

## A note on the rates of uniform approximation of fuzzy systems

Hoang Viet Long

*University of Transport and Communications,  
Lang Thuong, Dong Da, Hanoi, Vietnam  
E-mail: longhv08@gmail.com; longhvmath@uct.edu.vn*

Received 27 January 2011

Accepted 22 August 2011

### Abstract

For the fuzzy systems with the kernel-shaped fuzzy sets of if part, we estimate the rates of the uniform approximation for continuous functions. Results are given associatively with the rates of convergence of the sequence  $(\log k/k)^\alpha$ .

*Keywords:* fuzzy systems, kernel functions, universal approximation, uniform approximation rates.

### 1. Introduction

As an intelligent system of soft computing technique, fuzzy system (FS) is an efficient tool to deal with nonlinearly complex systems, in which there are linguistic information and data information, simultaneously. One most important advantage of using fuzzy systems is that linguistic fuzzy IF - THEN rules are naturally utilized in the systems. Linguistic fuzzy IF - THEN rules can be developed by human experts who are familiar with the process under consideration. Fuzzy systems have been developed to solve many different practical problems such as pattern recognition <sup>1</sup>, system identification <sup>2</sup>, adaptive control <sup>3,4,5,6</sup>, etc. Fuzzy systems offer a key advantage over traditional approaches, especially in fuzzy control systems. They offer model-free estimation of the control system. The user need not specify how the controller output mathematically depends on its input. Instead, the user provides a statistically representative set of numerical training samples. Even if a math-model controller is available, fuzzy systems

may prove more robustly and easily to modify. If experts provide structured knowledge of the control process or if sufficient numerical training samples are unavailable, the fuzzy approach may be preferable. We can use the numerical data to generate fuzzy rule base. Each fuzzy rule base defines a patch in the input-output state space, and the fuzzy system approximates the unknown function by covering its graph with fuzzy rule base patches.

In most of applications of the fuzzy systems, the main design objective can be transformed to find designed mapping from the input spaces to the output spaces, which may also be denoted as function. Thus, problems of designing fuzzy systems can be considered as approximation problems of functions. As an important research topic related to fuzzy systems, the approximation capability of fuzzy systems has been studied in theory and in application in the past few years (see <sup>7,8,9,10,11,12,13,14,15,16,17,18,19,20,21</sup> etc).

When those authors researching on the approximation accuracy theory of fuzzy systems, they deal

mainly with three questions: (1) universal approximation; (2) constructive approximation; (3) approximation rates. The first and the second questions are answered widely in the last few years (see 6,10,11,12,13,14,17,18). However, there is still a lack of the theoretical estimations of the approximation rates of fuzzy systems (see <sup>15</sup>). The approximation rates plays a key role in answering the question: what, if any, are the advantages of fuzzy systems as function approximators over the other methods such as the polynomial, spline, trigonometric, wavelets, neural networks (see <sup>22,23,24,25,26</sup>). So that, in the present paper, we prove that the fuzzy systems (with the kernel-shaped fuzzy sets of if part) have the universal approximation properties for any continuous function  $f$  defined on  $U \subset \mathbb{R}$ . Furthermore, we estimate the uniform approximation rates of the fuzzy system  $F_k$ , where the uniform approximation rates are defined as follows:

$$\|F_k(\circ) - f(\circ)\|_\infty = \sup_{x \in U} |F_k(x) - f(x)|.$$

One of the main features in the design of fuzzy systems for approximation functions is the flexibility in choosing their components such as membership functions, fuzzy inference engine, and defuzzification procedures, etc. Thus the problem of robustness of fuzzy systems arises. The choices of the input variables membership functions is quite important as it could affects substantially its output behavior. In this paper, the input membership functions of a fuzzy system are generated by the translations and scale transformations of a fixed kernel function. By using test functions methods in Fourier analysis, we explain how the shape of the input membership functions determines the approximating capability of the fuzzy systems. This choice has a number of advantages. One of them is that if the original function is intuitively meaningful, then the kernel constructed will be also intuitively meaningful. An other advantage is that the kernel can change their width and move on the real space to be able of representing different linguistic variables. But the main advantage of this construction is the flexibility and the ease of propagating properties of the original function to all other membership functions.

The paper is organized as follows. In section

2 some necessary notations and preliminaries are introduced. Fuzzy systems with kernel shapes of fuzzy sets are constructed in this section from training data. The universal approximation property of fuzzy system for a continuous function is proved in section 3. Our main results are given in section 4 with two methods to estimate the rates of approximation of fuzzy systems with kernel shapes of fuzzy sets to any continuous function. Some numerical examples will be stated in section 5 to illustrate our results. Finally, some conclusions are given in section 6.

## 2. Preliminaries

Assume that the analytical form of a continuous function  $f(x)$  defined on the closed interval  $U = [-1 + \delta_0, 1 - \delta_0] \subset \mathbb{R}$ , ( $0 < \delta_0 < 1$ ) is unknown, while the input - output behavior of  $f(x)$  for any  $x \in U$  is known. In this case,  $f(x)$  is similar to a black box. Assume that the input - output data of  $f$  on  $U$  are  $\{x_i, y_i\}_{i=1}^k$ , where  $\{x_i\}_{i=1}^k$  are considered as a random sample of size  $k$  from the absolute continuous distribution function  $H(x)$  with the density function  $h(x)$  and  $y_i = f(x_i)$ ,  $i = 1, \dots, k$ . Generally, the training data  $\{x_i\}_{i=1}^k$  could be a random sample from any distribution function; however,  $h(x)$  is determined in this paper by the denominator of a fuzzy system and is only required to be bounded away from 0 in  $U$  by some  $\varepsilon_0 > 0$ , i.e.,  $h(x) > \varepsilon_0$  for any  $x \in U$ . Theoretically, the size of a random sample could be arbitrarily (See in <sup>27</sup> for more details).

Let  $C(U)$  be the set of the continuous functions defined on  $U$  with the supremum norm

$$\|f\|_{C(U)} = \sup_{x \in U} |f(x)|,$$

and for any  $x \in U^c$ ,  $f(x)$  is defined to be zero. Recall that  $L^1(\mathbb{R})$  is the space of all integrable functions on  $\mathbb{R}$ . In this paper, the membership functions of each input variable are chosen as kernel shapes. Some related notations, concepts and properties are introduced below.

**Definition 1.** [see in <sup>28</sup>] A set of functions  $\{\chi(\circ, \sigma), \sigma > 0\}$  will be called a kernel on the real Euclid space  $\mathbb{R}$  if for any  $\sigma > 0$  we have  $\chi(\circ, \sigma) \in L^1(\mathbb{R})$  and  $\int_{\mathbb{R}} \chi(x, \sigma) dx = 1$ .

**Definition 2.** [see in <sup>28</sup>] A kernel  $\{\chi(\circ, \sigma), \sigma > 0\}$  is called an approximation identity kernel if there exists a constant  $M > 0$  such that

- i)  $\int_{\mathbb{R}} |\chi(x, \sigma)| dx \leq M, \forall \sigma > 0,$
- ii)  $\lim_{\sigma \rightarrow 0} \int_{\|x\| > \delta} |\chi(x, \sigma)| dx = 0, \forall \delta > 0.$

Following lemma gives a construction kernel from a integrable function in  $L^1(\mathbb{R})$  and can be proved similarly to Lemma 2.2 in <sup>12</sup>.

**Lemma 1.** Suppose that  $\chi \in L^1(\mathbb{R})$  and  $\int_{\mathbb{R}} \chi(x) dx \neq 0$  then  $\{\frac{1}{\sigma} \chi(\frac{x}{\sigma}), \sigma > 0\}$  confirms an approximation identity kernel.

Now, let  $\chi \in L^1(\mathbb{R}), \int_{\mathbb{R}} \chi(x) dx \neq 0$  and  $\chi(x) \geq 0, \forall x \in \mathbb{R}$ . A translation and scale transformation of  $\chi$  is defined as  $\frac{1}{\sigma} \chi(\frac{x-\alpha}{\sigma})$ , where  $x \in \mathbb{R}, \alpha$  is the translation factor and  $\sigma$  is the scale factor. Suppose that  $\{\sigma_k, k \in \mathbb{N}\}$  are regularizing scale factors, that is

- i)  $\sigma_k > 0, \forall k \in \mathbb{N}; \lim_{k \rightarrow +\infty} \sigma_k = 0$
- ii)  $\sum_{k=1}^{\infty} \exp(-\eta \sqrt{k} \sigma_k^2) < +\infty, \forall \eta > 0.$

Set

$$\chi_k(x) := \frac{1}{\sigma_k} \chi\left(\frac{x}{\sigma_k}\right). \tag{1}$$

It follows from Lemma 1 that  $\{\chi_k(x), \sigma_k\}$  confirms an approximation identity kernel on  $\mathbb{R}$ . Furthermore, we can suppose that  $0 \leq \chi_k(x) \leq 1$  for all  $x \in U$ , for all  $k$  and  $\{\chi_k(x - x_i)\}_{i=1}^k$  construct an  $\epsilon_0$ -complete partition on  $U, 0 < \epsilon_0 \leq 1$  (see <sup>21</sup>), i.e.,

$$\frac{1}{k} \sum_{i=1}^k \chi_k(x - x_i) \geq \epsilon_0, \forall x \in U,$$

where  $\{x_i\}_{i=1}^k \subset U$  is the random sample with the density function  $h(x) \in C(U)$ .

