Proceedings of the 2007 IEEE Symposium on
Foundations of Computational Intelligence (FOCI 2007)

Fuzzy Aggregation Techniques in Situations Without Experts: Towards A New Justification

Hung T. Nguyen
Department of Mathematical Sciences
New Mexico State University
Las Cruces, New Mexico, 88003, USA
Email: hunguyen@nmsu.edu

Vladik Kreinovich
Department of Computer Science
University of Texas at El Paso
El Paso, Texas 79968, USA
Email: vladik@utep.edu

***Abstract*— Fuzzy techniques have been originally invented as a methodology that transforms the knowledge of experts formulated in terms of natural language into a precise computer-implementable form. There are many successful applications of this methodology to situations in which expert knowledge exist, the most well known is an application to fuzzy control.**

In some cases, fuzzy methodology is applied even when no expert knowledge exists: instead of trying to approximate the unknown control function by splines, polynomials, or by any other traditional approximation technique, researchers try to approximate it by guessing and tuning the expert rules. Surprisingly, this approximation often works fine.

In this paper, we give a mathematical explanation for this phenomenon, and show that approximation by using fuzzy methodology is indeed (in some reasonable sense) the best.

I. Introduction

A. Fuzzy techniques: a brief reminder

Fuzzy techniques have been originally invented as a methodology that transforms the knowledge of experts formulated in terms of natural language into a precise computer-implementable form. There are many successful applications of this methodology to situations in which expert knowledge exist, the most well known is an application to fuzzy control; see, e.g., [2], [3], [10].

B. Universal approximation results

A guarantee of success comes from the fact that fuzzy systems are *universal approximators* in the sense that for every continuous function $f(x_1, \ldots, x_n)$ and for every $\varepsilon > 0$, there exists a set of rules for which the corresponding input-output function is ε-close to f; see, e.g., [1], [3], [4], [6], [7], [9], [10], [11], [12], [13], [14] and references therein.

C. Fuzzy methodology is sometimes successful without any expert knowledge

In some cases, fuzzy methodology is applied even when no expert knowledge exists: instead of trying to approximate the unknown control function by splines, polynomials, or by any other traditional approximation technique, researchers try to approximate it by guessing and tuning the expert rules. Surprisingly, this approximation often works fine.

D. What we plan to do

In this paper, we give a mathematical explanation for this phenomenon, and we show that approximation by using fuzzy methodology is indeed (in some reasonable sense) the best.

Comment. In this paper, we build upon our preliminary results published in [8].

II. In Many Practical Applications, Data Processing Speed Is Important

We have mentioned that one of the main applications of fuzzy methodology is to intelligent control.

In applications to automatic control, the computer must constantly compute the current values of control. The value of the control depends on the state of the controlled object (called *plant* in control theory). So, to get a high quality control, we must measure as many characteristics $x_1, \ldots, x_n$ of the current state as we can. The more characteristics we measure, the more numbers we have to process, so, the more computation steps we must perform. The results of these computations must be ready in no time, before we start the next round of measurements. So, automatic control, especially high-quality automatic control, is a real-time computation problem with a serious time pressure.

III. Parallel Computing Is An Answer

A natural way to increase the speed of the computations is to perform computations *in parallel* on several processors. To make the computations really fast, we must divide the algorithm into parallelizable steps, each of which requires a small amount of time.

What are these steps?

IV. Description of the Fastest Possible Control-Oriented Parallel Computer

A. The fewer variables, the faster

As we have already mentioned, the main reason why control algorithms are computationally complicated is that we must process many inputs. For example, controlling a car is easier than controlling a plane, because the plane (as a 3-D object) has more characteristics to take care of, more characteristics to measure and hence, more characteristics to process. Controlling a space shuttle, especially during the lift-off and landing, is even a more complicated task, usually performed by several groups of people who control the trajectory, temperature, rotation, etc. In short, the more numbers we need to process, the more complicated the algorithm. Therefore, if we want to decompose our algorithm into fastest possible modules, we must make each module to process as few numbers as possible.

B. Functions of one variable are not sufficient

Ideally, we should only use the modules that compute functions of one variable. However, if we only have functions of one variables (i.e., procedures with one input and one output), then, no matter how we combine them, we will always end up with functions of one variable. Since our ultimate goal is to compute the control function $u = f(x_1, \ldots, x_n)$ that depends on many variables $x_1, \ldots, x_n$, we must therefore enable our processors to compute at least one function of two or more variables.

