Possible New Directions in Mathematical Foundations of Fuzzy Technology: A Contribution to the Mathematics of Fuzzy Theory

Hung T. Nguyen[1] and Vladik Kreinovich[2]

[1]Department of Mathematical Sciences, New Mexico State University
Las Cruces, NM 88003, USA, email hunguyen@nmsu.edu
[2]Department of Computer Science, University of Texas at El Paso,
El Paso, TX 79968, USA, email vladik@cs.utep.edu

1 Introduction

Why mathematical foundations? Many successful applications of fuzzy logic and fuzzy set theory [20, 55], such as fuzzy control and many methods of fuzzy imaging, first appeared as *heuristics*, without any precise mathematical justification. In such heuristic applications, the choice of techniques or parameters is usually done empirically. After a sufficient amount of the corresponding empirical data becomes available, this data inspires *mathematical foundations* for the original heuristics and empirical choices:

- in some cases, these mathematical foundations simply confirm the already made choices; in such cases, the existence of mathematical foundations increased the *reliability* of the existing techniques;
- in other cases, the detailed mathematical analysis not only justified the existing techniques, but also provided better techniques, thus improving the *quality* of the resulting applications (e.g., the stability or smoothness of fuzzy control; see, e.g., [29, 44] and references therein).

Generalizations of traditional $[0,1]$-based fuzzy techniques. Most applications of fuzzy methodology are based on the traditional, $[0,1]$-based fuzzy logic. While this traditional fuzzy logic is still an important source of new applications, it has been noticed for some time already that there also exist important practical problems in which the traditional fuzzy techniques are not sufficient. Their existence was first noticed by L. Zadeh himself, and Zadeh has proposed new (heuristic) ideas which lead to successful new applications of fuzzy methodology:

- The first group of such ideas contains the ideas of Computing with Words and Granularity as a new basis for fuzzy techniques. Crudely speaking, these ideas take into consideration that the traditional interval $[0,1]$ for truth values is, in itself, only a useful first approximation, and in some applications, a more adequate set of truth values is needed.
- The second group of ideas is related to the concept of Soft Computing, an umbrella term which combines fuzzy, neural, genetic, interval, probabilistic, and other techniques, and which aims at a unified approach in which each of these techniques would be used for appropriate niche situations.

What we are planning to do. In this paper, we describe new possible application-oriented directions towards formalizing these new ideas:

- Re non-$[0,1]$ fuzzy logics: there exist many theoretical generalizations of $[0,1]$-fuzzy logic: lattice logics, second order fuzzy sets, etc. However, many of these notions are currently too general to be practically efficient. In this paper, we describe application-oriented efficient subcases of these general concepts such as:
  - multi-D generalizations of fuzzy logic (interval-valued, complex-valued, logics based on partially ordered linear and non-linear spaces, etc.), and
  - discrete modifications attempting to directly formalize the notion of computing with words.
- Re soft computing: there exist joint formalisms combining fuzzy with neural and genetic, but much fewer mathematical results combine fuzzy with *probability* and alternative AI methods such as *logic programming*. In this paper, we mention the corresponding results and directions.

We also outline the applications (existing and potential) of the new formalisms:

- to image processing (in particular, to justifying heuristic fuzzy and non-fuzzy methods in image processing and in data processing in general);

- to decision making (especially related to image processing);
- to foundations of physics; and
- to education (and applied psychology in general).

Most of our application results came from several collaboration projects with researchers from different fields; these application results are presented in the separate papers.

2 Multi-D generalizations of fuzzy logic: why

2.1 Why fuzzy technology: the main idea

Experts are often very skilled in solving real-life problems (e.g., in control). Therefore, when we design an automated systems for solving such problems, it is desirable to use expert's knowledge, skills, and experience.

- Sometimes, this expert knowledge is formulated in *precise* terms, so that it is reasonably easy to represent this knowledge in terms understandable to a computer.
- However, in many real-life situations, an expert cannot formulate his knowledge in precise mathematical terms, he can only describe his knowledge by using *imprecise* words from natural language.

So, for a computer to be able to use this knowledge, we must first represent this knowledge in a form which is understandable for a computer. The corresponding methodology was proposed by L. Zadeh under the name of *fuzzy* methodology [67].

One of the main ideas behind fuzzy methodology is as follows:

- When we represent the *precise* (*crisp*) knowledge, then every statement is either true or false. For example, a crisp control rule may use a condition "if x is positive" (i.e., "if $x > 0$"), which is true for all positive values x and false for all other values x.
- On the other hand, when we represent the imprecise knowledge, we may have a condition $P(x)$ of the type "if x is large". This condition is absolutely false for very small values x, absolutely true for truly large values of x, but for intermediate values of x, the expert is *uncertain* whether the corresponding value x is large or not.

So, to describe, for every x, the expert's knowledge about $P(x)$, it is not sufficient to have *two* possible "truth values" (*degrees of certainty*) "true" and "false", we must also allow *intermediate* degrees of certainty.

To implement this idea, we must be able:

- first, to *represent* these intermediate degrees of certainty inside a computer, and
- then, to *process* these values.

2.2 Traditional $[0,1]$-based fuzzy technology and its limitations

In the computer, "true" is usually represented as 1, and "false" as 0. Therefore, it is natural to represent *intermediate* degrees of certainty by real numbers between 0 and 1, i.e., by real numbers from the interval $[0,1]$. In this representation, the expert's degree of certainty $d(P(x))$ in each statement $P(x)$ is represented by a real number from the interval $[0,1]$; this number is often denoted by $\mu_P(x)$.

Most of the applications of fuzzy logic methodology are based on this $[0,1]$-based fuzzy logic (see, e.g., [20, 55]). However, from the applied viewpoint, the $[0,1]$-based methodology has a serious *drawback*:

- the very necessity for a fuzzy methodology comes from the fact that expert statements are *imprecise* and fuzzy;
- however, to apply the $[0,1]$-based formalism, we need to assign, to every (fuzzy) property P and to every value x, a *precise* real number $\mu_P(x)$.

If the expert was unable to tell whether a given value x (say 20) is large or not, we cannot expect from this expert that he would be able to tell us his *exact* degree of certainty that 20 is large. We can ask the expert to mark this degree on a scale from 0 to 1, but this will enable us to get an *approximate* value, not the *precise* one:

- an expert can probably meaningfully distinguish between degree of certainty 0.7 and 0.8 (or, at least, 0.6 and 0.8), but
- it is highly improbable that an expert would be able to meaningfully distinguish between, say degrees of certainty 0.7 and 0.701.

  It is known in psychology that, in general, humans are most comfortable with 5 to 9 items to choose from ("7 plus minus 2" law, see, e.g., [36, 37] and [16], where this law is used to describe the number of different membership functions in fuzzy control rules).

In other words, we arrive at the following problem:

- The traditional $[0,1]$-based fuzzy methodology deals with the degrees of certainty as if they were *exactly* known.
- However, in reality, these degrees are themselves known *imprecisely.*

So, to apply the $[0,1]$-fuzzy methodology, we must somehow transform the imprecise information into an exact number. Since the original information is imprecise, we can get somewhat *different* precise numbers (degree of certainty) to represent the *same* information. Different degrees of certainty, in their turn, may lead to different controls, different decisions, etc. How can we avoid this undesirable difference?

2.3 The first approach to solving the problem of $[0,1]$-based fuzzy methodology: choosing the least sensitive fuzzy techniques

The problem that we are trying to solve is that the same original fuzzy information can lead to slightly different degrees of certainty, and this can lead to slightly different controls, decisions, etc. We would like our decisions to depend *only* on the initial expert information and *not* on the exact way in which we represent this information in our computers. Therefore, we would like to choose a fuzzy methodology in which the results of applying fuzzy techniques are the least sensitive to the small changes in the initial degrees of certainty. Let us describe how this can be done, on the example of choosing "and" and "or"-operations.

As we have mentioned, elicitation methods are *approximate*, i.e., we can get two different values a and a' to represent the same degree of certainty of a statement A. Similarly, for some other statement B, elicitation can lead to two different values $b \neq b'$. As a result, when we want to estimate the degree of certainty in $A\&B$, we can apply the &-operation to different pairs of values, and get two different results: $f_\&(a,b)$ and $f_\&(a',b')$. We would like to choose an &-operation in such a way that this difference $f_\&(a',b') - f_\&(a,b)$ between these possible results is the smallest possible.

In [41, 49, 50], we assumed that the elicitation error is characterized by its *absolute* upper bound $\delta > 0$, i.e., that if two values a and a' correspond to the same degree of certainty, then $|a-a'| \leq \delta$. With this definition, we got the following results:

Definition 1.

- *By a binary operation (or operation for short) we mean a function $f(a,b)$ from $[0,1]\times[0,1]$ into $[0,1]$.*
- *A binary operation is called a &-operation if the following conditions are true:*
  - $f(0,a)=0$; $f(1,a)=a$;
  - $f(a,b)=f(b,a)$ *for all* a,b;
  - $f(a,b)\leq a$ *for all* a *and* b.
- *A binary operation is called an $\vee$-operation if the following conditions are true:*
  - $f(0,a)=a$; $f(1,a)=1$;
  - $f(a,b)=f(b,a)$ *for all* a,b;
  - $f(a,b)\geq a$ *for all* a *and* b.

Remark. The above binary operations are slightly more general than the usual t-norms and t-conorms in the literature [20, 55].

Definition 2. *Let $\delta > 0$ be a positive real number. We say that two real numbers are δ-close if $|a-a'|\leq\delta$.*

Definition 3. *Let f be a binary operation, and let $\delta > 0$ be a real number. By a $\delta-$sensitivity $r_f(\delta)$ of an operation f, we mean the smallest real number α for which for all a, a', b, and b', if a is δ-close to a' and b is δ-close to b', then $f(a,b)$ is α-close to $f(a',b')$.*

Definition 4.

- *We say that operations $f(a,b)$ and $g(a,b)$ are equally sensitive if for every δ, $r_f(\delta)=r_g(\delta)$.*
- *We say that an operation $f(a,b)$ is less sensitive than an operation $g(a,b)$, if for every δ, $r_f(\delta)\leq r_g(\delta)$, and at least for one $\delta>0$, $r_f(\delta)<r_g(\delta)$.*
- *We say that an &-operation $f(a,b)$ is the least sensitive &-operation, if it is either less sensitive, or equally sensitive than any other &-operation.*
- *We say that an $\vee$-operation $g(a,b)$ is the least sensitive $\vee$-operation, if it is either less sensitive, or equally sensitive than any other $\vee$-operation.*

With respect to absolute elicitation error, we get the following result:

Proposition 1. [41, 49, 50] *For absolute elicitation error:*

- $f(a,b)=\min(a,b)$ *is the least sensitive &-operation, and*
- $f(a,b)=\max(a,b)$ *is the least sensitive $\vee$-operation.*

Comments.

- Instead of comparing the *worst-case* sensitivity, we could compare the *average-case* sensitivity. With respect to the average-case sensitivity, the least sensitive operations are $f_\&(a,b)=a\cdot b$ and $f_\vee(a,b)=a+b-a\cdot b$ [41, 52, 55].
- Alternatively, instead of the *absolute* error, we may consider a more adequate *relative* error. Indeed:
  - if we try to give an expert estimate of a probability, then it may be reasonable to assume that, say, two estimates 0.7 and 0.75 are close (in the sense that they may describe the same probability);
  - this, however, does not necessarily mean that the values 0 and 0.05 (with the same absolute difference 0.05) are close in the above sense: there is a big intuitive difference between 0 (meaning that the event is impossible) and 0.05 (meaning that the event is possible but rare), the difference that did not exist between 0.7 and 0.75.

  It is much more intuitively reasonable to assume that when we consider 0.7 and 0.75 to be close, then we should consider 0.07 and 0.075 to be close too. In other words, it is more reasonable to believe that the above-described "closeness" is characterized more adequately not by *absolute* difference $a - a'$, but rather by a *relative* difference $(a - a')/a$. This leads us to the following definition:

Definition 2$'$. *Let $\delta > 0$ be a positive real number. We say that two real numbers are δ-close if $|a - a'| \le \delta \cdot |a|$ and $|a' - a| \le \delta \cdot |a'|$.*

Comments.

- We formulated the inequalities in the form $|a-a'| \le \delta\cdot|a|$, and not in the form $|a-a'|/|a| \le \delta$, because we want to also cover the case $a = 0$.

- In Definition 2, we only had *one* inequality, because the inequality $|a-a'| \le \delta$ does not change if we "swap" a and a'. For relative errors, the corresponding inequality is not symmetric and hence, we have to postulate *both* the original inequality relating a and a' and a similar inequality relating b and b'.

For relative errors, we get the same choice of "and" and "or"-operations:

Proposition 2. *For relative elicitation error:*

- *$f(a,b) = \min(a,b)$ is the least sensitive &-operation, and*

- *$f(a,b) = \max(a,b)$ is the least sensitive $\vee$-operation.*

Comments.

- For readers' convenience, all the proofs are placed in the last section.

- In the proof of Proposition 1 (as given in [41, 49, 50]), we did not use all the properties of "and" and "or"-operations from Definition 1.1. Namely:

  - For "and", it is sufficient, instead of requiring that $f(0,a) = 0$ and $f(1,a) = a$ be true for *all* a, to require these equalities only for $a = 0$ and $a = 1$.

  - Similarly, for "or", it is sufficient, instead of requiring that $f(0,a) = a$ and $f(1,a) = 1$ be true for *all* a, to require these equalities only for $a = 0$ and $a = 1$.

  For *relative* elicitation errors, we cannot make this restriction, because then, as one can easily see, the following crisp-valued binary operations will be the least sensitive: $f_\&(a,b) = \mathrm{sgn}(a) \,\&\, \mathrm{sgn}(b)$ and $f_\vee(a,b) = \mathrm{sgn}(a) \vee \mathrm{sgn}(b)$, where $\mathrm{sgn}(0) = 0$ and $\mathrm{sgn}(a) = 1$ for $a > 0$.

2.4 What if the least sensitive $[0,1]$-fuzzy methodology is still too sensitive

In the above text, we described how to choose the least sensitive $[0,1]$-fuzzy methodology, i.e., the methodology for which the initial imprecision in the degrees of certainty has the smallest effect on the result. If this smallest possible effect is negligible small, then we should apply this methodology. But what if, within $[0,1]$-fuzzy methodology, even this smallest possible effect is still unacceptably large?

