Fuzzy Systems Are Universal Approximators for a Smooth Function And Its Derivatives

Vladik Kreinovich[1], Hung T. Nguyen[2], and Yeung Yam[3]

[1]Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA
email vladik@utep.edu

[2]Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003, USA
email hunguyen@nmsu.edu

[3]Department of Mechanical &
Automation Engineering
The Chinese University of Hong Kong
Shatin, NT, Hong Kong, China
email yyam@mae.cuhk.edu.hk

Abstract

One of the reasons why fuzzy methodology is successful is that fuzzy systems are universal approximators, i.e., that we can approximate an arbitrary continuous function within any given accuracy by a fuzzy system. In some practical applications (e.g., in control), it is desirable to approximate not only the original function, but also its derivatives (so that, e.g., a fuzzy control approximating a smooth control will also be smooth). In our paper, we show that for any given accuracy, we can approximate an arbitrary smooth function by a fuzzy systems so that not only the function is approximated within this accuracy, but its derivatives are approximated as well. In other words, we prove that fuzzy systems are universal approximators for smooth functions and their derivatives.

Introduction. It is known that fuzzy systems are universal approximators; this result was proven almost simultaneously, in 1990–92 papers by J. Buckley, Z. Cao, E. Czogala, D. Dubois, M. Grabisch, J. Han, Y. Hayashi, C.-C. Jou, A. Kandel, B. Kosko, J. Mendel, H. Prade, and L.-X. Wang (see [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 19, 20, 22, 23, 24, 25, 26, 27, 28, 29, 31, 32, 36, 37, 38, 39, 40, 41, 42, 43, 44] and a survey [30]). To be more precise, it was proven that any input-output system can be approximated, within any given accuracy, by a system described by fuzzy rules. Such fuzzy-rule representation has two major advantages:

- first, fuzzy rules (in contrast to, say, differential equations) are intuitively clear;
- second, fuzzy rule representation is naturally parallelizable.

Due to these advantages, fuzzy rules work well in many practical applications.

However, in some applications, the existing fuzzy rule approximation techniques are not sufficient, because in these problems (e.g., in many control applications), derivatives of the approximated function are very important, and so, we want not only the approximating function be close to the approximated one, but we also want their derivatives to be close.

> For example, when we approximate a smooth control, we want the approximation to be smooth as well.

However, standard fuzzy approximation techniques do not guarantee the accuracy of approximating a derivative.

> It is worth mentioning that in contrast to fuzzy system, neural networks are known to be universal approximators which can approximate an arbitrary smooth function together with its derivatives (see, e.g., [17, 18]).

In this paper, we prove that (similar to neural networks) fuzzy systems are universal approximators not only for the approximated functions themselves, but also for their derivatives.

Basic formulas and notations. In this paper, we follow [36, 38], and consider fuzzy systems in which "and" is represented by an algebraic product, aggregation is represented by sum, membership functions for the input are Gaussian, and outputs are crisp. For such systems, the input-output function is given by a formula

$$g(x_1, \ldots, x_n) = \frac{N(x_1, \ldots, x_n)}{D(x_1, \ldots, x_n)}, \tag{1a}$$

where

$$N(x_1, \ldots, x_n) =$$
$$w_1 \cdot \mu_{1,1}(x_1) \cdot \ldots \cdot \mu_{1,n}(x_n) + \ldots + w_m \cdot \mu_{m,1}(x_1) \cdot \ldots \cdot \mu_{m,n}(x_n), \tag{1b}$$

$$D(x_1,\ldots,x_n) =$$
$$\mu_{1,1}(x_1)\cdot\ldots\cdot\mu_{1,n}(x_n)+\ldots+\mu_{m,1}(x_1)\cdot\ldots\cdot\mu_{m,n}(x_n), \tag{1c}$$

w_k are real numbers, and $\mu_{k,i}(x)$ are Gaussian functions.

In the following text, we follow a convenient notation (widely used in physics) of denoting a partial derivative of a function f with respect to a variable x_i by $f_{,i}$, and, correspondingly, a partial derivative with respect to x_i and x_j by $f_{,ij}$. A general derivative has the form $f_{,\mathcal{D}}$, where $\mathcal{D} = i_1 \ldots i_k$ is a sequence of indices (i.e., the form $f_{,i_1\ldots i_k}$).