What functions of two variables should we choose?

C. Choosing functions of two or more variables

Inside the computer, each function is represented as a sequence of hardware implemented operations. The fastest functions are those that are computed by a single hardware operation. The basic hardware supported operations are: arithmetic operations $a+b$, $a-b$, $a \cdot b$, a/b, and $\min(a,b)$ and $\max(a,b)$. The time required for each operation, crudely speaking, corresponds to the number of bits operations that have to be performed:

- Division is done by successive multiplication, comparison and subtraction (basically, in the same way as we do it manually), so, it is a much slower operation than $-$.
- Multiplication is implemented as a sequence of additions (again, basically in the same manner as we do it manually), so it is much slower than $+$.
- $-$ and $+$ are usually implemented in the same way. To add two n-bit binary numbers, we need n bit additions, and also potentially, n bit additions for carries. Totally, we need about $2n$ bit operations.
- min of two n-bit binary numbers can be done in n binary operations: we compare the bits from the highest to the lowest, and as soon as they differ, the number that has 0 as opposed to 1 is the desired minimum: e.g., the minimum of 0.10101 and 0.10011 is 0.10011, because in the third bit, this number has 0 as opposed to 1.
- Similarly, max is an n-bit operation.

So, the fastest possible functions of two variables are min and max. Similarly fast is computing the minimum and maximum of several (more than two) real numbers. Therefore, we will choose these functions for our control-oriented computer.

Summarizing the above-given analysis, we can conclude that our computer will contain modules of two type:

- modules that compute functions of one variable;
- modules that compute min and max of two or several numbers.

D. How to combine these modules?

We want to combine these modules in such a way that the resulting computations are as fast as possible. The time that is required for an algorithm is crudely proportional to the number of sequential steps that it takes. We can describe this number of steps in clear geometric terms:

- at the beginning, the input numbers are processed by some processors; these processors form the *first layer* of computations;
- the results of this processing may then go into different processors, that form the *second layer*;
- the results of the second layer of processing go into the *third layer*,
- etc.

In these terms, the fewer layers the computer has, the faster it is.

So, we would like to combine the processors into the smallest possible number of layers.

Now, we are ready for the formal definitions.

V. Definition and the Main Result

Let us first give an inductive definition of what it means for a function to be computable by a k-layer computer.

Definition.

- *We say that a function* $f(x_1, \ldots, x_n)$ *is* computable by a 1-layer computer *if either* $n = 1$, *or the function* f *coincides with* min *or with* max.
- *Let* $k \geq 1$ *be an integer. We say that a function* $f(x_1, \ldots, x_n)$ *is* computable by a $(k+1)$-layer computer *if one of the following three statements is true:*
  - $f = g(h(x_1, \ldots, x_n))$, *where* g *is a function of one variable, and* $h(x_1, \ldots, x_n)$ *is computable by a* k-*layer computer;*
  - $f = \min(g_1(x_1, \ldots, x_n), \ldots, g_m(x_1, \ldots, x_n))$, *where all functions* g_i *are computed by a* k-*layer computer;*
  - $f = \max(g_1(x_1, \ldots, x_n), \ldots, g_m(x_1, \ldots, x_n))$, *where all functions* g_i *are computed by a* k-*layer computer.*

Comment. A computer is a finite-precision machine, so, the results of the computations are never absolutely precise. Also, a computer is limited in the size of its numbers. So, we can only compute a function approximately, and only on a limited range. Therefore, when we say that we can compute an arbitrary function, we simply mean that for an arbitrary range T, for an arbitrary continuous function $f : [-T,T]^n \to R$, and for an arbitrary accuracy $\varepsilon > 0$, we can compute a function that is ε-close to f on the given range. In this sense, we will show that not every function can be computed on a 2-layer computer, but that 3 layers are already sufficient.

Proposition. *There exist real numbers* T *and* $\varepsilon > 0$, *and a continuous function* $f : [-T,T]^n \to R$ *such that no function* ε-*close to* f *on* $[-T,T]^n$ *can be computed on a 2-layer computer.*

Comment. To make the text more readable, we present both proofs in the last section. However, we will make one comment here. The function that will be proved to be not computable on a 2-layer computer is not exotic at all: it is $f(x_1, x_2) = x_1 + x_2$ on the domain $[-1, 1]^2$, and the Proposition is true for $\varepsilon = 0.4$.

