Novel fuzzy inference system (FIS) analysis and design based on lattice theory. Part I: Working principles

VASSILIS G. KABURLASOS *† and ATHANASIOS KEHAGIAS$\ddagger$

†Technological Educational Institution of Kavala

Dept. of Industrial Informatics, Div. of Computing Systems,

GR-65404 Kavala, Greece

*Corresponding author email: vgkabs@teikav.edu.gr

‡Aristotle University of Thessaloniki

Dept. of Math., Phys. & Comp. Sciences, Div. of Mathematics,

GR-54124 Thessaloniki, Greece

Email: kehagiat@gen.auth.gr

(15 September 2005)

This work substantiates novel perspectives and tools for analysis and design of Fuzzy Inference Systems (FIS). It is shown rigorously that the cardinality of the set F of fuzzy numbers equals $\aleph_1$, hence a FIS can implement "in principle" $\aleph_2$ functions, where $\aleph_2 = 2^{\aleph_1} > \aleph_1$ and $\aleph_1$ is the cardinality of the set R of real numbers; furthermore a FIS is endowed with a capacity for local generalization. A formulation in the context of lattice theory introduces a tunable metric distance d_K between fuzzy numbers. Implied advantages include: (1) an alleviation of the curse-of-dimensionality problem, regarding the number of rules, (2) a capacity to cope rigorously with heterogeneous data including (fuzzy) numbers and intervals, and (3) a capacity to introduce systematically useful nonlinearities. Extensive evidence from the literature appears to corroborate the proposed novel perspectives. Computational experiments demonstrate the utility of the proposed tools. A real-world industrial application is also described.

*Keyword*s: Fuzzy inference system, generalized interval analysis, lattice theory.

1 Introduction

Fuzzy sets have been proposed for processing non-numeric (linguistic) data and, ultimately, for *computing with perceptions* (Zadeh 2004). However, even though a fuzzy set can be defined on any universe of discourse, most often fuzzy sets are defined on the real number universe of discourse R where the name *fuzzy number/interval* has come to denote a convex, normal fuzzy set with bounded support (Zimmermann 1991). It turns out that fuzzy numbers are frequently involved in linguistic (fuzzy) rules representing knowledge in *Fuzzy Inference Systems* or *FIS* for short.

Knowledge representation is of fundamental significance in artificial intelligence applications. Several authors in different contexts have employed *mathematical lattice theory* for knowledge representation

1

(Ganter and Wille 1999, Sowa 2000). The authors of this work have published widely on a synergistic cross-fertilization of *(mathematical) lattice theory* with *fuzzy set theory* in clustering/classification applications involving lattice-ordered data (Kaburlasos and Petridis 1997, 2000, Petridis and Kaburlasos 1999, 2001, 2003, Kaburlasos 2004, Kaburlasos and Kehagias 2004, Kaburlasos and Papadakis 2005). It is worth pointing out that, even though an explicit connection was shown between *mathematical lattices* and *fuzzy sets* since the introduction of fuzzy set theory (Zadeh 1965), no tools have been established for FIS analysis and design based on lattice theory. This work engages mathematical lattice theory for enhancing conventional FIS analysis and design as explained below in this section.

A typical FIS, including both its *fuzzification* and *defuzzification* procedures, is a device for approximating a function $f : \mathsf{R}^N \longrightarrow \mathsf{R}^M$ in an "optimal" least square error sense (Wang and Mendel 1992, Kosko 1994, Dickerson and Kosko 1996, Xiao-Jun Zeng and Singh 1996, Puyin Liu 2002). Recall that alternative least square error methods for function approximation include polynomials, spline curves, ARMA models, statistical regressors, multilayer perceptrons, etc. It is shown in this work that the aforementioned "alternative methods" for function approximation, despite their capacity for generalization, can implement in principle one among a restrictive number of $\aleph_1$ functions, where $\aleph_1$ denotes the *cardinality* of the set R of real numbers.

A principal theoretical contribution of this work is in establishing that the *cardinality* of fuzzy numbers equals $\aleph_1$. It follows that a general FIS can implement "in principle" $\aleph_2 = 2^{\aleph_1} > \aleph_1$ functions[1]; moreover a (general) FIS is endowed with a capacity for local generalization. In other words, a FIS can implement a far larger number of functions $f : \mathsf{R}^N \longrightarrow \mathsf{R}^M$ than any alternative modeling method, while retaining a capacity for generalization. Furthermore, a tunable metric distance $d_K(.,.)$ is shown here between fuzzy numbers based on lattice-ordered *generalized intervals*. The metric $d_K(.,.)$ can deal rigorously with *heterogeneous* data including (fuzzy) numbers and intervals; moreover it is explained how $d_K(.,.)$ can potentially alleviate the curse-of-dimensionality problem regarding the number of rules in a FIS.

Preliminary results of this work have been presented in Kaburlasos and Kehagias (2004). This paper presents significant enhancements including a rigorous mathematical substantiation, useful novel perspectives, and convenient geometric interpretations. An industrial modeling application is outlined here, nevertheless it will be detailed elsewhere (Kaburlasos and Kehagias *submitted*) for lack of space here.

The layout is as follows. Section 2 gives mathematical preliminaries. Section 3 presents novel set-theoretic perspectives for Fuzzy Inference Systems (FIS); the main mathematical result is also presented. Section 4 describes generalized intervals including useful extensions. The capacity of novel techniques is demonstrated experimentally in section 5. Section 6 concludes by summarizing the contribution of this work; a real-world industrial application is also described. The Appendices contain the proofs of mathematical propositions.

2 Mathematical Preliminaries

This section summarizes rigorously useful definitions and propositions to be employed further below.

2.1 *Crisp Sets, Fuzzy Sets, Fuzzy Numbers*

Recall some well known facts regarding crisp sets.

---

[1] Where $\aleph_2$ is the *cardinality* of the power set of R; see Section 3 for details.

1. Given a set U (the *universe of discourse*) the *crisp powerset* of U is the family of all (crisp) subsets of U and is denoted by $\mathcal{P}(U)$.

2. The set of *natural numbers* $\{1, 2, ...\}$ is denoted by N; the set of real numbers $(-\infty, \infty)$ by R; the set of real nonnegative numbers $[0, \infty)$ by R_0^+.

3. The *cardinality* of a set U, denoted by $card(U)$, is informally defined to be the number of elements U contains[2]. Of particular interest in the context of this work are sets with infinite cardinalities. The following notation will be used

$$\aleph_0 = \text{card}(\mathsf{N}), \qquad \aleph_1 = \text{card}(\mathsf{R}) = \text{card}(\mathcal{P}(\mathsf{N})), \qquad \aleph_2 = \text{card}(\mathcal{P}(\mathsf{R})).$$

The use of subscripts in $\aleph_0$, $\aleph_1$, $\aleph_2$ is simply a matter of convenience and does *not* imply that $\aleph_1$ is the immediate successor of $\aleph_0$ or that $\aleph_2$ is the immediate successor of $\aleph_1$. However, it is well known that $\aleph_0 < \aleph_1 < \aleph_2$.

Basic facts about fuzzy sets are presented next (Zimmermann 1991, Nguyen and Walker 1999).

1. Given the *universe of discourse* set U, a *fuzzy (sub)set* F of U is a (*membership*) function $F : U \to [0, 1]$. The *fuzzy powerset* of U is the family of all fuzzy subsets of U and is denoted by $\mathcal{F}(U)$.

2. A *fuzzy relation* between elements of a set U is a fuzzy set $\mu : U \times U \to [0, 1]$, i.e. a fuzzy set of *pairs* of U elements.

3. Given a fuzzy set $F : U \to [0, 1]$ and a number $a \in [0, 1]$, the *a-cut* of F is the (crisp) set $F_a = \{x : F(x) \geq a\}$. Given a fuzzy set $F : U \to [0, 1]$, the following properties hold regarding *a*-cuts.

    A1 $F_0 = U$.

    A2 Take any $a, b \in [0, 1]$. Then: $a \leq b \Rightarrow F_b \subseteq F_a$.

    A3 Take any set $A \subseteq [0, 1]$ and let $b = \sup A$. Then: $\cap_{a \in A} F_a = F_b$.

In particular, *fuzzy numbers* are defined as follows (Zimmermann 1991).

Definition 2.1 *A* fuzzy number F *is a fuzzy set* $F : \mathsf{R} \to [0, 1]$*, which satisfies the following conditions.*

FN1 *The a-cut F_a is a closed interval for every $a \in (0, 1]$.*

FN2 *F is* normal*, i.e. $\max_{x \in \mathsf{R}} F(x) = 1$.*

FN3 *$\{x : F(x) > 0\}$ (the support of F) is a bounded interval.*

We will denote the set of all fuzzy numbers by F*.*

Conditions **FN1-FN3** are satisfied when, for instance, F is a triangular, a trapezoidal or a bounded support bell-shaped function. A popular representation of a fuzzy number is the LR-representation as described in the following.

---

[2] A rigorous definition of cardinality is rather involved; the interested reader can consult (Kamke 1950, Stoll 1979).

Proposition 2.2 *Let $F \in \mathsf{F}$. Then there exist real numbers p, q, r, s such that*

$$F(x) = \begin{cases} L(x) & for \ x \in (-\infty, q) \\ 1 & for \ x \in [q, r] \\ R(x) & for \ x \in (r, \infty) \end{cases} , \ and$$

1. *$L(x)$ is nondecreasing, continuous from the right and $L(x) = 0$ for $x < p$;*

2. *$R(x)$ is nonincreasing, continuous from the left and $R(x) = 0$ for $x > s$;*

3. *the numbers p, q, r, s satisfy $p \leq q \leq r \leq s$.*

Proof. In Zimmermann (1991). ∎

2.2 Lattices

The notion of a *partial order* is of fundamental importance below. Given a set P, a binary relation $\leq$ between elements of P is called a *partial order* if it satisfies the following conditions for all $x, y, z \in P$:

1. *Reflexivity*: $x \leq x$.

2. *Antisymmetry*: $(x \leq y$ and $y \leq x) \Rightarrow x = y$.

3. *Transitivity*: $(x \leq y$ and $y \leq z) \Rightarrow x \leq z$.

