12-2014

50 Years of Fuzzy: from Discrete to Continuous to -- Where?

Vladik Kreinovich
University of Texas at El Paso, vladik@utep.edu

Hung T. Nguyen
New Mexico State University - Main Campus, hunguyen@nmsu.edu

Olga Kosheleva
University of Texas at El Paso, olgak@utep.edu

Rujira Ouncharoen
Chiang Mai University, rujira.o@cmu.ac.th

Comments:
Technical Report: UTEP-CS-14-74

Undefined 0 (2014) 1–0
IOS Press

50 Years of Fuzzy: from Discrete to Continuous to – Where?

Vladik Kreinovich [a,*], Hung T. Nguyen [b], Olga Kosheleva [c], and Rujira Ouncharoen [d]
[a] *Department of Computer Science, University of Texas at El Paso, 500 W. University, El Paso, TX 79968, USA*
E-mail: vladik@utep.edu
[b] *Department of Mathematical Sciences, New Mexico State University, Las Cruces, NM 88003, USA, and*
Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand
E-mail: hunguyen@nmsu.edu
[c] *Department of Teacher Education, University of Texas at El Paso, 500 W. University, El Paso, TX 79968, USA*
E-mail: olgak@utep.edu
[d] *Department of Mathematics, Chiang Mai University, Chiang Mai, Thailand*
E-mail: rujira.o@cmu.ac.th

Abstract. While many objects and processes in the real world are discrete, from the computational viewpoint, discrete objects and processes are much more difficult to handle than continuous ones. As a result, a continuous approximation is often a useful way to describe discrete objects and processes. We show that the need for such an approximation explains many features of fuzzy techniques, and we speculate on to which promising future directions of fuzzy research this need can lead us.

Keywords: fuzzy techniques, discrete, continuous, interval-valued fuzzy, complex-valued fuzzy, computing with words, dynamical fuzzy logic, chemical kinetics, non-additive measures, symmetry

1. Fuzzy Techniques as an Easier-to-Compute Continuous Approximation for Difficult-to-Compute Discrete Objects and Processes

Discrete objects and processes are ubiquitous. Many real-life objects are processes are discrete. On the macro level, there is an abrupt transition in space between physical bodies, there is an abrupt transition in time when, e.g., a glass breaks or a person changes his/her opinion. On the micro level, matter consists of discrete atoms and molecules, with abrupt transitions between different states of an atom.

Continuous problems are easier to compute. While discrete objects and processes are ubiquitous in nature, from the computational viewpoint, it is often much easier to handle continuous problems. This may sound counter-intuitive, since intuitively, if we restrict our search or optimization to only integer values, the problem would become easier – but it is not.

For example, in the continuous case, it is relatively easy to find a solution $x_1, \ldots, x_n$ to a system of linear equations $\sum\limits_{j=1}^{n} a_{ij} \cdot x_j = b_i$, $1 \le i \le m$ (there are many known feasible algorithms for that), the problem becomes NP-hard (computationally intractable) if we only allow discrete values of x_i; see, e.g., [9,28].

Similarly, in the continuous case, it is relatively easy to find the values $x_1, \ldots, x_n$ that minimize a given quadratic function $f(x_1, \ldots, x_n)$: it is sufficient to solve the corresponding system of linear equations $\dfrac{\partial f}{\partial x_i} = 0$. However, optimization of quadratic functions for discrete inputs, e.g., for $x_i \in \{0, 1\}$, is NP-hard [9,28].

*Corresponding author. E-mail: vladik@utep.edu.

Continuous approximations of discrete objects and processes are ubiquitous in physics. Because dealing with discrete objects and processes is often computationally complicated, physicists often approximate discrete objects with continuous ones.

For example, it is not feasible to describe the changes in atmosphere by tracing all 10^{23} molecules, but approximate equations that describe the atmosphere as a continuous field leads to many useful weather predictions. Similar, a solid body – in effect, a collection of atoms – is well described by a continuous density field, and an atomic nucleus – a collection of protons and neutrons – is well described by a continuous (liquid) model; see, e.g., [8].

