mu}_n^2 + \cdots + \overline{b}_1^{N1}\dots\overline{b}_n^{N_n}\overline{\mu}_1^{N_1}\dots\overline{\mu}_n^{N_n}}{\left(\overline{\mu}_1^1\dots\overline{\mu}_n^1\right) + \left(\overline{\mu}_1^1\dots\overline{\mu}_{n-1}^1\overline{\mu}_n^2\right) + \cdots + \left(\overline{\mu}_1^{N_1}\dots\overline{\mu}_n^{N_n}\right)} \end{aligned}$$

$j = 1, 2, \ldots, n$, $\overline{b1}_j^{i_1}$, $\overline{b2}_j^{i_2}$, $\overline{\mu 1}_j^{i_1}$, and $\overline{\mu 2}_j^{i_2}$ are the parameters of the new fuzzy system which are defined like $\overline{b}_j^i$ and $\overline{\mu}_j^i$ in (4).

Note that as shown in the equation at the bottom of the page. Since

$$\overline{b1}_j^{i_1}\overline{\mu 1}_j^{i_1}\overline{\mu 2}_j^{i_2} + \overline{b2}_j^{i_2}\overline{\mu 1}_j^{i_1}\overline{\mu 2}_j^{i_2} = \left(\overline{b1}_j^{i_1} + \overline{b2}_j^{i_2}\right)\left(\overline{\mu 1}_j^{i_1}\overline{\mu 2}_j^{i_2}\right)$$

$$\overline{b1}_j^{i_1} + \overline{b2}_j^{i_2} = \overline{b}_j^i$$

and

$$\overline{\mu 1}_j^{i_1}\overline{\mu 2}_j^{i_2} = \overline{\mu}_j^i$$

we have

$$\overline{f}_1 + \overline{f}_2 \in Y.$$

Next, we show that

$$\overline{f}_1 \cdot \overline{f}_2 = \frac{\prod_{j=1}^{n}\left[\left(\sum_{i_1=1}^{N1_j}\overline{b1}_j^{i_1}\overline{\mu 1}_j^{i_1}\right)\left(\sum_{i_2=1}^{N2_j}\overline{b2}_j^{i_2}\overline{\mu 2}_j^{i_2}\right)\right]}{\prod_{j=1}^{n}\left[\sum_{i_1=1}^{N1_j}\sum_{i_2=1}^{N2_j}\left(\overline{\mu 1}_j^{i_1}\overline{\mu 2}_j^{i_2}\right)\right]}$$

$$= \frac{\prod_{j=1}^{n}\left[\sum_{i_1=1}^{N1_j}\sum_{i_2=1}^{N2_j}\left(\overline{b1}_j^{i_1}\overline{b2}_j^{i_2}\overline{\mu 1}_j^{i_1}\overline{\mu 2}_j^{i_2}\right)\right]}{\prod_{j=1}^{n}\left[\sum_{i_1=1}^{N1_j}\sum_{i_2=1}^{N2_j}\left(\overline{\mu 1}_j^{i_1}\overline{\mu 2}_j^{i_2}\right)\right]}.$$

Since $\overline{b1}_j^{i_1}\overline{b2}_j^{i_2}$ can be represented in the form of $\overline{b}_j^i$, and $\overline{\mu 1}_j^{i_1}\overline{\mu 2}_j^{i_2}$ can be represented in the form of $\overline{\mu}_j^i$,

$$\overline{f}_1\overline{f}_2 \in Y.$$

For $c \in \Re$,

$$c\overline{f} = \frac{\prod_{j=1}^{n}\left(\sum_{i=1}^{N_j} c\overline{b}_j^i\overline{\mu}_j^i\right)}{\prod_{j=1}^{n}\left(\sum_{i=1}^{N_j}\overline{\mu}_i(x)\right)}.$$

Since $c\overline{b}_j^i$ can be represented in the form of $\overline{b}_j^i$, $c\overline{f} \in Y$. Hence, Y is an algebra.