Definition 1. *Let d be an integer, let $\varepsilon > 0$ be a real numbers. We say that a function $g(x_1,\ldots,x_n)$ approximates a function $f(x_1,\ldots,x_n)$ and its derivatives of order $\le d$ with accuracy ε if for all $x_i \in [-\Delta,\Delta]$,*

$$|f(x_1,\ldots,x_n) - g(x_1,\ldots,x_n)| \le \varepsilon,$$

and for every derivative $\mathcal{D}$ of order $\le d$,

$$|f_{,\mathcal{D}}(x_1,\ldots,x_n) - g_{,\mathcal{D}}(x_1,\ldots,x_n)| \le \varepsilon.$$

Theorem. *Let d and n be integers, let $\Delta > 0$ and $\varepsilon > 0$ be real numbers, and let $f(x_1,\ldots,x_n)$ be a d-times differentiable function on $[-\Delta,\Delta]^n$. Then, there exists a function $g(x_1,\ldots,x_n)$ of type* (1a)–(1c) *which approximates $f(x_1,\ldots,x_n)$ and its derivatives of order $\le d$ with accuracy ε.*

In other words, *fuzzy systems are universal approximators for smooth functions and their derivatives.*

Proof. We want this proof to be natural. Therefore, at first, we start with the analysis of how this theorem can be proved, and then, we will transform this analysis into the actual proof of the theorem.

By definition, a general Gaussian function has the form

$$\mu(x) = \exp\left(-\frac{(x-a)^2}{\sigma^2}\right)$$

for some a and $\sigma > 0$. In this proof, we will fix the value σ and consider only the Gaussian functions with this particular value of σ. For such functions,

$$\mu_{k,1}(x_1)\cdot\ldots\cdot\mu_{k,n}(x_n) = \exp\left(-\frac{(x_1-a_{k,1})^2}{\sigma^2}\right)\cdot\ldots\cdot\exp\left(-\frac{(x_n-a_{k,n})^2}{\sigma^2}\right) =$$

$$\exp\left(-\frac{(\vec{x}-\vec{a}^{\,(k)})^2}{\sigma^2}\right) = G(\vec{x}-\vec{a}^{\,(k)}),$$

where

$$G(\vec{t}) = \exp\left(-\frac{(\vec{t})^2}{\sigma^2}\right),$$

and $\vec{a}^{\,(k)}$ is a vector with components $(a_{k,1},\ldots,a_{k,n})$. For such membership functions, the expressions (1b) and (1c) take the form:

$$N(\vec{x})=\sum_k w_k\cdot G(\vec{x}-\vec{a}^{(k)}), \tag{2b}$$

$$D(\vec{x})=\sum_k G(\vec{x}-\vec{a}^{(k)}). \tag{2c}$$

Both expressions look like integral sums: namely, if we take, as values $\vec{a}^{\,(k)}$, all points on a dense rectangular grid with linear step $\Delta a_1=\ldots=\Delta a_n=h$ filling a box $[-N,N]\times\ldots\times[-N,N]$, and as w_k, the values $w_k=w(\vec{a}^{\,(k)})$ of some function $w(\vec{a})$ on this grid, then after multiplying by $\Delta a_1\cdot\ldots\cdot\Delta a_n=h^n$, the expressions (2b)–(2c) become the integral sums:

$$N(\vec{x})\cdot h^n\approx\int_{-N}^{N}\ldots\int_{-N}^{N}w(\vec{a})\cdot G(\vec{x}-\vec{a})\,\mathrm{d}\vec{a}, \tag{3b}$$

$$D(\vec{x})\cdot h^n\approx\int_{-N}^{N}\ldots\int_{-N}^{N}G(\vec{x}-\vec{a})\,\mathrm{d}\vec{a}. \tag{3c}$$