Note that notation $y \geq x$ is equivalent to $x \leq y$. If $\leq$ is a partial order on P then $(P, \leq)$ is called a *partially ordered set* or, equivalently, a *poset*. A *lattice* is a poset $(\mathsf{L}, \leq)$ with the additional property that any two of its elements have a greatest lower bound (g.l.b.), and a least upper bound (l.u.b.) in L. Given a *lattice* $(\mathsf{L}, \leq)$, and any two elements $x, y \in \mathsf{L}$, their g.l.b. is *called the meet* of x and y and denoted by $x \wedge y$; their l.u.b. is *called the join* of x and y and denoted by $x \vee y$. Lattice elements x and y are *comparable* when at least one of the relations $x \leq y$, $y \leq x$ holds; otherwise, elements x and y are *incomparable*, symbolically $x \| y$. A lattice without incomparable elements is called *totally ordered* lattice. For example, a totally ordered lattice is the set R of real numbers.

Given a poset (lattice) $(\mathsf{L}, \leq)$ a few additional posets (lattices) can be derived as follows (Birkhoff 1967).

1. Let $\leq$ be a partial order on a set L. Define the relation $\leq_\partial$ as follows: $x \leq_\partial y \Leftrightarrow y \leq x$. Then $\leq_\partial$ is a partial order on L, namely the *dual order* of $\leq$. If $(\mathsf{L}, \leq)$ is a lattice then $(\mathsf{L}, \leq_\partial)$ is also a lattice.

2. Let $\leq$ be a partial order on a set L. Define the relation $\leq$ on the set $\mathsf{L} \times \mathsf{L}$ as follows: $(x, y) \leq (z, u) \Leftrightarrow (x \leq z$ and $y \leq u)$. Then $\leq$ is a partial order on the set $\mathsf{L} \times \mathsf{L}$. If $(\mathsf{L}, \leq)$ is a lattice then $(\mathsf{L} \times \mathsf{L}, \leq)$ is also a lattice.

3. Let $\leq$ be a partial order on a set L. Define the relation $\sqsubseteq$ on the set $\mathsf{L} \times \mathsf{L}$ as follows: $(x, y) \sqsubseteq (z, u) \Leftrightarrow (x \geq z$ and $y \leq u) \Leftrightarrow (x \leq_\partial z$ and $y \leq u)$. Then $\sqsubseteq$ is a partial order on the set $\mathsf{L} \times \mathsf{L}$. If $(\mathsf{L}, \leq)$ is a lattice then $(\mathsf{L} \times \mathsf{L}, \sqsubseteq)$ is also a lattice.

Recall the concepts of *metric distance* and *positive valuation* in the following.

Definition 2.3 *A metric distance in a set S is a nonnegative real function $d : S \times S \rightarrow \mathsf{R}_0^+$ which, for all $x, y, z \in S$, satisfies:*

D1a $d(x, x) = 0.$

D1b $d(x, y) = 0 \Rightarrow x = y.$

D2 $d(x, y) = d(y, x).$

D3 $d(x, y) \leq d(x, z) + d(z, y).$

*If only conditions **D1a, D2** and **D3** are satisfied, then d is called a pseudometric.*

Definition 2.4 *A* valuation *in a lattice* $(\mathsf{L}, \leq)$ *is a function* $v : \mathsf{L} \to \mathsf{R}$ *which, for all* $x, y \in \mathsf{L}$*, satisfies:*

$$v(x) + v(y) = v(x \wedge y) + v(x \vee y).$$

A valuation is called positive *if, for all* $x, y \in \mathsf{L}$

$$x < y \Rightarrow v(x) < v(y).$$

Proposition 2.5 *Let* $(\mathsf{L}, \leq)$ *be a lattice and* v *be a positive valuation; then*

$$d(x, y) = v(x \vee y) - v(x \wedge y)$$

is a metric distance.

3   Set Theoretic Perspectives for Fuzzy Inference Systems

This section, first, summarizes general principles of Fuzzy Inference Systems (FIS), second, it presents the main mathematical result of this work and, third, it introduces some novel perspectives.

A FIS includes a knowledge base of fuzzy rules. The antecedent (IF part) of a rule is typically a conjunction of N fuzzy statements involving N fuzzy sets; moreover the consequent (THEN part) of a rule will be (in a Mamdani type FIS) a fuzzy statement (Mamdani and Assilian 1975) or (in a Sugeno type FIS) an algebraic expression (Tagaki and Sugeno 1985). The fuzzy sets involved in the fuzzy rules of a FIS are typically *fuzzy numbers*, i.e. convex, normal fuzzy sets with bounded support defined on the real number universe of discourse R.

An input vector $x \in \mathsf{R}^N$ to a FIS applies in parallel to the rules in a FIS' knowledge base by a *fuzzification* procedure. An *inference mechanism* produces the consequents of all activated rules, then the partial results are combined and, finally, a single real number (vector) is produced by a *defuzzification* procedure. Hence, a FIS is practically used as a tool for implementing a function $f : \mathsf{R}^N \longrightarrow \mathsf{R}^M$, where (1) N and M are integers, and (2) function f is induced from n pairs $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ of training data vectors. The design of a FIS boils down to an estimation of the parameters involved in a FIS so as to minimize the least squares error (LSE) function LSE= $\sqrt{\sum_{i=1}^{n} ||f(x_i) - y_i||^2}$, (Xiao-Jun Zeng and Singh 1997, 2003). In the aforementioned sense a FIS may be regarded as a tool for optimal function estimation; moreover, a FIS retains a linguistic interpretation.

It is widely recognized that FISs can give better results than alternative function approximation methods in applications and, usually, a fuzzy logic explanation is sought. This work proposes, in addition, a set-theoretic explanation (Kaburlasos 2002, Kaburlasos and Kehagias 2004) by seeking an answer to the

following question: How many fuzzy numbers are there? Or, in other words, *what is the cardinality (card(F)) of the set F of fuzzy numbers*? It follows the main mathematical result of this work.

Proposition 3.1 *It holds* $card(\mathsf{F}) = \aleph_1$, *where $\aleph_1$ is the cardinality of the set R of real numbers.*

Proof. See in Appendix B. ∎

The above proposition claims the non-obvious result that there are as many fuzzy numbers as there are real numbers. Proposition 3.1 leads to novel perspectives regarding the capacity of FIS for function approximation as explained in the following.

In the first place it is interesting to calculate the cardinality of the set $\mathcal{F}$ of all functions $f : \mathsf{R}^N \longrightarrow \mathsf{R}^M$. Using standard *cardinal arithmetic* (Stoll 1979) it follows that $card(\mathcal{F}) = \aleph_1^{\aleph_1} = (2^{\aleph_0})^{\aleph_1} = 2^{\aleph_1} = \aleph_2 > \aleph_1$. In this sense $\aleph_2$ is the largest possible cardinality for a family of models $f : R^N \to R^M$. Unfortunately a general function f_0 in $\mathcal{F}$ is practically useless because it lacks a capacity for generalization. More specifically, knowledge of a function f_0 values $f_0(x_1), ... f_0(x_n)$ at a number of points $x_1, ..., x_n$ cannot imply the value of function f_0 at a different point $x_{n+1} \neq x_i$, $i = 1, \ldots, n$.

Consider now a parametric family of models, e.g. polynomials, ARMA models, statistical regressors, radial basis function (RBF) networks, multilayer perceptrons, etc. Any of the aforementioned families is characterized by a capacity for generalization. Due to the finite number p of parameters involved, the cardinality of any of the aforementioned families of models equals $\aleph_1^p = (2^{\aleph_0})^p = 2^{\aleph_o} = \aleph_1$ (Kaburlasos 2002).

It might be thought that $\aleph_1$ is an adequately large number of models to choose a "good" model from. Unfortunately the latter is not the case. Consider, for instance, the family of polynomials which includes $\aleph_1$ models. It is well known that a polynomial may not approximate usefully a set $(x_1, y_1), ..., (x_n, y_n)$ of training data points due to "overfitting"; hence a polynomial may not be useful for generalization. In the latter case a different family of models may be sought, e.g. an ARMA model, a multi-layer perceptron, etc. It appears that there is no "universally optimal" family of models. In the aforementioned sense the cardinality $\aleph_1$, of a family of models, is *inherently restrictive*. It turns out that the family of FIS models combines a cardinality of $\aleph_2$ with a capacity for (local) generalization as explained in the following.

It has been shown by proposition 3.1 that the cardinality of the set F of fuzzy numbers equals $card(\mathsf{F}) = \aleph_1$. A Mamdani type FIS can be regarded as a function $m : \mathsf{F}^N \longrightarrow \mathsf{F}^M$. Using standard cardinal arithmetic (Stoll 1979) it follows that the cardinality of the set M of Mamdani type FIS equals $card(\mathsf{M}) = \aleph_1^{\aleph_1} = \aleph_2 > \aleph_1$ – Recall that $\aleph_2$ also equals the cardinality of the set of functions $f : \mathsf{R}^N \to \mathsf{R}^M$; hence there exists a one-one correspondence between Mamdani type FIS models and real functions $f : \mathsf{R}^N \to \mathsf{R}^M$. Likewise, a Sugeno type FIS can be regarded as a function $s : \mathsf{F}^N \longrightarrow \mathcal{P}_p$, where $\mathcal{P}_p$ is a family of parametric models (e.g. polynomial models) with p parameters. Therefore it follows that the cardinality of the set S of Sugeno type FIS also equals $card(\mathsf{S}) = \aleph_1^{\aleph_1} = \aleph_2$. In conclusion, Mamdani- type FIS can implement, in principle, $\aleph_2$ functions; the same is true of Sugeno-type FIS.

It has been explained above that a general function $f : \mathsf{R}^N \to \mathsf{R}^M$ lacks a capacity for generalization. Fortunately this is not the case for a FIS of Mamdani- (or Sugeno-) type due to the non-trivial (interval) support of the fuzzy numbers involved in a FIS' fuzzy rule base. More specifically an input vector $x = (x_1, \ldots, x_n)$, within a fuzzy rule's interval of support, activates the corresponding rule; there follows a FIS' capacity for (local) generalization. In conclusion the family of FIS models combines "in principle" the cardinality $\aleph_2$ with a capacity for generalization in function $f : \mathsf{R}^N \to \mathsf{R}^M$ approximation problems.