Theorem. *For every real numbers T and $\varepsilon > 0$, and for every continuous function $f : [-T, T]^n \to R$, there exists a function $\widetilde{f}$ that is ε-close to f on $[-T, T]^n$ and that is computable on a 3-layer computer.*

Comment. In other words, *functions computed by a 3-layer computer are universal approximators.*

Relation to fuzzy control. As we will see from the proof, the approximating function $\widetilde{f}$ is of the type $\max(A_1, \ldots, A_m)$, where $A_j = \min(f_{j1}(x_1), \ldots, f_{jn}(x_n))$. These functions correspond the so-called *fuzzy control* [2], [3], [10]: Indeed, let us define

$$U = \max_{i,j,x_i \in [-T,T]} |f_{ji}(x_i)|,$$

and

$$\mu_{ji}(x_i) = \frac{f_{ji}(x_i) - (-U)}{U - (-U)}.$$

Let us now assume that the rules base that describes the expert recommendations for control consists of exactly two rules:

- "if one of the conditions C_j is true, then $u = U$";
- "else, $u = -U$",

where each condition C_j means that the following n conditions are satisfied:

- x_1 satisfies the property C_{j1} (described by a membership function $\mu_{j1}(x_1)$);
- x_2 satisfies the property C_{j2} (described by a membership function $\mu_{j2}(x_2)$);
- …
- x_n satisfies the property C_{jn} (described by a membership function $\mu_{jn}(x_n)$).

In logical terms, the condition C for $u = U$ has the form

$$(C_{11} \& \ldots \& C_{1n}) \vee \ldots \vee (C_{k1} \& \ldots \& C_{kn}).$$

If we use min for &, and max for $\vee$ (these are the simplest choices in fuzzy control methodology), then the degree μ_C with which we believe in a condition $C = C_1 \vee \ldots \vee C_k$ can be expressed as:

$$\mu_C =$$
$$\max[\min(\mu_{11}(x_1), \ldots, \mu_{1n}), \ldots, \min(\mu_{k1}, \ldots, \mu_{kn})].$$

Correspondingly, the degree of belief in a condition for $u = -U$ is $1 - \mu_C$. According to fuzzy control methodology, we must use a *defuzzification* to determine the actual control, which in this case leads to the choice of

$$u = \frac{U \cdot \mu_C + (-U) \cdot (1 - \mu_C)}{\mu_C + (1 - \mu_C)}.$$

Because of our choice of μ_{ji}, one can easily see that this expression coincides exactly with the function $\max(A_1, \ldots, A_m)$, where $A_j = \min(f_{j1}(x_1), \ldots, f_{jn}(x_n))$. So, we get exactly the expressions that stem from the fuzzy control methodology.

Conclusion. Since our 3-layer expression describes the fastest possible computation tool, we can conclude that *for control problems, the fastest possible universal computation scheme corresponds to using fuzzy methodology.*

This result explains why fuzzy methodology is sometimes used (and used successfully) without any expert knowledge being present, as an extrapolation tool for the (unknown) function.

Comment. We have considered *digital* parallel computers. If we use *analog* processors instead, then min and max stop being the simplest functions. Instead, the sum is the simplest: if we just join the two wires together, then the resulting current is equal to the sum of the two input currents. In this case, if we use a sum (and more general, linear combination) instead of min and max, 3-layer computers are also universal approximators; the corresponding computers correspond to *neural networks* [5].

VI. Discussion

Universal approximation property – traditional justification of fuzzy controllers. If we are looking for a general control methodology, i.e., a methodology that enables us to implement (within a given accuracy) an arbitrary control strategy, then the functions corresponding to this methodology must be universal approximators.

From this viewpoint, the known fact that fuzzy controllers are universal approximators is one of the reasons why fuzzy controllers are indeed used in many practical situations.

Several other types of controllers also have a universal approximation property. The universal approximation character of fuzzy controllers does not imply, however, that fuzzy controllers are the only possible class of controllers – indeed, there are many other universal approximators, e.g., polynomials, neural networks, etc.

Hence, if our only requirement on the control methodology is that this methodology be general (universal), we can also use, e.g., (more traditional) polynomial controllers or neural controllers.

Empirical fact: fuzzy controllers are often better. From the viewpoint of the universal approximation property, traditional or fuzzy controllers are as good as fuzzy controllers. However, in many practical situations, fuzzy controllers perform better.