The ratio $N(\vec{x})/D(\vec{x})$ does not change if we multiply both numerator and denominator by h^n. Therefore, when $h\to 0$ and $N\to\infty$, the ratio $g(\vec{x})=N(\vec{x})/D(\vec{x})$ becomes closer and closer to the ratio $N_\infty(\vec{x})/G_\infty(\vec{x})$ of the two (multi-dimensional) infinite integrals:

$$N_\infty(\vec{x})=\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}w(\vec{a})\cdot G(\vec{x}-\vec{a})\,\mathrm{d}\vec{a}, \tag{4b}$$

$$D_\infty(\vec{x})=\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}G(\vec{x}-\vec{a})\,\mathrm{d}\vec{a}. \tag{4c}$$

By introducing a new auxiliary vector variable $\vec{b}=\vec{x}-\vec{a}$, we can see that the denominator integral is equal to

$$D_\infty(\vec{x})=\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}G(\vec{b})\,\mathrm{d}\vec{b}; \tag{5}$$

therefore, this integral does not depend on $\vec{x}$ at all: the function $D_\infty(\vec{x})$ is a (known) constant C. So, to find the weights that approximate the desired function $f(\vec{x})$, it is sufficient to find a function $w(\vec{a})$ for which

$$C\cdot f(\vec{x})=\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}w(\vec{a})\cdot G(\vec{x}-\vec{a})\,\mathrm{d}\vec{a}. \tag{6}$$

The right-hand side of this equality is known as a *convolution* of the functions $w(\vec{x})$ and $G(\vec{x})$; it is well known that the Fourier transform of a convolution

$R * S$ of two functions $R(\vec{x})$ and $S(\vec{x})$ is equal to the product of their Fourier transforms: if $T(\vec{x}) = R(\vec{x}) * G(\vec{x})$, then $\hat{T}(\vec{\omega}) = \hat{R}(\vec{\omega}) \cdot \hat{S}(\vec{\omega})$ (see, e.g., [33, 34]). Therefore,

$$C \cdot \hat{f}(\vec{\omega}) = \hat{w}(\vec{\omega}) \cdot \hat{G}(\vec{\omega}), \tag{7}$$

and so, a natural way to find the desired function $w(\vec{a})$ is to first find its Fourier transform $\hat{w}(\vec{\omega})$ from the above equality, i.e., as

$$\hat{w}(\vec{\omega}) = \frac{C \cdot \hat{f}(\vec{\omega})}{\hat{G}(\vec{\omega})}, \tag{8}$$

and then apply the inverse Fourier transform and get the desired function $w(\vec{a})$.

> *Warning:* this is not yet the proof itself, just the idea of it, because a Fourier transform is not always defined; the actual proof follows.

For a first derivative $g_{,i}(\vec{x})$ of the function $g(\vec{x})$, from the formulas (1), (2b), and (2c), we can deduce a similar representation:

$$g_{,i}(\vec{x}) = \frac{N_{,i}(\vec{x})}{D(\vec{x})} - \frac{N(\vec{x}) \cdot D_{,i}(\vec{x})}{D^2(\vec{x})}, \tag{9}$$

where

$$N_{,i}(\vec{x}) = \sum_k w_k \cdot G_{,i}(\vec{x} - \vec{a}^{(k)}) \approx \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} w(\vec{a}) \cdot G_{,i}(\vec{x} - \vec{a})\, \mathrm{d}\vec{a}, \tag{10b}$$

and

$$D_{,i}(\vec{x}) = \sum_k G_{,i}(\vec{x} - \vec{a}^{(k)}) \approx \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} G_{,i}(\vec{x} - \vec{a})\, \mathrm{d}\vec{a}. \tag{10c}$$

The last integral is equal to 0, so, to guarantee that the corresponding expression $g_{,i}(\vec{x})$ is close to the corresponding derivative $f_{,i}(\vec{x})$ of the function $f(\vec{x})$, we must make sure that the

$$C \cdot f_{,i}(\vec{x}) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} w(\vec{a}) \cdot G_{,i}(\vec{x} - \vec{a})\, \mathrm{d}\vec{a}. \tag{11}$$

In principle, if the equality (6) is true, then we can get (11) simply by differentiating both sides of this equality.