4   Generalized Intervals and Extensions

While a fuzzy number is usually represented in terms of a membership function $F(x)$, it is well known that an alternative (and equivalent) representation of $F(x)$ can be given in terms of a family of a-cuts $\{F_a\}_{a \in (0,1]}$, where $F_a = \{x : F(x) \geq a\}$. Note that several practical advantages have been shown in fuzzy inference based on a-cuts (Uehara and Fujise 1993). It turns out that F_a is a closed interval for every value $a \in (0,1]$; hence a FIS is a mapping whose domain consists of families of closed intervals.

From a computational aspect there is a certain difference between the aforementioned two representations. In particular, in the first case one operates with numerical representations of functions, whereas in the second case one operates with numerical representations of families of closed intervals. Is it possible to enhance the advantages of the interval representations by operating with a wider class of sets? There is evidence that, *by an appropriate generalization of the concept of interval*, this is indeed the case. Hence, this section studies *generalized intervals* as well as useful extensions. Note that *generalized intervals* have been introduced elsewhere (Kaburlasos 2002, Petridis and Kaburlasos 2003, Kaburlasos 2004). This work introduces useful enhancements including: an improved mathematical notation, a correspondence with fuzzy set a-cuts, rigorous proofs of novel results and, finally, the introduction of instrumental *mass functions*. *Positive* generalized intervals are defined in the first place.

Definition 4.1 *A* positive generalized interval of height h is a mapping $m_{x_1,x_2}^h : \mathsf{R} \to \{0, h\}$ *(where $x_1 \leq x_2$ and $h \in (0,1]$) defined as follows*

$$m_{x_1,x_2}^h(x) = \begin{cases} h & when\ x_1 \leq x \leq x_2 \\ 0 & otherwise. \end{cases}$$

The family of all positive generalized intervals of height h will be denoted by M_+^h. For convenience of notation the abovementioned mapping m_{x_1,x_2}^h is denoted by $[x_1, x_2]^h$, where $x_1 \leq x_2$.

Positive generalized intervals are related to fuzzy numbers. More precisely, given a fuzzy number, the family of its a-cuts corresponds to a family of positive generalized intervals as shown in the following proposition (the proof is omitted, since it is immediate).

Proposition 4.2 *Take some fuzzy number $F \in \mathsf{F}$ and denote, for every $a \in (0,1]$, the indicator function of its a-cut F_a by $\widetilde{F}_a$. Then, for $a \in (0,1]$, the function $a\widetilde{F}_a$ is a positive generalized interval of height a.*

Negative generalized intervals are defined in the following.

Definition 4.3 *A* negative generalized interval of height h is a mapping $m_{x_1,x_2}^h : \mathsf{R} \to \{0, -h\}$ *(where $x_1 > x_2$ and $h \in (0,1]$) defined as follows*

$$m_{x_1,x_2}^h(x) = \begin{cases} -h & x_2 \leq x \leq x_1 \\ 0 & otherwise. \end{cases}$$

The family of all negative generalized intervals of height h will be denoted by M_-^h. For convenience of notation the abovementioned mapping m_{x_1,x_2}^h is denoted by $[x_1, x_2]^h$, where $x_1 > x_2$.

The rationale for introducing *negative* generalized intervals is the following. It is a well known fact that conventional intervals of real numbers form a mathematical lattice. In this lattice the infimum of two non-intersecting intervals is the *empty interval*. The latter is rather restrictive in practical applications due to

the absence of a positive valuation function. Hence, a lattice of intervals was sought where non-intersecting intervals have a nonempty infimum furthermore a positive valuation function exists. As will be seen in the sequel, negative generalized intervals serve this purpose well with rewarding results.

Notation 4.4 *The family of all generalized intervals of height h will be denoted by $\mathsf{M}^h = \mathsf{M}^h_+ \cup \mathsf{M}^h_-$.*

Notation 4.5 *The family of all positive generalized intervals (of any height) will be denoted by $\mathsf{M}_+ = \cup_{h\in(0,1]}\mathsf{M}^h_+$; likewise, the family of all negative generalized intervals will be denoted by $\mathsf{M}_- = \cup_{h\in(0,1]}\mathsf{M}^h_-$; in conclusion, the family of all generalized intervals will be denoted by $\mathsf{M} = \cup_{h\in(0,1]}\mathsf{M}^h = \mathsf{M}_+ \cup \mathsf{M}_-$.*

Our interest is only in generalized intervals of height $h \in (0,1]$, because the latter intervals may emerge from the a-cuts of fuzzy numbers. Now, for every $h \in (0,1]$, M^h is equipped with an ordering relation $\preceq$.

Definition 4.6 *Given $h \in (0,1]$, an ordering relation $\preceq$ is defined on $\mathsf{M}^h \times \mathsf{M}^h$ as follows:*

$$\begin{aligned}
&\text{if } [a,b]^h \in \mathsf{M}^h_+,\ [c,d]^h \in \mathsf{M}^h_+ \text{ then:} \quad && [a,b]^h \preceq [c,d]^h \Leftrightarrow [a,b] \subseteq [c,d] \\
&\text{if } [a,b]^h \in \mathsf{M}^h_-,\ [c,d]^h \in \mathsf{M}^h_- \text{ then:} \quad && [a,b]^h \preceq [c,d]^h \Leftrightarrow [d,c] \subseteq [b,a] \\
&\text{if } [a,b]^h \in \mathsf{M}^h_-,\ [c,d]^h \in \mathsf{M}^h_+ \text{ then:} \quad && [a,b]^h \preceq [c,d]^h \Leftrightarrow [b,a] \cap [c,d] \neq \varnothing.
\end{aligned}$$

In all other cases $[a,b]^h$ and $[c,d]^h$ are incomparable, symbolically $[a,b]^h || [c,d]^h$.

Proposition 4.7 $\left(\mathsf{M}^h, \preceq\right)$ *is a lattice. Let $\curlyvee$ and $\curlywedge$ denote the lattice join and meet, respectively; then*

$$[a,b]^h \curlyvee [c,d]^h = [a \wedge c, b \vee d]^h$$
$$[a,b]^h \curlywedge [c,d]^h = [a \vee c, b \wedge d]^h$$

where $a \wedge c = min\,(a,c)$ and $a \vee c = max\,(a,c)$.

Proof. See in Appendix B. ∎

The proof of proposition 4.7 establishes that lattices $(\mathsf{R} \times \mathsf{R}, \leq_\partial \times \leq)$ and $\left(\mathsf{M}^h, \preceq\right)$ are isomorphic[3]. The isomorphism between lattices $\left(\mathsf{M}^h, \preceq\right)$ and $(\mathsf{R} \times \mathsf{R}, \leq_\partial \times \leq)$ will be taken advantage of below as follows: First, lattice $\left(\mathsf{M}^h, \preceq\right)$ will be used for providing convenient geometric interpretations on the plane and, second, lattice $(\mathsf{R} \times \mathsf{R}, \leq_\partial \times \leq)$ will be used in algebraic computations.

The next proposition shows how a strictly increasing function ultimately implies a metric in $\left(\mathsf{M}^h, \preceq\right)$.

Proposition 4.8 *Let $f_h : \mathsf{R} \to \mathsf{R}$ be a strictly increasing function (f_h is called: the underlying positive valuation function). Then the function $v_h : \mathsf{M}^h \to \mathsf{R}$ given by*

$$v_h([a,b]^h) = f_h(b) - f_h(a).$$

is a positive valuation in lattice $\left(\mathsf{M}^h, \preceq\right)$. It follows that the function $d_h : \mathsf{M}^h \times \mathsf{M}^h \to \mathsf{R}^+_0$ given by

$$d_h([a,b]^h, [c,d]^h) = [f_h(a \vee c) - f_h(a \wedge c)] + [f_h(b \vee d) - f_h(b \wedge d)]$$

is a metric distance in lattice $\left(\mathsf{M}^h, \preceq\right)$.

---

[3] Two partially ordered sets $(P, \leq)$ and $(Q, \leq)$ are called *isomorphic*, symbolically $(P, \leq) \cong (Q, \leq)$, if there exists a mapping $\psi : P \to Q$ such that both '$x \leq y$ in $P \Leftrightarrow \psi(x) \leq \psi(y)$ in Q' and 'ψ is onto Q'.

Proof. See in Appendix B. ∎

Note that the strictly increasing function f_h above was named *underlying positive valuation* because f_h is a positive valuation in the totally ordered lattice R of real numbers.

A *very large number* of metric distances can be defined in the above manner. Let $\mathbf{D}_h$ denote the family of all metrics in M^h. Consider the following result.

Proposition 4.9 $card(\mathbf{D}_h) \geq \aleph_1$.

Proof. See in Appendix B. ∎

An underlying positive valuation f_h can be constructed from an integrable *mass function* $m_h : \mathsf{R} \to \mathsf{R}_0^+$ using the formula

$$f_h(x) = \int_0^x m_h(t)dt.$$

Note that the above integral is positive (negative) for $x > 0$ ($x < 0$). One may regard a mass function $m_h(x)$ as an instrument for attaching "a weight of significance" to a real number x. Various mass functions can be employed in applications. Typical FIS applications in the literature employ solely (and implicitly) mass function $m_h(x) = 1$. This work has shown analytically how alternative mass functions can be employed. For example the mass function $m_h(x) = h$ implies (all other things being equal) that a proportionally larger value of $v_h(.)$ is assigned to a generalized interval of a larger height; in the latter case the corresponding metric distance between two generalized intervals $[a, b]^h$ and $[c, d]^h$ is given by $d_h([a, b]^h, [c, d]^h) = h(|a - c| + |b - d|)$. Alternative mass functions include *probability density functions* (*pdfs*), etc. as demonstrated below.

Based on a metric distance $d_h(.,.)$ in M^h there follows a metric distance $d_K(.,.)$ in the set F of fuzzy numbers given by $d_K(F_1, F_2) = \int_0^1 d_h(F_1(h), F_2(h))dh$, where $F_1(h)$ and $F_2(h)$ are positive intervals of height h corresponding to fuzzy numbers F_1 and F_2, respectively (Kaburlasos 2004). Recall that the aforementioned definition of d_K is based on metric d_h between generalized intervals, where the metric d_h requires the employment of both positive and negative generalized intervals; hence, negative generalized intervals are instrumental for defining d_K.