In this paper, we introduce the mathematical formula of fuzzy systems in the single input - single output (SISO) case. With  $k$  rules, a fuzzy system is

denoted by  $F_k : \mathbb{R} \rightarrow \mathbb{R}$ , which comprises four principal components: singleton fuzzifier, product inference engine, center-average defuzzifier (see <sup>4,29</sup> for more details) and the fuzzy rule base. The fuzzy rule base stores  $k$  rules

$$R_i : \text{IF } x \text{ is } A_i \text{ THEN } y \text{ is } B_i$$

( $i = 1, 2, \dots, k$ ), where  $A_i(i = 1, 2, \dots, k)$  are the if-part sets and  $B_i(i = 1, 2, \dots, k)$  are the then-part sets,  $x \in U$  is the input variable and  $y \in \mathbb{R}$  is the output variable. Assuming that the then-part sets are singleton fuzzy sets, the rules can be rewritten in the following form

$$R_i : \text{IF } x \text{ is } A_i \text{ THEN } y = y_i$$

( $i = 1, 2, \dots, k$ ), where  $y_i$  is the point in  $\mathbb{R}$  at which  $B_i(y)$  achieves its maximum value. Let  $\mu_i(x)$  be the fuzzy membership function corresponding to the if-part set  $A_i$ . Clearly, in the SISO case, the fuzzy system can be expressed as follows:

$$F_k(x) = \sum_{i=1}^k \left[ \frac{\mu_i(x)}{\sum_{i=1}^k \mu_i(x)} \right] y_i. \tag{2}$$

Therefore, with the kernel (1), the fuzzy system (2) can be rewritten in the following form:

$$F_k(x) = \sum_{i=1}^k \left[ \frac{\chi_k(x - x_i)}{\sum_{j=1}^k \chi_k(x - x_j)} \right] y_i, \tag{3}$$

where  $x \in U$  and  $y_i$  are the corresponding outputs of  $x_i, i = 1, \dots, k$ .

Assume that  $f \in C(U)$ . For simplicity, we introduce some following notations that are used fluently in the last.

Set  $(-\infty, x] := \{u \in U | u \leq x\}$ . Denote

$$H_k(x) = \frac{1}{k} \sum_{i=1}^k I_{(-\infty, x]}(x_i).$$

$$\begin{aligned} f_k(x) &= \frac{1}{k \sigma_k} \sum_{i=1}^k f(x_i) \chi\left(\frac{x - x_i}{\sigma_k}\right) \\ &= \int_{\mathbb{R}} f(u) \chi_k(x - u) dH_k(u). \end{aligned}$$

$$E[f_k(x)] = \int_{\mathbb{R}} f(u) \chi_k(x - u) dH(u).$$

$$g_k(x) = \frac{1}{k\sigma_k} \sum_{i=1}^k \chi\left(\frac{x-x_i}{\sigma_k}\right) = \int_{\mathbb{R}} \chi_k(x-u)dH_k(u).$$

$$E[g_k(x)] = \int_{\mathbb{R}} \chi_k(x-u)dH(u).$$

$$\widehat{F}_k(x) = \frac{E[f_k(x)]}{E[g_k(x)]}.$$

### 3. Approximation property

In this section, by using Fourier analysis technology, we explain how the fuzzy systems (3) with the kernel-shaped if part of fuzzy sets can approximate for a continuous function. At first, we have following lemma, that can be proved similarly to Theorem 3.1.6 in <sup>28</sup>.

**Lemma 2.** *Suppose that  $\{\chi(\circ, \sigma), \sigma > 0\}$  is an approximation identity kernel, then for any  $f \in C(U)$  we have*

$$\lim_{\sigma \rightarrow 0} \|I(f, \circ, \sigma) - f(\circ)\|_{\infty} = 0,$$

where

$$I(f, x, \sigma) = \int_{\mathbb{R}} f(u)\chi(x-u, \sigma)du.$$

Following theorem gives the universal approximation of the fuzzy system (3) for any function  $f \in C(U)$ .

**Theorem 3.** *Suppose that  $f \in C(U)$ . Then for every  $\varepsilon > 0$ , there exists  $N_0 \in \mathbb{N}$  such that*

$$\|F_k(\circ) - f(\circ)\|_{\infty} < \varepsilon$$

for all  $k \geq N_0$  almost sure (a.s.).

**Proof.** We have

$$\|F_k - f\|_{\infty} \leq \|F_k - \widehat{F}_k\|_{\infty} + \|\widehat{F}_k - f\|_{\infty} = I_1 + I_2. \tag{4}$$

At first, we consider  $I_1$ . One has

$$\begin{aligned} |F_k(x) - \widehat{F}_k(x)| &= \left| \frac{f_k(x)}{g_k(x)} - \frac{E[f_k(x)]}{E[g_k(x)]} \right| \\ &\leq \frac{1}{|g_k(x)|} \left| f_k(x) - E[f_k(x)] \right| \\ &\quad + \frac{|E[f_k(x)]|}{|g_k(x)||E[g_k(x)]|} |E[g_k(x)] - g_k(x)| \end{aligned}$$

Since  $\{\chi_k(x-x_i), \sigma_k\}_{i=1}^k$  is an  $\varepsilon_0$ -partition complete of  $U$ , then we have

$$|g_k(x)| \geq \varepsilon_0, \forall x \in U, \forall k, \varepsilon_0 > 0.$$

It follows from (10) that  $|E[g_k(x)]| \geq \varepsilon_0$  for all  $x \in U$  and with  $k$  large enough. So we get

$$\frac{1}{|g_k(x)|} \leq \frac{1}{\varepsilon_0}, \quad \frac{1}{|g_k(x)| \cdot |E[g_k(x)]|} \leq \frac{1}{\varepsilon_0^2}, \quad \forall x \in U. \tag{5}$$

Because  $\{\chi_k(x-x_i), \sigma_k\}$  is an approximation identity kernel and  $f \in C(U)$ , then one gets

$$\begin{aligned} |f_k(x)| &= \left| \int_{\mathbb{R}} f(u)\chi_k(x-u)dH_k(u) \right| \\ &\leq \sup_U |f(u)| \int_{\mathbb{R}} \chi_k(x-u)dH_k(u) \leq M_1. \end{aligned} \tag{6}$$

In the same manner we can see that

$$|E[f_k(x)]| \leq M_2. \tag{7}$$

It follows from (5)-(7) that

$$\begin{aligned} |F_k(x) - \widehat{F}_k(x)| &\leq \frac{1}{\varepsilon_0} |f_k(x) - E[f_k(x)]| \\ &\quad + \frac{M_2}{\varepsilon_0^2} |g_k(x) - E[g_k(x)]| \quad (\forall x \in U). \end{aligned}$$

Therefore

$$I_1 \leq \frac{1}{\varepsilon_0} \|f_k - E[f_k]\|_{\infty} + \frac{M_2}{\varepsilon_0^2} \|g_k - E[g_k]\|_{\infty}. \tag{8}$$

It is easy to see that

$$\begin{aligned} f_k(x) - E[f_k(x)] &= \int_{\mathbb{R}} f(u)\chi_k(x-u)d(H_k(u) - H(u)) \\ &= -\frac{1}{\sigma_k} \int_{\mathbb{R}} (H_k(u) - H(u))d(f(u)\chi_k\left(\frac{x-u}{\sigma_k}\right)). \end{aligned}$$

This implies

$$|f_k(x) - E[f_k(x)]| \leq \frac{M_3}{\sigma_k} \sup_U |H_k(u) - H(u)|,$$

where  $M_3$  is the total variation of  $f, \chi$  on  $U$ . From the results of <sup>30</sup> we have

$$P\{\sup |H_k(u) - H(u)| > \frac{\sigma_k \eta_k}{M_0}\} \leq M_4 \exp(-\frac{2k\sigma_k^2 \eta_k^2}{M_0^2}),$$

where  $M_4$  is an absolute constant. By choosing  $\eta_k = \frac{M_0}{\sigma_k} (\frac{\log k}{k})^{1/2}$  we have

$$\sum_{k=1}^{\infty} P\{\sup_U |H_k(x) - H(x)| > (\frac{\log k}{k})^{1/2}\} < +\infty.$$