Similarly, to guarantee that this approximation approximates derivatives $f_{,\mathcal{D}}$ of higher order, we must guarantee that

$$C \cdot f_{,\mathcal{D}}(\vec{x}) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} w(\vec{a}) \cdot G_{,\mathcal{D}}(\vec{x} - \vec{a})\, \mathrm{d}\vec{a} \tag{12}$$

for this derivative $\mathcal{D}$. If we have attained the formula (6), then (12) follows by differentiating both sides of that formula.

Now, let us transform the above idea into the exact proof. We start with a smooth function $f(\vec{x})$ which is defined on an n-dimensional box $[-\Delta, \Delta] \times \ldots \times [-\Delta, \Delta]$. It is known (see, e.g., [21]), that we can extend this function f to a smooth function which is equal to 0 outside a larger box

$$[-2\Delta, 2\Delta] \times \ldots \times [-2\Delta, 2\Delta]. \tag{13}$$

So, without losing generality, we can assume that our function $f(\vec{x})$ is everywhere defined, everywhere smooth, and is equal to 0 outside the box (13).

Based on this function, we can now construct a new infinitely differentiable function $f_1(\vec{x})$ which is close to $f(\vec{x})$ and for which all the derivatives of order $\le d$ are close to the corresponding derivatives of $f(\vec{x})$. It is known, from the theory of generalized functions (also called *Schwartz distributions*) (see. e.g., [14, 35]), that as such a function, we can take

$$f_1(\vec{x}) = \frac{\int G_0(\vec{x} - \vec{y}) \cdot f(\vec{y})\, \mathrm{d}\vec{y}}{\int G_0(\vec{y})\, \mathrm{d}\vec{y}}, \tag{14}$$

where

$$G_0(\vec{y}) = \exp\left(-\frac{(\vec{y}\,)^2}{\sigma_0^2}\right) \tag{15}$$

and $\sigma_0 > 0$ is sufficiently small. (To be more precise, as $\sigma_0 \to 0$, we have $G_0(\vec{y}) \to \delta(\vec{y})$ (modulo a multiplicative constant) and thus, $f_1 = G_0 * f \to \delta * f = f$.) The new function $f_1(\vec{x})$ is infinitely differentiable: indeed, its derivative of each order $\mathcal{D}$ can be computed as

$$f_{1,\mathcal{D}}(\vec{x}) = \frac{\int G_{0,\mathcal{D}}(\vec{x} - \vec{y}) \cdot f(\vec{y})\, \mathrm{d}\vec{y}}{\int G_0(\vec{y})\, \mathrm{d}\vec{y}}, \tag{16}$$

and the convergence of these integrals follows from the fact that $f(\vec{y})$ is different from 0 only within a box. Since the values of the function f_1 and of its derivatives can be made (by choosing appropriate σ_0) arbitrarily close to the values of function f and its derivatives, it is sufficient to be able to approximate the function f_1, then this approximation will approximate the original function f as well.

One can also easily check that the function $f_1(\vec{x})$ tends to 0 as $\vec{x} \to \infty$ (and decreases fast), so that its Fourier transform can be defined for all $\vec{\omega}$:

$$\hat{f}_1(\vec{\omega}) = \frac{1}{(2\pi)^{n/2}} \cdot \int f_1(\vec{x}) \cdot \exp(-\mathrm{i}\,\vec{\omega} \cdot \vec{x})\, \mathrm{d}\vec{x}. \tag{17}$$

From this Fourier transform, we can reconstruct the function $f_1(\vec{x})$ back by applying the inverse Fourier transform:

$$f_1(\vec{x}) = \frac{1}{(2\pi)^{n/2}} \cdot \int \hat{f}_1(\vec{\omega}) \cdot \exp(\mathrm{i}\,\vec{\omega} \cdot \vec{x})\, \mathrm{d}\vec{\omega}. \tag{18}$$

In the formula (18), we consider all possible values of $\vec{\omega}$. Let us show that we can restrict the integration by values $\vec{\omega}$ of a bounded length ($|\vec{\omega}| \le R$ for some R), and get a new function $f_2(\vec{x})$ which is itself close to $f_1(\vec{x})$, and its derivatives of orders $\le d$ are close to the corresponding derivatives of $f_1(\vec{x})$. This function $f_2(\vec{x})$ is, thus, defined by the following formula:

$$f_2(\vec{x}) = \frac{1}{(2\pi)^{n/2}} \cdot \int_{|\vec{\omega}|\le R} \hat{f}_1(\vec{\omega}) \cdot \exp(\mathrm{i}\,\vec{\omega}\cdot\vec{x})\,\mathrm{d}\vec{\omega}, \tag{19}$$

or, in terms of Fourier transforms, by the formulas $\hat{f}_2(\vec{\omega}) = \hat{f}_1(\vec{\omega})$ if $|\vec{\omega}| \le R$ and $f_2(\vec{\omega}) = 0$ if $|\vec{\omega}| > R$.