The set F of fuzzy numbers can be easily extended so as to include both intervals and real numbers. More specifically, a conventional interval $[a, b]$, $a \leq b$ can be represented as $\bigcup\limits_{h \in (0,1]} \{[a, b]^h\}$, whereas a single real number x can be represented, in particular, as $\bigcup\limits_{h \in (0,1]} \{[x, x]^h\}$. An element of the extended set (F^*) is called *Fuzzy Interval Number*, of *FIN* for short (Kaburlasos 2004). The latter set may accommodate jointly *heterogeneous fuzzy data* including real numbers, intervals, and fuzzy numbers (Pedrycz *et al.* 1998, Paul and Kumar 2002).

The above results can enhance conventional FIS analysis and design by carrying out FIS fuzzification in F^* based on metric distance $d_K(.,.)$ using the formula $\mu_F(H) = 1/(1 + d_K(F, H))$ (Krishnapuram and Keller 1993). Hence, a "metric FIS design" can be proposed as outlined in Kaburlasos and Kehagias (2004) and will be detailed elsewhere (Kaburlasos and Kehagias *submitted*). Some advantages of the metric distance $d_K(.,.)$ are demonstrated experimentally in the following.

5 Experiments

First, the metric distances $d_1([2, 2]^1, [5, 5]^1)$ and $d_1([-6, -5]^1, [-1, 1]^1)$ are computed between (positive) generalized intervals for the four choices of mass functions shown in the following table:

Figure 1(a): $m_1(x) = 1$.

Figure 1(b): $m_2(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ (with $\mu = -1$ and $\sigma = 1$).

Figure 1(c): $m_3(x) = 2e^{-(x-1)}/(1 + e^{-(x-1)})^2$.

Figure 1(d): $m_4(x) = 0.01x^4 + 0.02x^3 - 0.41x^2 - 0.42x + 4.44$.

The corresponding positive valuation functions, given by $f_i(x) = \int_0^x m_i(t)dt$, $i = 1, \ldots, 4$, are shown in figures 2(a), 2(b), 2(c) and 2(d), respectively.

Using the formula

$$d_h([a, b]^h, [c, d]^h) = [f_h(a \vee c) - f_h(a \wedge c)] + [f_h(b \vee d) - f_h(b \wedge d)]$$

the following metric distances are computed:

Figure 2(a): $d_1([2, 2]^1, [5, 5]^1) = 2 \cdot (5 - 2) = 6$

            $d_1([-6, -5]^1, [-1, 1]^1) = (-1 - (-6)) + (1 - (-5)) = 11$.

Figure 2(b): $d_1([2, 2]^1, [5, 5]^1) \approx 2 \cdot (0.1587 - 0.1573) = 0.0028$

            $d_1([-6, -5]^1, [-1, 1]^1) \approx (-0.3414 - (-0.8413)) + (0.1359 - (-0.8413)) = 1.4771$.

Figure 2(c): $d_1([2, 2]^1, [5, 5]^1) \approx 2 \cdot (1.4260 - 0.9239) = 1.0042$

            $d_1([-6, -5]^1, [-1, 1]^1) \approx (-0.2989 - (-0.5361)) + (0.4621 - (-0.5329)) = 1.2322$.

Figure 2(d): $d_1([2, 2]^1, [5, 5]^1) \approx 2 \cdot (9.2491 - 7.0851) = 4.3280$

            $d_1([-6, -5]^1, [-1, 1]^1) \approx (-4.5091 - (-13.7585)) + (4.0967 - (-13.4909)) = 26.8370$.

Hence, different mass functions and, subsequently, different positive valuation functions may yield different distances between "fixed" intervals. We remark that the underlying positive valuation function in figure 2(a) is the classic (Lebesque) measure in R. Furthermore, the underlying positive valuation in figure 2(b) was meant to demonstrate the employment of a probability distribution function as an underlying positive valuation. In particular, note that since the mass function in figure 1(b) is a normal (Gaussian) probability density function, with expected value $\mu = -1$ and standard deviation $\sigma = 1$, it follows that the implied positive valuation in figure 2(b) is a normal cumulative distribution function; the latter is an example of a saturated positive valuation function. Another saturated positive valuation function is shown in figure 2(c), that is the logistic (sigmoid) function $f_1(x) = \frac{2}{1+e^{-(x-1)}} - 1$. Recall that sigmoids are often employed as activation functions in the neurons of neural networks. Finally, figure 2(d) demonstrates the use of an unbounded, non-linear (polynomial) underlying positive valuation.

[Insert figure 1 about here]

[Insert figure 2 about here]

We next compute the metric distance d_K between fuzzy numbers. Consider the three fuzzy numbers $F1, F2$, and E, with piecewise linear membership functions, shown in figure 3(a). Note that the left sides of $F1$ (solid line) and $F2$ (dashed line) coincide, nevertheless the corresponding right sides are clearly different; both fuzzy numbers $F1$ and $F2$ attain their unique maximum value at $x = 1$. Moreover fuzzy

number E has an isosceles triangular membership function centered at $x = 4$. In figure 3(a) the membership functions of fuzzy numbers $F1$ and $F2$ are denoted explicitly by $f1(x)$ and $f2(x)$, respectively.

Two different mass functions are shown, respectively, in figures 3(b) and 3(c). On the one hand, the mass function $m_h(t) = h$ (shown in figure 3(b) for $h = 1$) assumes that all the real numbers are equally important; the corresponding positive valuation function is given by $f_h(x) = hx$. On the other hand, the mass function $m_h(t) = 4he^{-7(t-1.5)}/(1 + e^{-7(t-1.5)})^2$ (shown in figure 3(c) for $h = 1$) emphasizes symmetrically the numbers around $t = 1.5$; the corresponding positive valuation function, namely *logistic function* (in statistics) or *sigmoid function* (in neural computing), is given by $f_h(x) = (4h/7)/(1 + e^{-7(x-1.5)})$.

[Insert figure 3 about here]

Figure 4 displays the metric distances $d_K(F1(h), E(h))$ and $d_K(F2(h), E(h))$ in solid and dashed lines, respectively. In particular the mass function $m_h(t) = h$ (figure 3(b)) has been employed for computing the curves shown in figure 4(a), whereas the mass function $m_h(t) = 4he^{-7(t-1.5)}/(1 + e^{-7(t-1.5)})^2$ (figure 3(c)) has been employed for computing the curves shown in figure 4(b). From figure 4(a) it follows $d_K(F1, E) \approx 3.0 > 2.9754 \approx d_K(F2, E)$, whereas from figure 4(b) it follows $d_K(F1, E) \approx 0.3587 < 0.3811 \approx d_K(F2, E)$.

[Insert figure 4 about here]

The example above was meant to demonstrate that a mass function can be used as an instrument for tuning, non-linearly, the distance between two fuzzy numbers. The above example has also demonstrated that it is not necessary to cover a FIS input data domain with (fuzzy) rules because a rule can be activated, even when the input data fall outside all fuzzy rule supports, using a fuzzy activation function such as $\mu_F(H) = 1/(1 + d_K(F, H))$. The latter potentially implies an alleviation of the curse-of-dimensionality problem regarding the number of rules in a FIS.

6 Conclusion

The thrust of this work has been in introducing sound tools and novel perspectives for enhanced Fuzzy Inference System (FIS) analysis and design. Using a combination of analytic- and set-theoretic results it was established a theoretical $\aleph_2$ bound capacity of FISs to implement real functions $f : \mathsf{R}^N \longrightarrow \mathsf{R}^M$; furthermore a general FIS has a capacity for (local) generalization. Based on generalized interval analysis, a tunable metric distance d_K was presented between fuzzy numbers; more specifically, an underlying *mass function* can be used for "tuning" by attaching a *weight of significance* to individual real numbers.

There is evidence that fuzzy numbers can produce better results than real numbers in classification problems (Petridis and Kaburlasos 2003, Kaburlasos and Papadakis 2005); in particular, improvements have been reported based on an optimal estimation of the underlying mass functions (Kaburlasos and Papadakis 2005). Furthermore, the proposed tools could be employed elsewhere. For instance, metric $d_K(.,.)$ could be employed for calculating a metric distance between type-2 fuzzy sets (Karnik *et al.* 1999) by the calculation of an additional integral. Further applications could be in other fuzzy classifiers (Ishibuchi *et al.* 1999) as well as in various decision support systems (Cassaigne and Singh 2001). Automatic control (Passino and Yurkovich 1998) is also a promising domain for application of the novel tools presented here. A real-world application is described next regarding system modeling for industrial quality control.

The industrial production of nitrogenous fertilizers proceeds as follows. Highly concentrated hot Ammonium Nitrate (AN) melt is sprayed from a spraying nozzle manifold to a rotating *pan granulator* mill.

The fertilizer end-product consists of small fertilizer granules each having size in the range of a few millimeters. Control of granule size is important for maintaining high fertilizer quality and is obtained by tuning the values of several *operating parameters/variables* including: AN melt flow, AN melt pressure, pan speed/inclination, volume of recycled fertilizer, nozzle location, etc. Optimal parameter values are constantly sought, especially after switching production from one fertilizer type to a different one; various disturbances during the industrial production may also call for additional tuning.

Tuning is currently effected in the industry by trial-and-error; a dependable open loop model of the pan granulator as well as a feedback automatic control mechanism are desirable. A FIS (open loop) model of the average diameter $d(x)$ of produced fertilizer granules has been developed, where x is a N-dimensional vector of "important" pan granulator operating parameters. Using a recent "variable selection method" (Papadakis *et al.* 2005) the most important variables have been selected, furthermore a FIS classifier model was developed for categories: *small*, *medium-small*, *medium-large*, and *large*. The classifier gave a 97% percentage of correct classification of granule diameter; furthermore, descriptive decision-making knowledge (fuzzy rules) was induced from the training data. The aforementioned performance was highly satisfactory for industrial production purposes and much better than the performance of competing classification methods as it will be detailed in a companion paper (Kaburlasos and Kehagias *submitted*).

A   Appendix: The cardinality of nondecreasing functions

In this Appendix the following result is proven (it is stated more formally in Proposition A.6 below):

"there are uncountably many nondecreasing functions from $[a, b]$ to $[c, d]$".

This result will be used in the next section to prove that there are uncountably many fuzzy numbers. Before proving this basic result some preliminaries are necessary.