In this case, we have no other choice but to replace the $[0,1]$-methodology with a more adequate approach.

3 Second order (and higher order) fuzzy logic, L-fuzzy logic: description and problems

3.1 Second order and higher order fuzzy logic: why and how

Second order fuzzy logic: the main idea. In the original $[0,1]$-based fuzzy formalism, to represent an arbitrary property, we must describe, for every object x from the universe of discourse, a real number $\mu_P(x) \in [0,1]$ which characterizes our degree of certainty that this object x has the desired property. This function which assigns a number to each object is called a *membership function*, or a *fuzzy set*.

One of the main problems of the traditional ($[0,1]$-based) fuzzy logic is that according to this logic, we must use *exact* numbers from the interval $[0,1]$ to represent experts' degrees of certainty, while in reality, these degrees of certainty are, by themselves, a rather *fuzzy* notion. It is, therefore, natural to represent our degree of certainty in a statement A not by a *single* (crisp) number $d(A) \in [0,1]$ (as in the $[0,1]$-based fuzzy logic), but rather by a *fuzzy* number $\mathbf{d}(A)$, i.e., by a *function* $\mu_{\mathbf{d}(A)}$ which assigns, to each possible real number $d \in [0,1]$, a *degree* $\mu_{\mathbf{d}(A)}(d)$ with which this number d can be the (desired) degree of certainty of A.

In the resulting formalism, if we want to describe an arbitrary property $P(x)$, then, instead of a fuzzy set, we must describe a complicated version of a fuzzy set, in which, for each x, we get not a *number* $\mu_P(x)$ but a *fuzzy set* $\mu_P(x)$. In other words, instead of a $[0,1]$-valued fuzzy set, we need a *fuzzy set*-valued fuzzy set. Such a set is usually called a *second order* fuzzy set.

Third and higher order fuzzy logic. In second-order fuzzy logic, to describe a degree with which a given number $d \in [0,1]$ can be a degree of certainty of a statement A, we use a (crisp) *real number* $\mu_{\mathbf{d}(A)}(d)$. As we have already mentioned, it is difficult to describe our degree of certainty by a single number. Therefore, to make this description even more realistic, we can use, instead of a single real number $\mu_{\mathbf{d}(A)}(d)$, a *fuzzy number*. In other words, we can represent each degree of certainty $d(P(x))$ not by a regular ($[0,1]$-based) fuzzy set, but by a *second order* fuzzy set.

As a result, to represent the property P, we use a *second order fuzzy set*-valued fuzzy set. Such fuzzy sets are naturally called *third order* fuzzy sets.

Similarly, to make our description even more realistic, we can use the third order fuzzy sets to describe degrees of certainty; then, we get *fourth order* fuzzy sets, etc.

3.2 Is the third order fuzzy logic really necessary?

Theoretically, we can define third, fourth order, etc., fuzzy sets, but in practical applications, only second order fuzzy sets were used. Based on this empirical fact, it is natural to conclude that third and higher order fuzzy sets are not really necessary. We will show that this conclusion can be theoretically justified.

Let us first describe the problem formally. An expert uses words from a natural language to describe his degrees of certainty. In every language, there are only finitely many words, so we have a finite set of words that needs to be interpreted. We will denote this set of words by W.

Then, if we have any property P on a universe of discourse U, an expert can describe, for each element $x \in U$, his degree of certainty $d(x) \in W$ that the element x has the property P.

Our ultimate goal is to provide a computer representation for each word $w \in W$. In the traditional $[0,1]$-fuzzy logic, this computer representation assigns, to every word, a *real number* from the interval $[0,1]$; in general, we may have some other computer representations (examples will be given later). Let us denote the set of all possible computer representations by S.

In the first approximation, i.e., in the first order fuzzy set, we represent each word $w \in W$, which describes a degree of uncertainty, by an element $s \in S$ (e.g., by a real number from the interval $[0,1]$). In this section, we will denote this first-approximation computer representation of a word w by $s = \|w\|$.

If the set S is too small, then it may not contain enough elements to distinguish between different expert's degree of belief: this was exactly the problem with classical $\{0,1\}$-based logic, in which we only have two possible computer representations – "true" and "false" – that are not enough to adequately describe the different degrees of certainty. We will therefore assume that the set S is rich enough to represent different degrees of certainty. In particular, the set $[0,1]$ contains infinitely many points, so it should be sufficient; even if we only consider computer-representable real numbers, there are still much more of them (millions and billions) than words in a language (which is usually in hundreds of thousands at most), so we can safely make this "richness" assumption. In mathematical terms, it means that two different degrees of belief are represented by different computer terms, i.e., that if $w_1 \neq w_2$, then $\|w_1\| \neq \|w_2\|$.

The problem with this first-order representation is that the relation between words $w \in W$ and computer representation $s \in S$ is, in reality, also imprecise. Typically, when we have a word $w \in W$, we cannot pick a single corresponding representative $s \in S$; instead, we may have *several* possible representatives, with different degrees of adequacy. In other words, instead of a *single* value $s = \|w\|$ assigned to a word w, we have *several* values $s \in S$, each with its own degree of adequacy; this degree of adequacy can also be described by an expert, who uses an appropriate word $w \in W$ from the natural language. In other words, for every word $w \in W$ and for ever representation $s \in S$, we have a degree $w' \in W$ describing to what extent s is adequate in representing w. Let us represent this degree of adequacy by $a(w,s)$; the symbol a represents a function $a : W \times S \to W$, i.e., a function that maps every pair (w,s) into a new word $a(w,s)$.

So, the meaning of a word $w \in W$ is represented by a *function* a which assigns, to every element $s \in S$, a degree of adequacy $a(w,s) \in W$. We want to represent this degree of adequacy in a computer; therefore, instead of using the word $a(w,s)$ itself, we will use the computer representation $\|a(w,s)\|$ of this word. Hence, we get a *second-order* representation, in which a degree of certainty corresponding to a word $w \in W$ is represented not by a *single* element $\|w\| \in S$, but by a *function* $\mu_w : S \to S$, a function which is defined as $\mu_w(s) = \|a(w,s)\|$.

This second-order representation is also not absolutely adequate, because, to represent the degree $a(w,s)$, we used a single number $\|a(w,s)\|$. To get a more adequate representation, instead of this single value, we can use, for each element $s' \in S$, a degree of adequacy with which the element s' represents the word $a(w,s)$. This degree of adequacy is also a word $a(a(w,s),s')$, so we can represent it by an appropriate element $\|a(a(w,s),s')\|$. Thus, we get a *third-order* representation, in which to every element s, we assign a second-order fuzzy set. To get an even more adequate representation, we can use fourth- and higher order fuzzy set. Let us express this scheme formally.

Definition 5.

- *Let W be a finite set; element of this set will be called words.*
- *Let U be set called a universe of discourse. By a fuzzy property P, we mean a mapping which maps each element $x \in U$ into a word $P(x) \in W$; we say that this word described the degree of certainty that x satisfies the property P.*
- *By a first-approximation uncertainty representation, we mean a pair $\langle S, \|.\|\rangle$, where:*
  - *S is a set; elements of this set will be called computer representations; and*
  - *$\|.\|$ is a function from W to S; we say that an element $\|w\| \in S$ represents the word w.*
- *We say that an uncertainty representation is sufficiently rich if for every two words $w_1, w_2 \in W$, $w_1 \neq w_2$ implies $\|w_1\| \neq \|w_2\|$.*

Definition 6. *Let W be a set of words, and let S be a set of computer representations. By an adequacy function, we mean a function $a : W \times S \to W$; for each word $w \in W$, and for each representation $s \in S$, we say that $a(w, s)$ describes the degree to which the element s adequately describes the word w.*

Definition 7. *Let U be a universe of discourse, and let S be a set of computer representations. For each $n = 1, 2, \ldots$, we define the notions of n-th order degree of certainty and of a n-th order fuzzy set, by the following induction over n:*

- *By a first-order degree of certainty, we mean an element $s \in S$ (i.e., the set S_1 of all first-order degrees of certainty is exactly S).*
- *For every n, by a n-th order fuzzy set, we mean a function $\mu : U \to S_n$ from the universe of discourse U to the set S_n of all n-th order degrees of certainty.*
- *For every $n > 1$, by a n-th order degree of certainty, we mean a function s_n which maps every value $s \in S$ into an $(n-1)$-th order degree of certainty (i.e., a function $s_n : S \to S_{n-1}$).*

Definition 8. *Let W be a set of words, let $\langle S, \|.\|\rangle$ be an uncertainty representation, and let a be an adequacy function. For every $n > 1$, and for every word $w \in W$, we define the n-th order degree of uncertainty $\|w\|_{a,n} \in S_n$ corresponding to the word w as follows:*

- *As a first order degree of uncertainty $\|w\|_{a,1}$ corresponding to the word w, we simply take $\|w\|_{a,1} = \|w\|$.*
- *If we have already defined degrees of orders $1, \ldots, n-1$, then, as an n-th order degree of uncertainty $\|w\|_{a,n} \in S_n$ corresponding to the word w, we take a function s_n which maps every value $s \in S$ into a $(n-1)$-th order degree $\|a(w, s)\|_{a,n-1}$.*

Definition 9. *Let W be a set of words, let $\langle S, \|.\|\rangle$ be an uncertainty representation, let a be an adequacy function, and let P be a fuzzy property on a universe of discourse P. Then, by a n-th order fuzzy set (or a n-th order membership function) $\mu^{(n)}_{P,a}(x)$ corresponding to P, we mean a function which maps every value $x \in U$ into an n-th order degree of certainty $\|P(x)\|_{a,n}$ which corresponds to the word $P(x) \in W$.*

We will prove that for fuzzy properties which are *non-degenerate* in some reasonable sense, it is sufficient to know the *first* and *second* order membership functions, and then the others can be uniquely reconstructed. Moreover, if we know the membership functions of first two orders for a non-degenerate *class* of fuzzy properties, then we will be able to reconstruct the higher order membership functions for *all* fuzzy properties from this class.

Definition 10.

- *We say that a fuzzy property P on a universe of discourse U is non-degenerate if for every $w \in W$, there exists an element $x \in U$ for which $P(x) = w$.*
- *We say that a class $\mathcal{P}$ of fuzzy properties P on a universe of discourse U is non-degenerate if for every $w \in W$, there exists a property $P \in \mathcal{P}$ and an element $x \in U$ for which $P(x) = w$.*

Comment. For example, if $W \neq \{0, 1\}$, then every crisp property, i.e., every property for which $P(x) \in \{0, 1\}$ for all x, is *not* non-degenerate (i.e., degenerate).

Proposition 3. *Let W be a set of words, let $\langle S, \|.\|\rangle$ be a sufficiently rich uncertainty representation, let U be a universe of discourse. Let P and P' be fuzzy properties, so that P is non-degenerate, and let a and a' be adequacy functions. Then, from $\mu^{(1)}_{P,a} = \mu^{(1)}_{P',a'}$ and $\mu^{(2)}_{P,a} = \mu^{(2)}_{P',a'}$, we can conclude that $\mu^{(n)}_{P,a} = \mu^{(n)}_{P',a'}$ for all n.*

Comments.

- In other words, under reasonable assumptions, for each property, the information contained in the first and second order fuzzy sets is sufficient to reconstruct all higher order fuzzy sets as well; therefore, in a computer representation, it is sufficient to keep only first and second order fuzzy sets.
- This result is somewhat similar to the well-known result that a Gaussian distribution can be uniquely determined by its moments of first and second orders, and all higher order moments can be uniquely reconstructed from the moments of the first two orders.
- It is possible to show that the non-degeneracy condition is needed, because if a property P is

not non-degenerate, then there exist adequacy functions $a \neq a'$ for which $\mu^{(1)}_{P,a} = \mu^{(1)}_{P,a'}$ and $\mu^{(2)}_{P,a} = \mu^{(2)}_{P,a'}$, but $\mu^{(3)}_{P,a} \neq \mu^{(3)}_{P,a'}$ already for $n = 3$.

Proposition 4. *Let W be a set of words, let $\langle S, \|.\|\rangle$ be a sufficiently rich uncertainty representation, let U be a universe of discourse. Let $\mathcal{P}$ and $\mathcal{P}'$ be classes of fuzzy properties, so that the class $\mathcal{P}$ is non-degenerate, and let $\varphi : \mathcal{P} \to \mathcal{P}'$ be a 1-1-transformation, and let a and a' be adequacy functions. Then, if for every $P \in \mathcal{P}$, we have $\mu^{(1)}_{P,a} = \mu^{(1)}_{\varphi(P),a'}$ and $\mu^{(2)}_{P,a} = \mu^{(2)}_{\varphi(P),a'}$, we can conclude that $\mu^{(n)}_{P,a} = \mu^{(n)}_{\varphi(P),a'}$ for all n.*

Comment. So, even if we do not know the adequacy function (and we do not know the corresponding fuzzy properties $P \in \mathcal{P}$), we can still uniquely reconstruct fuzzy sets of all orders which correspond to all fuzzy properties P.

3.3 Alternative approach: L-fuzzy logics

We started this section by saying that the original $[0,1]$-based fuzzy logic is sometimes inadequate in describing the expert's uncertainty. One possibility to make it more adequate is to consider the $[0,1]$-based logic as a first approximation to true logic, and to use the same idea which underlies the fuzzy logic itself, to come up with a *second* approximation. This idea has led to *second order* fuzzy logic.

A natural alternative is not to *patch up* the $[0,1]$-based fuzzy logic, but to *abandon* the set $[0,1]$ altogether, and to consider more general algebraic structures instead. The most widely used specific case of this idea is the so-called L-fuzzy logic (introduced in [17, 18]), in which instead of the interval $[0,1]$, we consider an arbitrary *lattice* of degrees of certainty, i.e., an ordered set with operations "and" ($\wedge$) and "or" ($\vee$) which satisfy certain reasonable conditions.