It is known that if a function $h(\vec{x})$ is differentiable with respect to one of the variables x_i, then, differentiating by parts, we conclude that

$$\hat{h}(\vec{\omega}) = \frac{1}{(2\pi)^{n/2}} \cdot \int h(\vec{x}) \cdot \exp(-\mathrm{i}\,\vec{\omega}\cdot\vec{x})\,\mathrm{d}\vec{x} =$$

$$-\frac{1}{(2\pi)^{n/2}} \cdot \int h_{,i}(\vec{x}) \cdot \frac{\exp(-\mathrm{i}\,\vec{\omega}\cdot\vec{x})}{-\mathrm{i}\,\omega_i}\,\mathrm{d}\vec{x}, \tag{20}$$

and therefore,

$$|\hat{h}(\omega)| \le \frac{1}{(2\pi)^{n/2}} \cdot \frac{1}{|\omega_i|} \cdot \left| \int h_{,i}(\vec{x}) \cdot \exp(-\mathrm{i}\,\vec{\omega}\cdot\vec{x})\,\mathrm{d}\vec{x} \right| \le$$

$$\frac{1}{(2\pi)^{n/2}} \cdot \frac{1}{|\omega_i|} \cdot \int |h_{,i}(\vec{x})|\,\mathrm{d}\vec{x}. \tag{21}$$

So, if the integral $\int |h_{,i}(\vec{x})|\,\mathrm{d}\vec{x}$ converges, the Fourier transform goes to 0 at least as fast as $|\omega_i|^{-1}$ when $\vec{\omega} \to \infty$. Similarly, if a function is twice differentiable, and the corresponding integral converges, the Fourier transform goes to 0 as $|\omega_i|^{-1} \cdot |\omega_j|^{-1}$, etc. Our function $f_1(\vec{x})$ is differentiable infinitely many times, and the corresponding integrals do converge. Therefore, its Fourier transform goes to 0 faster than $|\vec{\omega}|^{-N}$ for any integer N, i.e., for some constant $C_1 > 0$, we have

$$|\hat{f}_1(\vec{\omega})| \le \frac{C_1}{|\vec{\omega}|^N}. \tag{22}$$

Therefore, for every $R > 0$, the difference between the full integral (18) (which defines $f_1(\vec{x})$) and its restriction (19) (which defines $f_2(\vec{x})$) can be bounded as follows:

$$f_1(\vec{x}) - f_2(\vec{x}) = \frac{1}{(2\pi)^{n/2}} \cdot \int_{|\vec{\omega}|>R} \hat{f}_1(\vec{\omega}) \cdot \exp(\mathrm{i}\,\vec{\omega}\cdot\vec{x})\,\mathrm{d}\vec{\omega}, \tag{23}$$

hence

$$|f_1(\vec{x}) - f_2(\vec{x})| \le \frac{1}{(2\pi)^{n/2}} \cdot \int_{|\vec{\omega}|>R} |\hat{f}_1(\vec{\omega})|\,\mathrm{d}\vec{\omega} \le \frac{1}{(2\pi)^{n/2}} \cdot \int_{|\vec{\omega}|>R} \frac{C_1}{|\vec{\omega}|^N}\,\mathrm{d}\vec{\omega}. \tag{24}$$

If we turn to radial coordinates, with radius r and angles $\theta_1, \ldots$, then the bounding integral in the right-hand side of (24) turns into

$$|f_1(\vec{x}) - f_2(\vec{x})| \le \text{const} \cdot \int_R^\infty \frac{r^{n-1}\,\mathrm{d}r}{r^N} = \frac{\text{const}}{R^{N-n}}. \tag{25}$$

So, if we take $N > n$, then, for sufficiently large R, we can get a sufficiently small upper bound which bounds the difference between the values of the functions $f_1(\vec{x})$ and $f_2(\vec{x})$ for all $\vec{x}$.