A number of facts will be used about cardinal numbers (cardinalities) and their algebra. For proofs see in Kamke (1950) and Stoll (1979).

As already mentioned, the cardinality of the set of natural numbers $N=\{1, 2, ...\}$ is $\aleph_0$ (*aleph-zero*). The cardinality of the set of all subsets of $\{1, 2, ...\}$ is $\aleph_1 = 2^{\aleph_0}$. Rational numbers will be denoted by Q; given any interval $[a, b] \subseteq R$, the rational numbers in $[a, b]$ will be denoted by $Q_{a,b}$. It holds $card(Q)= card(Q_{a,b})= \aleph_0$. Furthermore, $card([a, b]) = card(R) = \aleph_1$. A set is called *countable* if its cardinality is at most $\aleph_0$ (it can also be finite); it is called *uncountable* if its cardinality is $\aleph_1$ or higher.

Functions $f : [a, b] \to [c, d]$ are examined next, where $[a, b]$ and $[c, d]$ are bounded closed intervals of real numbers. Denote

1. the set of all functions from $[a, b]$ to $[c, d]$ by $\mathcal{F}_{a,b}^{c,d}$;

2. the set of all *nondecreasing* functions from $[a, b]$ to $[c, d]$ by $\mathcal{I}_{a,b}^{c,d}$;

3. the set of all *continuous* functions from $[a, b]$ to $[c, d]$ by $\mathcal{C}_{a,b}^{c,d}$;

4. the set of all functions from $[a, b]$ to $[c, d]$ which are *continuous except possibly at a set* $X \subseteq [a, b]$ by ${}^X\overline{\mathcal{C}}_{a,b}^{c,d}$;

5. the set of all functions from $[a, b]$ to $[c, d]$ which have *at most a countable number of discontinuities* by $\widetilde{\mathcal{C}}_{a,b}^{c,d}$.

Let $a, b, c, d \in \mathsf{R}$ be fixed. Here are some useful facts.

Lemma A.1 *Every nondecreasing function $f \in \mathcal{I}_{a,b}^{c,d}$ can have at most a countable number of discontinuities.*

Proof. Take some $f \in \mathcal{I}_{a,b}^{c,d}$ and denote the set of the points where it is discontinuous by X. Define for $n = 1, 2, ...$ the sets

$$\mathsf{A}_n = \left\{ x : \frac{1}{n+1} \leq f\left(x^+\right) - f\left(x^-\right) < \frac{1}{n} \right\}$$
$$\mathsf{B}_n = \left\{ x : n \leq f\left(x^+\right) - f\left(x^-\right) < n+1 \right\}.$$

Clearly

$$X = \cup_{n=1}^{\infty} \left(\mathsf{A}_n \cup \mathsf{B}_n \right).$$

Now, for every n the cardinality of A_n is finite (otherwise it would be $f(b) - f(a) = \infty$) and hence $card\left(\mathsf{A}_n\right) < \aleph_0$. Similarly $card\left(\mathsf{B}_n\right) < \aleph_0$. Hence

$$card\left(X\right) = \sum_{n=1}^{\infty} \left(\mathsf{A}_n \cup \mathsf{B}_n \right) \leq \aleph_0 \aleph_0 = \aleph_0$$

(the proof of $\aleph_0 \aleph_0 = \aleph_0$ can be found in Stoll (1979)) and this completes the proof. ∎

Lemma A.2 *Take any countable set $X = \{x_1, x_2, ...\} \subseteq [a, b]$ and any function $f \in{}^X \overline{\mathcal{C}}_{a,b}^{c,d}$. Then f is specified by its values on $\mathsf{Q}_{a,b} \cup X$.*

Proof. Choose any $x \in [a, b]$; there are two possibilities.

1. Suppose that $x \in \mathsf{Q}_{a,b} \cup X$, then the value $f(x)$ is known.

2. Suppose that $x \in [a, b] \backslash (\mathsf{Q}_{a,b} \cup X)$. Since $\mathsf{Q}_{a,b} \cup X$ contains all the rationals in $[a, b]$ and these form a *dense* subset of $[a, b]$, there is a sequence $z_1, z_2, ...$ such that: (a) $\{z_1, z_2, ...\} \subseteq \mathsf{Q}_{a,b} \cup X$ and (b) $\lim_{n \to \infty} z_n = x$. And, since f is continuous in $[a, b] \backslash (\mathsf{Q}_{a,b} \cup X) \subseteq [a, b] \backslash X$, it follows that $f(x) = \lim_{n \to \infty} f(z_n)$, i.e. the value $f(x)$ is determined by the values of f on $\mathsf{Q}_{a,b} \cup X$.

Hence, for every $x \in [a, b]$ the value $f(x)$ is determined by the values of f on $\mathbf{Q}_{a,b} \cup X$ and the proof is complete. ∎

Lemma A.3 *Take any countable set $X = \{x_1, x_2, ...\} \subseteq [a, b]$. Then $card\left(\mathbf{Q}_{a,b} \cup X\right) = \aleph_0$.*

Proof. We have

$$\mathbf{Q}_{a,b} \subseteq \mathbf{Q}_{a,b} \cup X$$

hence

$$\aleph_0 = card\left(\mathbf{Q}_{a,b}\right) \leq card\left(\mathbf{Q}_{a,b} \cup X\right) \leq card\left(\mathbf{Q}_{a,b}\right) + card\left(X\right) = \aleph_0 + \aleph_0 = \aleph_0$$

and the proof is complete. ∎

The next Lemma says that the number of functions which are continuous everywhere except (possibly) at the points of a *specified*, countable set X is $\aleph_1$.

Lemma A.4 *Take any countable set $X = \{x_1, x_2, ...\} \subseteq [a, b]$. Then $card\left(^X\overline{\mathcal{C}}_{a,b}^{c,d}\right) = \aleph_1$.*

Proof. (i) According to Lemma A.2, every function $f \in^X \overline{\mathcal{C}}_{a,b}^{c,d}$ is specified by its values on $\mathbf{Q}_{a,b} \cup X$. For each of these values some element of $[c, d]$ can be chosen, i.e. one of $\aleph_1$ numbers. There may be further restrictions in the choice of some of these values, but the total number of choices for the function values cannot be more than

$$\underbrace{\aleph_1 \times \aleph_1 \times ... \times \aleph_1}_{\aleph_0 \text{ times}} = \aleph_1^{\aleph_0} = \left(2^{\aleph_0}\right)^{\aleph_0} = 2^{\aleph_0 \cdot \aleph_0} = 2^{\aleph_0} = \aleph_1 \tag{1}$$

(in (1) some standard facts from the algebra of cardinals have been used, see in Stoll (1979)) which shows that

$$card\left(^X\overline{\mathcal{C}}_{a,b}^{c,d}\right) \leq \aleph_1. \tag{2}$$

(ii) On the other hand, $^X\overline{\mathcal{C}}_{a,b}^{c,d}$ contains all the constant functions which form a set of cardinality $\aleph_1$ (since $card\left([c, d]\right) = \aleph_1$). Hence

$$\aleph_1 \leq card\left(^X\overline{\mathcal{C}}_{a,b}^{c,d}\right) \tag{3}$$

(iii) From (2) and (3) it follows $card(\mathbf{C}_{[a,b]\backslash \mathbf{X}}) = \aleph_1$, and the proof is complete. ∎

The next Lemma says that the number of functions which have (at most) a countable number of discontinuities is $\aleph_1$.

Lemma A.5 $card\left(\widetilde{\mathcal{C}}_{a,b}^{c,d}\right) = \aleph_1$.

Proof. (i) We have

$$\widetilde{\mathcal{C}}_{a,b}^{c,d} = \cup_{X \subseteq [a,b], X \text{ countable}} \left(^X\overline{\mathcal{C}}_{a,b}^{c,d}\right).$$

Hence

$$card\left(\widetilde{\mathcal{C}}_{a,b}^{c,d}\right) \leq \sum_{X \subseteq [a,b], X \text{ countable}} card\left(^X\overline{\mathcal{C}}_{a,b}^{c,d}\right) = \sum_{X \subseteq [a,b], X \text{ countable}} \aleph_1 = \aleph_1 \aleph_1 = \aleph_1 \tag{4}$$

where the following facts have been used: (i) the number of all countable subsets of $[a, b]$ is $\aleph_1$ (i.e. that $\aleph_1^{\aleph_0} = \aleph_1$) and (ii) $\aleph_1 \aleph_1 = \aleph_1$. The proofs of both these facts can be found in Stoll (1979).

(ii) On the other hand, the set of constant functions $f(x) = y$ (with $y \in [c, d]$) is a subset of $\widetilde{\mathcal{C}}_{a,b}^{c,d}$ and, clearly, there are $\aleph_1$ such functions. Hence

$$\aleph_1 \leq card\left(\widetilde{\mathcal{C}}_{a,b}^{c,d}\right). \tag{5}$$

(iii) From (4) and (5) it follows $card\left(\widetilde{\mathcal{C}}_{a,b}^{c,d}\right) = \aleph_1$ and the proof is complete. $\blacksquare$

The main result can now be proved.

Proposition A.6 $card(\mathcal{I}_{a,b}^{c,d}) = \aleph_1$.

Proof. (i) Since every nondecreasing function can have at most a countable number of discontinuities (Lemma A.1) it follows

$$\mathcal{I}_{a,b}^{c,d} \subseteq \widetilde{\mathcal{C}}_{a,b}^{c,d} \Rightarrow card(\mathcal{I}_{a,b}^{c,d}) \leq card(\widetilde{\mathcal{C}}_{a,b}^{c,d}) = \aleph_1. \tag{6}$$

(ii) On the other hand, the set of constant functions $f(x) = y$ is a subset of $\mathcal{I}_{a,b}^{c,d}$ and, as already mentioned, there are $\aleph_1$ such functions. Hence

$$\aleph_1 \leq card(\mathcal{I}_{a,b}^{c,d}). \tag{7}$$

(iii) From (6) and (7) it follows $card\left(\mathcal{I}_{a,b}^{c,d}\right) = \aleph_1$ and the proof is complete. $\blacksquare$

B   Appendix: Additional proofs

This Appendix shows the proofs of Propositions 3.1, 4.7, 4.8, and 4.9.

Proof of Proposition 3.1 It is shown that $card(\mathsf{F}) = \aleph_1$.