By using Borel-Cantelli Lemma <sup>31</sup> we receive

$$\sup_U |H_k(x) - H(x)| \leq (\frac{\log k}{k})^{1/2}, \text{ (a.s.)}$$

with  $k$  is large enough. Hence,

$$\|f_k(\circ) - E[f_k(\circ)]\|_{\infty} \leq \frac{M_3}{\sigma_k} (\frac{\log k}{k})^{1/2} \text{ (a.s.)} \quad (9)$$

Using similar arguments we have

$$\|g_k(\circ) - E[g_k(\circ)]\|_{\infty} \leq \frac{M_5}{\sigma_k} (\frac{\log k}{k})^{1/2} \text{ (a.s.)} \quad (10)$$

Combining (9), (10) with (8) we get

$$I_1 \leq \left[ \frac{M_3}{\varepsilon_0 \sigma_k} + \frac{M_5 M_2}{\varepsilon_0^2 \sigma_k} \right] (\frac{\log k}{k})^{1/2} \rightarrow 0, k \rightarrow \infty \text{ (a.s.)} \quad (11)$$

According to Lemma 2, since  $\{\chi_k(x - x_i), \sigma_k\}$  is an approximation identity kernel then we have

$$\lim_{\sigma_k \rightarrow 0} \|I(fh, \circ, \sigma_k) - f(\circ)h(\circ)\|_{\infty} = 0$$

and

$$\lim_{\sigma_k \rightarrow 0} \|I(h, \circ, \sigma_k) - h(\circ)\|_{\infty} = 0.$$

That follows

$$\lim_{\sigma_k \rightarrow 0} \|E[f_k(\circ)] - f(\circ)h(\circ)\|_{\infty} = 0, \quad (12)$$

$$\lim_{\sigma_k \rightarrow 0} \|E[g_k(\circ)] - h(\circ)\|_{\infty} = 0. \quad (13)$$

We have

$$\begin{aligned} |\widehat{F}_k(x) - f(x)| &= \left| \frac{E[f_k(x)]}{E[g_k(x)]} - f(x) \right| \\ &\leq \frac{1}{E[g_k(x)]} \left( |E[f_k(x)] - f(x)h(x)| + \right. \\ &\quad \left. |f(x)||E[g_k(x)] - h(x)| \right), \forall x \in U. \end{aligned}$$

Then by using  $f \in C(U)$  and (5)-(7) one gets

$$I_2 \leq \frac{1}{\varepsilon_0} \left( \|E[f_k] - f.h\|_{\infty} + M_6 \|E[g_k] - h\|_{\infty} \right). \quad (14)$$

From (12), (13) and (14) we have

$$\lim_{k \rightarrow \infty} I_2 = 0 \text{ (a.s.)} \quad (15)$$

Combining (4), (11) and (15) we have  $\lim_{k \rightarrow \infty} \|F_k(\circ) - f(\circ)\|_{\infty} = 0$ . So that, for every  $\varepsilon > 0$ , there exists  $N_0 \in \mathbb{N}$  such that

$$\|F_k(\circ) - f(\circ)\|_{\infty} < \varepsilon$$

for all  $k \geq N_0$  (a.s.). The theorem is proved.  $\square$

#### 4. The uniform approximation rates of fuzzy systems

As we have shown in <sup>12</sup>, in the  $n$ -dimensional case (in  $\mathbb{R}^n$ ), the uniform approximation rates of fuzzy systems had estimated with  $f \in L^B(\mathbb{R}^n)$ . However, this class of functions is a strict subset of the class of continuous functions but not equal to. So that there are some parts of the class of continuous functions, although can be approximated by a fuzzy system, do not have the estimations of the uniform approximation rates (section 3). But in one-dimensional case (in  $\mathbb{R}$ ), we receive results more better than those of in the  $n$ -dimensional. Concretely, the uniform approximation rates of the kernel fuzzy systems for continuous functions are estimated as the rates of convergence of the sequence  $(\log k/k)^{\beta}$ .

In this part of the paper, we introduce two ways to estimate the rates of uniform approximation of the fuzzy systems for continuous functions. These base on two following different methods.

1. The first method is based on technique of the test function methods (see 1.5.3 in <sup>28</sup>). In this way, the singular integrals of functions are used to approximate. And then a kernel fuzzy system is constructed to approximate the singular integrals. Our result based on this method is given in Theorem 4.
2. The second method is based on the  $\alpha$ th absolute moment of the kernel function. For some hypotheses of the kernel, the rates of uniform approximation is equivalent to the rates of uniformly converge of sequence  $\{(\ln k/k)^\beta\}$ . The main result is given in Theorem 5.

At first, we consider the approach of the first method. For  $\delta \geq 0$  arbitrary, the modulus continuity of  $f \in C(U)$  is defined

$$\omega(C(U), f, \delta) := \sup_{|h| < \delta} \|f(\circ + h) - f(\circ)\|_\infty$$

and the generalized modulus continuity of  $f \in C(U)$  is defined

$$\omega^*(C(U), f, \delta) := \sup_{|h| < \delta} \|f(\circ + h) + f(\circ - h) - 2f\|_\infty.$$

**Definition 3.** [see in <sup>28</sup>] A function  $f \in C(U)$  is said to satisfy a Lipschitz condition of order  $\alpha$  ( $\alpha > 0$ ), in notation  $f \in Lip(C(U), \alpha)$ , if

$$\omega(C(U), f, \delta) = O(\delta^\alpha).$$

A function  $f \in C(U)$  is said to satisfy a generalized Lipschitz condition of order  $\alpha$  ( $\alpha > 0$ ), in notation  $f \in Lip^*(C(U), \alpha)$ , if

$$\omega^*(C(U), f, \delta) = O(\delta^\alpha).$$

**Theorem 4.** Assume that

1.  $\chi \in L^1(\mathbb{R})$ ,  $\int_{\mathbb{R}} \chi(t)dt \neq 0$  and  $\chi$  is an even, positive function on  $\mathbb{R}$ ,
2. there exists  $x_0 \in \mathbb{R}$  such that the singular integral can be represented in following form

$$I(u^2, x_0, \sigma_k) = x_0^2 + \delta(x_0, \sigma_k), \quad (16)$$

where  $\lim_{k \rightarrow \infty} \delta(x_0, \sigma_k) = 0$ .

Then for all  $f \in C(U)$  there is a fuzzy system  $F_k(x)$ , defined in (3), satisfying

$$\begin{aligned} \|F_k - f\|_\infty &\leq \frac{C_1}{\varepsilon_0 \sigma_k} \left(\frac{\log k}{k}\right)^{1/2} \\ &+ \frac{1}{\varepsilon_0} O(\omega(C(U), fh, \sqrt{\delta(x_0, \sigma_k)})) \\ &+ \frac{C_2}{\varepsilon_0} O(\omega(C(U), h, \sqrt{\delta(x_0, \sigma_k)})) \quad (a.s.) \end{aligned} \quad (17)$$

for all  $k$  is large enough, where  $C_1, C_2$  are positive constants.

**Proof.** Recall that

$$\|F_k - f\|_\infty \leq \|F_k - \widehat{F}_k\|_\infty + \|\widehat{F}_k - f\|_\infty = I_1 + I_2.$$

From (11) we have

$$I_1 \leq \frac{C_1}{\varepsilon_0 \sigma_k} \left(\frac{\log k}{k}\right)^{1/2}, \quad (18)$$

where  $C_1$  is positive number.

In another way, from hypotheses of the theorem and using Proposition 3.3.1 in <sup>28</sup> we have for all  $\phi \in C(U)$

$$\|I(\phi, \circ, \sigma_k) - \phi\|_\infty = O(\omega(C(U), \phi, \sqrt{\delta(x_0, \sigma_k)})),$$

when  $k \rightarrow \infty$ . Then when  $k \rightarrow \infty$  we have

$$\begin{aligned} \|E[a_k(\circ)] - fh(\circ)\|_\infty &= \|I(fh, \circ, \sigma_k) - fh(\circ)\|_\infty \\ &= O(\omega(C(U), fh, \sqrt{\delta(x_0, \sigma_k)})) \end{aligned} \quad (19)$$

and

$$\begin{aligned} \|E[g_k(\circ)] - h(\circ)\|_\infty &= \|I(h, \circ, \sigma_k) - h(\circ)\|_\infty \\ &= O(\omega(C(U), h, \sqrt{\delta(x_0, \sigma_k)})). \end{aligned} \quad (20)$$

From (14), (19) and (20) we have

$$\begin{aligned} I_2 &\leq \frac{1}{\varepsilon_0} O(\omega(C(U), fh, \sqrt{\delta(x_0, \sigma_k)})) \\ &+ \frac{C_2}{\varepsilon_0} O(\omega(C(U), h, \sqrt{\delta(x_0, \sigma_k)})) \end{aligned} \quad (21)$$

when  $k \rightarrow \infty$ .