The Fourier transform of a derivative $f_{1,i}$ is equal to $\mathrm{i}\,\omega_i$ times the Fourier transform of the original function. Therefore, the Fourier transforms of all the derivatives of the function f_1 also tend to 0 faster than $|\vec{\omega}|^{-N}$ for any integer N. Therefore, by choosing R sufficiently large, we can guarantee that not only the values of the function $f_2(\vec{x})$ are close to the values of the function $f_1(\vec{x})$, but also that the values of all derivatives of $f_2(\vec{x})$ of order $\le d$ are close to the corresponding derivatives of $f_1(\vec{x})$.

Therefore, to prove that we can approximate $f_1(\vec{x})$, it is sufficient to be able to approximate the close function $f_2(\vec{x})$. This function $f_2(\vec{x})$ has an advantage – that its Fourier transform is equal to 0 outside a sphere $|\vec{\omega}| \le R$. We can therefore define the "weight" function $w(\vec{x})$ by applying the above-described idea of defining $w(\vec{x})$ to the function $f_2(\vec{x})$: namely, we first define its Fourier transform

$$\hat{w}(\vec{\omega}) = \frac{\hat{f}_2(\vec{\omega})}{\hat{G}(\vec{\omega})}, \tag{26}$$

and then reconstruct $w(\vec{x})$ by applying inverse Fourier transform:

$$w(\vec{x}) = \frac{1}{(2\pi)^{n/2}} \cdot \int \hat{w}(\vec{\omega}) \cdot \exp(\mathrm{i}\,\vec{\omega} \cdot \vec{x})\,\mathrm{d}\vec{\omega}. \tag{27}$$

For this function $w(\vec{x})$, we have $f_2 = w * G$.

Since the Fourier transform of a function $w(\vec{x})$ is equal to 0 outside the bounded area, this function $w(\vec{x})$ is infinitely differentiable and therefore, all integrals of the type $\int w(\vec{a}) \cdot G(\vec{x} - \vec{a})\,\mathrm{d}\vec{a}$ can be approximated by grid-based integral sums, for appropriate h and N. Thus, for appropriately small h and large N, we get a fuzzy system $g(\vec{x})$ of the type (1a)–(1c) which approximates the function $f_2(\vec{x})$ together with all its derivatives of orders $\le d$. Since the function $f_2(\vec{x})$, in its turns, approximates the function $f_1(\vec{x})$ and its derivatives, and the function $f_1(\vec{x})$ approximates the original function $f(\vec{x})$ and its derivatives, we conclude that the fuzzy system $g(\vec{x})$ approximates the original smooth function $f(\vec{x})$ and its derivatives. The theorem is proven.

Remaining open problem. We have shown that if we use *Gaussian* membership functions, then we can approximate an arbitrary smooth function together with its derivatives. Our proof make an essential use of the fact that the membership functions are Gaussian. An interesting open question is: if we use use smooth *non-Gaussian* membership functions, will we still be able to get a universal approximation for a function and for its derivatives?

Acknowledgments. This work was supported in part by NASA under cooperative agreement NCC5-209, by NSF grant No. DUE-9750858, by the United Space Alliance, grant No. NAS 9-20000 (PWO C0C67713A6), by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant number F49620-95-1-0518, by the National Security Agency under Grant No. MDA904-98-1-0564, and by the Hong Kong RGC grant 4138/97E.

Part of this work was conducted while one of the authors (V.K.) was visiting the Department of Mechanical and Automation Engineering at the Chinese University of Hong Kong under the support of the Hong Kong RGC grant 4138/97E.

References

[1] J. J. Buckley, "Universal Fuzzy Controllers", *Automatica*, 1992, Vol. 28, pp. 1245–1248.

[2] J. J. Buckley, "Controllable processes and the fuzzy controller", *Fuzzy Sets and Systems*, 1993, Vol. 53, pp. 27–31.