It has been mentioned (Proposition 2.2) that every fuzzy number can be written in a representation which involves numbers p, q, r, s and functions $L(x)$, $R(x)$. Hence to every fuzzy number F there corresponds a representation $(p, q, r, s, L(x), R(x))$. This correspondence is not 1-to-1; there may be two different representations for the same fuzzy number. However, certainly the total of fuzzy numbers cannot be more than the total of such representations. Denote the set of all sextuples $(p, q, r, s, L(x), R(x))$ by $\mathcal{S}$; it follows

$$card(\mathsf{F}) \leq card(\mathcal{S}).$$

What is the cardinality of $\mathcal{S}$? p, q, r, s can be any real numbers as long as they satisfy $p \leq q \leq r \leq s$; hence there is a total of $\aleph_1 \times \aleph_1 \times \aleph_1 \times \aleph_1$ quadruples (p, q, r, s). Also, for every choice of (p, q, r, s), $L(x)$ is a nondecreasing function with domain $[p, q]$ and range $[0, 1]$, i.e. an element of $\mathcal{I}_{p,q}^{0,1}$ and $card\left(\mathcal{I}_{p,q}^{0,1}\right) = \aleph_1$ by Proposition A.6. Finally, $R(x)$ is a nonincreasing function with domain $[r, s]$ and range $[0, 1]$. It is easy to say that there is a 1-to-1 correspondence between such functions and the functions from $\mathcal{I}_{r,s}^{0,1}$ and, again, $card\left(\mathcal{I}_{r,s}^{0,1}\right) = \aleph_1$. Hence, the set $\mathcal{S}$ of all admissible sextuples $(p, q, r, s, L(x), R(x))$ satisfies

$$card(\mathcal{S}) \leq \aleph_1 \times \aleph_1 \times \aleph_1 \times \aleph_1 \times \aleph_1 \times \aleph_1 = \aleph_1^6 = \aleph_1$$

and so it follows

$$card(\mathsf{F}) \leq card(\mathcal{S}) \leq \aleph_1. \tag{8}$$

On the other hand consider the set F_1 of all functions of the form

$$F(x) = \begin{cases} 0 & \text{for } x < -1 \\ y & \text{for } x \in [-1, 1] \\ 0 & \text{for } x > 1 \end{cases}$$

where y is any constant from the interval $[0, 1]$. Clearly $\mathsf{F}_1 \subseteq \mathsf{F}$ and $card\,(\mathsf{F}_1) = \aleph_1$. Hence

$$\aleph_1 \leq card\,(\mathsf{F}_1) \leq card\,(\mathsf{F}). \tag{9}$$

From (8) and (9) follows that $card\,(\mathsf{F}) = \aleph_1$ and the proof of Proposition 3.1 is complete.∎

Proof of Proposition 4.7 It is shown that $\left(\mathsf{M}^h, \preceq\right)$ is a lattice and that

$$[a, b]^h \curlyvee [c, d]^h = [a \wedge c, b \vee d]^h, \qquad [a, b]^h \curlywedge [c, d]^h = [a \vee c, b \wedge d]^h.$$

Choose some $h \in (0, 1]$ and consider it fixed. For simplicity, in this proof the superscript h will be dropped from all generalized intervals.

To prove the above facts, consider the product lattice $(L \times L, \leq_\partial \times \leq)$. Elements of this lattice are pairs (a, b) with $a, b \in L$. For brevity $\sqsubseteq$ will be used instead of $\leq_\partial \times \leq$. Also, the meet operation in $(L \times L, \sqsubseteq)$ will be denoted by $\sqcap$, and the join operation will be denoted by $\sqcup$.

It will be shown that the algebras $\left(\mathsf{M}^h, \preceq\right)$ and $(L \times L, \sqsubseteq)$ are isomorphic. To this end, it suffices to show that, for every $a, b, c, d \in L$, it holds

$$(a, b) \sqsubseteq (c, d) \Leftrightarrow [a, b] \preceq [c, d].$$

1. It is shown first that $(a, b) \sqsubseteq (c, d) \Rightarrow [a, b] \preceq [c, d]$. Assume then that $(a, b) \sqsubseteq (c, d)$, in other words that

$$c \leq a \text{ and } b \leq d.$$

   Consider now several cases.

   (a) $c \leq a$, $b \leq d$ **and** $a \leq b$. Then

   $$c \leq a \leq b \leq d \Rightarrow \left. \begin{cases} [a, b] \in \mathsf{M}^h_+ \\ [c, d] \in \mathsf{M}^h_+ \\ [a, b] \subseteq [c, d] \end{cases} \right\} \Rightarrow [a, b] \preceq [c, d].$$

   (b) $c \leq a$, $b \leq d$ **and** $b < a$ **and** $d < c$. Then

   $$b \leq d < c \leq a \Rightarrow \left. \begin{cases} [a, b] \in \mathsf{M}^h_- \\ [c, d] \in \mathsf{M}^h_- \\ [d, c] \subseteq [b, a] \end{cases} \right\} \Rightarrow [a, b] \preceq [c, d].$$

16

(c) $c \le a$, $b \le d$ **and** $b < a$ **and** $c \le d$. Then

$$\left.\begin{array}{c} b \le b \vee c \le a \\ c \le b \vee c \le d \end{array}\right\} \Rightarrow \left.\begin{array}{c} [a,b] \in \mathsf{M}^h_- \\ [c,d] \in \mathsf{M}^h_+ \\ [c,d] \cap [b,a] \ne \emptyset \end{array}\right\} \Rightarrow [a,b] \preceq [c,d].$$

All possibilities have been exhausted by (a), (b), (c) above, hence $(a,b) \sqsubseteq (c,d) \Rightarrow [a,b] \preceq [c,d]$.

2. Next it is shown that $[a,b] \preceq [c,d] \Rightarrow (a,b) \sqsubseteq (c,d)$. Assume then that $[a,b] \preceq [c,d]$, and now consider several cases.

(a) $[a,b], [c,d] \in \mathsf{M}^+_h$. Then $a \le b$, $c \le d$ and $[a,b] \subseteq [c,d]$. Hence $c \le a \le b \le d$ and so $(a,b) \sqsubseteq (c,d)$.

(b) $[a,b], [c,d] \in \mathsf{M}^-_h$. Then $b \le a$, $d \le c$ and $[d,c] \subseteq [b,a]$. Hence $b \le d \le c \le a$ and so $(a,b) \sqsubseteq (c,d)$.

(c) $[a,b] \in \mathsf{M}^-_h$ and $[c,d] \in \mathsf{M}^+_h$. Then $b \le a$, $c \le d$ and $[c,d] \cap [b,a] \ne \emptyset$. The latter implies

$$c \vee b \le d \wedge a$$

hence $b \le d$ and $c \le a$ which implies $(a,b) \sqsubseteq (c,d)$.

All possibilities have been exhausted by (a), (b), (c) above, hence $[a,b] \preceq [c,d] \Rightarrow (a,b) \sqsubseteq (c,d)$.

3. In short, it has been shown, as required, that $(a,b) \sqsubseteq (c,d) \Leftrightarrow [a,b] \preceq [c,d]$. Hence $\left(\mathsf{M}^h, \preceq\right)$ and $(L \times L, \sqsubseteq)$ are isomorphic, and hence $\left(\mathsf{M}^h, \preceq\right)$ is a lattice. Furthermore,

$$(a,b) \sqcap (c,d) \leftrightarrow [a,b] \curlywedge [c,d] \ \text{and} \ (a,b) \sqcup (c,d) \leftrightarrow [a,b] \curlyvee [c,d].$$

Hence

$$(a,b) \sqcap (c,d) = (a \vee b, c \wedge d) \Rightarrow [a,b] \curlywedge [c,d] = [a \vee c, b \wedge d]$$
$$(a,b) \sqcup (c,d) = (a \wedge c, b \vee d) \Rightarrow [a,b] \curlyvee [c,d] = [a \wedge c, b \vee d].$$

and the proof of the proposition is complete.∎

Proof of Proposition 4.8 Let $f_h : \mathsf{R} \to \mathsf{R}$ be a strictly increasing function. It will be shown that the function $v_h : \mathsf{M}^h \to \mathsf{R}$ defined by

$$v_h([a,b]^h) = f_h(b) - f_h(a).$$

is a positive valuation in $\left(\mathsf{M}^h, \preceq\right)$, that is function v_h satisfies the two conditions of definition 2.4.

First, let $[a,b]^h$ and $[c,d]^h$ be generalized intervals in M^h. It follows

$$v_h([a,b]^h) + v_h([c,d]^h) = [f_h(b) - f_h(a)] + [f_h(d) - f_h(c)]$$
$$= [f_h(b) + f_h(d)] - [f_h(a) + f_h(c)]$$
$$= [f_h(b \vee d) + f_h(b \wedge d)] - [f_h(a \vee c) + f_h(a \wedge c)]$$
$$= [f_h(b \vee d) - f_h(a \wedge c)] + [f_h(b \wedge d) - f_h(a \vee c)]$$
$$= v_h([a \wedge c, b \vee d]^h) + v_h([a \vee c, b \wedge d]^h)$$
$$= v_h([a,b]^h \vee [c,d]^h) + v_h([a,b]^h \wedge [c,d]^h).$$

That is, v_h is a valuation function.

Second, let $[a,b]^h < [c,d]^h$ for generalized intervals $[a,b]^h$ and $[c,d]^h$ in M^h. There are three cases:

1. Both $[a,b]^h$ and $[c,d]^h$ are in M_+^h. It follows $((c \leq a)$ and $(b < d))$ or $((c < a)$ and $(b \leq d))$.

2. Both $[a,b]^h$ and $[c,d]^h$ are in M_-^h. It follows $((b \leq d)$ and $(c < a))$ or $((b < d)$ and $(c \leq a))$.

3. $[a,b]^h$ is in M_-^h and $[c,d]^h$ is in M_+^h. Since always is $b < a$, there is a point $x \in [c,d]$ such that at least one of $(b < x)$ or $(x < a)$ holds; it follows either $(b < d)$ or $(c < a)$, respectively.