Fuzzy controllers not only lead to better control, they usually enable us to faster compute the desired control. In many practical situations, fuzzy controllers perform better. Better in what sense? In different practical situations, we may have different requirements to a controller and thus, different criteria for gauging how good a controller is. For example, we may want to look for a control which is smoother or which is more robust or which is more stable.

In some practical situations, fuzzy controllers do have these advantages: e.g., in many cases, a fuzzy controller is more robust than the traditional one. However, in most case, fuzzy controller is also computationally simpler and thus, its computations are much faster. For example, for non-linear systems, computing a fuzzy control requires an explicit use of simple functions, while, e.g., to apply a more traditional non-linear controller we may need to solve systems of equations.

We provide a theoretical explanation for this empirical phenomenon. In this paper, we explain that fuzzy controllers are indeed, in some reasonable sense, faster. This result explains the above empirical fact – that fuzzy controllers often enables us to compute control faster.

There may be other classes of fast controllers. The fact that fuzzy controllers are *among* the fastest does not necessarily mean that the class of fuzzy controllers is *the only* fastest class: there may be other non-fuzzy controllers which are also computable by a 3-layer computers.

In our proof of the theorem, we use 3-layer computers corresponding to fuzzy control, but there could be different proofs of the universal approximation property of 3-layer computers, proofs which would use different types of controllers.

In other words, while we prove that fuzzy controllers are a reasonable class, there may be other classes of controllers which are as reasonable (and as fast).

For example, in our proof – similarly to most proofs that fuzzy systems are universal approximators – the construction fuzzy system is based on the values of the function $f(x_1,\ldots,x_n)$ at different tuples $x^{(k)} = (x_1^{(k)},\ldots,x_n^{(k)})$. Fuzzy controllers provide a continuous transition between the corresponding values; however, instead, we can simply use the look-up table and assign, to each tuple x, the value $f(x^{(k)})$ at the nearest selected tuple $x^{(k)}$. This look-up table idea also leads to a computationally low lost – although the discontinuity of the resulting approximating piece-wise function is, in many practical applications, a definite disadvantage.

VII. Proofs

A. Proof of the Proposition

0°. Let us proof (by reduction to a contradiction) that if a function $\widetilde{f}(x_1,x_2)$ is 0.4–close to $f(x_1,x_2) = x_1 + x_2$ on $[-1,1]^2$, then $\widetilde{f}$ cannot be computed on a 2-layer computer. Indeed, suppose that it is. Then, according to the Definition, the function $\widetilde{f}(x_1,x_2)$ is of one of the following three forms:

- $g(h(x_1,x_2))$, where h is computable on a 1-layer computer;
- $\min(g_1(x_1,x_2),\ldots,g_m(x_1,x_2))$, where all the functions g_i are computable on a 1-layer computer;
- $\max(g_1(x_1,x_2),\ldots,g_m(x_1,x_2))$, where all the functions g_i are computable on a 1-layer computer.

Let us show case-by-case that all these three cases are impossible.

1°. In the first case, $\widetilde{f}(x_1,x_2) = g(h(x_1,x_2))$, where h is computable on a 1-layer computer. Be definition, this means that h is either a function of one variable, or min, or max. Let us consider all these three sub-cases.

1.1°. If $\widetilde{f}(x_1,x_2) = g(h(x_1))$, then the function $\widetilde{f}$ depends only on x_1. In particular,

$$\widetilde{f}(0,-1) = \widetilde{f}(0,1). \tag{1}$$

But since $\widetilde{f}$ is ε-close to $f(x_1+x_2) = x_1+x_2$, we get

$$\widetilde{f}(0,-1) \le f(0,-1) + \varepsilon = -1 + 0.4 = -0.6,$$

and

$$\widetilde{f}(0,1) \ge f(0,1) - \varepsilon = 1 - 0.4 > 0.6 > -0.6.$$

So, $\widetilde{f}(0,-1) \le -0.6 < \widetilde{f}(0,1)$, hence, $\widetilde{f}(0,-1) \ne \widetilde{f}(0,1)$, which contradicts to (1). So, this sub-case is impossible. Similarly, it is impossible to have h depending only on x_2.