[3] J. J. Buckley, "Sugeno Type Controllers are Universal Controllers", *Fuzzy Sets and Systems,* 1993, Vol. 53, pp. 299–303.

[4] J. J. Buckley, "Approximation paper: Part I", *Proceedings of the Third International Workshop on Neural Networks and Fuzzy Logic, Houston, TX, June 1–3, 1992*, NASA, January 1993, Vol. I (NASA Conference Publication No. 10111), pp. 170–173.

[5] J. J. Buckley, "Applicability of the fuzzy controller", In: P. Z. Wang and K. F. Loe (eds.), *Advances in Fuzzy Systems: Application and Theory*, World Scientific, Singapore, 1993.

[6] J. J. Buckley and E. Czogala, "Fuzzy models, fuzzy controllers, and neural nets", *Proc. Polish Academy of Sciences*, 1993.

[7] J. J. Buckley and Y. Hayashi, "Fuzzy input-output controllers are universal approximators", *Fuzzy Sets and Systems,* 1993, Vol. 58, pp. 273–278.

[8] J. J. Buckley, Y. Hayashi, and E. Czogala, "On the equivalence of neural nets and fuzzy expert systems", *Proc. of Int. Joint Conf. on Neural Networks*, June 7–11, 1992, Baltimore, MD, Vol. 2, pp. 691–695.

[9] J. J. Buckley, Y. Hayashi, and E. Czogala, "On the equivalence of neural nets and fuzzy expert systems", *Fuzzy Sets and Systems,* 1993, Vol. 53, pp. 129–134.

[10] Z. Cao, "Mathematical principle of fuzzy reasoning", *Proceedings NAFIPS'90*, Toronto, Canada, June 1990, pp. 362–365.

[11] Z. Cao, A. Kandel, and J. Han, "Mechanism of fuzzy logic controller", *Proceedings of ISUMA'90*, Univ. of Maryland, December 1990, pp. 603–607.

[12] J. L. Castro, "Fuzzy logic controllers are universal approximators", *IEEE Trans. Syst., Man, Cybern.*, 1995, Vol. 25, No. 4, pp. 629–635.

[13] D. Dubois, M. Grabisch, and H. Prade, "Gradual rules and the approximation of functions", *Proceedings of the 2nd International Conference on Fuzzy Logic and Neural Networks*, Iizuka, Japan, July 17–22, 1992, pp. 629–632.

[14] I. M. Gelfand and G. E. Shilov, *Generalized functions*, Vol. 1, Academic Press, N. Y. and London, 1964.

[15] Y. Hayashi, J. J. Buckley, and E. Czogala, "Fuzzy expert systems versus neural networks", *Proc. of Int. Joint Conf. on Neural Networks*, June 7–11, 1992, Baltimore, MD, Vol. 2, pp. 720–726.

[16] Y. Hayashi, J. J. Buckley, and E. Czogala, "Approximations between fuzzy expert systems and neural networks", *Proc. of the 2nd Int. Conf. on Fuzzy Logic and Neural Networks*, July 17-22, 1992, Iizuka, Japan, pp. 135–139.

[17] K. Hornik, "Approximation capabilities of multilayer feedforward networks", *Neural Networks*, 1991, Vol. 4, pp. 251–257.

[18] K. Hornik, M. Stinchcombe, H. White, "Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks", *Neural Networks*, 1990, Vol. 3, pp. 551–560.

[19] C.-C. Jou, "On the mapping capabilities of fuzzy inference systems", *Proceedings of the International Joint Conference on Neural Networks*, Baltimore, Maryland, June 7–11, 1992, Vol. 2, pp. 708–713.

[20] S. Kawamoto, K. Tada, N. Onoe, A. Ishigame, A., and T. Taniguchi, "Construction of exact fuzzy system for nonlinear system and its stability analysis", *8th Fuzzy System Symposium*, Hiroshima, 1992, pp. 517–520 (in Japanese).

[21] J. Kelley, *General topology*, Van Nostrand, Princeton, NJ, 1955.

[22] B. Kosko, *Neural networks and fuzzy systems: a dynamical systems approach to machine intelligence*, Prentice Hall, 1991.