We now show that Y separates points on U by constructing a required fuzzy system $\overline{f}(x|\overline{\theta})$. Let $x_1^\circ, x_2^\circ \in U$ be two arbitrary points, $x_1^\circ \neq x_2^\circ$. We then choose the parameters of $\overline{f}(x|\overline{\theta})$ as follows: $N_1 = 1$ and $N_j = 1$ where $j = 2, 3, \ldots, n$, $\overline{b}_1^2 = 0$ and $\overline{b}_j^1 = 1$ where $j = 2, 3, \ldots, n$, $\overline{c}_1^1 = x_1^\circ$, $\overline{c}_1^2 = x_2^\circ$ and $\overline{c}_j^i = x_1^\circ$ where $j = 2, 3, \ldots, n$ and $i = 1, 2, \ldots, N_j$, and $\overline{\sigma}_j^i = 1$ where $j = 1, 2, \ldots, n$ and $i = 1, 2, \ldots, N_j$.

For $\overline{f}(x_1^\circ)$,

$$\overline{\mu}_1^1(x_1^\circ) = \exp\left(-\left(\frac{x_1^\circ - x_1^\circ}{1}\right)^2\right) = 1$$

$$\overline{\mu}_1^2(x_1^\circ) = \exp\left(-\left(\frac{x_1^\circ - x_2^\circ}{1}\right)^2\right)$$

$$\overline{\mu}_1^j(x_1^\circ) = 1$$

where $j = 2, 3, \ldots, n$. So

$$\overline{f}(x_1^\circ) = \frac{1}{1 + \exp -(x_1^\circ - x_2^\circ)^2}.$$

$$\overline{f}_1 + \overline{f}_2 = \frac{\prod_{j=1}^{n}\left(\sum_{i_1=1}^{N1_j}\overline{b1}_j^{i_1}\overline{\mu 1}_j^{i_1}\right)}{\prod_{j=1}^{n}\left(\sum_{i_1=1}^{N1_j}\overline{\mu 1}_j^{i_1}\right)} + \frac{\prod_{j=1}^{n}\left(\sum_{i_2=1}^{N2_j}\overline{b2}_j^{i_2}\overline{\mu 2}_j^{i_2}\right)}{\prod_{j=1}^{n}\left(\sum_{i_2=1}^{N2_j}\overline{\mu 2}_j^{i_2}\right)}$$

$$= \frac{\prod_{j=1}^{n}\left[\left(\sum_{i_1=1}^{N1_j}\overline{b1}_j^{i_1}\overline{\mu 1}_j^{i_1}\right)\left(\sum_{i_2=1}^{N2_j}\overline{\mu 2}_j^{i_2}\right) + \left(\sum_{i_2=1}^{N2_j}\overline{b2}_j^{i_2}\overline{\mu 2}_j^{i_2}\right)\left(\sum_{i_1=1}^{N1_j}\overline{\mu 1}_j^{i_1}\right)\right]}{\prod_{j=1}^{n}\left[\left(\sum_{i_1=1}^{N1_j}\overline{\mu 1}_j^{i_1}\right)\left(\sum_{i_2=1}^{N2_j}\overline{\mu 2}_j^{i_2}\right)\right]}$$

$$= \frac{\prod_{j=1}^{n}\left[\sum_{i_1=1}^{N1_j}\sum_{i_2=1}^{N2_j}\left(\overline{b1}_j^{i_1}\overline{\mu 1}_j^{i_1}\overline{\mu 2}_j^{i_2} + \overline{b2}_j^{i_2}\overline{\mu 1}_j^{i_1}\overline{\mu 2}_j^{i_2}\right)\right]}{\prod_{j=1}^{n}\left[\sum_{i_1=1}^{N1_j}\sum_{i_2=1}^{N2_j}\left(\overline{\mu 1}_j^{i_1}\overline{\mu 2}_j^{i_2}\right)\right]}$$