1.2°. Let us consider the sub-case when

$$\widetilde{f}(x_1,x_2) = g(\min(x_1,x_2)).$$

In this sub-case,

$$\widetilde{f}(-1,-1) = g(\min(-1,-1)) = g(-1) =$$
$$g(\min(-1,1)) = \widetilde{f}(-1,1),$$

and

$$\widetilde{f}(-1,-1) = \widetilde{f}(-1,1). \tag{2}$$

But

$$\widetilde{f}(-1,-1) \le f(-1,-1) + \varepsilon = -2 + 0.4 = -1.6,$$

and

$$\widetilde{f}(-1,1) \ge f(-1,1) - \varepsilon = 0 - 0.4 = -0.4 > -1.6,$$

so, the equality (2) is also impossible.

1.3°. Let us now consider the sub-case

$$\widetilde{f}(x_1,x_2) = g(\max(x_1,x_2)).$$

In this sub-case,

$$\widetilde{f}(-1,1) = g(\max(-1,1)) = g(1) =$$
$$g(\max(1,1)) = \widetilde{f}(1,1),$$

and

$$\widetilde{f}(-1,1) = \widetilde{f}(1,1). \tag{3}$$

But

$$\widetilde{f}(-1,1) \le f(-1,1) + \varepsilon = 0 + 0.4 = 0.4,$$

and

$$\widetilde{f}(1,1) \ge f(1,1) - \varepsilon = 2 - 0.4 = 1.6 > 0.4,$$

so, the equality (3) is also impossible.

2°. In the second case,

$$\widetilde{f}(x_1,x_2)=\min(g_1(x_1,x_2),\ldots,g_m(x_1,x_2)),$$

where all the functions g_i are computable on a 1-layer computer. For this case, the impossibility follows from the following sequence of steps:

2.1°. If one of the functions g_i is of the type $\min(x_1,x_2)$, then we can rewrite

$$\min(g_1,\ldots,g_{i-1},\min(x_1,x_2),g_{i+1},\ldots,g_m)$$

as

$$\min(g_1,\ldots,g_{i-1},g_i^{(1)},g_i^{(2)},g_{i+1},\ldots,g_m),$$

where $g^{(i)}(x_1,x_2)=x_i$ is a function that is clearly computable on a 1-layer computer. After we make such transformations, we get an expression for $\widetilde{f}$ that only contains max and functions of one variable.

2.2°. Let us show that this expression cannot contain max. Indeed, if it does, then

$$\widetilde{f}(x_1,x_2)=\min(\ldots,\max(x_1,x_2))\le\max(x_1,x_2).$$

In particular, $\widetilde{f}(1,1)\le\max(1,1)=1$. But we must have

$$\widetilde{f}(1,1)\ge f(1,1)-\varepsilon=2-0.4=1.6>1.$$

The contradiction shows that max cannot be one of the functions g_i.

2.3°. So, each function g_i depends only on one variable. If all of them depend on one and the same variable, say, x_1, then the entire function $\widetilde{f}$ depends only on one variable, and we have already proved (in the proof of the first case) that it is impossible. So, some functions g_i depend on x_1, and some of the functions g_i depend on x_2. Let us denote by $h_1(x_1)$ the minimum of all functions g_i that depend on x_1, and by $h_2(x_2)$, the minimum of all the functions g_i that depend on x_2. Then, we can represent $\widetilde{f}$ as $\widetilde{f}(x_1,x_2)=\min(h_1(x_1),h_2(x_2))$.

2.4°. To get a contradiction, let us first take $x_1=1$ and $x_2=1$. Then,

$$\widetilde{f}(1,1)=\min(h_1(1),h_2(1))\ge f(1,1)-\varepsilon=2-0.4=1.6.$$

Since the minimum of the two numbers is ≥ 1.6, we can conclude that each of them is ≥ 1.6, i.e., that $h_1(1)\ge 1.6$ and $h_2(1)\ge 1.6$. For $x_1=1$ and $x_2=-1$, we have

$$\widetilde{f}(1,-1)=\min(h_1(1),h_2(-1))\le f(1,-1)+\varepsilon=0.4.$$

Since $h_1(1)\ge 1.6$, we conclude that $\widetilde{f}(1,-1)=h_2(-1)$. From

$$\widetilde{f}(1,-1)\ge f(1,-1)-\varepsilon=-0.4, \qquad (4)$$

we can now conclude that $h_2(-1)\ge -0.4$. Similarly, one can prove that $h_1(-1)\ge -0.4$. Hence,

$$\widetilde{f}(-1,-1)=\min(h_1(-1),h_2(-1))\ge -0.4.$$

But

$$\widetilde{f}(-1,-1)\le f(-1,-1)+\varepsilon=-2+0.4=-1.6<-0.4:$$

a contradiction with (4).