Such approximations are also useful in analyzing social phenomena. For example, in analyzing how epidemics spread, is it not feasible to trace the interactions of all the people, but it is possible to make good predictions by using a few parameters to describe the current state of the epidemics: e.g., the number of people who are sick and the number of people who are infected but not yet sick.

Fuzzy techniques as a continuous approximation of discrete objects and processes. Many applications of fuzzy techniques [12,25,40] can be viewed as a similar continuous approximation of discrete objects and processes. For example, when a person changes his mind about a voting issue, then, instead of describing this change as a discrete transition, we can view it as continuous change, in which the person's confidence in the previous candidate gradually decreases.

What we do in this paper. In this paper, we show that this approximation interpretation can explain many features of fuzzy techniques, and that we can use this interpretation to predict which possible modifications of fuzzy techniques are the most promising.

The structure of this paper is as follows. In Section 2, we consider fuzziness as a way to overcome discreteness in time; some of its results first appeared in [19]. Section 3 – expanding [16] – deals with spatial discreteness. In Section 4 (expanding [34]) we show how these interpretation can help understand different fuzzy operations. The following sections describe the future of fuzzy: Section 5 explains how spatial discreteness naturally leads to interval values. Section 6 (expanding [15]) explains how the hierarchical character of discrete data leads to computing with words. Section 7 (expanding [14]) shows that an approximate character of fuzzy values leads to complex-valued fuzzy sets. Finally, Section 8 (expanding [17]) speculates on how fuzzy techniques can properly take into account the dynamic character of objects and processes.

2. Fuzzy Techniques as an Approximation to Discreteness in Time

Qualitative idea. Let us consider a simple transition from small to medium to large. When the value of the quantity q is small, we are sure that this value is small and not medium or large. As we increase q, what was originally small starts slowly transforming into medium, then what was originally medium starts slowly transforming into large, etc. On the qualitative level, this can be described by the following transitions: $s \to m$, $m \to \ell$.

How to describe this idea in numerical terms? There are many physical situations when we have a similar transition. Historically the first such transitions were observed in chemistry; as a result, the corresponding equations were first designed in chemistry and are thus known as equations of chemical kinetics. Let us use these equations to provide a numerical description of the transitions (1).

Equations of chemical kinetics: brief reminder. Chemical kinetics describes how the concentration of different chemical substances change when the chemical reactions occur. It is usually assumed that the rate of a chemical reaction is proportional to the production of the concentrations of all the substances which are needed for the reaction to take place. For example, for a reaction $A + B \to C$, the reaction rate is proportional to the product $a \cdot b$ of the concentrations of the substances A and B. Due to this reaction rate $k \cdot a \cdot b$:

- the amounts a and b of substances A and B decrease with this rate, while
- the amount c of the substance C increases with this rate,

i.e., we have

$$\frac{da}{dt} = -k \cdot a \cdot b, \quad \frac{db}{dt} = -k \cdot a \cdot b, \quad \frac{dc}{dt} = k \cdot a \cdot b.$$

If we have several different chemical reactions, then, to describe the resulting rate of change of each concentration, we add the rates corresponding to different reactions.

Resulting equations. If we assume that both reactions $s \to m$ and $m \to \ell$ have the same rate k, then we conclude that for this system of reactions, the corresponding system of equations has the form

$$\frac{ds}{dq} = -k \cdot s; \ \ \frac{dm}{dq} = k \cdot s - k \cdot m; \ \ \frac{d\ell}{dq} = k \cdot m.$$

(Note that since we increase q, not time, the derivatives are with respect to q.) Instead of three consequent membership functions corresponding to "small", "medium", and "large", we can consider an arbitrary number n of such functions $\mu_1(q)$, ..., $\mu_n(q)$. In this case, the chemical reactions have the form

$$\mu_1 \to \mu_2, \ \ \mu_2 \to \mu_3, \ldots, \mu_{n-1} \to \mu_n.$$

If we assume that all these reactions have the same rate λ, then the corresponding chemical kinetic equations have the form:

$$\frac{d\mu_1(q)}{dq} = -\lambda \cdot \mu_1(q), \ \ \frac{d\mu_2(q)}{dq} = \lambda \cdot \mu_1(q) - \lambda \cdot \mu_2(q),$$