All three cases above imply the following strict inequality

$$f_h(b) + f_h(c) < f_h(a) + f_h(d).$$

Hence, $[a,b]^h < [c,d]^h$ implies

$$f_h(b) + f_h(c) < f_h(a) + f_h(d) \implies f_h(b) - f_h(a) < f_h(d) - f_h(c) \implies v_h([a,b]^h) < v_h([c,d]^h).$$

That is, v_h is a positive valuation function.$\blacksquare$

Proof of Proposition 4.9 It will be shown that $card(\mathbf{D}_h) \geq \aleph_1$. This is actually rather easy. From Proposition 4.8 it is known that every strictly increasing function $f_h(x)$ yields a metric d_h by the formula

$$d_h\left([a,b]^h, [c,d]^h\right) = f_h(a \vee c) - f_h(a \wedge c) + f_h(b \vee d) - f_h(b \wedge d).$$

Now, choose any $\kappa \in \mathsf{R}_0^+$ and define $f_h^\kappa(x) = \kappa x$. Evidently,

$$d_h^\kappa\left([a,b]^h, [c,d]^h\right) = \kappa \cdot ((a \vee c) - (a \wedge c) + (b \vee d) - (b \wedge d))$$

is a metric; furthermore when $\kappa \neq \lambda$, it follows $d_h^\kappa \neq d_h^\lambda$. Since $card\left(\mathsf{R}_0^+\right) = \aleph_1$, there are $\aleph_1$ distances d_h^κ and, since the set $\{d_h^\kappa\}_{\kappa \in \mathsf{R}_0^+} \subseteq \mathbf{D}_h$ it follows that $\aleph_1 = card\left(\{d_h^\kappa\}_{\kappa \in \mathsf{R}_0^+}\right) \leq card(\mathbf{D}_h).\blacksquare$

References

G. Birkhoff, *Lattice Theory*, Providence, RI: American Mathematical Society, Colloquium Publications 25, 1967.

N. Cassaigne and M.G. Singh, "Intelligent decision support for the pricing of products and services in competitive consumer markets", *IEEE Trans. Systems, Man and Cybernetics – Part C*, 31, pp. 96-106, Feb. 2001.

J.A. Dickerson and B. Kosko, "Fuzzy function approximation with ellipsoidal rules", *IEEE Trans. Systems, Man and Cybernetics – Part B*, 26, pp. 542-560, Aug. 1996.

B. Ganter and R. Wille, *Formal Concept Analysis: Mathematical Foundations*, Heidelberg, Germany: Springer-Verlag, 1999.

H. Ishibuchi, T. Nakashima and T. Murata, "Performance evaluation of fuzzy classifier systems for multidimensional pattern classification problems", *IEEE Trans. Systems, Man and Cybernetics - Part B*, 29, pp. 601-618, Oct. 1999.

V.G. Kaburlasos, "Novel fuzzy system modeling for automatic control applications", in *4th Intl. Conf. Technology & Automation*, 2002, pp. 268-275.

V.G. Kaburlasos, "FINs: Lattice theoretic tools for improving prediction of sugar production from populations of measurements", *IEEE Trans. Systems, Man and Cybernetics – Part B*, 34, pp. 1017-1030, Apr. 2004.

V.G. Kaburlasos and A. Kehagias, "Novel analysis and design of fuzzy inference systems based on lattice theory", in *IEEE Intl. Conf. Fuzzy Systems* (*FUZZ-IEEE*), 2004, pp. 281-286.

V.G. Kaburlasos and A. Kehagias, "Novel fuzzy inference system (FIS) analysis and design based on lattice theory. Part II: Industrial system modeling", (*submitted*).

V.G. Kaburlasos and S.E. Papadakis, "Granular self-organizing map (grSOM) for structure identification", *Neural Networks*, 2005 (*in press*).

V.G. Kaburlasos and V. Petridis, "Fuzzy lattice neurocomputing (FLN): A novel connectionist scheme for versatile learning and decision making by clustering", *Intl. J. Computers and Their Applications*, 4, pp. 31-43, Dec. 1997.

V.G. Kaburlasos and V. Petridis, "Fuzzy lattice neurocomputing (FLN) models", *Neural Networks*, 13, pp. 1145-1170, Dec. 2000.

E. Kamke, *Theory of Sets*, New York, NY: Dover Publications, 1950.

N.N. Karnik, J.M. Mendel and Qilian Liang, "Type-2 fuzzy logic systems", *IEEE Trans. Fuzzy Systems*, 7, pp. 643-658, Dec. 1999.

B. Kosko, "Fuzzy systems as universal approximators", *IEEE Trans. Computers*, 43, pp. 1329-1333, Nov. 1994.

R. Krishnapuram and J.M. Keller, "A possibilistic approach to clustering", *IEEE Trans. Fuzzy Systems*, 1, pp. 98-110, May 1993.

E.H. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller", *Intl. J. Man-Machine Studies*, 7, pp. 1-13, 1975.

H.T. Nguyen and E.A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall, 1999.

S.E. Papadakis, P. Tzionas, V.G. Kaburlasos and J.B. Theocharis, "A genetic based approach to the Type I structure identification problem", *Informatica*, 16, pp. 365-382, 2005.

K. Passino and S. Yurkovich, *Fuzzy Control*, Reading, MA: Addison Wesley Longman, 1998.

S. Paul and S. Kumar, "Subsethood-product fuzzy neural inference system (SuPFuNIS)", *IEEE Trans. Neural Networks*, 13, pp. 578-599, May 2002.

W. Pedrycz, J.C. Bezdek, R.J. Hathaway and G.W. Rogers, "Two parametric models for fusing heterogeneous fuzzy data", *IEEE Trans. Fuzzy Systems*, 6, pp. 411-425, Aug. 1998.

V. Petridis and V.G. Kaburlasos, "Learning in the framework of fuzzy lattices", *IEEE Trans. Fuzzy Systems*, 7, pp. 422-440, Aug. 1999.

V. Petridis and V.G. Kaburlasos, "Clustering and classification in structured data domains using Fuzzy Lattice Neurocomputing (FLN)", *IEEE Trans. Knowledge and Data Engineering*, 13, pp. 245-260, Mar/Apr. 2001.

V. Petridis and V.G. Kaburlasos, "FINkNN: A fuzzy interval number k-nearest neighbor classifier for prediction of sugar production from populations of samples", *J. Machine Learning Research*, 4, pp. 17-37, Apr. 2003.

Puyin Liu, "Mamdani fuzzy system: Universal approximator to a class of random processes", *IEEE Trans. Fuzzy Systems*, 10, pp. 756-766, Dec. 2002.

J.F. Sowa, *Knowledge Representation: Logical, Philosophical, and Computational Foundations*, Pacific Grove, CA: Brooks Cole Publishing, 2000.

R.R. Stoll, *Set Theory and Logic*, New York, NY: Dover Publications, 1979.

T. Tagaki and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control", *IEEE Trans. Systems, Man, and Cybernetics*, 15, pp. 116-132, Jan/Feb. 1985.

K. Uehara and M. Fujise, "Fuzzy inference based on families of α-level sets", *IEEE Trans. Fuzzy Systems*, 1, pp. 111-124, May 1993.

L.-X. Wang and J.M. Mendel, "Fuzzy basis functions, universal approximation, and orthogonal least-squares learning", *IEEE Trans. Neural Networks*, 3, pp. 807-814, Sep. 1992.

Xiao-Jun Zeng and M.G. Singh, "Approximation accuracy analysis of fuzzy systems as function approximators", *IEEE Trans. Fuzzy Systems*, 4, pp. 44-63, Feb. 1996.

Xiao-Jun Zeng and M.G. Singh, "Fuzzy bounded least-squares method for the identification of linear systems", *IEEE Trans. Systems, Man and Cybernetics – Part A*, 27, pp. 624-635, Sep. 1997.

Xiao-Jun Zeng and M.G. Singh, "Knowledge bounded least squares method for the identification of fuzzy systems", *IEEE Trans. Systems, Man and Cybernetics – Part C*, 33, pp. 24-32, Feb. 2003.

L.A. Zadeh, "Fuzzy sets", *Information and Control*, 8, pp. 338-353, 1965.

L.A. Zadeh, "Precisiated natural language (PNL) - Toward an enlargement of the role of natural languages in scientific theories", in *IEEE Intl. Conf. Fuzzy Systems (FUZZ-IEEE)*, 2004, Keynote Talk I.