The contradiction shows that the second case is also impossible.

3°. In the third case,

$$\widetilde{f}(x_1,x_2)=\max(g_1(x_1,x_2),\ldots,g_m(x_1,x_2)),$$

where all the functions g_i are computable on a 1-layer computer. For this case, the impossibility (similarly to the second case) follows from the following sequence of steps:

3.1°. If one of the functions g_i is of the type $\max(x_1,x_2)$, then we can rewrite

$$\max(g_1,\ldots,g_{i-1},\max(x_1,x_2),g_{i+1},\ldots,g_m)$$

as

$$\max(g_1,\ldots,g_{i-1},g_i^{(1)},g_i^{(2)},g_{i+1},\ldots,g_m),$$

where $g^{(i)}(x_1,x_2)=x_i$ is a function that is clearly computable on a 1-layer computer. After we make such transformations, we get an expression for $\widetilde{f}$ that only contains min and functions of one variable.

3.2°. Let us show that this expression cannot contain min. Indeed, if it does, then

$$\widetilde{f}(x_1,x_2)=\max(\ldots,\min(x_1,x_2))\ge\min(x_1,x_2).$$

In particular,

$$\widetilde{f}(-1,-1)\ge\min(-1,-1)=-1.$$

But we must have

$$\widetilde{f}(-1,-1)\le f(-1,-1)+\varepsilon=$$

$$-2+0.4=-1.6<-1.$$

The contradiction shows that min cannot be one of the functions g_i.

3.3°. So, each function g_i depends only on one variable. If all of them depend on one and the same variable, say, x_1, then the entire function $\widetilde{f}$ depends only on one variable, and we have already proved (in the proof of the first case) that it is impossible. So, some functions g_i depend on x_1, and some of the functions g_i depend on x_2. Let us denote by $h_1(x_1)$ the maximum of all functions g_i that depend on x_1, and by $h_2(x_2)$, the maximum of all the functions g_i that depend on x_2. Then, we can represent $\widetilde{f}$ as

$$\widetilde{f}(x_1,x_2)=\max(h_1(x_1),h_2(x_2)).$$

3.4°. To get a contradiction, let us first take $x_1=-1$ and $x_2=-1$. Then,

$$\widetilde{f}(-1,-1)=\max(h_1(-1),h_2(-1))\le$$

$$f(-1,-1)+\varepsilon=-2+0.4=-1.6.$$

Since the maximum of the two numbers is ≤ -1.6, we can conclude that each of them is ≤ -1.6, i.e., that $h_1(-1) \le -1.6$ and $h_2(-1) \le -1.6$. For $x_1 = 1$ and $x_2 = -1$, we have

$$\widetilde{f}(1,-1) = \max(h_1(1), h_2(-1)) \ge$$
$$f(1,-1) - \varepsilon = -0.4.$$

Since $h_2(-1) \le -1.6$, we conclude that $\widetilde{f}(1,-1) = h_1(1)$. From

$$\widetilde{f}(1,-1) \le f(1,-1) + \varepsilon = 0.4,$$

we can now conclude that $h_1(1) \le 0.4$. Similarly, one can prove that $h_2(1) \le 0.4$. Hence,

$$\widetilde{f}(1,1) = \max(h_1(1), h_2(1)) \ge 0.4. \qquad (5)$$

But

$$\widetilde{f}(1,1) \ge f(1,1) - \varepsilon = 2 - 0.4 = 1.6 > 0.4,$$

which contradicts to (5).

The contradiction shows that the third case is also impossible.

4°. In all there cases, we have shown that the assumption that $\widetilde{f}$ can be computed on a 2-layer computer leads to a contradiction. So, $\widetilde{f}$ cannot be thus computed. Q.E.D.

B. Proof of the Theorem

Since the function f is continuous, there exists a $\delta > 0$ such that if $|x_i - y_i| \le \delta$, then

$$|f(x_1,\ldots,x_n) - f(y_1,\ldots,y_n)| \le \varepsilon.$$

Let us mark the grid points on the grid of size δ, i.e., all the points for which each coordinate $x_1,\ldots,x_n$ has the form $q_i \cdot \delta$ for integer q_i (i.e., we mark the points with coordinates 0, $\pm\delta$, $\pm 2\delta$, …, $\pm T$).