Combining (18) and (21) we receive the assertion of the theorem.  $\square$

In second method, we recall the notation of  $\alpha$ th absolute moment of a function, it can be stated as follows.

**Definition 4.** [see in <sup>28</sup>] Let  $\alpha > 0$ . The  $\alpha$ th absolute moment of  $\chi$  is defined by

$$m(\chi, \alpha) := \int_{\mathbb{R}} |u|^\alpha |\chi(u)| du$$

if the right integral exists.

**Theorem 5.** Assume that  $\chi \in L^1(\mathbb{R})$ ,  $\int_{\mathbb{R}} \chi(t) dt \neq 0$  and  $\chi$  is an even, positive function. Moreover, we suppose  $\chi$  having finite  $\alpha$ th absolute moment ( $\alpha > 0$ ). Then for each  $f \in C(U)$ , there exists a fuzzy system  $F_k$  defined by (3) such that

$$\begin{aligned} \|F_k(\circ) - f(\circ)\|_\infty &\leq \frac{C_1}{\varepsilon_0 \sigma_k} \left(\frac{\log k}{k}\right)^{1/2} \\ &+ \frac{1}{\varepsilon_0} O(\omega^*(C(U), fh, \sigma_k)) \\ &+ \frac{C_2}{\varepsilon_0} O(\omega^*(C(U), h, \sigma_k)) \quad (a.s.) \quad (22) \end{aligned}$$

when  $k \rightarrow \infty$ , where  $C_1, C_2$  are positive numbers.

**Proof.** From the hypotheses of theorem and Proposition 3.4.1 in <sup>28</sup> we have for all  $\phi \in C(U)$

$$\|I(\phi, \circ, \sigma_k) - \phi\|_\infty = O(\omega^*(C(U), \phi, \sigma_k)), \quad k \rightarrow \infty.$$

By using similar arguments to those of Theorem 4 one gets

$$\|E[a_k(\circ)] - fh(\circ)\|_\infty = O(\omega^*(C(U), fh, \sigma_k)) \quad (23)$$

and

$$\|E[g_k(\circ)] - h(\circ)\|_\infty = O(\omega^*(C(U), h, \sigma_k)) \quad (24)$$

when  $k \rightarrow \infty$ . From (23)-(24) and using (14) we have

$$I_2 \leq \frac{1}{\varepsilon_0} O(\omega^*(C(U), fh, \sigma_k)) + \frac{C_2}{\varepsilon_0} O(\omega^*(C(U), h, \sigma_k)). \quad (25)$$

Combining (18) and (25) we receive (22). The theorem is proved.  $\square$

If the density function is assumed to be uniform on  $U$  then we receive following corollary.

**Corollary 6.** Assume that  $h(x)$  is an uniformly density function on  $U$  and the conditions of Theorem 5 hold. Then for any  $f(\circ) \in C(U)$ , there is a fuzzy system  $F_k(x)$  defined by (3) such that

$$\begin{aligned} \|F_k(\circ) - f(\circ)\|_\infty &\leq \frac{C_3}{\varepsilon_0 \sigma_k} \left(\frac{\log k}{k}\right)^{1/2} \\ &+ \frac{1}{\varepsilon_0} O(\omega^*(C(U), f, \sigma_k)) \quad (a.s.) \end{aligned}$$

when  $k \rightarrow \infty$ , where  $C_3$  is positive number.

**Proof.** The uniformly density function  $h(x)$  on  $U$  has the form

$$h(x) = \begin{cases} \frac{1}{2-2\delta_0} & \text{if } x \in U \\ 0 & \text{if } x \notin U. \end{cases}$$

Then

$$O(\omega^*(C(U), h, \sigma_k)) = 0,$$

$$O(\omega^*(C(U), fh, \sigma_k)) = \frac{1}{2-2\delta_0} O(\omega^*(C(U), f, \sigma_k)).$$

It follows from Theorem 5 that the assertion of corollary holds.  $\square$

In the class of functions satisfying generalized Lipschitz condition of order  $\alpha$ ,  $\alpha > 0$ , we receive better results on the estimation of the rates of approximation.

**Corollary 7.** Assume that  $h(x)$  is an uniformly density function on  $U$  and the conditions of Theorem 5 hold. Moreover, suppose that  $\chi$  has finite  $\alpha$ th absolute moment ( $0 < \alpha \leq 2$ ). Then for all  $f \in Lip^*(C(U), \alpha)$ , there exists a fuzzy system  $F_k$  defined by (3) such that

$$\begin{aligned} \|F_k(\circ) - f(\circ)\|_\infty &\leq \frac{C_1}{\varepsilon_0 \sigma_k} \left(\frac{\log k}{k}\right)^{1/2} \\ &+ \frac{1}{\varepsilon_0} O(\sigma_k^\alpha) \quad (a.s.) \quad (26) \end{aligned}$$

when  $k \rightarrow \infty$ .

Furthermore, when  $\sigma_k \approx (\log k/k)^{1/2(1+\alpha)}$  then we receive the following estimation

$$\|F_k - f\|_\infty \leq O((\log k/k)^{\alpha/2(1+\alpha)}) \quad (a.s.). \quad (27)$$

**Proof.** Since  $h(x)$  is uniformly density function on  $U$  and  $h \in Lip^*(C(U), \alpha)$  then by Corollary 6 we receive easily the estimation (26).

For  $\sigma_k \approx (\log k/k)^{1/2(1+\alpha)}$  in (26) we receive (27). The corollary is proved completely.  $\square$

The rates of uniform approximation are usually measured in terms of the order of error  $\|F(\circ) - f(\circ)\|_\infty$ . The determination of the optimal approximation rates is called the saturation problem. We remark that classical approximation processes (polynomial, spline, trigonometric) are saturated with order  $O(\log k/k)^\alpha$  (when  $f \in C[\mathbb{R}]$ ), where  $k$  is the number of basis functions (see <sup>32,33</sup>). From <sup>34,6</sup> we can see that fuzzy systems (3) can be viewed as interpolation functions of input-output data. These interpolation functions are also combinations of  $k$  basis functions  $\left\{ \frac{\chi_k(x-x_i)}{\sum_{j=1}^k \chi_k(x-x_j)} \right\}_{i=1}^k$ . From above results we can see that the approximation rates of fuzzy systems are also saturated with order  $O(\log k/k)^\alpha$ .

### 5. Numerical example

**Example 1.** In this example, we introduce some essential functions such as: Triangular set function, Gaussian set function, Parabol set function, Laplace set function and Spline set function. These functions are used to construct approximation identity kernels and to be input membership functions of fuzzy systems. Here are some of them:

#### 1. Triangular set function

$$\chi_T(x) = \begin{cases} x + 1, & -1 \leq x < 0, \\ -x + 1, & 0 \leq x \leq 1, \\ 0, & |x| > 1. \end{cases}$$

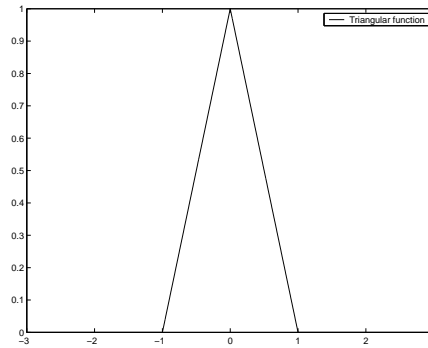


Fig. 1. Triangular set function.