3.4 Second order fuzzy logic and L-fuzzy logics: the main problem

From the practical viewpoint, the main problem with the second order fuzzy logic and of the L-fuzzy logic is that these notions are too general to be practically efficient; namely:

- in the *traditional* $[0,1]$-based fuzzy logic, to represent the degree of belief in a single statement, we must use a *single real number* from the interval $[0,1]$; a real number does not take much computer space, and it is relatively easy to process (it is worth mentioning, however, that some problems are very computationally complicated even in the $[0,1]$-based fuzzy methodology [40]);
- in the *second order* fuzzy logic, to represent the degree of certainty in a single statement, we must use a membership function, i.e., a *function* from $[0,1]$ to $[0,1]$.

From the purely mathematical viewpoint, to uniquely determine a function, we must select the values of *infinitely many* parameters. In real computations, of course, due to limited accuracy of all computer calculations, a *finite* number of parameters is sufficient to describe all possible computer-representable functions; however, the number of parameters is so huge that storing and processing such functions is practically impossible.

With L-fuzzy logics, the general case is even worse: for second-order fuzzy logics, we at least know that we must consider *functions*, but general L-fuzzy logics can contain objects of much *higher order*, which require even more space to store and even more time to process.

We therefore need to somehow limit our description; in other words, instead of considering the most *general* notions of second order logics and L-fuzzy logics, we must select *computational efficient* subclasses of these general notions. In this paper, we will describe several mathematical results oriented towards such a selection.

4 Application-oriented multi-D generalizations of fuzzy logic

4.1 Two main approaches to eliciting degrees of certainty

In order to describe possible practically useful multi-D generalizations of fuzzy logic, let us recall how in $[0,1]$-based fuzzy logic, we can elicit the degrees of certainty. There are two basic classes of methods for eliciting these degrees:

- First, we can ask an expert to estimate his degree of certainty by marking a point on a *scale*. Then, if, e.g., an experts marks his degree of confidence as 7 on a scale from 0 to 10, it is natural to take $7/10$ as the desired degree of certainty.
- Alternatively, we can use elicitation methods which are closer to *statistics*. Namely:
  - if possible, we can collect records about the occurrence of an event A, and take the frequency of A in these records as our degree of certainty in A;
  - alternatively, we can collect records about things about which the experts said, say, that they were "very possible", and take the frequency of those of these events which really happened later as a degree of certainty describing the corresponding word (such as "very possible");
  - we can also directly ask several (N) experts about the statement A, and take, as the degree of certainty in A, the ratio $Y(N,A)/N$, where $Y(N,A)$ denotes the number of experts who answered "yes".

These two approaches lead to two directions in generalizing fuzzy logic.

4.2 Scaling approach leads to interval-valued fuzzy logic

In scaling elicitation, each scale has finitely many points which an expert can mark. Therefore, each scale leads to only finitely many possible values of degree of certainty: if we use a scale from 0 to 2, we get only three values 0, 1/2, and 1; if we use a scale from 0 to 10, we get only 11 values $0, 0.1, 0.2, \ldots, 1$. The more elements a scale contains, the more detailed information about the degree of certainty we can provide. Therefore, it is natural to consider values obtained from a scale as *approximations* to the true degree of certainty: the more elements on a scale, the better this approximation. When an expert picks 7 on a scale from 0 to 10, it does not necessarily mean that his degree of certainty is exactly 7/10=0.7; it rather means that the actual expert's degree of certainty is closer to 7/10=0.7 than to 6/10=0.6 or to 8/10=0.8, i.e., that this degree of certainty belongs to the interval $[0.65, 0.75]$.

Theoretically, we can design a scale with 100 points, which corresponds to the accuracy of 0.01, a scale with 1000 values which corresponds to the accuracy of 0.001, etc. The more elements on a scale, the narrower the corresponding interval. If all these scales were feasible, then we would be able to determine the actual degree of certainty as a unique real number which belongs to the intersection of this sequence of narrower and narrower intervals.

In reality, however, as we have already mentioned, if a scale is too large, people cannot easily describe their preferences by marking the points on this scale. Therefore, instead of a sequence of narrowing intervals which tends to a single point, we get a finite sequence of narrowing intervals which stops after a while, ending in an *interval* $[d^-(A), d^+(A)]$ of possible values of degrees of certainty.

The resulting *interval-valued* fuzzy logic, in which degrees of certainty are *intervals* (namely, subintervals of the interval $[0,1]$) is the most widely used particular case of second order fuzzy logic. It was originally proposed by J. A. Goguen and actively developed by I. B. Türkşen and L. Kohout. In interval-valued fuzzy logic, truth values are This logic has indeed led to many useful practical applications; see, e.g., interval sections of NAFIPS 1994 [38], or surveys [41, 54].

4.3 Multi-D generalizations of fuzzy logic which stem from frequency-based elicitation methods: Fuzzy/Probability $\sim$ Fractal/Smooth

Multi-D: why. One of the most natural ways to describe a degree of belief $d(A)$ in a statement A is by asking N experts and by taking the ratio $d_N(A) = Y(N,A)/N$ of those who believe in A as the desired degree $d(A)$ [4, 3, 10, 20]. Ideally, the more experts we ask (i.e., the larger N), the better estimates we get; in mathematical terms, as N increases, the estimates $d_N(A)$ converge to the *actual* (*limit*) value $d(A)$. However, in real life, there are problems with this definition of degree of confidence.

The *first problem* is that in some situations, this definition assigns *the same* degree of confidence $d(A) = d(A')$ to two different statements while it is intuitively clear that our confidence in the first statement is much larger than our confidence in the second one. Let us give an example. As A, we can take a statement which is clearly false; then, $d(A) = 0$. As A', we take a statement on the "cutting edge" of science, a statement the truth of which has just been discovered, and which is still only known to the top experts in the field. For this statement:

- while N is smaller than the number of these top researchers, the value $Y(N, A')$ grows, but
- as soon as N exceeds the number of these top researchers, the value $Y(N, A')$ stays constant, does not increase with N and, therefore, the ratio $d_N(A') = Y(N, A')/N$ tends to 0 as $N \to \infty$.

If we simply use the limit as $d(A')$, then we would have $d(A') = 0 = d(A)$, while intuitively, our degree of confidence in $d(A')$ is much higher than in A.

The *second problem* is that in some real-life situations, the sequence $d_N(A)$ does not seem to tend to any limit at all. For example, we may have a statement A which:

- seems intuitively true (e.g., that "optimism helps to fight a disease"),
- contradicts to the accepted science, but
- has been recently re-discovered and experimentally confirmed (so recently that it is not yet common knowledge among experts).

Then, as we increase the number N of experts, the ratio $d_N(A)$ exhibits the following oscillating behavior:

- at first, when N is reasonably small, so that we only consider top experts in the field, we have $Y(N, A) \approx N$, and $d_N(A) \approx 1$;
- then, as we start including more and more experts who are not yet at the top research level, the number $Y(N, A)$ stays fixed, while N increases, so we get $d_N(A) \approx 0$;
- finally, when we increase N to such an extent that our list of experts starts including people with commonsense reasoning, the value $Y(N, A)$ again starts increasing as $Y(N, A) \approx N$, and the ratio $d_N(A)$ shoots back to 1.

In more sophisticated examples, we may have even more oscillations. For example, L. Zadeh gives an example of such oscillating behavior in estimating the probability that he (or any other person) will have a tax audit:

- first, we can consider all people in the US, and get a certain probability;
- as we go from the US as a whole to California, the probability of an audit increases;
- as we further narrow down the list to only people from Berkeley (thus, excluding Silicon Valley and Los Angeles), the probability goes down again;
- if we only consider middle-class people from Berkeley, the probability goes up again;
- as we further restrict ourselves to university professors, this probability goes down, etc.

The point that Zadeh makes is that it is very difficult to describe such an oscillating process by using methods from probability theory, which presumes a convergence. A natural next question is: how can we describe this oscillating behavior? Since we cannot describe it by using a *single* parameter (such as probability), we need to use a *multi-D* formalism.

Multi-D: how. We want to describe possible asymptotics of $Y(N,A)$ (and of the ration $d_N(A)$) as N increases. In the traditional probability approach, we have a one-parametric family of asymptotics $Y(N,A) \sim p \cdot N$, with a parameter p (which leads to $d_N(A) \sim p$). In a more general multi-D case, it is natural to consider families with *several* parameters, i.e., families of the type $\{C_1 \cdot f_1(N) + \ldots + C_n \cdot f_n(N)\}$, where $f_1(N), \ldots, f_n(N)$ are given functions, and C_i are arbitrary constants. We would like to describe the families which are *the best* in describing expert estimates. Since we do not have a precise formalization of what "the best" means, the problem of choosing the best family is the problem of optimization under an uncertain criterion. In [42], we have described a general formalism for solving such problems, and we have shown that this formalism is in good accordance with the empirical optimality of different fuzzy, neural, genetic, and other techniques. So, we will use this approach to describe the best families.

Optimal in what sense? The main idea. We are looking for the *best* (*optimal*) choice of a potential function.

Normally, the word "best" is understood in the sense of some *numerical* optimality criterion. However, in our case of *fuzzy* choice, it is often difficult to formulate the exact *numerical* criterion. Instead, we assume that there is an *ordinal* criterion, i.e., that we can compare arbitrary two choices, but that we cannot assign numerical values to these choices.

It turns out that in many cases, there are reasonable *symmetries*, and it is natural to assume that the (ordinal) optimality criterion is invariant with respect to these symmetries. Then, we are able to describe all choices that are optimal with respect to some invariant ordinal optimality criteria.

This general approach was described and used in [5, 32, 33, 42, 59, 61], in particular, for fuzzy control. In this section, we will show that this approach is applicable to fuzzy elicitation as well.

Let us borrow from the experience of modern physics and use symmetries. In modern physics, symmetry groups are a tool that enables to compress complicated differential equations into compact form (see, e.g., [65]). Moreover, the very differential equations themselves can be uniquely deduced from the corresponding symmetry requirements (see, e.g., [11, 12]).

It is possible to use symmetry. As we have mentioned, in our previous papers, we have shown that the symmetry group approach can be used to find optimal membership functions, optimal t-norms and t-conorms, and optimal defuzzification procedures.

It is therefore reasonable to expect that the same approach can also be used to choose the best potential function for fuzzy elicitation.

What is a criterion for choosing a family of functions? What does it mean to choose a *best* family of functions? It means that we have some *criterion* that enables us to choose between the two families.

Traditionally, optimality criteria are *numerical*, i.e., to every family F, we assign some value $J(F)$ expressing its quality, and choose a family for which this value is maximal (i.e., when $J(F) \geq J(G)$ for every other alternative G). However, it is not necessary to restrict ourselves to such numeric criteria only.

For example, if we have several different families F that have the same adequacy $P(F)$, we can choose between them the one that has the minimal computational complexity $C(F)$. In this case, the actual criterion that we use to compare two families is not numeric, but more complicated:

> *A family F_1 is better than the family F_2 if and only if*
>
> – *either $P(F_1) > P(F_2)$,*
>
> – *or $P(F_1) = P(F_2)$ and $C(F_1) < C(F_2)$.*

A criterion can be even more complicated.

The only thing that a criterion *must* do is to allow us, for every pair of families (F_1, F_2), to make one of the following conclusions:

- the first family is better with respect to this criterion (we'll denote it by $F_1 \succ F_2$, or $F_2 \prec F_1$);
- with respect to the given criterion, the second

family is better ($F_2 \succ F_1$);

- with respect to this criterion, the two families have the same quality (we'll denote it by $F_1 \sim F_2$);
- this criterion does not allow us to compare the two families.

Of course, it is necessary to demand that these choices be consistent.

> For example, if $F_1 \succ F_2$ and $F_2 \succ F_3$ then $F_1 \succ F_3$.

The criterion must be final, i.e., it must pick the unique family as the best one. A natural demand is that this criterion must choose a *unique* optimal family (i.e., a family that is better with respect to this criterion than any other family).

The reason for this demand is very simple: If a criterion *does not choose* any family at all, then it is of no use. If *several* different families are the best according to this criterion, then we still have the problem of choosing the best among them. Therefore we need some additional criterion for that choice, like in the above example:

> If several families $F_1, F_2, \ldots$ turn out to have the same adequacy ($P(F_1) = P(F_2) = \ldots$), we can choose among them a family with minimal computational complexity ($C(F_i) \rightarrow \min$).

So what we actually do in this case is abandon that criterion for which there were several "best" families, and consider a new "composite" criterion instead: F_1 is better than F_2 according to this new criterion if either it was better according to the old criterion, or they had the same quality according to the old criterion and F_1 is better than F_2 according to the additional criterion.

In other words, if a criterion does not allow us to choose a unique best family, it means that this criterion is not final, we'll have to modify it until we come to a final criterion that will have that property.

The criterion must not change whether we count experts or schools of experts. When we talk about counting experts, we can literally count them. However, this may not always be the best approach, because the whole idea of increasing N is to increase the *diversity* of the experts, and so, if, e.g., two experts belong to a single school of researchers (and therefore, have similar views), it may not make big sense to interview both. Instead, we should interview people from different schools, and count these *schools*, not individual researchers.

When we count *researchers*, the value $Y(N,A)$ is always an integer. When we count *schools*, and the school is divided on this particular issue (the larger schools we take, the more frequent such a situation will be), then it is natural to add $1/2$ (or whatever ratio is in this school) to the total number of schools that believe in A. In this case, the value of $Y(N,A)$ is not necessarily an integer: it can be an arbitrary *rational* number. In this case, it is natural to assume that the function $f(N)$ that approximates $Y(N,A)$ can take arbitrary real values.

The notion of the "school of researchers" may mean different things; for example:

- we can divide the researchers into large groups whose views are more or less similar, but can be different in details;
- we can also divide the researchers into very small groups with practically identical views.

The exact mathematical form of an approximating function $f(N)$ depends on how we count these "schools of scientists". If we re-define a school, and in our new definition, a school is λ times smaller, then N old schools correspond to $N' = \lambda \cdot N$ new schools, and, correspondingly, the original number $b = f(N) = Y(N,A)$ of schools who believed in A changes to $b' = \lambda \cdot Y(N,A)$. In terms of the new number of school $N' = \lambda \cdot N$, this new dependence takes the form $b' = f_\lambda(N')$, where $f_\lambda(N) = \lambda \cdot f(N/\lambda)$.

It is reasonable to assume that the relative quality of different families should not change if we simply change the units, i.e., if the family F is better than a family G, then the transformed family F_λ should also be better than the family G_λ. Now, we are ready for the formal definition.

Definition 11.