On each coordinate, we thus mark $\approx 2T/\delta$ points. So, totally, we mark $\approx (2T/\delta)^n$ grid points. Let us denote the total number of grid points by k, and the points themselves by $P_j = (x_{j1},\ldots,x_{jn})$, $1 \le j \le k$.

By m_f, let us denote the minimum of f:

$$m_f = \min_{x_1\in[-T,T],\ldots,x_n\in[-T,T]} f(x_1,\ldots,x_n).$$

For each grid point P_j, we will form piece-wise linear functions $f_{ji}(x_i)$ as follows:

- if $|x_i - x_{ji}| \le 0.6\cdot\delta$, then
$$f_{ji}(x_i) = f(P_j)(\ge m_f);$$
- if $|x_i - x_{ji}| \ge 0.7\cdot\delta$, then
$$f_{ji}(x_i) = m_f;$$
- if $0.6\cdot\delta \le |x_i - x_{ji}| \le 0.7\cdot\delta$, then
$$f_{ji}(x_i) = m_f + (f(P_j) - m_f)\cdot\frac{0.7\cdot\delta - |x_i - x_{ji}|}{0.7\cdot\delta - 0.6\cdot\delta}.$$

Let us show that for these functions f_{ji}, the function

$$\widetilde{f}(x_1,\ldots,x_n) = \max(A_1,\ldots,A_m),$$

where

$$A_j = \min(f_{j1}(x_1),\ldots,f_{jn}(x_n)),$$

is ε-close to f.

To prove that, we will prove the following two inequalities:

- For all $x_1,\ldots,x_n$,
$$\widetilde{f}(x_1,\ldots,i_n) \ge f(x_1,\ldots,x_n) - \varepsilon.$$
- For all $x_1,\ldots,x_n$,
$$\widetilde{f}(x_1,\ldots,i_n) \le f(x_1,\ldots,x_n) + \varepsilon.$$

Let us first prove the first inequality. Assume that we have a point $(x_1,\ldots,x_n)$. For every $i = 1,\ldots,n$, by q_i, we will denote the integer that is the closest to x_i/δ. Then,

$$|x_i - q_i\cdot\delta| \le 0.5\cdot\delta.$$

These values q_i determine a grid point $P_j = (x_{j1},\ldots,x_{jn})$ with coordinates $x_{ji} = q_i\cdot\delta$. For this j, and for every i,

$$|x_i - x_{ji}| \le 0.5\cdot\delta < 0.6\cdot\delta,$$

therefore, by definition of f_{ji}, we have $f_{ji}(x_i) = f(P_j)$. Hence,

$$A_j = \min(f_{j1}(x_1),\ldots,f_{jn}(x_n)) =$$
$$\min(f(P_j),\ldots,f(P_j)) = f(P_j).$$

Therefore,

$$\widetilde{f}(x_1,..,x_n) = \max(A_1,\ldots,A_m) \ge A_j = f(P_j).$$

But since $|x_{ji} - x_i| \le 0.5\cdot\delta < \delta$, by the choice of δ, we have $|f(x_1,\ldots,x_n) - f(P_j)| \le \varepsilon$. Therefore, $f(P_j) \ge f(x_1,\ldots,x_n) - \varepsilon$, and hence,

$$\widetilde{f}(x_1,\ldots,x_n) \ge f(P_j) \ge f(x_1,\ldots,x_n) - \varepsilon.$$

Let us now prove the second inequality. According to our definition of f_{ji}, the value of $f_{ji}(x_i)$ is always in between m_f and P_j, and this value is different from m_f only for the grid points P_j for which $|x_{ji} - x_i| \le 0.7\cdot\delta$. The value

$$A_j = \min(f_{j1}(x_1),\ldots,f_{jn}(x_n))$$

is thus different from m only if all the values $f_{ji}(x_i)$ are different from m, i.e., when $|x_{ji} - x_i| \le 0.7\cdot\delta$ for all i. For this grid point, $|x_{ji} - x_i| \le 0.7\cdot\delta < \delta$; therefore,

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

and hence, $f(P_j) \le f(x_1,\ldots,x_n) + \varepsilon$. By definition of f_{ji}, we have $f_{ji}(x_i) \le f(P_j)$. Since this is true for all i, we have

$$A_j = \min(f_{j1}(x_1),\ldots,f_{jn}(x_n)) \le$$
$$f(P_j) \le f(x_1,\ldots,x_n) + \varepsilon.$$