We have  $I(u^2, x, \sigma_k) = \int_{-\infty}^{+\infty} u^2 \frac{1}{\sigma_k} \chi_T\left(\frac{x-u}{\sigma_k}\right) du$ . By setting  $t = \frac{x-u}{\sigma_k}$ , we have

$$\begin{aligned} I(u^2, x, \sigma_k) &= \int_{-\infty}^{+\infty} (x-t\sigma_k)^2 \chi_T(t) dt \\ &= \int_{-1}^0 (x-t\sigma_k)^2 (1+t) dt + \int_0^1 (x-t\sigma_k)^2 (1-t) dt \\ &= x^2 + \frac{\sigma_k^2}{6}. \end{aligned}$$

In other words  $m(\chi_T, \alpha) = \int_{-\infty}^{+\infty} |u|^\alpha |\chi_T(u)| du = 2 \int_0^{+\infty} u^\alpha \chi_T(u) du$ . Setting  $t = \frac{u}{\sigma_k}$ , we get

$$\begin{aligned} m(\chi_T, \alpha) &= 2\sigma_k^\alpha \int_0^{+\infty} t^\alpha \chi_T(t) dt \\ &= 2\sigma_k^\alpha \int_0^1 t^\alpha (1-t) dt = \frac{2\sigma_k^\alpha}{\alpha^2 + 3\alpha + 2}. \end{aligned}$$

#### 2. Gauss set function

$$\chi_G(x) = e^{-\frac{1}{2}x^2}.$$

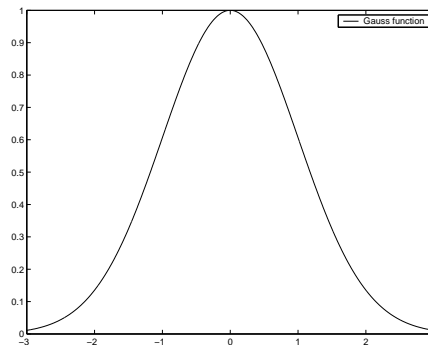


Fig. 2. Gauss set function.



By the same arguments we have

$$I(u^2, x, \sigma_k) = x^2 + \sigma_k^2.$$

$$m(\chi_G, \alpha) = 2^{\frac{\alpha-1}{2}} \Gamma\left(\frac{\alpha+1}{2}\right) \sigma_k^\alpha.$$

$$I(u^2, x, \sigma_k) = x^2 + 2\sigma_k^2.$$

$$m(\chi_L, \alpha) = 2\sigma_k^\alpha \Gamma(\alpha + 1).$$

where  $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt$  is Gamma integral.

### 3. Parabol set function

$$\chi_P(x) = \begin{cases} \frac{3}{4}(1-x^2), & -1 \leq x \leq 1, \\ 0, & |x| > 1. \end{cases}$$

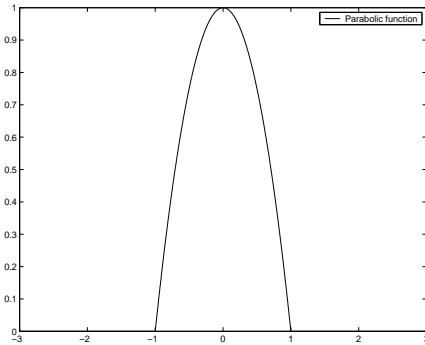


Fig. 3. Parabol set function.

$$I(u^2, x, \sigma_k) = x^2 + \sigma_k^2/5.$$

$$m(\chi_P, \alpha) = \frac{4\sigma_k^\alpha (1 + (-1)^\alpha)}{(\alpha + 1)(\alpha + 3)}.$$

### 4. Laplace set function

$$\mu_L(x) = \frac{1}{2} e^{-|x|}.$$

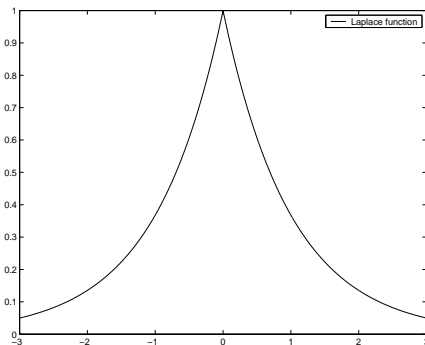


Fig. 4. Laplace set function.

### 5. Spline set function

$$\chi_S(x) = \begin{cases} 0, & x \leq -\frac{3}{2}, \\ \frac{8}{9}(x + \frac{3}{2})^2, & -\frac{3}{2} < x \leq -\frac{3}{4}, \\ 1 - \frac{8}{9}x^2, & -\frac{3}{4} < x \leq \frac{3}{4}, \\ \frac{8}{9}(x - \frac{3}{2})^2, & \frac{3}{4} < x \leq \frac{3}{2}, \\ 0, & x > \frac{3}{2}. \end{cases}$$

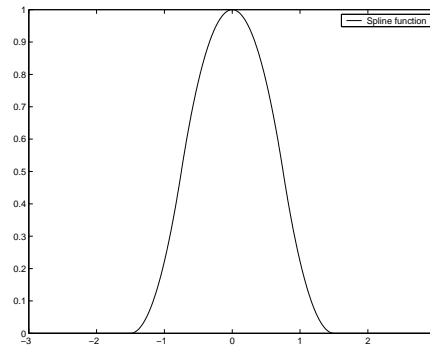


Fig. 5. Spline set function.

$$I(u^2, x, \sigma_k) = x^2 + 9\sigma_k^2/32.$$

$$m(\chi_S, \alpha) = \frac{3^{1+\alpha} (2^{2+\alpha} - 1) \sigma_k^\alpha}{4^\alpha (\alpha^3 + 6\alpha^2 + 11\alpha + 6)}.$$

### 6. Trapezoidal set functions

$$\chi_H(x) = \begin{cases} 2x + 2, & -1 \leq x \leq -\frac{1}{2}, \\ 1, & -\frac{1}{2} < x \leq \frac{1}{2}, \\ -2x + 2, & \frac{1}{2} < x \leq 1, \\ 0, & |x| > 1. \end{cases}$$

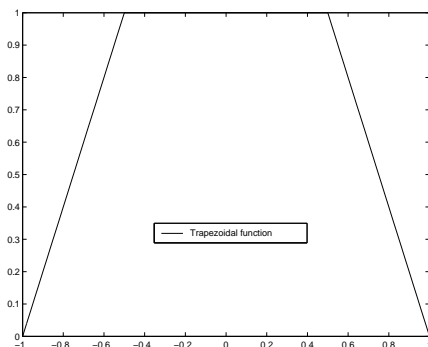


Fig. 6. Trapezoidal set function.

$$I(u^2, x, \sigma_k) = 5\sigma_k^2/24 + x^2.$$

$$m(\chi_S, \alpha) = \frac{(4 - 2^{-\alpha})\sigma_k^\alpha}{\alpha^2 + 3\alpha + 2}.$$

It is easy to see that these functions is valued on  $[0, 1]$ . So that they all are input membership functions of fuzzy systems. Moreover, by using a translation and scale transformation, we can construct approximation identity kernels based on these functions, corresponding.

From  $I(u^2, x, \sigma_k)$  and  $m(\chi, \alpha)$ , we can see that they all satisfy hypotheses of Theorem 4 and Theorem 5. The integral  $I(u^2, x, \sigma_k)$  is always represented in form  $x^2 + \delta(x_0, \sigma_k)$ , where  $\delta(x_0, \sigma_k)$  depends only on  $\sigma_k^2$ . So that we can construct a fuzzy system, based on the kernel constructed on those functions, such that it can approximate a continuous function on  $U$  with the rates of approximation is estimated on Theorem 4 or Theorem 5.

On the other hand, all those functions have finite  $\alpha$ th absolute moments and these are represented by  $\sigma_k^\alpha$ . Hence, if  $f$  satisfies generalized Lipschitz condition of order  $\alpha$  and if we choose  $\sigma_k \approx (\log k/k)^{1/2(1+\alpha)}$ , then from Corollary 7 we can see that the rates of approximation of fuzzy systems is  $O((\log k/k)^{\alpha/2(1+\alpha)})$  for all  $0 < \alpha \leq 2$ .