- *By a family F, we mean a finite set of differentiable functions $f_1(N), \ldots, f_n(N)$ from R to R. This family will also be denoted as*
$$\{C_1 \cdot f_1(N) + \ldots + C_n \cdot f_n(N)\}.$$
- *We say that a function $f(N)$ belongs to the family if $f(N) = C_1 \cdot f_1(N) + \ldots + C_n \cdot f_n(N)$ for some real numbers C_i.*
- *Two families F and G are considered equal if they contain the same functions.*

Definition 12. *A pair of relations $(\prec, \sim)$ on a set Φ is called consistent if it satisfies the following conditions, for every $F, G, H \in \Phi$:*

(1) *if $F \prec G$ and $G \prec H$ then $F \prec H$;*
(2) *$F \sim F$;*
(3) *if $F \sim G$ then $G \sim F$;*
(4) *if $F \sim G$ and $G \sim H$ then $F \sim H$;*
(5) *if $F \prec G$ and $G \sim H$ then $F \prec H$;*
(6) *if $F \sim G$ and $G \prec H$ then $F \prec H$;*
(7) *if $F \prec G$ then it is not true that $G \prec F$, and it is not true that $F \sim G$.*

Definition 13. *Assume a set Φ is given. Its elements will be called alternatives.*

- *By an optimality criterion, we mean a consistent pair $(\prec, \sim)$ of relations on the set Φ of all alternatives.*
  - *If $F \succ G$ we say that F is better than G;*
  - *if $F \sim G$ we say that the alternatives F and G are equivalent with respect to this criterion.*
- *We say that an alternative F is optimal (or best) with respect to a criterion $(\prec, \sim)$ if for every other alternative G either $F \succ G$ or $F \sim G$.*
- *We say that a criterion is final if there exists an optimal alternative, and this optimal alternative is unique.*

Comment. In this paper, we will consider optimality criteria on the set Φ of all families.

Definition 14. *Let $\lambda > 0$ be a positive real number.*

- *By a λ-rescaling of a function $f(N)$ we mean a function $f_\lambda(N) = \lambda \cdot f(N/\lambda)$.*
- *By a λ-rescaling F_λ of a family of functions F we mean the family consisting of λ-rescalings of all functions from F.*

Definition 15. *We say that an optimality criterion on Φ is unit-invariant if for every two families F and G and for every number $\lambda > 0$, the following two conditions are true:*

i) if F is better than G in the sense of this criterion (i.e., $F \succ G$), then $F_\lambda \succ G_\lambda$;

ii) if F is equivalent to G in the sense of this criterion (i.e., $F \sim G$), then $F_\lambda \sim G_\lambda$.

Comment. As we have already remarked, the demands that the optimality criterion is final, unit-invariant, and rotation invariant are quite reasonable. At first glance they may seem rather trivial and therefore weak, because these demands do not specify the exact optimality criterion. However, these demands are strong enough, as the following theorem shows:

Proposition 5. [42] *If a family F is optimal in the sense of some optimality criterion that is final and unit-invariant, then every function $f(N)$ from this family F is equal to a linear combination of the functions of the type $f(N) = N^\alpha \cdot (\ln(N))^p \cdot \sin(\beta \cdot \ln(N) + \varphi)$, where p is a non-negative integer, and α, β and φ are real numbers.*

Comments.

- This result (to be more precise, the appearance of sin) explains the above-described "oscillating" behavior of $Y(N, A)$.
- As shown in [42], for $n = 1$, the only possible families are $\{C_1 \cdot N^\alpha\}$ considered in [34, 41]. For $\alpha = 1$, we get functions corresponding to a probability approach; in [34, 41], we also give an interpretation for families corresponding to $\alpha \neq 1$.
- For $n = 2$, we already have a possibility of an oscillating function

  $$f(N) = C_1 \cdot N + C_2 \cdot N \cdot \sin(\beta \cdot \ln(N)).$$

  For this function, as sin oscillates between -1 and 1, the ratio $d_N(A) = f(N)/N$ oscillates between $C_1 - C_2$ and $C_1 + C_2$. Thus, it is natural to say that the corresponding degree of certainty is an *interval* $[C_1 - C_2, C_1 + C_2]$. However, the exact form gives us more information than the interval: namely, it also describes the "oscillation rate" β.

Analogy between fuzzy and fractal. The difference between the situations which are easily describable by methods of *probability* theory, and more complicated situations which require *non-probabilistic* (*fuzzy*) description can be described in terms of the asymptotics of $Y(N, A)$:

- in the probabilistic case, $Y(N, A) \sim p \cdot N$, while
- in the fuzzy case, we have a more general asymptotics, e.g., $Y(N, A) \sim p \cdot N^\alpha$.

This difference is similar to the difference between *smooth* curves (or surfaces) and *fractal* curves (surfaces) [35]. Indeed, according to the definition of a fractal, a *fractal* is a set of a fractal (non-integer) dimension, and dimension of a set is defined in terms of finite approximations. follows. For a given real number ε, we say that a finite set $S \subseteq A$ is an *ε-approximation* to a set A if every point $a \in A$ is ε-close to one of the points from S. The smaller ε, the more points we need to approximate a given set A. Thus, as a natural measure of complexity of a set A, we can take, for every ε, the smallest number $N_\varepsilon(A)$ of points which are necessary to approximate the set A with an accuracy ε.

- In the simplest case when the set A is the interval $[0, 1]$, we can explicitly compute $N_\varepsilon(A)$.
  - Indeed, in this set S, we need a point which is ε-close to 0, i.e., which is $\leq \varepsilon$. This point s_1 covers everything from 0 from $s_1 + \varepsilon$. The further this point from 0, the less area is left for other points to cover, so the smallest possible number of points occurs when s_1 is at the farthest possible location, i.e., when $s_1 = \varepsilon$.
  - The above-chosen point s_1 covers all points from 0 to $s_1 + \varepsilon = 2\varepsilon$, so to cover points right after 2ε, we need a next point at a location $s_2 \leq 2\varepsilon + \varepsilon = 3\varepsilon$; a similar argument shows that the smallest number of points is when we take $s_2 = 3\varepsilon$.
  - Similarly, $s_k = (2k - 1) \cdot \varepsilon$.

So, we need $N_\varepsilon([0,1]) \sim 1/2\varepsilon$ points to cover the entire interval $[0,1]$. Similarly, for any other smooth curve γ, $N_\varepsilon(\gamma) \sim \text{const}/\varepsilon$.

- Similarly, for a 2-D smooth *surface* A, we have $N_\varepsilon(A) \sim \text{const}/\varepsilon^2$.
- For a 3-D domain with a smooth boundary, we have $N_\varepsilon(A) \sim \text{const}/\varepsilon^3$.

In all these smooth cases, $N_\varepsilon(A) \sim \text{const}/\varepsilon^D$, where D is the dimension of the set A. In view of this fact, we can define *dimension* for *non-smooth* sets A as a parameter α for which $N_\varepsilon(A) \sim \text{const}/\varepsilon^\alpha$. Then, a set is a *fractal* if either α is not an integer, or if the asymptotics of $N_\varepsilon(A)$ is more complicated than $\text{const}/\varepsilon^\alpha$.

Similarly, in our case:

- when $Y(N,A) \sim p \cdot N$, we get a *probabilistic* situation, and
- when $Y(N,A) \sim \text{const} \cdot N^\alpha$ for some $\alpha \neq 1$, or if the asymptotics of $Y(N,A)$ is more complicated than N^α, we have a non-probabilistic (*fuzzy*) situations.

Thus, we can say that fuzzy is a generalization of probability in the same sense in which fractals are a generalization of smooth surfaces:

$$\frac{\text{Fuzzy}}{\text{Probability}} = \frac{\text{Fractal}}{\text{Smooth}}.$$

4.4 General description of multi-D fuzzy logics: case of crisp preferences

In the above text, we have described two possible approaches to multi-D fuzzy logics which correspond to two major elicitation techniques. Since there exist several other elicitation techniques (see, e.g., [8]), it is conceivable that other approaches to multi-D fuzzy logic may emerge. With this possibility in mind, it is desirable to find a *general* description of multi-D fuzzy logics.

Under the (simplifying) assumption that we have a crisp preference relation between different degrees of certainty, such a general description is obtained and analyzed in [25].

4.5 General description of multi-D fuzzy logics: case of fuzzy preferences

Our main goal is to describe the preferences of an expert whose knowledge is imprecise. Since the knowledge itself is imprecise (fuzzy), it is natural to assume that the preference relation between the expert's degrees of certainty is also fuzzy: although an expert can sometimes order words describing his degrees of certainty, in some other cases, an expert can only say that, e.g., "very probable" is *most probably* more certain than "quite possible". In other words, it is quite possible that the order on the expert's degrees of certainty is actually not a crisp order, but rather a *fuzzy* order. So, to describe this more general situation, we must analyze fuzzy orders.

In this section, we will show that under some reasonable assumptions, a fuzzy ordering can be described in terms of nested sequence of crisp orders. Indeed, in classical mathematics, an *ordering* is defined as a relation $>$ for which the following conditions are true:

- for all a and b, if $a > b$, then $b \not> a$;
- for all a, b, and c, if $a > b$ and $b > c$, then $a > c$.

The second condition (transitivity) can be also expressed in the following two equivalent forms:

- for all a, b, and c, if $a \not> c$ and $a > b$, then $b \not> c$;
- for all a, b, and c, if $a \not> c$ and $b > c$, then $a \not> c$.

We want to describe the fact that the decision maker is not sure whether a is preferable to b. To describe this fact, the fuzzy relation is usually defined as a mapping from the set of all pairs $X \times X$ to the interval [0,1] of possible degrees of belief. Let's denote the degree of belief that $a > b$ by $d(a,b)$.

To generalize the above definition to the fuzzy case, we need to be able to generalize negation and &. As an analogue of negation, it is natural to take $x \to 1-x$. As an analogue of the classical &, we will take the original Zadeh's $\min(x,y)$ [67]. The fact that A implies B can be expressed as meaning that the degree of belief $d(B)$ in B is either greater, or the same, as the degree of belief in A, i.e., that $d(B) \geq d(A)$. As a result, we arrive at the following definition:

Definition 16. *We say that a fuzzy relation $d : X \times X \to [0,1]$ is a fuzzy ordering if the following four conditions are true:*

- *for all a and b, $1 - d(b,a) \geq d(a,b)$;*
- *for all a, b, and c,*
$$d(a,c) \geq \min(d(a,b), d(b,c));$$
- *for all a, b, and c,*
$$1 - d(b,c) \geq \min(1 - d(a,c), d(a,b));$$
- *for all a, b, and c,*
$$1 - d(a,b) \geq \min(1 - d(a,c), d(b,c)).$$

Comment. In the crisp case, the last two conditions can be deduced from the first two. However, in the fuzzy case, we need to state them separately.

Suppose that a fuzzy ordering is given. If we need to make a choice between a and b, then a natural idea is to choose a if our belief that $a > b$ exceeds our belief that $a \not> b$, i.e., if $d(a,b) > 1 - d(a,b)$. This inequality is equivalent to $d(a,b) > 1/2$. If $d(a,b)$ is close to

$1 - d(a,b)$, then we are not certain whether this choice is reasonable. So, we may want to be more cautious, and choose a only if our degree of belief that $a > b$ exceeds our degree of belief that $a \not> b$ by a certain amount $\varepsilon > 0$, i.e., if $d(a,b) > 1 - d(a,b) + \varepsilon$. This inequality, in its turn, is equivalent to $d(a,b) > \alpha$ for $\alpha = (1/2)(1+\varepsilon)$. In other words, the actual preference relation that we will use for decision making is an $\alpha-$cut of the fuzzy relation d. In view of this remark, it is interesting to study these $\alpha-$cuts; for example, we would like to know whether these $\alpha-$cuts are real orderings (if not, then this seemingly reasonable strategy will enable us to prefer a to b, b to c, and c to a). First, we prove that $\alpha-$cuts do define crisp ordering:

Proposition 6. *If $\alpha \geq 1/2$, then the relation $a > b$ defined as $d(a,b) > \alpha$ is a crisp ordering.*

Comment. A similar result was proven in [15] (see also references therein).

The following result shows the relationship between crisp orderings that correspond to different α:

Definition 17. *Let $1/2 \leq \alpha < \beta < 1$.*

- *We say that a is preferable to b, and denote it by $a > b$, if $d(a,b) > \alpha$.*
- *We say that a is strongly preferable to b, and denote it by $a \gg b$, if $d(a,b) > \beta$.*

Proposition 7.

- *If $a \gg b$, then $a > b$;*
- *if $a \gg b$, and $b > c$, then $a \gg c$;*
- *if $a > b$, and $b \gg c$, then $a \gg c$.*

4.6 Towards general description of multi-D fuzzy logics: complex-valued fuzzy logics and beyond

The above-described methods were based on the assumptions that experts are *rational* agents, whose beliefs and arguments are flawlessly *consistent*. It is well known that real-life humans (and experts are no exception) are not flawless; people are prone to inconsistent behavior and inconsistent preferences (when, e.g., they prefer A to B, B to C, but C to A). This paradoxical behavior is an important part of our behavior and of our decision making (see, e.g., [13, 23]).

If we take these inconsistencies into consideration when describing degrees of certainty, then we do not have a consistent *ordering* on the set of these degrees. What we still have is *operations* which correspond to "and" and "or". These operations must satisfy certain algebraic properties: e.g., it is reasonable to assume that our degree of certainty in a statement $A \,\&\, B \,\&\, C$ should not depend on the order in which we apply the "and" operation, i.e., whether we compute the degree of belief by representing the statement as $(A \,\&\, B) \,\&\, C$ or as $A \,\&\, (B \,\&\, C)$. So, instead of considering only *ordered* algebraic structures (as in the previous sections), we must consider *general* (not necessarily ordered) algebraic structures.

The simplest multi-D algebraic structures are 2-D ones, and the most well-known (and well-analyzed) non-ordered 2-D algebraic structure is the set of all *complex numbers*. In [51], we show that complex-valued fuzzy sets can indeed capture some important inconsistencies in human decision making. This success makes us believe that further generalizations can indeed lead to a general description of multi-D fuzzy logics.