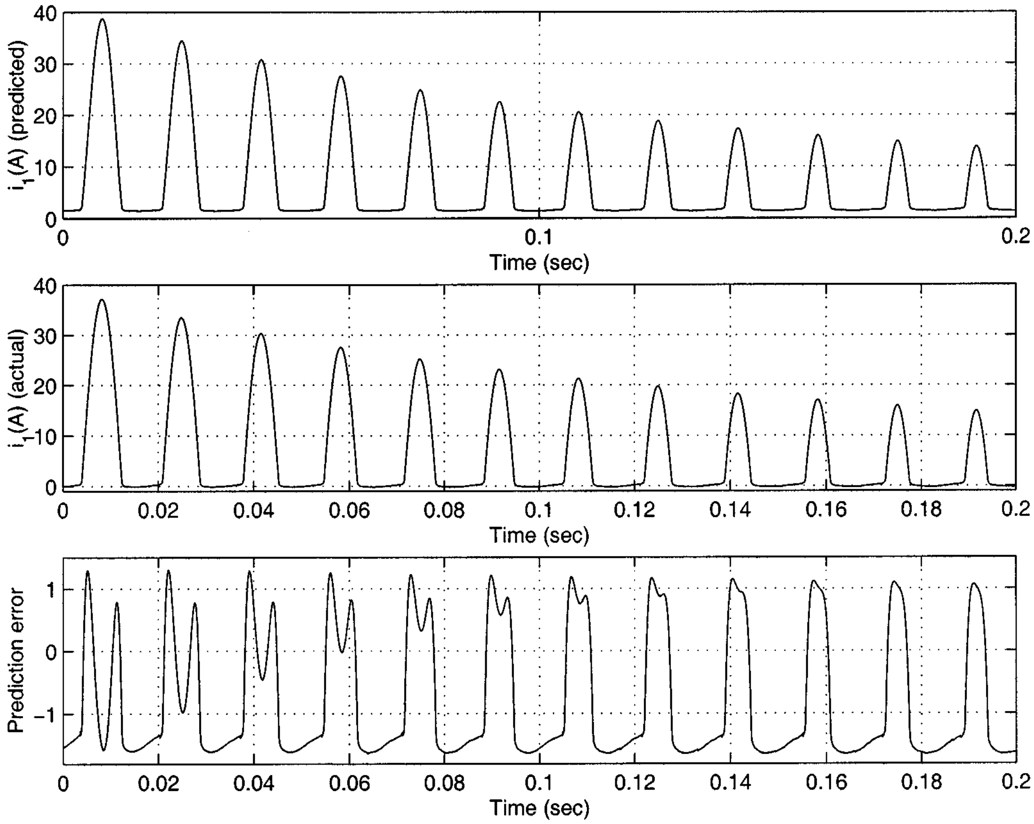

Fig. 1. Fuzzy system output (i_1 predicted), the test data Γ (i_1 actual) and the approximation error are plotted.

For $\overline{f}(x_2^\circ)$,

$$\begin{aligned}\overline{\mu}_1^1(x_2^\circ) &= \exp(-(x_2^\circ - x_1^\circ)^2)\\ \overline{\mu}_1^2(x_2^\circ) &= 1,\\ \overline{\mu}_2^1(x_2^\circ) &= \exp(-(x_2^\circ - x_1^\circ)^2)\\ &\vdots\\ \overline{\mu}_n^1(x_2^\circ) &= \exp(-(x_2^\circ - x_1^\circ)^2)\end{aligned}$$

and

$$\begin{aligned}\overline{f}(x_2^\circ) &= \frac{\exp(-n(x_2^\circ - x_1^\circ)^2)}{(1+\exp(-n(x_2^\circ - x_1^\circ)^2))\exp(-(n-1)(x_2^\circ - x_1^\circ)^2)}\\ &= \frac{\exp(-n(x_2^\circ - x_1^\circ)^2)}{1+\exp(-n(x_2^\circ - x_1^\circ)^2)}.\end{aligned}$$

Obviously, $\overline{f}(x_1^\circ) \neq \overline{f}(x_2^\circ)$. Hence, Y separates points on U.

Finally, to show that Y vanishes at no point of U, we simply observe that any fuzzy system $\overline{f}(x|\overline{\theta})$ with all $\overline{b}_j^i > 0$ has the property that $\overline{f}(x|\overline{\theta}) > 0$, $\forall x \in U$. Hence, Y vanishes at no point on U.

IV. Sample Simulation: Identification of the Transient Model of a Transformer

In this section, an inrush model of a transformer will be produced by using the new fuzzy system as defined earlier. Since the new fuzzy system is not linear in parameters, gradient method will be used for training. The update formulas of the gradient method for the new fuzzy system can be found easily by following the procedure for the standard fuzzy system [7] and therefore the tedious algebra is omitted here.