$$\ldots \frac{d\mu_i(q)}{dq} = \lambda \cdot \mu_{i-1}(q) - \lambda \cdot \mu_i(q), \ldots$$

$$\frac{d\mu_{n-1}(q)}{dq} = \lambda \cdot \mu_{n-2}(q) - \lambda \cdot \mu_{n-1}(q),$$

$$\frac{d\mu_n(q)}{dq} = \lambda \cdot \mu_{n-1}(q).$$

Initially, the value is small, so $\mu_1(0) = 1$ and $\mu_2(0) = \ldots = \mu_n(0) = 0$. This system of reactions is well-studied in chemical kinetics [38,39]. Its solution has the form

$$\mu_i(q) = \frac{\lambda^{i-1}}{(i-1)!} \cdot q^{i-1} \cdot \exp(-\lambda \cdot q).$$

Comment. It should be noticed that the corresponding functions are not normalized, in the sense that the maximum of each such function is not equal to 1. If we want normalized membership functions, we must multiply each such function by a corresponding normalizing factor.

Why these membership functions? The above formula provides possible membership functions, but why necessarily these ones? Let us show that exactly these functions emerge if we take into account an important physical property of symmetry.

Shift-invariant quantities: a brief reminder. In many physical theories, there is no fixed starting point for measuring the corresponding physical quantities; see, e.g., [8]. For example, we can measure time based on the current calendar or – as the French Revolution suggested – starting with the year 1789. If instead of the original starting point, we select a new one which is q_0 units smaller, then the original numerical value q changes into $q' = q + q_0$. For such quantities, all the properties do not change if we simply change this starting point, i.e., if we replace each value q by a shifted value $q + q_0$.

Comment. Strictly speaking, according to cosmology, there is the absolute starting point for measuring time: namely, the time of the Big Bang. However, for most practical applications, the physical equations remains the same if we simply change a starting point for time.

Ideally, membership functions should reflect this symmetry. For shift-invariant quantities, it is desirable that our selection of the corresponding membership functions $\mu(q)$ reflect the corresponding shift-invariance.

How to formalize this idea. A seemingly natural idea is to require that each membership function is shift-invariant, i.e., that $\mu(q) = \mu(q + q_0)$ for all q and q_0. However, the only membership functions $\mu(q)$ which satisfies this condition are constant functions – and such functions do not carry any knowledge about the quantity q.

Since we cannot require that a *single* membership function is shift-invariant, it is reasonable to require that a collection of *several* membership functions is shift-invariant.

The idea of multiple membership functions is natural in applications of fuzzy techniques: we usually have several rules containing different membership functions $\mu_i(q)$ [12,25]. If we want to predict the values of a quantity q, then, in some versions of fuzzy system modeling and fuzzy control, we first generate an appropriate linear combination $\sum c_i \cdot \mu_i(q)$ of these membership functions.

Thus, since we cannot require that each membership function $\mu_i(q)$ is shift-invariant, we can require that the set of all such linear combinations is shift-invariant.

From the mathematical viewpoint, this set of all linear combinations is closed under linear combination and thus, forms a *linear space* of functions. Thus, we arrive at the following definition.

Definition 2.1.

- *By a* finite-dimensional linear space of functions *(or simply* linear space*, for short), we mean the class of all functions of the type* $\sum_{i=1}^{n} c_i \cdot \mu_i(q)$*, where:*
  - $n \geq 1$,
  - *differentiable functions* $\mu_1(q)$, ..., $\mu_n(x)$ *are fixed (and assumed to be linearly independent), and*
  - *the coefficients* $c_1, \ldots, c_n$ *can take any real values.*
- *We say that a linear space* L *is* shift-invariant *if for every function* $f(q)$ *from the space* L *and for every real number* q_0*, the function* $f(q + q_0)$ *also belongs to the class* L*.*

The main objective of this paper is to describe all fuzzy-related shift-invariant linear spaces of functions. This description can be simplified if we take into account that if we have two disjoint linear spaces L_1 and L_2 each of which is shift-invariant, then the set L of linear combinations of all functions from L_1 and L_2 is also shift-invariant. Thus, to describe all shift-invariant families, it is sufficient to describe all *basic* families, i.e., all linear spaces which cannot be decomposed into smaller spaces L_1 and L_2.