5 Discrete modifications of fuzzy logic: towards computing with words

5.1 Computing with words. Granularity

Experts describe their degrees of certainty by *words*. In the traditional $[0,1]$-based fuzzy logic, these words are translated into real numbers from the interval $[0,1]$. We have already mentioned that this translation is *approximate*; in some cases, this approximate character of the representation is permissible, in some other cases, it is not. For these cases, L. Zadeh proposed to use the original *words* themselves instead of the real numbers. The main problem here is then developing ways of *computing with words*.

There is an additional *computational* advantage of using words instead of real numbers: there are much fewer words than real numbers, so if we use words, we need fewer bits to store the information about uncertainty, and hopefully, smaller computation time to process it. The reason why there are fewer words than real numbers is that a description in terms of words is *granular* – granules correspond to different words – while the description in terms of real numbers is not granular. It can shown that granularity indeed helps in solving complex problems [14].

5.2 Towards a fuzzy logic appropriate for computing with words: continuous case

We must define & and $\vee$-operations. Representing the truth values (= degrees of certainty) inside a computer is not all: we must be able to process these values. For example, suppose that we know the truth values $d(A)$ and $d(B)$ of two statements A and B, and the user asks a query "$A \& B$?". Since we are not sure whether A and B are true, we are also not sure whether $A \& B$ is true or not. Therefore, the only possible answer that we can give to this query is to describe a (reasonable) degree of belief $d(A \& B)$ in $A \& B$. If the only information that we have about A and B consists of their truth values, then we must somehow produce this reasonable estimate $d(A \& B)$ based on the known values $d(A)$ and $d(B)$. In other words, we must have a function (moreover, an algorithm) that would transform $d(A)$ and $d(B)$ into $d(A\&B)$. If we denote this function by $f_\&$, then we can describe the resulting "reasonable" estimate for $d(A\&B)$ as $f_\&(d(A), d(B))$.

In case both $d(A)$ and $d(B)$ coincide with "true" or "false", this function must coincide with the usual $\&-$operation that is defined on a classical set of truth values $\{0, 1\}$. Therefore, this function $f_\&$ is called an *&-operation.*

Likewise, there must exist a function $f_\vee$ that corresponds to $\vee$ and is therefore called an $\vee$-operation, and a function $f_\neg$ (an $\neg$-operation) that generalizes "not" to the bigger set of truth values.

A set with logical operations on it ("and", "or", and "not") is usually called a *logic.* A logic that is a finite set is called a *finite logic.* Our finite set of truth values has all these operations, and is therefore a finite logic.

Therefore, an ideal representation of degrees of uncertainty must form a finite logic.

How to choose &- and $\vee$-operations for finite logics: empirical solution. Since our main objective is to represent experts' beliefs in the most adequate manner, it is reasonable to choose &- and $\vee$-operations so as to provide the best description of the human reasoning with uncertainty. To do this, we must first ask the experts to estimate their degrees of belief in different statements and their logical combinations. Then, we choose a function $f_\&$ as follows: For every pair of degrees of belief a and b, we find all the statements in our record for which the degree of belief was a ($d(A) = a$), and all the statement B for which $d(B) = b$. For different A and B, we look for the truth values that the experts assigned to the statements $A\&B$. For different A and B, these truth values may be different; we find the "average" one (e.g., the one that is most frequent) and use it as $f_\&(a, b)$.

In a similar way, we can experimentally determine $f_\vee(a, b)$.

This is (in essence) the method that was originally used to choose &- and $\vee$-operations in one the first successful expert systems MYCIN (see, e.g., [7]). More recently, a similar method was efficiently used to produce &- and $\vee$-operations on finite logics in a MILORD system [1, 58].

If we can afford to perform the above-described procedure, fine, this procedure is the ideal solution to the choice problem. However, already the authors of MYCIN noticed that it is a very expensive and time-consuming procedure [7]. So, what to do if we cannot afford it, but still have to choose &- and $\vee$-operations? In this case, we need to develop theoretical methods to choose these operations. The authors of MILORD formulated reasonable conditions that &- and $\vee$-operations must satisfy [1, 58]. However, these are several different operations that satisfy all these conditions. Hence, the problem of choice remains.

At present, this choice problem is solved in the following manner. In the majority of actual expert systems the set of possible truth values is infinite (see, e.g., [7, 60]; MILORD is one of the few exceptions). Usually, the numbers from the interval [0.1] are used to represent degrees of belief. The reason for choosing this interval is very simple: inside the computer, "true" is usually represented as 1, and "false" as 0. So, it is reasonable to represent all intermediate degrees of belief by real numbers that are intermediate between 0 and 1.

If we assume that all numbers from [0,1] are possible, then we need to define &- and $\vee$-operations as functions from $[0, 1] \times [0, 1]$ to [0,1]. There exist several reasonable approaches that enable us to make a choice of such a function (see, e.g., [42]).

Formulation of a problem. These approaches provide us with reasonable &- and $\vee$-operations, but they essentially depend on the assumption that *all* numbers from the interval [0,1] can be truth values. Strictly speaking, this assumption is not true. Therefore, it is reasonable to formulate the following problem: if we are unable to elicit these operations from the experts, can we still choose them using only the actual truth values?

How we are going to solve this problem. In order to solve this problem, we will assume that both &- and $\vee$-operations $f_\&(a, b)$ and $f_\vee(a, b)$ are "continuous" in the following sense. If we gradually (= without skipping any intermediate values) increase our degrees of belief $a = d(A)$ and $b = d(B)$, then the resulting degrees of belief $d(A\&B) = f_\&(a, b)$ and $d(A \vee B) = f_\vee(a, b)$ must also change gradually.

It turns out that this reasonable demand is satisfied by only one pair of operations: min and max, that were originally proposed by L. Zadeh [67]. This result is in good accordance with the known experiments [19, 56, 68], according to which in many situations, min and max describe human reasoning better than other possible &- and $\vee$-operations.

Definition 18. *By a finite logic, we understand a (partially) ordered finite set L that contains two elements T and F such that $F \leq a \leq T$ for every $a \in L$. The elements of L will be called truth values, or degrees of belief.*

Motivation. We consider finitely many truth values, that represent different degrees of belief. Sometimes, we are certain that belief expressed by a degree a is stronger than the belief that is expressed by a degree b. For example, $a =$"for certain" is stronger than $b =$"maybe'. We will denote this by $a > b$. So, on our set of truth values, there is a ordering relation.

In particular, if we denote the degree of belief that expresses our absolute certainty in A, by T (T from "true"), and the degree of belief that expresses the absolute belief in $\neg A$ by F (from "false"), then $F \le a \le T$ for an arbitrary degree of belief a.

It is possible that for some words that describe uncertainty, there is no clear understanding which of them corresponds to greater belief (e.g., it is difficult to compare "probable" and "possible"). Therefore, we do not require that this ordering is a total (linear) order, it can be only partial.

Definition 19. *Let L be a finite logic. By an &-operation on L we mean a function $f_\&: L \times L \to L$ with the following properties:*

- $f_\&(a, b) \le a$;
- $f_\&(a, b) = f_\&(b, a)$;
- $f_\&(a, F) = F$;
- *if $a \le a'$ and $b \le b'$, then $f_\&(a, b) \le f_\&(a', b')$.*

Motivations.

- The first property is motivated by the following: if we believe in A *and* B, then we must believe in both statements A and B; therefore, our belief in $A\&B$ is either of the same strength or less strong than our belief in A.
- The second property is motivated by the fact that "$A\&B$" and "$B\&A$" are equivalent statements, so it is reasonable to demand that our estimated degree of belief in $A\&B$ ($= f_\&(d(A), d(B))$) is the same as the estimated degree of belief in $B\&A$ ($= f_\&(d(B), d(A))$).
- The third property expresses the following: if B is false, then "A and B" is false for all A.
- The fourth property means that if the degree of belief in A and B increases (i.e., if we found additional reasons to believe in A or B), then the resulting degree of belief in $A\&B$ must either increase, or stay the same.

Comment. This definition is similar to the usual definition of a t-norm (see, e.g., [20, 55]) and to the definition of an &-operation on a finite logic from [1, 58]. The reader may notice, however, that we do not require some additional properties that are usually required for a t-norm, like associativity ($f_\&(a, f_\&(b, c)) = f_\&(f_\&(a, b), c)$)). The reason is that in our case, as we will see later, associativity automatically follows from the other properties.

Definition 20. *Let L be a finite logic. By an $\vee$-operation on L we mean a function $f_\vee : L \times L \to L$ with the following properties:*

- $f_\vee(a, b) \ge a$;
- $f_\vee(a, b) = f_\vee(b, a)$;
- $f_\vee(a, T) = T$;
- *if $a \le a'$, and $b \le b'$, then $f_\vee(a, b) \le f_\vee(a', b')$.*

Motivations for these demands are similar to the ones given for an &-operation.

Definition 21. *We say that an element $a' \in L$ immediately follows a (and denote it by $a \ll b$, or $b \gg a$) if $a < a'$, and there exists no c such that $a < c < a'$. We say that a function $f : L \to L$ is discontinuous if there exist elements a, a', c such that $a \ll a'$, and either $f(a) < c < f(a')$, or $f(a') < c < f(a)$.*

Motivation. If such values a, a', c exist, this means that when we gradually increase our degree of belief from a to a' (gradually in the sense that we do not skip any intermediate values), then the resulting value of f "jumps" from $f(a)$ to $f(a')$, skipping an intermediate value c. So, in this sense, the function f is discontinuous.

We can use the same definition for a function of two variables.

Definition 22. *A function $f : L \times L \to L$ is called discontinuous is there exist the values a, a', b, b', c for which the following three conditions are true:*

- $a \ll a'$, $a' \ll a$, *or* $a = a'$;
- $b \ll b'$, $b' \ll b$, *or* $b = b'$;
- $f(a, b) < c < f(a', b')$, *or* $f(a', b') < c < f(a, b)$.

Comment. The first condition means that a gradually changes into a' (i.e., either a' immediately follows a, or a immediately follows a', or a' equals a). The second condition means that b gradually changes into b'. The third condition means that there is a "gap" between $f(a, b)$ and $f(a', b')$.

Definition 23. *A function is called continuous if it is not discontinuous.*

Comments.

- If a function f is continuous in the intuitive sense of this word, then it cannot have discontinuities in the sense of Definitions 21 and 22, and therefore it will be continuous in the sense of Definition 23. We do not claim, however, that an arbitrary function that satisfies Definition 23 is intuitively continuous, because there may be other types of discontinuity. We will prove that this weak continuity is sufficient to select &- and $\vee$-operations.
- It is worth mentioning that usually in mathematics, continuity is understood as continuity with

respect to some topology. For finite sets, however, this notion is not applicable: on a finite set, we either have a discrete topology (in which case all functions are continuous), or a topology that is reduced to an ordering relation, in which case monotonic functions and only they are continuous (see, e.g., [6]). This monotonicity is not enough for us: we have already included monotonicity in our definitions of &- and ∨-operations, and we want to formalize the evident fact that some monotonic operations are "continuous" (in intuitive sense), and some are not. Hence, we had to use new definitions of continuity.

Now, we are ready to formulate the main results of this section.

Proposition 8. *If f is a continuous &-operation on a finite logic L, then L is linearly ordered, and $f(a,b)=\min(a,b)$.*

Comment. For a linearly ordered set, $\min(a,b)$ is defined as the smallest of a and b.

Proposition 9. *If f is a continuous ∨-operation on a finite logic L, then L is linearly ordered, and $f(a,b)=\max(a,b)$.*

Example. Let us give an example of an &-operation that is different from min, and show that it is really discontinuous. As a finite logic, let us take the set of 11 numbers $\{0,0.1,0.2,\ldots,0.9,1.0\}$ with natural order. We thus defined L as a subset of the interval [0,1]. In the original paper of L. Zadeh [67], another operation on the interval [0,1] has been proposed for &: $f(a,b)=a\cdot b$. This operation, unlike min, cannot be directly applied to the chosen values, because, e.g., $0.6\cdot 0.6=0.36$ and 0.36 does not belong to the set of 11 chosen values. This difficulty is, however, easy to overcome: we can take as $f(a,b)$ the number from L that is the closest to $a\cdot b$ (and if there are two closest numbers, like 0.2 and 0.3 for $0.25=0.5\cdot 0.5$, choose the biggest of these two). For this operation, we will have $f(0.6,0.6)=0.4$, $f(0.3,0.5)=0.2$, etc.

Let us now consider the case when we have two statements A and B, and our degree of belief in each of them is equal to 0.9. Then, our degree of belief in $A\&B$ is equal to $f(0.9,0.9)=0.8$. In the chosen set L, 1.0 immediately follows 0.9, which means that an increase in the degree of belief from 0.9 to 1.0 can be called gradual. So, we can consider the possibility that our degrees of belief in both A and B gradually increase from 0.9 to 1.0. After this increase, the degree of belief in $A\&B$ becomes equal to $f(1.0,1.0)=1.0$. So, we gradually increased our degrees of belief in A and B, but the resulting degree of belief in $A\&B$ "jumped" from 0.8 to 1.0, skipping the value 0.9. Hence, this function f is discontinuous.

In Definition 22, we can thus take $a=b=0.9$, $a'=b'=1.0$, and $c=0.9$. *End of example.*

Let us now describe continuous operations with degrees of belief that correspond to other logical connectives.

Definition 24. *By a ¬-operation on L we mean a function $f:L\to L$ such that $f(T)=F$ and $f(F)=T$.*

Motivation. This condition simply means that if A is absolutely true, then $\neg A$ is absolutely false, and vice versa.

Proposition 10. *If $L=\{F=a_0<a_1<a_2<\ldots<a_n=T\}$ is a linearly ordered finite logic, and f is a continuous ¬-operation on L, then $f(a_i)=a_{n-i}$.*

Comment. We can represent this result in a manner that is closer to the traditional representation of uncertainty, if we describe each degree of belief a_i by a real number i/n. Then, for each truth value a, $f_\neg(a)=1-a$. This is exactly the operation originally proposed by Zadeh. In other words, not only the &- and ∨-operations initially proposed by Zadeh are the only continuous &- and ∨-operations, but his negation operation is the only continuous "not"-operation on a finite logic.

Let us now describe the implication operations.