**Example 2.** Consider 6 following functions

1.  $f1(x) = \begin{cases} x \sin \frac{1}{x}, & x \neq 0, \\ 0, & x = 0. \end{cases}$
2.  $f2(x) = 10(e^{-\frac{|x|}{0.2}} + e^{-\frac{|x-0.8|}{0.3}} + e^{-\frac{|x+0.6|}{0.1}}).$
3.  $f3(x) = 8 \sin(10x^2 + 5x + 1).$
4.  $f4(x) = 5 \arctan \left( \frac{2000g1(x)g2(x)}{x^2 + 1.5x + 1} \right)$  where

$$g1(x) = (x - 0.1)(x - 0.3)(x - 0.5)(x - 0.9)(x - 1.1)$$

$$g2(x) = (x + 0.2)(x + 0.4)(x + 0.6)(x + 0.8)(x + 1)$$

$$5. f5(x) = \frac{x}{2000}$$

$$6. f6(x) = 10[\sin(4x + 0.1) + \sin(14x) + \sin(11x - 0.2) + \sin(17x + 0.3)];$$

With following training data

$$\{(x_i = -0.9 + \frac{i}{1000}, y_i = f2(x_i))\}_0^{1800},$$

we approximate functions  $f1, f2, f3, f4, f5$  and  $F6$  by fuzzy systems constructed directly as (3) by using approximation identity kernels from  $\chi_T, \chi_G, \chi_P, \chi_L, \chi_S$  and  $\chi_H$  in Example 1, respectively. The results are listed below

Fig.	The curve of target function and fuzzy system	Maximum error
7	Function $f1$ and triangular fuzzy system	5.641030677038259e-002
8	Function $f2$ and Gauss fuzzy system	1.144553440354985e-001
9	Function $f3$ and Parabolic fuzzy system	8.442668133852935e-001
10	Function $f4$ and Laplace fuzzy system	2.540585179966536e+000
11	Function $f5$ and Spline fuzzy system	1.626303258728257e-019
12	Function $f6$ and Trapezoidal fuzzy system	3.157647797971208e-002

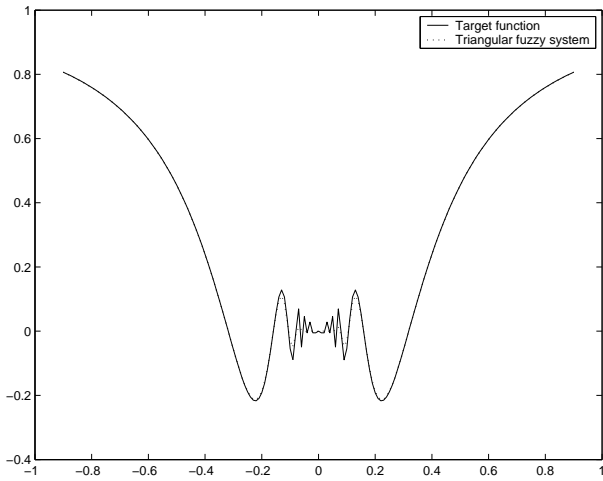


Fig. 7. The curve of the target function, Triangular FS.

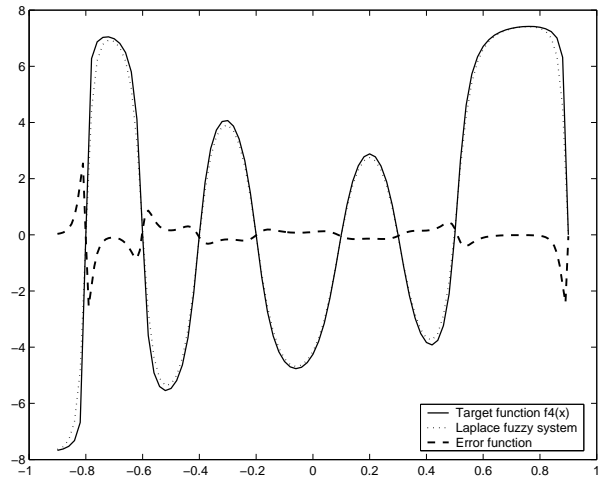


Fig. 10. The curve of the target function, Laplace FS.

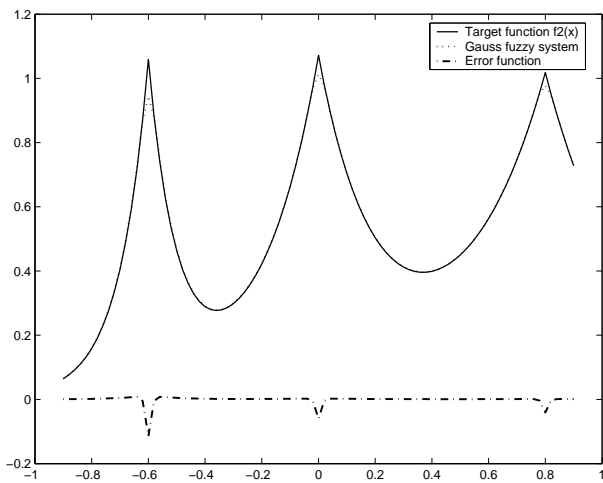


Fig. 8. The curve of the target function, Gauss FS.

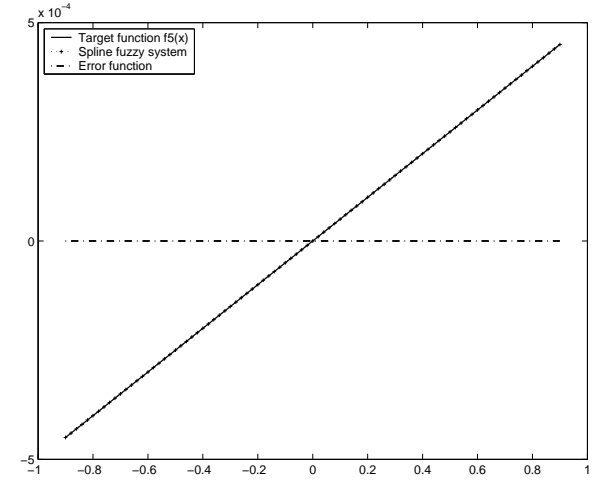


Fig. 11. The curve of the target function, Spline FS.

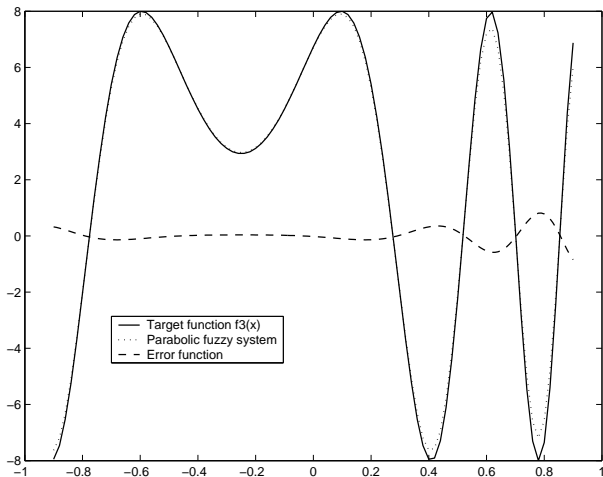


Fig. 9. The curve of the target function, Parabolic FS.

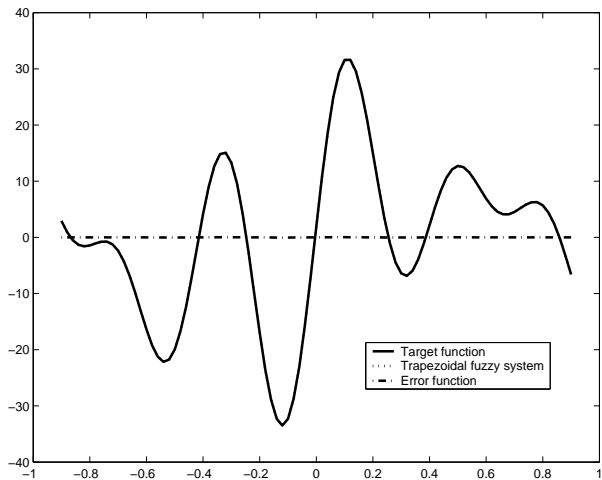


Fig. 12. The curve of the target function, Trapezoidal FS.

**Example 3.**

Consider  $\chi(x) = \frac{1}{2\sqrt{\pi}}e^{-x^2/4}$  and regularizing scale factors  $\{\sigma_k = \frac{1}{\sqrt{k}}, k \in \mathbb{N}\}$ . It is easy to see that  $\chi(x)$  is an even, positive function and  $\chi \in L^1(\mathbb{R})$ ,  $\int_{\mathbb{R}} \chi(x)dx \neq 0$ . In another way, we get

$$I(u^2, x, \sigma_k) = \frac{1}{\sqrt{4\pi\sigma_k}} \int_{-\infty}^{\infty} u^2 e^{-\left(\frac{x-u}{2\sigma_k}\right)^2} du.$$