Definition 2.2.

- *If* L_1 *and* L_2 *are two linear spaces, then their* linear envelope *is a space of all functions of the type* $f_1(q) + f_2(q)$*, where* $f_1(q) \in L_1$ *and* $f_2(q) \in L_2$*.*
- *We say that a shift-invariant linear space* L *is* basic *if it cannot be represented as a linear envelope of two shift-invariant linear spaces* L_1 *and* L_2*.*

We are not just interested in general functions, we are interested in membership functions, i.e., in functions whose values are from the interval $[0, 1]$. Let us therefore make the following additional requirement that these values are from the interval $[0, 1]$ – at least when q is non-negative.

We also want to exclude the trivial membership function $\mu(q) \equiv 1$ for all q.

Definition 2.3. *We say that a linear space of functions* L *is* fuzzy-related *if the following two conditions hold:*

- L *is the set of all linear combinations of functions* $\mu_1(q)$, ..., $\mu_n(q)$ *for each of which* $\mu_i(q) \in [0, 1]$ *for all* $q \geq 0$*.*
- L *does not include the constant functions* $f(q) \equiv 1$ *for all* q*.*

Proposition 2.1. [19,23] *Each basic shift-invariant fuzzy-related linear space* L *is a linear combination of functions* $\mu_i(q) = q^{i-1} \cdot \exp(-\lambda \cdot q)$*,* $i = 1, \ldots, n$*, for some* $\lambda > 0$*.*

Comment. For reader's convenience, all the proofs are placed in the special Appendix.

Discussion. These are exactly the functions that we obtained by using the equations of chemical kinetics!

3. Fuzzy Techniques as an Approximation to Spatial Discreteness

Discrete data. How can we find the meaning of imprecise ("fuzzy") words like "small", "medium", and "large" – e.g., with respect to size? One way is to ask the experts' opinion about known objects of different size. Even when each expert classifies each object exactly into one of the three categories, we get a list of values marked by experts as "small", list of values corresponding to "medium", and list of values corresponding to "large".

Need for a continuous approximation. The more labelled values we collect, the more accurate is our description of the experts' opinions. On the other hand, the more values we collect, the more time-consuming is to process all these values. To speed up the corresponding computations, it is therefore reasonable to develop a simpler-to-process continuous approximation to this discrete data.

A natural continuous description of discrete data. If we had an expert opinion about each value q, we would then have several objects of size q (maybe none) classified as small, several as medium, and several as large. It would then make sense to take the proportion of the objects classified as small as a degree to which q is small.

In practice, we only have expert's opinions about finitely many objects. In such a situation, to estimate the degree to which a given value q is small, we cannot

simply use objects of size exactly q, we also need to use objects of close size. Thus, we can fix $\varepsilon > 0$, and, as the degree $\mu(q)$ that q is small, we can take the ratio $\frac{m}{n}$, where n is the total number of labelled values q' within an ε-vicinity $[q-\varepsilon, q+\varepsilon]$ of q, and m is the total number of those values from this vicinity that are labelled as small.

Instead of treating all values within the vicinity equally, we can, alternatively, set up weights $w(d)$ depending on the distance $d = |q'-q|$, and estimate $\mu(q)$ as the ratio $\frac{\sum\limits_{\text{small}} w(|q'-q|)}{\sum w(|q'-q|)}$, where in the denominator, we take all the labelled values q', while in the numerator, we only take values classified as small.

Comment. The more labelled values we collect, the more accurate our description of the expert opinion. Thus, the "true" description of the expert opinion can be viewed as a limit, when the overall number of values tends to infinity, both overall and in each subinterval, and, correspondingly, the size ε of the neighborhood tends to 0. From this viewpoint, we can view fuzzy sets as *limits* of crisp sets [16] – as opposed to a usual view of fuzzy sets as *generalizations* of crisp sets.