Definition 25. *Let L be a finite logic. By an →-operation on L we mean a function $f_\to : L\times L\to L$ with the following properties:*

- *$f_\to(F,a)=T$;*
- *$f_\to(T,a)=a$;*
- *$f_\to(a,T)=T$;*
- *$f_\to(a,a)=1$;*
- *if $a\le a'$, then $f_\to(a,b)\ge f_\to(a',b)$.*

Motivations. The intended meaning of the function $f_\to(a,b)$ is as follows: if we know the degrees of belief $a=d(A)$ and $b=d(B)$ in some statements A and B, then $f_\to(a,b)$ is a reasonable degree of belief in the statement $A\to B$ ("A implies B"). With this interpretation in mind:

- The first of the above properties states that anything follows from a false statement.
- The second property states that to believe that A follows from an an absolutely true statement is the same as to believe that A is true, and therefore, the corresponding degrees of belief must coincide.
- The third condition means that a true statement follows from everything.
- The fourth condition means that for any statement A, A follows from A (and therefore, the degree of belief in $A\to A$ must be equal to T).

- The last condition is related to the third one: Namely, the third one says that if A is false, then $A \to B$ is always true. Therefore, if for some reason our degree of belief in a statement A decreases (from a' to a), then our belief that A can be false will correspondingly increase. Therefore, our degree of belief that $A \to B$ is true, will also increase. Hence, it is reasonable to demand that $f_\to(a', b) \le f_\to(a, b)$.

Proposition 11. *If $L = \{F = a_0 < a_1 < \ldots < a_n = T\}$ is a linearly ordered finite logic, and f is a continuous $\to$-operation on L, then $f(a_i, a_j) = a_{\min(n, n+j-i)}$.*

Comment. If we describe a_i by a real number i/n, then this $\to$-operation turns into $f(a, b) = \min(1, 1 + b - a)$.

Conclusion. In this section, we formalized the natural demand that gradual changes in $d(A)$ and $d(B)$ must lead to gradual changes in our estimate for $d(A\&B)$ (we called it continuity). We show that the only continuous &-operation is $\min(a, b)$. Likewise, the only continuous $\vee$-operation is $\max(a, b)$, the only continuous "not"−operation corresponds to $f(a) = 1 - a$, etc.

5.3 Towards a fuzzy arithmetic appropriate for computing with words

A techniques formalizing the computing-with-words approach to fuzzy arithmetic was described in [26].

6 Towards a combination of fuzzy logic and other soft computing formalisms

Another natural generalization of fuzzy logic is *Soft Computing*, an umbrella term which combines fuzzy, neural, genetic, interval, probabilistic, and other techniques, and which aims at a unified approach in which each of these techniques would be used for appropriate niche situations.

There exist joint formalisms combining fuzzy with neural and genetic, but much fewer mathematical results combine fuzzy with *probability* and alternative AI methods such as *logic programming*. The corresponding results and directions are described in [28, 43, 48]

7 Existing and potential applications: a brief overview

The main goal of this paper is to provide mathematical foundations for the application-oriented modifications and generalizations of fuzzy technology. Some of the described ideas are still being investigated, and they are not yet ready for practical applications. However, some other ideas have already made it to the application level. In this short section, we will briefly overview the main areas of these applications, with references to the papers in which these applications are described in more detail:

- applications to *control* [46, 47];
- applications to *decision making* [45], especially related to image processing [30];
- applications to *image processing* [24, 30, 62, 63] and *pattern recognition* [30, 31];
- applications to *data processing*: justification of known heuristic methods [43, 63] and design of *new* improved data processing techniques [9, 63];
- applications to *non-destructive testing* [66];
- applications to *medicine* [2, 39, 66, 64];
- applications to *foundations of physics* [22]; and
- applications to *education* [27].

Most of our application results came from several collaboration projects with researchers from different fields; these application results are presented in the separate papers presented at this workshop.

8 Proofs

Proof of Proposition 2. To prove this proposition, we will prove the following two statements:

- that for every "and" and "or" operation $f(a, b)$, and for every $\delta >$, we have $r_f(\delta) \ge \delta$; and that
- for min and max, we have $r_{\min}(\delta) = r_{\max}(\delta) = \delta$.

Let us prove the first statement first.

- To prove it for &, we consider $b = b' = 1$. Then, $f_\&(a, b) = a$ and $f_\&(a', b') = a'$; hence, according to Definition 3, if a and a' are δ-close, they must be α-close as well, where $\alpha = r_f(\delta)$. Thus, we must have $\alpha = r_f(\delta) \ge \delta$.
- Similarly, to prove this statement for $\vee$, we consider $b = b' = 0$. Then, $f_\vee(a, b) = a$ and $f_\vee(a', b') = a'$; hence, according to Definition 3, if a and a' are δ-close, they must be α-close as well, where $\alpha = r_f(\delta)$. Thus, we must have $\alpha = r_f(\delta) \ge \delta$.

Let us now show that for every δ, $r_{\min}(\delta) = r_{\max}(\delta) = \delta$, i.e., that if a is δ-close to a', and b is δ-close to b', then

- $\min(a, b)$ is δ-close to $\min(a', b')$, and
- $\max(a, b)$ is δ-close to $\max(a', b')$.

To prove these two statements, let us first describe what it means that two non-negative real numbers are δ-close.

If one of the two numbers (e.g., a) is equal to 0, then we have $|a - a'| \le \delta \cdot |a| = 0$, i.e., $|a - a'| = 0$ and therefore, the second number (a') is also equal to 0.

If both numbers are different from 0, i.e., if they are both positive, then the first inequality means that $\delta \cdot a \leq a' - a \leq \delta \cdot a$. Adding a to all three sides of this inequality, we get

$$(1-\delta) \cdot a \leq a' \leq (1+\delta) \cdot a. \qquad (1)$$

Similarly, the second inequality leads to $(1-\delta) \cdot a' \leq a$ and $a \leq (1+\delta) \cdot a'$. We can divide each part of each of these inequalities by the coefficient at a' and thus, get an equivalent inequality in terms of a':

$$\frac{1}{1+\delta} \cdot a \leq a' \leq \frac{1}{1-\delta} \cdot a. \qquad (2)$$

Since $(1-\delta) \cdot (1+\delta) = 1 - \delta^2 < 1$, we conclude that

$$1 - \delta < \frac{1}{1+\delta} \text{ and } 1 + \delta < \frac{1}{1-\delta}.$$

Thus, the two inequalities (1) and (2) are equivalent to a single two-sided inequality

$$\frac{1}{1+\delta} \cdot a \leq a' \leq (1+\delta) \cdot a.$$

This inequality is, in its turn, equivalent to $-\Delta \leq A - A' \leq \Delta$ (i.e., to $|A - A'| \leq \Delta$), where we denoted $a = \ln(a)$, $a' = \ln(a')$, and $\Delta = \ln(1+\delta)$.

For $c = f(a,b)$, since $\ln(x)$ is a monotonic function, we have $C = \min(A,B)$, where $C = \ln(c)$. So, in terms of A and B, the desired statement is that

if A and A' are absolutely Δ-close (i.e., $|A - A'| \leq \Delta$),

and B and B' are absolutely Δ-close (i.e., $|B - B'| \leq \Delta$),

then $C = \min(A,B)$ and $C' = \min(A',B')$ are also absolutely Δ-close (i.e., $|C - C'| \leq \Delta$).

This statement is indeed known to be true. Similarly, we can prove a similar statement for max. The proposition is proven.

Proof of Propositions 3 and 4. Proposition 3 can be viewed as particular case of Proposition 4, when $\mathcal{P} = \{P\}$, $\mathcal{P}' = \{P'\}$, and φ maps P onto P'. Therefore, to prove both Propositions 3 and 4, it is sufficient to prove Proposition 4.

We will show that under the conditions of Proposition 4, from $\mu^{(1)}_{P,a} = \mu^{(1)}_{\varphi(P),a'}$ and $\mu^{(2)}_{P,a} = \mu^{(2)}_{\varphi(P),a'}$, we will be able to conclude that $\varphi(P) = P$ for all $P \in \mathcal{P}$, and that $a = a'$; therefore, we will easily conclude that $\mu^{(n)}_{P,a} = \mu^{(n)}_{\varphi(P),a'}$ for all n.

Indeed, by definition of the first membership function, for every $x \in U$, we have $\mu^{(1)}_{P,a}(x) = \|P(x)\|$. Thus, from the equality $\mu^{(1)}_{P,a} = \mu^{(1)}_{\varphi(P),a'}$, we conclude that for every $P \in \mathcal{P}$, we have $\|P(x)\| = \|\varphi(P)(x)\|$ for all $x \in U$. Since the uncertainty representation is assumed to be sufficiently rich, we can conclude that $\varphi(P)(x) = P(x)$ for all $x \in U$, i.e., that $\varphi(P) = P$ for every $P \in \mathcal{P}$.

Let us now show that $a = a'$, i.e., that for every $w \in W$ and for every $s \in S$, we have $a(w,s) = a'(w,s)$. Indeed, since $\mathcal{P}$ is a non-degenerate class, there exists a value $x \in U$ and a property $P \in \mathcal{P}$ for which $P(x) = w$. Let us consider the equality of the second order membership functions for this very P. Since $\varphi(P) = P$, the given equality $\mu^{(2)}_{P,a} = \mu^{(2)}_{\varphi(P),a'}$ can be simplified into the following form: $\mu^{(2)}_{P,a} = \mu^{(2)}_{P,a'}$. Let us consider this equality for the above-chosen value x (for which $P(x) = w$). For this x, by definition of the second-order membership function, $\mu^{(2)}_{P,a}(x) = \|P(x)\|_{a,2} = \|w\|_{a,2}$; and similarly, $\mu^{(2)}_{P,a'}(x) = \|P(x)\|_{a,2} = \|w\|_{a',2}$; thus, $\|w\|_{a,2} = \|w\|_{a',2}$.

By definition, $\|w\|_{a,2}$ is a function which maps every value $s \in S$ into a 1-st order degree $\|a(w,s)\|_{a,1} = \|a(w,s)\|$. Thus, from the equality of the functions $\|w\|_{a,2}$ and $\|w\|_{a',2}$, we can conclude that their values at a given s are also equal, i.e., that $\|a(w,s)\| = \|a'(w,s)\|$. Since the uncertainty structure is sufficiently rich, we conclude that $a(w,s) = a'(w,s)$. The proposition is proven.

Proof of a comment after Proposition 3. Since P is *not* non-degenerate, there exists a value $w_0 \in W$ which cannot be represented as $P(x)$ for any $x \in U$. Let us pick arbitrary elements $x_0 \in U$ and $s_0 \in S$, and define $a(w,s)$ and $a'(w,s)$ as follows:

- first, we define $a(w,s) = a'(w,s)$ for all words w of the type $w = P(x)$: namely, we take $a(P(x_0), s_0) = a'(P(x_0), s_0) = w_0$ and take arbitrary other values for different pairs (w,s) with $w = P(x)$;
- then, we define $a(w,s)$ and $a'(w,s)$ for the remaining pairs (w,s): namely, we take $a(w_0, s_0) = w_0$, $a'(w_0, s_0) = P(x_0) \neq w_0$, and we define a and a' arbitrarily for all other pairs (w,s).

Let us show that for thus chosen adequacy functions, the membership functions of first and second order coincide, but the membership functions of the third order differ. Indeed:

- For the *first* order, we have, for every x, $\mu^{(1)}_{P,a}(x) = \|P(x)\|$ and similarly, $\mu^{(1)}_{P,a'}(x) = \|P(x)\|$; therefore, $\mu^{(1)}_{P,a}(x) = \mu^{(1)}_{P,a'}(x)$ for all x. Hence, $\mu^{(1)}_{P,a} = \mu^{(1)}_{P,a'}$.
- For the *second* order, for every x, $\mu^{(2)}_{P,a}(x)$ is a function which maps $s \in S$ into a value $\|a(P(x),s)\|_{a,1} = \|a(P(x),s)\|$. Similarly, $\mu^{(2)}_{P,a'}(x)$ is a function which maps $s \in S$ into a value $\|a'(P(x),s)\|_{a',1} = \|a'(P(x),s)\|$. For

words w of the type $P(x)$, we have defined a and a' in such a way that $a(w,s) = a'(w,s)$; therefore, $\|a(P(x),s)\| = \|a'(P(x),s)\|$ for all x and s. Thus, $\mu^{(2)}_{P,a} = \mu^{(2)}_{P,a'}$.

- Finally, let us show that the *third* order membership functions differ. We will show that the values of the functions $\mu^{(3)}_{P,a}$ and $\mu^{(3)}_{P,a'}$ differ for $x = x_0$. Indeed, by definition of the third order membership function,
  - $\mu^{(3)}_{P,a}(x_0)$ is a function which maps every s into the value $\|a(P(x_0),s)\|_{a,2}$, and
  - $\mu^{(3)}_{P,a'}(x_0)$ is a function which maps every s into the value $\|a'(P(x_0),s)\|_{a',2}$.

  To prove that these function are different, it is sufficient to show that their values differ for *some* values s; we will show that they differ for $s = s_0$, i.e., that $\|a(P(x_0),s_0)\|_{a,2} \neq \|a'(P(x_0),s_0)\|_{a',2}$. By our construction of a, we have $a(P(x_0),s_0) = a'(P(x_0),s_0) = w_0$, so the inequality that we need to prove takes the form $\|w_0\|_{a,2} \neq \|w_0\|_{a',2}$.
  By definition, $\|w_0\|_{a,2}$ is a function which maps every value $s \in S$ into $\|a(w_0,s)\|_{a,1} = \|a(w_0,s)\|$. Similarly, $\|w_0\|_{a',2}$ is a function which maps every value $s \in S$ into $\|a'(w_0,s)\|_{a,1} = \|a'(w_0,s)\|$. For s_0, according to our construction of a and a', we have $a(w_0,s_0) = w_0 \neq P(x_0) = a'(w_0,s_0)$. Thus, since the uncertainty representation is sufficiently rich, we conclude that $\|a(w_0,s_0)\| \neq \|a'(w_0,s_0)\|$, and therefore, that $\|w_0\|_{a,2} \neq \|w_0\|_{a',2}$ and $\mu^{(3)}_{P,a} = \mu^{(3)}_{P,a'}$.

The statement is proven.

Proof of Proposition 6. To prove this Proposition, we must prove the two properties of a crisp ordering.