In this simulation, a 2-input–1-output fuzzy model of the inrush behavior of a transformer is produced by using the new fuzzy system. The inputs are chosen as the current primary and secondary currents, $i_1(k)$, $i_2(k)$ and the output is chosen as the next value of the primary current, $i_1(k+1)$. The transient behavior of the transformer is represented by a data set, a portion of which is used for training, which will be called G, and the remaining portion is used for testing, which will be called Γ. Note that the training and testing data sets contains different data points and the number of data in $|\Gamma|$ is greater than the one in $|G|$.

For each input universe of discourse 3 membership functions are used and the centers $\overline{c}_i^j$ and relative widths $\overline{\sigma}_i^j$ are defined as

$$\begin{aligned}\overline{c}_i^j &= \min(G) + (i-1)\,\frac{\max(G) - \min(G)}{2}\\ \overline{\sigma}_i^j &= (\max(G) - \min(G))/3\end{aligned}$$

for $i = 1, 2, 3$, where G is the training data set, max and min are maximum and minimum values of G, so that the membership functions are distributed uniformly on each universe of discourse.

The simulation results are given in Fig. 1. In this figure, first the predicted primary current $\hat{i}_1(k+1)$ is plotted. Second, the actual primary current $i_1(k+1)$, which is given in the test data set Γ, is plotted, and last the approximation error, $e(k+1)$ is plotted. As can be seen, a model for the transient behavior of a transformer can be produced by the new fuzzy system. It should be noted that it may be possible to have a better approximation by using different initial parameters, step sizes or training method (e.g., Levenberg–Marquardt). Also, noted that the purpose of the given application is to show that introduced fuzzy system can be trained. The purpose of this example is not to compare the performance of the two types of fuzzy systems.

V. Conclusion

The new fuzzy system defined earlier avoids exponential growth in parameters and, at the same time, still satisfies the Universal Approximation Property. The new fuzzy system in (4) becomes very useful if one would like to increase the number of inputs or rules; since the growth of the number of parameters is linear, one may be able to avoid problems with computational complexity. On the other hand, this new fuzzy system is not linear in parameters anymore. Therefore, for some applications, it may not be useful. For instance, linear least-squares methods cannot be used to train this fuzzy system. However, a steepest descent gradient method or the Levenberg–Marquardt method can be used to train it. So, for systems with a high number of inputs, this fuzzy system can be very useful. It should also be noted that the new fuzzy system (4) is not local in the same sense as the standard fuzzy system (1). When one of the $\bar{b}_j^i$ in (4) changes, then the centers of each membership in the output universe of discourse will change, which is obvious from (6). This new fuzzy system has been applied to the modeling of transformers for inrush behavior. In the process, a gradient method has been used and satisfactory results have been produced.

Acknowledgment

The authors would like to thank the reviewers for their useful remarks.

References

[1] S. Chiu, "Method and software for extracting fuzzy classification rules by substructive clustering," *N. Amer. Fuzzy Information Processing Society, Initiative Soft Computing IEEE Neural Networks Council—IEEE Syst., Man, Cybern. Soc.*, pp. 19–22, June 1996.

[2] S. Halmagamuge and M. Glesner, "Neural networks in designing fuzzy systems for real world applications," *Fuzzy Sets Syst.*, vol. 65, no. 1, pp. 1–12, July 1994.

[3] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka, "Selecting fuzzy if–then rules for classification using genetic algorithms," *Fuzzy Sets Syst.*, vol. 3, pp. 260–270, Aug. 1995.

[4] W. E. Combs and J. E. Andrews, "Combinatorial rule explosion eliminated by a fuzzy rule configuration," *IEEE Trans. Fuzzy Syst.*, vol. 6, pp. 1–11, Feb. 1998.

[5] G. V. S. Raju and R. A. Kisner, "Hierarchical fuzzy systems," *Int. J. Contr.*, vol. 54, no. 12, pp. 1201–1216, 1991.

[6] G. V. S. Raju and J. Zhou, "Adaptive hierarchical fuzzy controller," *IEEE Trans. Syst., Man, Cybern.*, vol. 23, no. 10, pp. 973–980, Oct. 1993.

[7] K. M. Passino and S. Yurkovich, *Fuzzy Control*. Menlo Park, CA: Addison Wesley Longman, 1998.

[8] L. X. Wang, *Fuzzy Systems and Control: Design and Stability Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1994.