Setting  $t = \frac{x-u}{2\sigma_k}$  with note that

$$\int_{-\infty}^{+\infty} e^{-t^2} dt = \sqrt{\pi}, \quad \int_{-\infty}^{+\infty} te^{-t^2} dt = 0$$

and

$$\int_{-\infty}^{+\infty} t^2 e^{-t^2} dt = \frac{\sqrt{\pi}}{2},$$

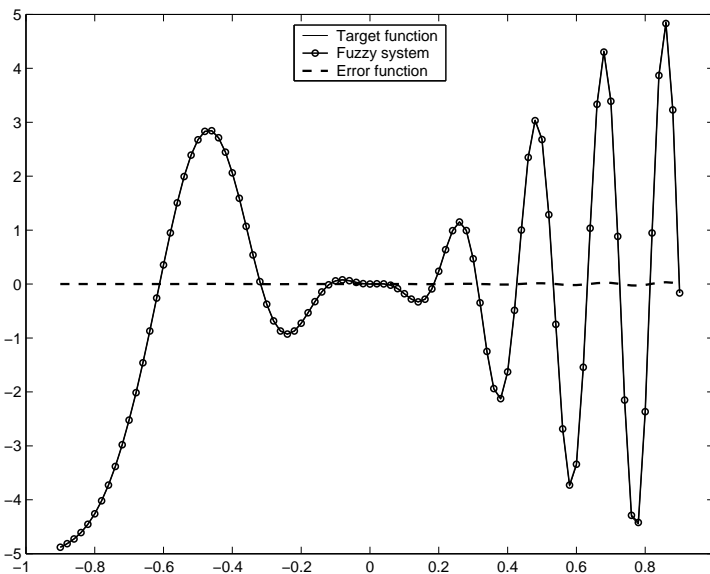
we have

$$I(u^2, x, \sigma_k) = x^2 + 2\sigma_k^2.$$

That shows the hypotheses of Theorem 4 are satisfied with  $\delta(x_0, \sigma_k) = 2\sigma_k^2$ . Then the fuzzy system can be stated as follows

$$F_k(x) = \sum_{i=1}^k \left[ \frac{\frac{1}{\sqrt{4\pi\sigma_k^2}} e^{-(x-x_i)^2/4\sigma_k^2}}{\sum_{j=1}^k \frac{1}{\sqrt{4\pi\sigma_k^2}} e^{-(x-x_j)^2/4\sigma_k^2}} \right] y_i, \quad (28)$$

From Figure 13 we can see that the curve of target function  $f(x)$  (the solid line (—)) and the curve of fuzzy system  $F(x)$  (the circle line (o)) are practically overlapped.



where  $y_i = f(x_i)$  and satisfies the estimation (17). Moreover, if we assume that  $f, h$  satisfy a Lipschitz condition of order  $\alpha$ , ( $0 < \alpha \leq 2$ ) then we receive

$$\|F_k(\circ) - f(\circ)\|_{\infty} \leq \frac{C_1}{\epsilon_0 \sigma_k} \left(\frac{\log k}{k}\right)^{1/2} + O(\sigma_k^{\alpha/2}) \text{ (a.s.)}$$

when  $k$  is large enough, where  $F_k$  is a fuzzy system defined by (28).

In this example we use fuzzy systems (28) in order to show the simultaneously approximation capability of this fuzzy systems for not only target function but also its derivative (see Theorem 5.1 in [12]). We consider following target function

$$f(x) = -20 \left( \frac{x}{x^2 + 1} \right)^2 \cos(10x^2 + 20x - 191)$$

and its first order derivative

$$f'(x) = \frac{40x(x+1)(10x \sin u(x^2+1) + \cos u(x-1))}{(x^2+1)^3},$$

where  $u = 10x^2 + 20x - 191$ .

Let  $\delta_0 = 0.1$  and the input space  $U = [-0.9, 0.9]$ . We uniformly sampled 18001 point of the target function  $f$  on  $U$  to give the training data  $\{(x_i = -0.9 + \frac{i}{10000}, y_i = f(x_i))\}_{i=0}^{18000}$ .

Fig. 13. The curve of target function  $f(x)$  and fuzzy system  $F(x)$ .

From Figure 14 we can see that, the derivative function  $f'$  varies very fast when  $x \in U$  but the curve of  $f'(x)$  (the solid line (—)) and the curve of  $F'(x)$  (the circle line (o)) are still practically overlapped.

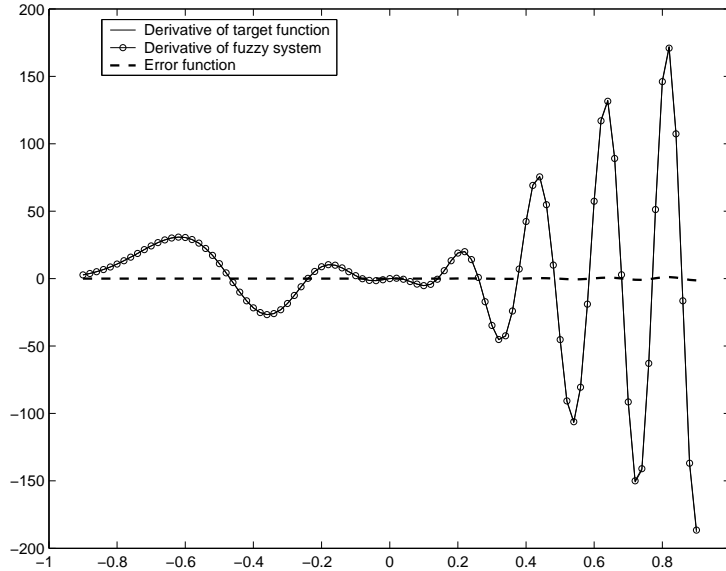


Fig. 14. The curve of  $f'$  and  $F'(x)$ .

**Example 4.** In this example we will make some comparisons about the capability approximation of kernel fuzzy systems with the capability approximation of Hermite interpolation polynomial for continuous functions.

Let  $\{(x_i = -1 + \frac{2i}{k}, y_i = f(x_i))\}_{i=0}^k$  be a training data of function  $f$  on  $U$  and

$$\chi(x) := \begin{cases} 0 & \text{if } x < -\frac{2}{k}, \\ 2(\frac{kx+2}{2})^2 & \text{if } -\frac{2}{k} \leq x \leq -\frac{2}{2k}, \\ 1 - 2(\frac{k}{2}x)^2 & \text{if } -\frac{2}{2k} \leq x \leq \frac{2}{2k}, \\ 2(\frac{kx-2}{2})^2 & \text{if } \frac{2}{2k} \leq x \leq \frac{2}{k}, \\ 0 & \text{if } x > \frac{2}{k}. \end{cases} \quad (29)$$

be a central Spline kernel.

Consider the target function

$$f(x) = \frac{100(x + 0.95)(x + 0.6)(x + 0.4)(x - 0.1)(x - 0.4)(x - 0.8)(x - 0.9)}{(x + 1.7)(x - 2)^2};$$

By setting

$$\chi_j(x) = \chi(x + 1 - j\frac{2}{k}), \quad j = 0, 1, \dots, k$$

we have  $\{\chi_j(x)\}_{j=0,1,\dots,k}$  is an approximation identity kernel on  $\mathbb{R}$ .

From the properties of Spline functions (see in <sup>11</sup>) we can see that

$$\sum_{j=0}^k \chi_j(x) = 1, \quad \text{for all } x \in [-1, 1]. \quad (30)$$

So fuzzy systems (3) can be rewritten in reduced form as follows

$$F_k(x) = \sum_{i=1}^k \chi_j(x) \cdot y_i, \quad (31)$$

With 100 knots, Figure 15 shows the curve of target function, fuzzy system and Hermite interpolation polynomial.

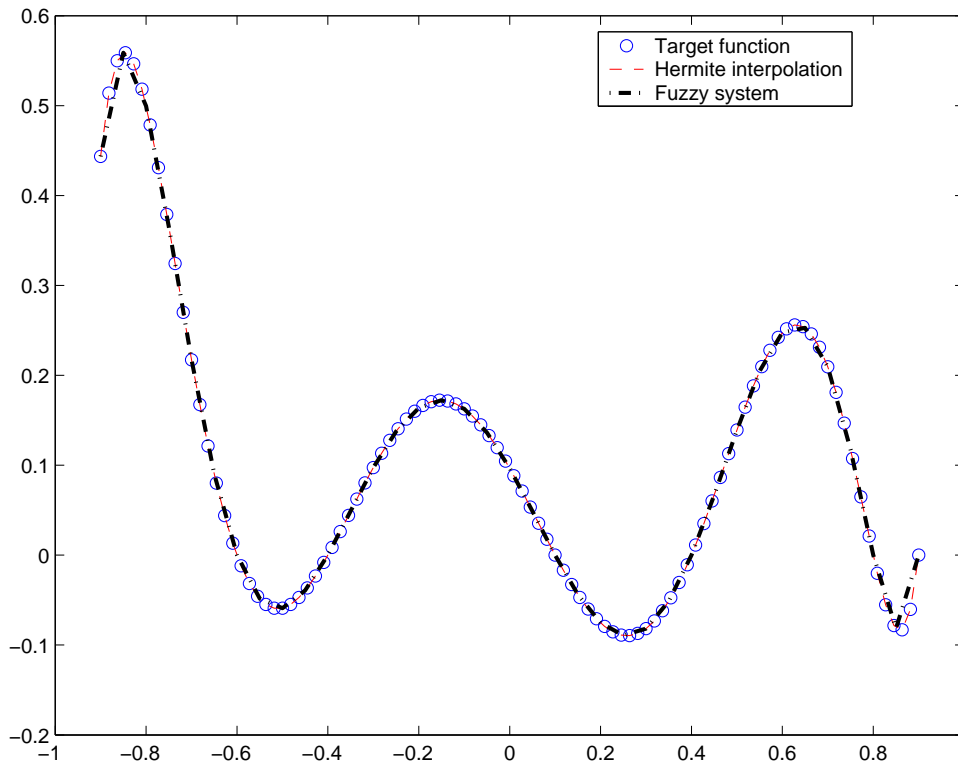


Fig. 15. The curve of target function, Hermite interpolation polynomial and fuzzy systems