If $a > b$, i.e., if $d(a,b) > \alpha$, then due to Definition 16, $1 - d(b,a) \geq d(a,b) > \alpha$. Hence, $d(b,a) < 1 - \alpha$. But $\alpha \geq 1/2$, so, $1-\alpha \leq 1/2$, and $d(b,a) < 1-\alpha \leq 1/2 \leq \alpha$, $d(b,a) < \alpha$, and $b \not> a$.

If $a > b$ and $b > c$, then $d(a,b) > \alpha$, $d(b,c) > \alpha$, and, due to Definition 16, $d(a,c) \geq \min(d(a,b), d(b,c)) > \alpha$. The proposition is proven.

Proof of Proposition 7. The first part of this proposition trivially follows from Definition 17. So, let us prove the second and the third parts. We will prove the second part; the third is proven similarly.

Assume that $d(a,b) > \beta$ and $d(b,c) > \alpha$, and let us prove that $d(a,c) > \beta$. Indeed, from Proposition 6, it follows that $d(a,c) > \alpha \geq 1/2$. So all three values $d(a,b)$, $d(b,c)$, and $d(a,c)$ are > 0.5. Due to Definition 16, $1-d(a,b) \geq \min(1-d(a,c), d(b,c))$. Since $d(a,c) > 1/2$, we conclude that $1-d(a,c) < 1/2 < d(b,c)$; therefore, $\min(1-d(a,c), d(b,c)) = 1-d(a,c)$, and the above inequality turns into $1 - d(a,b) \geq 1 - d(a,c)$, which is equivalent to $d(a,c) \geq d(a,b)$. But $d(a,b) > \beta$; therefore, $d(a,c) > \beta$, i.e., $a \gg c$. The proposition is proven.

Proof of Proposition 8.

1^o. Let us first prove that every element $a \in L$ can be connected to T by a finite chain $T = a_0 \gg a_1 \gg \ldots \gg a_k = a$ $(k \geq 0)$.

Indeed, if $a = T$, then we already have a chain, with $k = 0$.

If $a \neq T$, then according to our definition of a finite logic, we have $a < T$. If $a \ll T$, then we have a chain $a_0 = T$, $a_1 = a$. If $a \not\ll T$, then, according to the definition of $\ll$, it means that there exists a c such that $T > c > a$. If $T \gg c$, and $c \gg a$, then we have a desired chain. Else, we can insert additional elements in between them, etc.

On each step of this procedure, we either have a chain, or we can insert more elements into a sequence $T = a_0 > a_1 > \ldots > a_n = a$. Since there are only finitely many elements in the set L, and all a_i are different, this insertion cannot go on forever. Therefore, sooner or later, it will stop, and we will get the desired chain.

2^o. Let us now prove that $f(a,a) = a$ for every $a \in L$.

Indeed, suppose that $a \in L$ is given. According to 1^o, there exists a chain $T = a_0 \gg a_1 \gg \ldots \gg a_k = a$ that connects T and a.

If $k = 0$, then $a = T$, and $f(T,T) = T$ follows from the properties of an &-operation.

So, we can assume that $k > 0$. We will prove that $f(a,a) = a$ by reduction to a contradiction. Indeed, suppose that $f(a,a) \neq a$. Hence, $f(a_0,a_0) = a_0$, and $f(a_k,a_k) \neq a_k$. Let us denote by p the smallest integer for which $f(a_p,a_p) \neq a_p$. From this definition of p it follows, in particular, that $f(a_{p-1},a_{p-1}) = a_{p-1}$.

Since f is an &-operation, we can conclude that $f(a_p,a_p) \leq a_p$. Since $f(a_p,a_p) \neq a_p$ (by the choice of p), we conclude that $f(a_p,a_p) < a_p$.

Therefore, we have $a_p \ll a_{p-1}$, and $f(a_p,a_p) < a_p < a_{p-1} = f(a_{p-1},a_{p-1})$, i.e., f is discontinuous (here, $a = b = a_p$, $a' = b' = a_{p-1}$, and $c = a_p$). However, we assumed that f is continuous.

This contradiction proves that $f(a,a)$ cannot be different from a, so $f(a,a) = a$ for all a.

3^o. Let us prove that L is linearly ordered, i.e., for every two elements $a,b \in L$, either $a = b$, or $a < b$, or $b < a$.

Indeed, let us take $a, b \in L$. Following 1^o, we will form chains $T = a_0 \gg a_1 \gg \ldots \gg a_k = a$, and $T = b_0 \gg b_1 \gg \ldots \gg b_l = b$. Let us denote by p the biggest integer for which a_p and b_p are both defined and equal to each other ($a_p = b_p$).

If $p = k = l$, then $a = a_k = a_p = b_p = b_l = b$, i.e., $a = b$.

If $p = k \neq l$, then $a = a_k = b_p \gg b_{p+1} \gg \ldots \gg b_l = b$, therefore $a > b_{p+1} > \ldots > b_l = b$, and $a > b$.

Likewise, if $p = l \neq k$, then $b > a$.

Let us prove that the remaining case when $p < k$ and $p < l$, is impossible. Indeed, in this case, both a_{p+1} and b_{p+1} are defined and different from each other. Since f is an &-operation, we can conclude that $f(a_{p+1}, b_{p+1}) \leq a_{p+1}$ and $f(a_{p+1}, b_{p+1}) = f(b_{p+1}, a_{p+1}) \leq b_{p+1}$.

The first inequality means that we have two possibilities: $f(a_{p+1}, b_{p+1}) = a_{p+1}$, and $f(a_{p+1}, b_{p+1}) < a_{p+1}$. We will show that in both cases, we have a contradiction.

Suppose first that $f(a_{p+1}, b_{p+1}) = a_{p+1}$. We already know that $f(a_{p+1}, b_{p+1}) \leq b_{p+1}$, so $a_{p+1} \leq b_{p+1}$. We chose p in such a way that $a_{p+1} \neq b_{p+1}$ (and $a_p = b_p$), therefore $a_{p+1} < b_{p+1}$. So, $a_{p+1} < b_{p+1} < b_p = a_p$. The existence of the intermediate value b_{p+1} contradicts the assumption that $a_{p+1} \ll a_p$. So, in this case, we have a contradiction.

Let us now consider the case when $f(a_{p+1}, b_{p+1}) < a_{p+1}$. Since $a_p = b_p$ (because of our choice of p), and $f(a, a) = a$ for all a (this we have proved), we have $f(a_{p+1}, b_{p+1}) < a_{p+1} < a_p = f(a_p, a_p) = f(a_p, b_p)$. Therefore, in this case, $a_{p+1} \ll a_p$, $b_{p+1} \ll a_p$, and $f(a_{p+1}, b_{p+1}) < a_{p+1} < f(a_p, b_p)$. Hence, we have a proof that f is discontinuous (with $a = a_{p+1}, b = b_{p+1}, a' = a_p, b' = b_p$, and $a_{p+1} = c$). This contradicts to our assumption that f is continuous.

Summarizing: in both cases the assumption that $p < k$ and $p < l$ led us to a contradiction. So, either $p = k$, or $p = l$, in which cases, as we have already proved, either $a = b$, or $a < b$, or $b < a$. We have thus proved that L is linearly ordered.

4^o. It now remains to prove that $f(a, b) = \min(a, b)$ for all a, b.

Since L is finite and linearly ordered, we can order all its elements into a sequence $F = a_0 < a_1 < \ldots < a_{n-1} < a_n = T$. So, each element of L has the form a_i, and $a_i < a_j$ if and only if $i < j$.

In these terms, it is necessary to prove that $f(a_i, a_j) = a_{\min(i,j)}$. If $i = j$, this follows from 2^o. Let us now consider the case, when $i < j$, and prove that in this case, $f(a_i, a_j) = a_i$.

Let us fix j. For every i, the value of $f(a_i, a_j) \in L$ is equal to a_k for some k. Let us denote this k by $\phi(i)$. So, in these denotations, $f(a_i, a_j) = a_{\phi(i)}$. The desired equality can be then expressed as $\phi(i) = i$ for all $i \leq j$.

We already know the value of this function $\phi(i)$ for $i = 0$ and $i = j$: Indeed, since f is an &-operation, we have $f(T, a_j) = T$, i.e., in our notations, $f(a_0, a_j) = a_0$, hence $\phi(0) = 0$. From 2^o, it follows that $f(a_j, a_j) = a_j$, so $\phi(j) = j$.

Since f is an &-operation, it is monotonically non-decreasing, hence ϕ is also non-decreasing: $0 = \phi(0) \leq \phi(1) \leq \phi(2) \leq \ldots \leq \phi(j) = j$.

Since $a_i \ll a_{i+1}$, and f is continuous, there cannot be a gap between $F(a_i)$ and $F(a_{i+1})$. Therefore, for each i, we must either have $\phi(i+1) = \phi(i)$, or $\phi(i+1) = \phi(i) + 1$. Since

$$j = j - 0 = \phi(j) - \phi(0) =$$

$$(\phi(j) - \phi(j-1)) + \ldots + (\phi(2) - \phi(1)) + (\phi(1) - \phi(0)),$$

the number j is the sum of j differences, each of which is ≤ 1. If one of these differences was smaller than 1, then the entire sum would be smaller than j. Since this sum is equal to j, none of these differences can be smaller than 1. Therefore, $\phi(i+1) - \phi(i) = 1$ for all i. This equality is equivalent to $\phi(i+1) = \phi(i) + 1$.

So, we have $\phi(0) = 0$, and $\phi(i+1) = \phi(i) + 1$ for all $i < j$. From this, we can conclude (using mathematical induction), that $\phi(i) = i$ for all $i < j$. By definition of ϕ this means that $f(a_i, a_j) = a_{\phi(i)} = a_i$, i.e., that $f(a, b) = \min(a, b)$.

If $i > j$, then the desired equality follows from the fact that f is commutative ($f(a_i, a_j) = f(a_j, a_i)$), and so this case is reduced to the previous one. The proposition is proven.

Comment. The ideas of this proof are similar to the proofs from [1, 58].

Proof of Proposition 9 is similar, with the only difference that we must use F instead of T, $>$ instead of $<$, and $\ll$ instead of $\gg$.

Proof of Proposition 10. For every $a_i \in L$, $f(a_i) = a_k$ for some k. Let us denote this k by $\psi(i)$. In these terms, $f(a_i) = a_{\psi(i)}$. The definition of a negation operation means that $\psi(0) = n$, and $\psi(n) = 0$. Continuity means that for each i, since $a_i \ll a_{i+1}$, there cannot be anything in between $a_{\psi(i)} = f(a_i)$ and $a_{\psi(i+1)} = f(a_{i+1})$. In other words, there cannot be anything in between $\psi(i)$ and $\psi(i+1)$. So, $\psi(i)$

and $\psi(i+1)$ must either coincide, or be neighbors: $|\psi(i+1)-\psi(i)| \le 1$. In particular, $\psi(i+1)-\psi(i) \ge -1$.

Now, the difference $\psi(n) - \psi(0) = 0 - n = -n$ can be represented as

$$-n = \psi(n) - \psi(0) = (\psi(n) - \psi(n-1)) + \ldots + (\psi(2) - \psi(1)) + (\psi(1) - \psi(0)).$$

So, $-n$ is represented as the sum of n terms each of which is ≥ -1. If one of them was greater than -1, then the entire sum would have been greater than $-n$. Since this sum is equal to $-n$, we can conclude that all the terms in this sum are exactly equal to -1: $\psi(i+1) - \psi(i) = -1$. Therefore, $\psi(0) = n$, and $\psi(i+1) = \psi(i) - 1$ for all i. From these two conditions, one can easily conclude that $\psi(i) = n - i$. Hence, $f(a_i) = a_{\psi(i)} = a_{n-i}$. The proposition is proven.

Proof of Proposition 11. For every i and j, the value $f(a_i, a_j)$ belongs to L and is, therefore, equal to a_k for some k. Let us denote this k by $h(i,j)$, so that $f(a_i, a_j) = a_{h(i,j)}$.

We will consider two cases: $i \le j$, and $i > j$.

Let us first assume that $i \le j$. According to the definition of an $\rightarrow$-operation, $f(a_j, a_j) = T = a_n$, and $f(F, a_j) = f(a_0, a_j) = T = a_n$. In terms of h, it means that $h(j,j) = n$, and $h(0,j) = n$. From the fifth property of an $\rightarrow$-operation, we can conclude that $h(0,j) \ge h(1,j) \ge \ldots \ge h(j-1,j) \ge h(j,j)$. Since $h(0,j) = h(j,j) = n$, we can conclude that all the terms in this inequality are equal to n, i.e., $h(i,j) = n$ if $i \le j$.

Let us now consider the case, when $i > j$. According to the definition of an $\rightarrow$-operation, for every j, we have $f(T, a_j) = a_j$, and $f(a_j, a_j) = 1$. In terms of h, this turns into $h(n,j) = j$ and $h(j,j) = n$. Since f is continuous, we can conclude (just like we did in the proofs of Theorems 1 and 3) that $|h(i+1,j) - h(i,j)| \le 1$. So, the difference between $h(n,j)$ and $h(j,j)$ that is equal to $j - n = -(n-j)$, can be represented as the sum of $n-j$ differences $h(i+1,j) - h(i,j)$ $(j \le i < n)$, each of which is ≥ -1. If one of these differences was > -1, then the entire sum would be $> -(n-j)$. Therefore, all these difference are equal to -1. So, $h(j,j) = n$, and for $i \ge j$, $h(i+1,j) = h(i,j) - 1$. Therefore, for $i \ge j$, we have $h(i,j) = n - (i-j) = n + j - i$.

Combining the cases $i \le j$ and $i > j$, we get the desired formula. The proposition is proven.

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

References

[1] J. Agustií et al. "Structured local fuzzy logics in MILORD", In: L. Zadeh, J. Kacrpzyk (eds.) *Fuzzy Logic for the Management of Uncertainty*, Wiley, N.Y., 1992, pp. 523–551.

[2] K. M. Aló, R. Aló, A. de Korvin, and V. Kreinovich, "Spinal cord stimulation for chronic pain management: towards an expert system", *Proceedings of the 4th World Congress on Expert Systems, Mexico City, March 16–20, 1998*, Vol. 1, pp. 156–164.

[3] J. M. Blin, "Fuzzy relations in group decision theory", *J. of Cybernetics*, 1974, Vol. 4, pp. 17–22.