4. How This Interpretation Helps to Understand Fuzzy Operations

Need for fuzzy operations. One of the main motivations behind fuzzy and other non-probabilistic uncertainty is that the traditional probability theory is sometimes not very adequate for describing uncertainty. Traditional probabilistic description of uncertainty is based on *additive* probability measures. To describe non-probabilistic uncertainty, it is therefore reasonable to consider *non-additive* measures.

For an additive measure, if two sets A and B do not have common elements, we have $\mu(A \cup B) = \mu(A) + \mu(B)$. Thus, if we know $\mu(A)$ and $\mu(B)$, then we can determine $\mu(A \cup B)$. In logical terms, this means that if know the probability of A and the probability of B, then we can estimate the probability of disjunction $A \vee B$.

For general non-additive measures, we can no longer use addition to estimate the value $\mu(A \cup B)$. So, how can we estimate this value? We need an operation that, given our degrees of belief $\mu(A)$ and $\mu(B)$ in statements A and B, returns a good estimate for our degree of belief in $A \vee B$. In fuzzy logic, such an operation is known as an *"or"-operation*, or, alternatively as a *t-conorm*.

Often, the $\max$ *operation is used.* Often, the $\max$ t-conorm is used, in which we estimate our degree of belief in $A \vee B$ as $\max(\mu(A), \mu(B))$.

Sometimes, these estimates are exact. Non-additive measures for which this estimate is always exact are known as *possibility* measures [7,10,22,41]. For such measures, we have $\mu(A \cup B) = \max(\mu(A), \mu(B))$ for all sets A and B.

Why $\max$ *operations?* In many other situations, however, we have non-additive measures which are different from possibility ones. So why are possibility measures so successful?

In this section, we use the approximation interpretation of fuzzy techniques to answer this question. Specifically, we show that possibility measures are, in some reasonable sense, universal approximators: for every $\varepsilon > 0$, every non-additive measure which satisfies a certain reasonable boundedness property is equivalent to a measure which is ε-close to a possibility measure.

To formulate this result, we need to describe what is a general measure, which are the reasonable properties, what we mean by equivalence, and what we mean by an "almost" possibility measure. Let us define these notions one by one.

Definition 4.1. *Let X be a set called a* universal set. *By an* algebra of sets *(or* algebra*, for short) $\mathcal{A}$, we mean a non-empty class of subsets $A \subseteq X$ which is closed under complement and union and intersection, i.e.:*

- *if $A \in \mathcal{A}$, then its complement $-A$ also belongs to $\mathcal{A}$;*
- *if $A \in \mathcal{A}$ and $B \in \mathcal{A}$, then $A \cup B \in \mathcal{A}$;*
- *if $A \in \mathcal{A}$ and $B \in \mathcal{A}$, then $A \cap B \in \mathcal{A}$.*

Definition 4.2. *By a* non-additive measure *on the set X, we mean a function μ which assigns, to some subsets $A \subseteq X$ from a certain algebra $\mathcal{A}$, a real number $\mu(A) \geq 0$.*

Motivations for the definition of reasonable (r-) boundedness. In general, a measure $\mu(A)$ describes how important is the set A: the larger the measure, the more important is the set A.

From this viewpoint, if we take the union $A \cup B$ of two sets of bounded size, then the size of the union cannot be arbitrarily large, it should be limited by some bound depending on the bound on $\mu(A)$ and $\mu(B)$.

Similarly, if the sizes of A and B are sufficiently small, then the size of the union is also small – i.e., for sufficiently small bounds on $\mu(A)$ and $\mu(B)$ it should be smaller than any given size.

In precise terms, we arrive at the following definition.

Definition 4.3. *We say that a non-additive measure μ is* r-bounded *if it satisfies the following two properties:*

- *for every $\Gamma > 0$, there exists a $\Delta > 0$ such that if $\mu(A) \le \Gamma$ and $\mu(B) \le \Gamma$, then $\mu(A \cup B) \le \Delta$;*
- *for every $\eta > 0$, there exists a $\nu > 0$ such that if $\mu(A) \le \nu$ and $\mu(B) \le \nu$ then $\mu(A \cup B) \le \eta$.*

Comment. One can easily see that every additive measure is r-bounded. Indeed, it is known all additive measures have the property $\mu(A \cup B) \le \mu(A) + \mu(B)$. Thus, we can take $\Delta = 2 \cdot \Gamma$ and $\nu = \dfrac{\eta}{2}$.