For all other grid points P_j, we have

$$A_j(x_1,\ldots,x_n) = m_f$$

for a given $(x_1,\ldots,x_n)$. Since m_f has been defined as the minimum of f, we have

$$A_j = m_f \le f(x_1,\ldots,x_n) < f(x_1,\ldots,x_n) + \varepsilon.$$

So, for all grid points, we have

$$A_j \le f(x_1, \ldots, x_n) + \varepsilon,$$

and therefore,

$$\widetilde{f}(x_1, \ldots, x_n) = \max(A_1, \ldots, A_m) \le f(x_1, \ldots, x_n) + \varepsilon.$$

The second inequality is also proven.

So, both inequalities are true, and hence, $\widetilde{f}$ is ε-close to f. Q.E.D.

VIII. Conclusion

Fuzzy techniques have been originally invented as a methodology that transforms the knowledge of experts (formulated in terms of natural language) into a precise computer-implementable form. There are many successful applications of this methodology to situations in which expert knowledge exist; the most well known (and most successful) are applications to fuzzy control.

In some cases, fuzzy methodology is applied even when no expert knowledge exists. In such cases, instead of trying to approximate the unknown control function by splines, polynomials, or by any other traditional approximation technique, researchers try to approximate it by guessing and tuning the expert rules. Surprisingly, this approximation often works fine.

In this paper, we give a mathematical explanation for this empirical phenomenon. Specifically, we show that approximation by using fuzzy methodology is indeed the best (in some reasonable sense).

Acknowledgments

This work was supported in part by NASA under cooperative agreement NCC5-209, NSF grants EAR-0225670 and DMS-0532645, Star Award from the University of Texas System, and Texas Department of Transportation grant No. 0-5453.

The authors are thankful to the anonymous referees for valuable suggestions.

References

[1] J. J. Buckley, "Sugeno type controllers are universal controllers", *Fuzzy Sets and Systems*, 1993, pp. 299–303.

[2] A. Kandel and G. Langholtz (eds.), *Fuzzy Control Systems*, CRC Press, Boca Raton, FL, 1994.

[3] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications* (Prentice Hall, Upper Saddle River, NJ, 1995).

[4] B. Kosko, "Fuzzy systems as universal approximators", *Proceedings of the 1st IEEE International Conference on Fuzzy Systems*, San Diego, CA, 1992, pp. 1153–1162.

[5] V. Kreinovich and A. Bernat, "Parallel algorithms for interval computations: an introduction", *Interval Computations*, 1994, No. 3, pp. 6–62.

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

[7] V. Kreinovich, H. T. Nguyen, and Y. Yam, "Fuzzy Systems Are Universal Approximators for a Smooth Function And Its Derivatives", *International Journal of Intelligent Systems*, 2000, Vol. 15, No. 6, pp. 565–574.

[8] R. N. Lea and V. Kreinovich, "Intelligent Control Makes Sense Even Without Expert Knowledge: an Explanation", *Reliable Computing*, 1995, Supplement (Extended Abstracts of APIC'95: International Workshop on Applications of Interval Computations, El Paso, TX, Febr. 23–25, 1995), pp. 140–145.

[9] H. T. Nguyen and V. Kreinovich, "On approximation of controls by fuzzy systems", *Proceedings of the Fifth International Fuzzy Systems Association World Congress*, Seoul, Korea, July 1993, pp. 1414–1417.

[10] H. T. Nguyen and E. A. Walker, *A first course in fuzzy logic*, CRC Press, Boca Raton, Florida, 2005.

[11] I. Perfilieva and V. Kreinovich, "A New Universal Approximation Result For Fuzzy Systems, Which Reflects CNF-DNF Duality", *International Journal of Intelligent Systems*, 2002, Vol. 17. No. 12, pp. 1121–1130.

[12] L.-X. Wang, "Fuzzy systems are universal approximators", *Proceedings of the IEEE International Conference on Fuzzy Systems*, San Diego, CA, 1992, pp. 1163–1169.

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

[14] R. R. Yager and V. Kreinovich, "Universal Approximation Theorem for Uninorm-Based Fuzzy Systems Modeling", *Fuzzy Sets and Systems*, 2003, Vol. 140, No. 2, pp. 331–339.