[4] J. M. Blin and A. B. Whinston, "Fuzzy sets and social choice", *J. of Cybernetics*, 1973, Vol. 3, pp. 28–36.

[5] B. Bouchon-Meunier *et al.*, "On the formulation of optimization under elastic constraints (with control in mind)", *Fuzzy Sets and Systems,* 1996, Vol. 81, pp. 5–29.

[6] N. Bourbaki. *General Topology*, Addison-Wesley, Reading, MA, 1967.

[7] B. G. Buchanan and E. H. Shortliffe, *Rule-Based Expert Systems*, Addison-Wesley, Reading, MA, Menlo Park, CA, 1984.

[8] B. Cloteaux, C. Eick, B. Bouchon-Meunier, and V. Kreinovich, "From ordered beliefs to numbers: how to elicit numbers without asking for them (doable but computationally difficult)", *International Journal of Intelligent Systems*, 1998 (to appear).

[9] D. E. Cooke, V. Kreinovich, and S. A. Starks, "ALPS: A Logic for Program Synthesis (motivated by fuzzy logic)", *Proceedings of the FUZZ-IEEE'98 International Conference on Fuzzy Systems*, Anchorage, Alaska, May 4–9, 1998, Vol. 1, pp. 779–784.

[10] D. Dubois and H. Prade, *Fuzzy sets and systems: theory and applications,* Academic Press, N.Y., London, 1980.

[11] A. Finkelstein, O. Kosheleva, and V. Kreinovich, "Astrogeometry: towards mathematical foundations", *International Journal of Theoretical Physics*, 1997, Vol. 36, No. 4, pp. 1009–1020.

[12] A. M. Finkelstein, V. Kreinovich, and R. R. Zapatrin, "Fundamental physical equations uniquely determined by their symmetry groups," *Lecture Notes in Mathematics*, Springer-Verlag, Berlin-Heidelberg-N.Y., Vol. 1214, 1986, pp. 159–170.

[13] P. C. Fishburn, *Nonlinear preferences and utility theory*, John Hopkins University Press, Baltimore, 1988.

[14] O. N. Garcia, V. Kreinovich, L. Longpré, and H. T. Nguyen, "Complex problems: granularity is necessary, granularity helps", *These Proceedings.*

[15] V. B. Gisin, "On transitivity of strict preference relation", *Fuzzy Sets and Systems*, 1994, Vol. 67, pp. 293–301.

[16] L. Godo, R. Lopez de Mantaras, C. Sierra, and A. Verdaguer, "MILORD: The Architecture and management of Linguistically expressed Uncertainty", *International Journal of Intelligent Systems*, 1989, Vol. 4, pp. 471–501.

[17] J. A. Goguen, "L–fuzzy sets", *Journal of Mathematical Analysis and Applications*, 1967, Vol. 18, pp. 145–174.

[18] J. A. Goguen, "The logic of inexact reasoning", *Synthese*, 1969, Vol. 19, pp. 325–373 (reprinted in D. Dubois, H. Prade, and R. Yager (eds.), *Reading in Fuzzy Sets for Intelligent Systems*, Morgan Kaufmann, San Mateo, CA, 1994, pp. 417–441).

[19] H. M. Hersch and A. Camarazza. "A fuzzy-set approach to modifiers and vagueness in natural languages", *J. Exp. Psychology: General*, 1976, Vol. 105, pp. 254–276.

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

[21] M. Koshelev and V. Kreinovich, "Towards computers of generation omega – non-equilibrium thermodynamics, granularity, and acausal processes: a brief survey", *Proceedings of the International Conference on Intelligent Systems and Semiotics (ISAS'97)*, National Institute of Standards and Technology Publ., Gaithersburg, MD, 1997, pp. 383–388.

[22] M. Koshelev, V. Kreinovich, H. T. Nguyen, and "Uncertainty representation explains and helps methodology of physics and science in general", *These Proceedings.*

[23] O. M. Kosheleva and V. Kreinovich, "Algorithmic problems of nontransitive (SSB) utilities," *Mathematical Social Sciences*, 1991, Vol. 21, pp. 95–100.

[24] O. Kosheleva, V. Kreinovich, B. Bouchon-Meunier, and R. Mesiar, "Operations with fuzzy numbers explain heuristic methods in image processing", *Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'98),* Paris, France, July 6–10, 1998.

[25] O. Kosheleva, V. Kreinovich, H. T. Nguyen, and B. Bouchon-Meunier, "How to describe partially ordered preferences: mathematical foundations", *These Proceedings.*

[26] V. Kreinovich and B. Bouchon-Meunier, "Granularity via Non-Deterministic Computations: What We Gain and What We Lose", *International Journal of Intelligent Systems*, 1997, Vol. 12, pp. 469–481.

[27] V. Kreinovich, E. Johnson-Holubec, L. K. Reznik, and M. Koshelev, "Cooperative learning is better: explanation using dynamical systems, fuzzy logic, and geometric symmetries" *These Proceedings.*

[28] V. Kreinovich, L. Longpré, and H. T. Nguyen, "Towards formalization of feasibility, randomness, and commonsense implication: Kolmogorov complexity, and the necessity of considering (fuzzy) degrees", *These Proceedings.*

[29] 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.

[30] V. Kreinovich, H. T. Nguyen, S. A. Starks, and Y. Yam, "Decision making based on satellite images: optimal fuzzy clustering approach", *Proceedings of the 37th IEEE Conference on Decision and Control CDC'98,* Tampa, Florida, December 16–18, 1998 (to appear).

[31] V. Kreinovich, H. T. Nguyen, and Y. Yam, "Optimal choices of potential functions in fuzzy clustering", *The Chinese University of Hong Kong, Department of Mechanical & Automation Engineering*, Technical Report CUHK-MAE-98-001, January 1998.

[32] V. Kreinovich, C. Quintana, and R. Lea, "What procedure to choose while designing a fuzzy control? Towards mathematical foundations of fuzzy control", *Working Notes of the 1st International Workshop on Industrial Applications of Fuzzy Control and Intelligent Systems*, College Station, TX, 1991, pp. 123–130.

[33] V. Kreinovich *et al.*, "What non-linearity to choose? Mathematical foundations of fuzzy control", *Proceedings of the 1992 International Conference on Fuzzy Systems and Intelligent Control*, Louisville, KY, 1992, pp. 349–412.

[34] C. Langrand, V. Kreinovich and H. T. Nguyen, "Two-dimensional fuzzy logic for expert systems", *Sixth International Fuzzy Systems Association World Congress*, San Paulo, Brazil, July 22–28, 1995, Vol. 1, pp. 221–224.

[35] B. B. Mandelbrot, *The fractal geometry of Nature*, Freeman, San Francisco, 1982.

[36] G. A. Miller, "The magical number seven plus or minus two: some limits on our capacity for processing information", *Psychological Review*, 1956, Vol. 63, pp. 81–97.

[37] P. M. Milner, *Physiological psychology,* Holt, NY, 1970.

[38] *NAFIPS/IFIS/NASA'94, San Antonio, December 1994*, IEEE, Piscataway, NJ, 1994.

[39] H. P. Nguyen, V. Kreinovich, and S. A. Starks, "An expert system for acupuncture", *These Proceedings.*

[40] H. T. Nguyen, M. Koshelev, O. Kosheleva, V. Kreinovich, and R. Mesiar, "Computational complexity and feasibility of fuzzy data processing: why fuzzy numbers, which fuzzy numbers, which operations with fuzzy numbers", *Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'98),* Paris, France, July 6–10, 1998.

[41] H. T. Nguyen and V. Kreinovich, "Towards theoretical foundations of soft computing applications", *Int'l J. on Uncertainty, Fuzziness, and Knowledge-Based Systems,* 1995, Vol. 3, No. 3, pp. 341–373.

[42] H. T. Nguyen and V. Kreinovich, *Applications of continuous mathematics to computer science*, Kluwer, Dordrecht, 1997.

[43] H. T. Nguyen and V. Kreinovich, "Using Gelfond-Przymusinska's epistemic specifications to justify (some) heuristic methods used in expert systems and intelligent control", *Soft Computing*, 1997, Vol. 1, No. 4, pp. 198–209.

[44] H. T. Nguyen and V. Kreinovich, "Methodology of fuzzy control: an introduction", In: H. T. Nguyen and M. Sugeno (eds.), *Fuzzy Systems: Modeling and Control*, Kluwer, Boston, MA, 1998, pp. 19–62.

[45] H. T. Nguyen and V. Kreinovich, "Multi-criteria optimization – an important foundation of fuzzy system design", In: L. Reznik, V. Dimitrov, and J. Kacprzyk, *Fuzzy system design: social and engineering applications*, Physica Verlag, Heidelberg, 1998, pp. 24–35.

[46] H. T. Nguyen and V. Kreinovich, "A modification of Sugeno integral describes stability and smoothness of fuzzy control", *Proceedings of the FUZZ-IEEE'98 International Conference on Fuzzy Systems*, Anchorage, Alaska, May 4–9, 1998, Vol. 1, pp. 360–365.

[47] H. T. Nguyen, V. Kreinovich, and R. Aló, "Adding fuzzy integral to fuzzy control", *Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU'98)*, Paris, France, July 6–10, 1998.

[48] H. T. Nguyen, V. Kreinovich, D. E. Cooke, Luqi, and O. Kosheleva, "Towards combining fuzzy and logic programming techniques", *These Proceedings.*

[49] H. T. Nguyen, V. Kreinovich, R. N. Lea, and D. Tolbert. "Interpolation that leads to the narrowest intervals, and its application to expert systems and intelligent control", *Reliable Computing*, 1995, Vol. 1 No. 3, pp. 299–316.

[50] H. T. Nguyen, V. Kreinovich, D. Tolbert, "On robustness of fuzzy logics". *Proceedings of IEEE-FUZZ International Conference*, San Francisco, CA, March 1993, Vol. 1, pp. 543–547.

[51] H. T. Nguyen, V. Kreinovich, and V. Shekhter, "On the possibility of using complex values in fuzzy logic for representing inconsistencies", *International Journal of Intelligent Systems*, 1998, Vol. 13, No. 8, pp. 683–714.

[52] H. T. Nguyen, V. Kreinovich, and D. Tolbert, "A measure of average sensitivity for fuzzy logics", *International Journal on Uncertainty, Fuzziness, and Knowledge-Based Systems*, 1994, Vol. 2, No. 4, pp. 361–375.

[53] H. T. Nguyen, V. Kreinovich, and P. Wojciechowski, "Strict Archimedean t-norms and t-conorms as universal approximators", *International Journal of Approximate Reasoning*, 1998, Vol. 18, pp. 239–249.

[54] H. T. Nguyen, V. Kreinovich, and Q. Zuo, "Interval-valued degrees of belief: applications of interval computations to expert systems and intelligent control", *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems (IJUFKS)*, 1997, Vol. 5, No. 3, pp. 317–358.

[55] H. T. Nguyen and E. A. Walker, *A first course in fuzzy logic,* CRC Press, Boca Raton, Florida, 1997.

[56] G. C. Oden, "Integration of fuzzy logical information", *Journal of Experimental Psychology: Human Perception Perform.*, 1977, Vol. 3, No. 4, pp. 565–575.

[57] A. T. Popov, H. T. Nguyen, and L. K. Reznik, "An application of fuzzy mathematical morphology to interval-valued knowledge representation: a remark", *Reliable Computing* (to appear).

[58] J. Puyol-Gruart, L. Godo, and C. Sierra. *A specialization calculus to improve expert systems communication*, Research Report IIIA 92/8, Institut d'Investigació en Intelligènicia Artificial, Spain, May 1992.

[59] E. R. Scerri, V. Kreinovich, P. Wojciechowski, and R. R. Yager, "Ordinal explanation of the periodic system of chemical elements", *International Journal of Uncertainty, Fuzziness, and Knowledge-Based Systems (IJUFKS)* (to appear).

[60] G. Shafer and J. Pearl (eds.) *Readings in Uncertain Reasoning*, M. Kaufmann, San Mateo, CA, 1990.

[61] M. H. Smith and V. Kreinovich, "Optimal strategy of switching reasoning methods in fuzzy control", Chapter 6 in H. T. Nguyen, M. Sugeno, R. Tong, and R. Yager (eds.), *Theoretical aspects of fuzzy control,* J. Wiley, N.Y., 1995, pp. 117–146.

[62] S. A. Starks and V. Kreinovich, "Soft computing: frontiers? A case study of hyper-spectral satellite imaging", *Working Notes of the AAAI Symposium on Frontiers in Soft Computing and Decision Systems,* Boston, MA, November 8–10, 1997, p. 66–71.

[63] S. A. Starks, V. Kreinovich, and A. Meystel, "Multi-resolution data processing: it is necessary, it is possible, it is fundamental", *Proceedings of the International Conference on Intelligent Systems and Semiotics (ISAS'97),* National Institute of Standards and Technology Publ., Gaithersburg, MD, 1997, pp. 145–150.

[64] S. A. Starks, H. T. Nguyen, V. Kreinovich, H. P. Nguyen, and M. Navara, "Strong negation: its relation to intervals and its use in expert systems", In: G. Alefeld and R. A. Trejo (eds.), *Interval Computations and its Applications to Reasoning Under Uncertainty, Knowledge Representation, and Control Theory. Proceedings of MEXICON'98, Workshop on Interval Computations, 4th World Congress on Expert Systems*, México City, México, 1998.

[65] *Symmetries in physics*, Springer-Verlag, Berlin, N.Y., 1992.

[66] K. Worden, R. Osegueda, C. Ferregut, S. Nazarian, E. Rodriguez, D. L. George, M. J. George, V. Kreinovich, O. Kosheleva, and S. Cabrera, "Interval approach to non-destructive testing of aerospace structures and to mammography", In: G. Alefeld and R. A. Trejo (eds.), *Interval Computations and its Applications to Reasoning Under Uncertainty, Knowledge Representation, and Control Theory. Proceedings of MEXICON'98, Workshop on Interval Computations, 4th World Congress on Expert Systems*, México City, México, 1998.

[67] L. Zadeh, "Fuzzy sets", *Information and control*, 1965, Vol. 8, pp. 338–353.

[68] H. J. Zimmerman. "Results of empirical studies in fuzzy set theory", In: G. J. Klir (ed.), *Applied General Systems Research*, Plenum, N.Y., 1978, pp. 303–312.