[23] B. Kosko, "Fuzzy Systems as Universal Approximators", *IEEE Int. Conf. on Fuzzy Systems,* San Diego, CA, March 1992, pp. 1143–1162.

[24] B. Kosko, "Fuzzy function approximation", *Proceedings of the International Joint Conference on Neural Networks*, Baltimore, Maryland, June 7–11, 1992, Vol. 1, pp. 209–213.

[25] B. Kosko, "Fuzzy Systems as Universal Approximators", *IEEE Trans. on Computers,* 1994, Vol. 43, No. 11, pp. 1329–1333.

[26] B. Kosko, "Optimal fuzzy rules cover extrema", *Proceedings of the World Congress on Neural Networks WCNN'94*, 1994.

[27] B. Kosko, "Optimal fuzzy rules cover extrema", *International Journal of Intelligent Systems*, 1995, Vol. 10, No. 2, pp. 249–255.

[28] B. Kosko, "Additive fuzzy systems: from function approximation to learning", In: C. H. Chen (ed.), *Fuzzy Logic and Neural Network Handbook*, McGraw-Hill, N.Y., 1996, pp. 9-1–9-22.

[29] B. Kosko and J. A. Dickerson, "Function approximation with additive fuzzy systems", Chapter 12 in: H. T. Nguyen, M. Sugeno, R. Tong, and R. Yager (eds.), *Theoretical aspects of fuzzy control*, J. Wiley, N.Y., 1995, pp. 313–347.

[30] V. Kreinovich, G. C. Mouzouris, and H. T. Nguyen, "Fuzzy rule based modeling as a universal control tool", In: H. T. Nguyen and M. Sugeno (eds.), Fuzzy Systems: Modeling and Control, Kluwer, Boston, MA, 1998, pp. 135-195.

[31] H. T. Nguyen and V. Kreinovich, *On approximation of controls by fuzzy systems*, Technical Report 92-93/302, LIFE Chair of Fuzzy Theory, Tokyo Institute of Technology, 1992.

[32] H. T. Nguyen and V. Kreinovich, "On Approximation of Controls by Fuzzy Systems", *Proceedings of Fifth IFSA Congress,* Seoul, Korea, 1993, Vol. 2, pp. 1414–1417.

[33] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1989.

[34] C. Van Loan, *Computational Frameworks for the Fast Fourier Transform*, SIAM, Philadelphia, 1992.

[35] V. S. Vladimirov, *Equations of Mathematical Physics,* Marcel Dekker, N.Y., 1971.

[36] L.-X. Wang, “Fuzzy Systems Are Universal Approximators”, *Proceedings of Second IEEE International Conference on Fuzzy Systems,* San Diego, CA, March 1992, pp. 1163–1170.

[37] L.-X. Wang, *Adaptive Fuzzy Systems and Control,* Prentice-Hall, Englewood-Cliffs, NJ, 1994.

[38] L.-X. Wang and J. L. Mendel, *Generating fuzzy rules from numerical data with applications*, University of Southern California, Signal and Image Processing Institute, Technical Report USC-SIPI # 169, 1991.

[39] L.-X. Wang and J. M. Mendel, “Fuzzy basis functions, universal approximation, and orthogonal least-squares learning”, *IEEE Transactions on Neural Networks*, 1992, Vol. 3, pp. 807–814.

[40] L.-X. Wang and J. M. Mendel, “Generating fuzzy rules by learning from examples”, *IEEE Transactions on Systems, Man, and Cybernetics*, 1992, Vol. 22, pp. 1414–1417.

[41] R. R. Yager and D. P. Filev, *Essentials of fuzzy modeling and control*, J. Wiley & Sons, 1994.

[42] H. Ying, “Sufficient conditions on general fuzzy systems as function approximators”, *Automatica*, 1994, Vol. 30, No. 3, pp. 521–525.

[43] X.-J. Zeng and M. G. Singh, “Approximation theory of fuzzy systems - SISO case”, *IEEE Trans. on Fuzzy Systems,* 1994, Vol. 2, No. 2, pp. 162–176.

[44] X.-J. Zeng and M. G. Singh, “Approximation theory of fuzzy systems - MIMO case”, *IEEE Trans. on Fuzzy Systems,* 1995, Vol. 3, pp. 219–235.