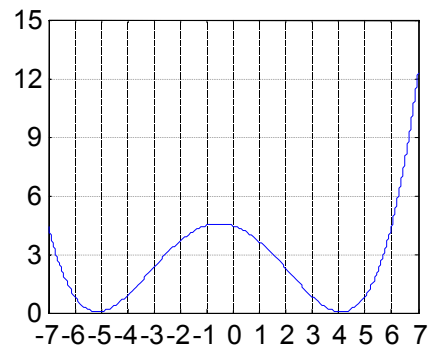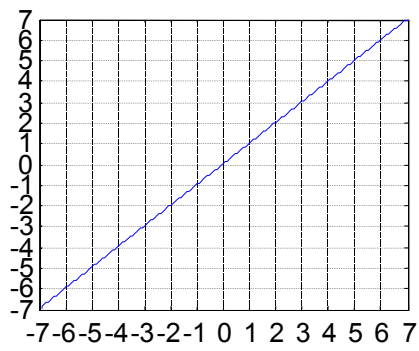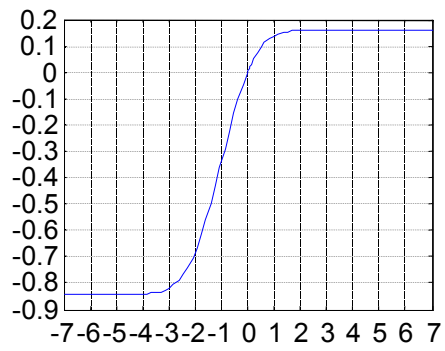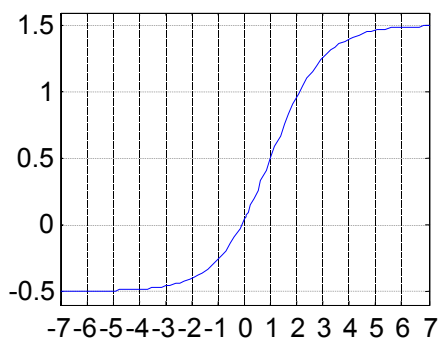H.J. Zimmermann, *Fuzzy Set Theory and Its Applications*, 2nd ed., Boston, MA: Kluwer, 1991.
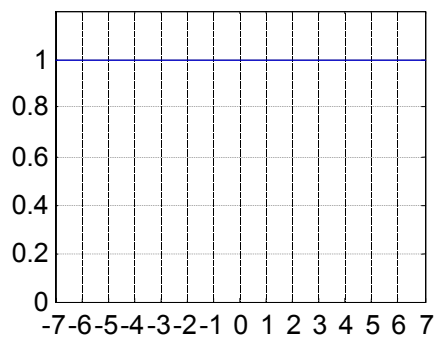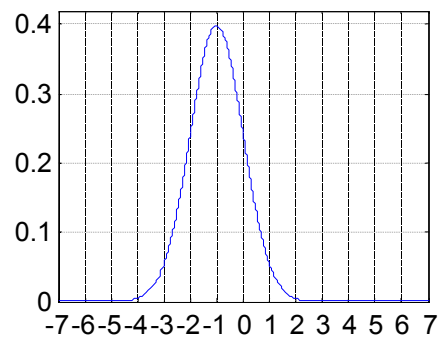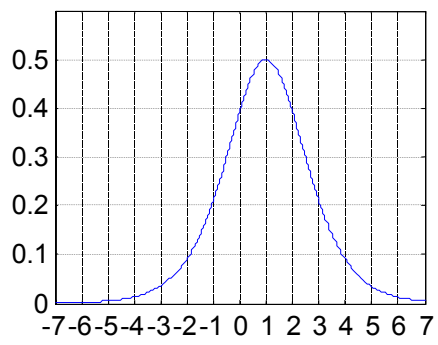
Figure 1:
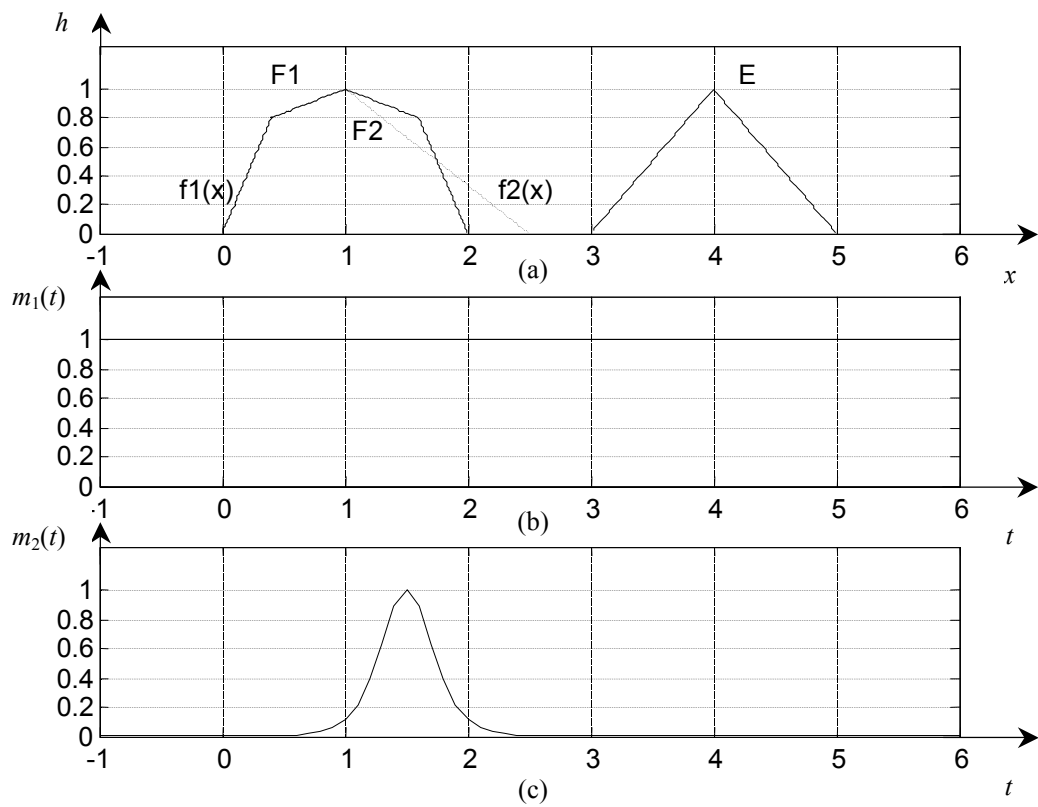
(a)

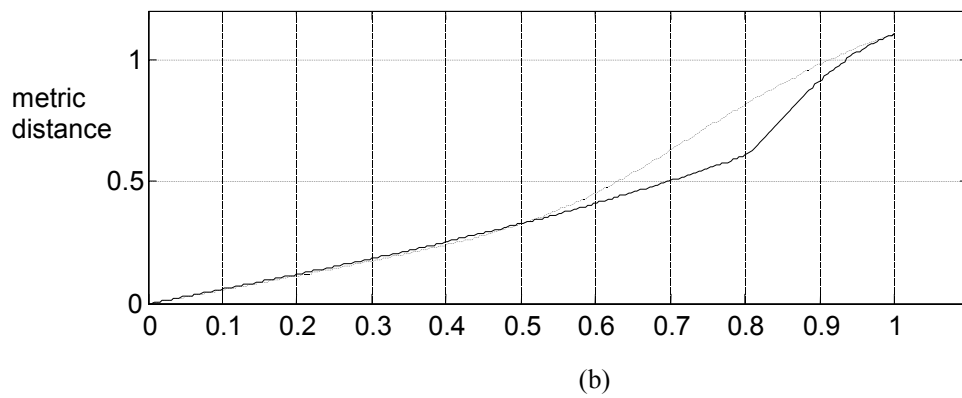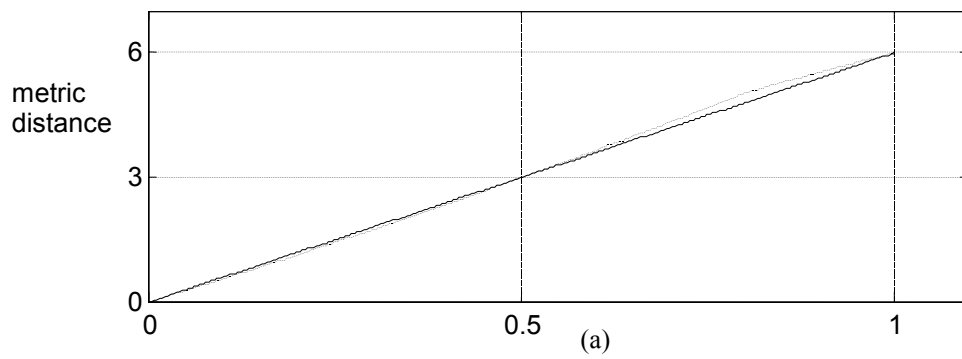(b)

(c)

(d)

Figure 2:

Figure 3:

23

(a)

(b)
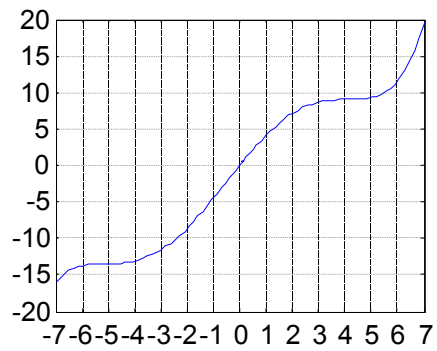
Figure 4:

Figure Captions

Figure 1:

(a) Mass function $m_1(x) = 1$.

(b) The Gaussian probability density function $m_2(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, with $\mu = -1$ and $\sigma = 1$, can be used as a mass function.

(c) Mass function $m_3(x) = 2e^{-(x-1)}/(1 + e^{-(x-1)})^2$.

(d) Polynomial mass function $m_4(x) = 0.01x^4 + 0.02x^3 - 0.41x^2 - 0.42x + 4.44$.

Figure 2:

(a) The linear underlying positive valuation function $f_1(x) = \int_0^x m_1(t)dt = x$ derived from the mass function $m_1(t)$ in Fig.1(a).

(b) A normal cumulative distribution function can be used as an underlying positive valuation function. This positive valuation function is derived from the mass function $m_2(t)$ in Fig.1(b).

(c) The saturated *logistic* (*sigmoid*) underlying positive valuation function $f_3(x) = \int_0^x m_3(t)dt = \frac{2}{1+e^{-(x-1)}} - 1$ derived from the mass function $m_3(t)$ in Fig.1(c).

(d) An unbounded polynomial underlying positive valuation function $f_4(x) = \int_0^x m_4(t)dt = 0.002x^5 + 0.005x^4 - 0.1367x^3 - 0.21x^2 + 4.44x$ derived from the mass function $m_4(t)$ in Fig.1(d).

Figure 3:

(a) Three fuzzy numbers F1, F2, and E, whose domain interval $[0,1]$ is shown on the vertical axis. The left sides of F1 (solid line) and F2 (dashed line) coincide. The membership functions of fuzzy numbers F1 and F2 are denoted explicitly by f1(x) and f2(x), respectively.

(b) The mass function $m_h(t) = h$, for h=1.

(c) The mass function $m_h(t) = 4he^{-7(t-1.5)}/(1 + e^{-7(t-1.5)})^2$, for h=1.

Figure 4:

The fuzzy numbers $F1$, $F2$, and E mentioned below are shown in Fig.3(a).

(a) The metric distance functions $d_K(F1(h), E(h))$ and $d_K(F2(h), E(h))$ are plotted in solid and dashed lines, respectively, using the mass function $m_h(t) = h$ shown in Fig.3(b) for $h = 1$. The area underneath a curve equals the corresponding distance between two fuzzy numbers; it turns out $d_K(F1, E) \approx 3.0 > 2.9754 \approx d_K(F2, E)$.

(b) The metric functions $d_K(F1(h), E(h))$ and $d_K(F2(h), E(h))$ are plotted, respectively, in solid and dashed lines using the mass function $m_h(t) = 4he^{-7(t-1.5)}/(1 + e^{-7(t-1.5)})^2$ shown in Fig.3(c) for $h = 1$. The area underneath a curve equals the distance between two fuzzy numbers; it turns out $d_K(F1, E) \approx 0.3587 < 0.3811 \approx d_K(F2, E)$.