Set

$$Error := \|f - F\|_2 = \sqrt{\sum_{j=1}^k (y_j - F(x_j))^2},$$

where Error is the error of approximation methods,  $F$  is the fuzzy system or the interpolation polynomial Hermite. Then the error of fuzzy systems is

$$Error_{Fuzzy} = 1.407491319767667e - 016$$

and the error of Hermite interpolation is

$$Error_{Hermite} = 4.278987665021929e - 017.$$

From above results we can see that with the special choice of kernel shapes, the error of fuzzy systems are the same order as the error of Hermite interpolation polynomial. While Hermite interpolation requires the target functions must be continuously differentiable, fuzzy systems can be constructed directly from learning data of functions.

## 6. Conclusion

By using Fourier technique, we have proved the capability of approximation of fuzzy systems for continuous one-variables functions. Moreover, the rates of uniform approximation is estimated as the rates of convergence of the sequence  $(\log k/k)^\alpha$ . Some approaches to deal with this problem are given. One base on the technique of the test function methods, another is based on the  $\alpha$ th absolute moment of the kernel function. The estimations of the rates of uniform approximation depend on our kernel function satisfying what conditions in Theorem 4 or Theorem 5. Further results for continuous multi-variables functions will be given in our future works.

## Acknowledgments

The author would like to thank Professor Bui Cong Cuong for his valuable discussions at the beginning of this study. This paper was partially supported by grants from the Sciences Research Program of University of Transport and Communications (No. T2011-KHCB 08).

1. J. C. Bezdek, J. Keller, R. Krisnapuram and N. R. Pal, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*, Springer, 2005.
2. T. Takagi and M. Sugeno, "Fuzzy identification of systems and its application to modeling and control", *IEEE Transactions on Systems - Man - Cybernetics*, **15**, 116 - 135 (1985).
3. Kairui Cao, Kairui Cao, X. Z. Gao, X. Z. Gao, Xianlin Huang, Xiaojun Ban, "Stability Analysis of a Type of Takagi-Sugeno PI Fuzzy Control Systems Using Circle Criterion", *International Journal of Computational Intelligence Systems*, **4**, 196-207 (2011).
4. H. T. Nguyen and M. Sugeno, *Fuzzy Systems: Modeling and Control*, Kluwer, Boston, MA, 1998.
5. M. Sugeno, *Industrial applications of fuzzy control*, Elsevier Science Pub. Co, 1985.
6. L. X. Wang, J. M. Mendel, "Fuzzy basis functions, universal approximation, and orthogonal least-squares learning", *IEEE Transactions on Neural Networks*, **3** 807 - 814 (1992).
7. J. L. Castro, "Fuzzy logic controllers are universal approximators", *IEEE Trans. SMC*, **25**, 629-635 (1995).
8. E. H. Mamdani, "Application of fuzzy algorithms for control of simple dynamic farm", *Proceedings of the Institution of Electrical Engineers London*, **121**, 1585 - 1588 (1974).
9. E. H. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller", *International Journal of Man-Machine Studies*, **7**, 1 - 13 (1975).
10. B. C. Cuong and H. V. Long, "An approach to the functions approximation problems by Mamdani fuzzy system", *Proc. of IEEE, 10th Intl. Conf. on Control, Automation, Robotics and Vision*, 850 - 855 (2008).
11. B. C. Cuong and H. V. Long, "On the approximate realization of a class of stochastic processes by spline functions fuzzy systems", *Advances in Fuzzy Mathematics*, **5**, 47 - 64 (2010).
12. B. C. Cuong, N. C. Luong and H. V. Long, "Approximation properties of fuzzy systems for multi-variables functions", *Pan - American Mathematical Journal*, **20**, 97 - 113 (2010).
13. Y. Ding, H. Ying, "Necessary conditions on minimal system configuration for general MISO Mamdani fuzzy systems as universal approximators", *IEEE Transactions Fuzzy Systems*, **30**, 857 - 863 (2000).
14. S. Mitaim, B. Kosko, "The shape of fuzzy sets in adaptive function approximation", *IEEE Transactions on Fuzzy Systems*, **9**, 637 - 656 (2001).
15. D. Tikk, L. T. Koczy, T.D. Gedeon, "A survey on the universal approximation and its limits in soft computing techniques", *International Journal of Approximate Reasoning*, **33**, 185 - 202 (2003).
16. L. X. Wang, "Fuzzy systems are universal approximators", *Proc. of the IEEE Int. Conf. on Fuzzy Systems, San Diego*, 1163-1169 (1992).
17. L. X. Wang, "Universal approximation by hierarchical fuzzy systems", *Fuzzy Sets and Systems*, **93**, 223 - 230 (1998).
18. L. X. Wang, "Analysis and design of hierarchical fuzzy systems", *IEEE Transactions on Fuzzy Systems*, **7**, 617-624 (1999).
19. Hao Ying, "General SISO Takagi-Sugeno fuzzy systems with linear rule consequent are universal approximators", *IEEE Transactions on Fuzzy Systems*, **6**, 582 - 587 (1998).
20. X. J. Zeng, M. G. Sing, "Approximation theory of fuzzy systems - MIMO case", *IEEE Transactions on Fuzzy Systems*, **3**, 219-235 (1995).
21. X. J. Zeng, M. G. Singh, "Approximation accuracy analysis of fuzzy systems as function approximators", *IEEE Transactions on Fuzzy Systems*, **4**, 44 - 63 (1996).
22. J. L. Castro, C. J. Mantas, J. M. Benítez, "Neural networks with a continuous squashing function in the output are universal approximators", *Neural Networks*, **13**, 561-563 (2000).
23. V. Kůrková, "Kolmogorov's theorem and multi-layer neural networks", *Neural Networks*, **5**, 501-506 (1992).
24. V. Kůrková, "Approximation of functions by perceptron networks with bounded number of hidden units", *Neural Networks*, **8**, 745-750 (1995).
25. A. E. Ruano, C. Cabrita, J. V. Oliveira, L. T. Kóczy, "Supervised training algorithms for B-Spline neural networks and neuro-fuzzy systems", *International Journal of Systems Science*, **33**, 1464-5319 (2002).
26. W. Rudin, *Real and Complex Analysis*, New York: McGraw-Hill, 1987.
27. J. R. Thompson and R. A. Tapia, *Nonparametric function, Modeling and Simulation*, Society for Industrial and Applied Mathematics - Philadelphia, USA, 1990.
28. P. L. Butzer, R. J. Nessel, *Fourier Analysis and Approximation*, Vol. 1, Birkhäuser Press, West Germany, 1971
29. E. H. Ruspini, P. P. Bonissone and W. Pedrycz, *Handbook of Fuzzy Computation*, Institute of Physics Publishing, Bristol and Philadelphia, 1986.
30. A. Dvoretzky, J. Kiefer, J. Wolfowitz, "Asymptotic minimax character of the sample distribution function

- and of the classical multinomial estimator”, *Annals of Mathematical Statistics*, **27**, 642 - 669 (1956).
31. A. N. Shiryaev, *Probability*, Second Edition, Springer-Verlag, 1996.
  32. A. I. Stepanets, *Methods of Approximation Theory*, V.S.P. Intl Science, 2005.
  33. J. Szabados, “Direct and converse approximation the-  
orems for Shepard operator”, *J. Approx. Theory Appl.*, **7**, 63-76 (1991).
  34. R. Rovatti, “Fuzzy Piecewise Multilinear and Piece-  
wise Linear Systems as Universal Approximations in  
Sobolev Norms”, *IEEE Transactions on fuzzy systems*, **6**, 235-249 (1998).