Motivation for the definition of equivalence. The numerical values of probabilities have observable sense – the probability of an event E can be defined as the limit of the frequency with which the event E occurs. This is how, e.g., we can define the probability of a rain at a certain location: by dividing the number of days when it rained by the total number of days for which we had observations.

In contrast, e.g., possibility values do not have direct meaning, the only important thing is which values are larger and which are smaller – this describes which events are more possible and which are less possible. From this viewpoint, if two measures can be obtained from each other by a transformation that preserves the order, such measures can be considered to be equivalent.

Definition 4.4. *Two non-additive measures $\mu(A)$ and $\mu'(A)$ are called* equivalent *if there exists a 1-1 monotonic function $f(x)$ such that for every set A, we have $\mu'(A) = f(\mu(A))$.*

Definition 4.5. *Let $\varepsilon > 0$ be a real number. We will call a non-additive measure $\mu(A)$ an* ε-possibility measure *if for every two sets A and B, we have $\mu(A \cup B) \le (1 + \varepsilon) \cdot \max(\mu(A), \mu(B))$.*

Comment. It is reasonable to consider *monotonic* measures, for which $A \subseteq B$ implies $\mu(A) \le \mu(B)$ [2,37]. For monotonic measures, due to $A \subseteq A \cup B$ and $B \subseteq A \cup B$, we have $\mu(A) \le \mu(A \cup B)$ and $\mu(B) \le \mu(A \cup B)$, thus, $\max(\mu(A), \mu(B)) \le \mu(A \cup B)$. So, if μ is a monotonic ε-probabilistic measure, we have

$$\max(\mu(A), \mu(B)) \le \mu(A \cup B) \le (1 + \varepsilon) \cdot \max(\mu(A), \mu(B)).$$

Thus, the value $\mu(A \cup B)$ is almost equal – with relative accuracy $\varepsilon > 0$ – to the value $\max(\mu(A), \mu(B))$ corresponding to the possibility measure.

Proposition 4.1. *For every $\varepsilon > 0$, every r-bounded non-additive measure is equivalent to an ε-possibility measure.*

Discussion. Sometimes, we have several measures. A natural question is: can we re-scale all of them by using the same re-scaling function $f(x)$ so that all of them become ε-possibility measures?

Proposition 4.2. *For every $\varepsilon > 0$, and for every finite set of r-bounded non-additive measures $\mu_1(A)$, ..., $\mu_n(A)$, there exists a 1-1 function $f(x)$ for which all n measures $\mu_i'(A) \stackrel{\text{def}}{=} f(\mu_i(A))$ are ε-possibility measure.*

5. First Promising Future Direction: Interval-Valued Fuzzy Sets

Let us start the analysis of future directions. Let us analyze how the approximation interpretation of fuzzy techniques can help us predict promising future directions of fuzzy research and applications. Let us start with the idea of fuzzy as approximating spatial discreteness, as described by Section 3.

Fuzzy as an approximation of spatial discreteness: reminder. In this interpretation, for each value q, the actual (final) fuzzy value $\mu(q)$ is a limit of the corresponding ratios when the number N of labelled data points tends to ∞ and $\varepsilon \to 0$.

Measuring fuzzy degrees and measuring physical quantities: similarities and differences. At first glance, the situation is similar to measuring a physical quantity: the more measurements we perform and the more accurate the corresponding measurements, the closer we are to the actual value of the corresponding physical quantity. In the measurement case, we thus always have convergence.

However, in the fuzzy case, there is a big difference: in contrast to measurement of physical quantities, when there is such a thing as "actual value", there is no such thing as the "actual membership degree". As a result, there is no guarantees that the sequence of ratios will converge to a single limit value. Instead, this sequence of ratios may have several limit points.

Describing the set of limit points. A natural way to describe $\mu(x)$ is to first tend N to infinity, and then tend ε to 0. In this case, the next ratio is different from the previous one when we add a new labelled value; in this case, we move from the previous ratio $\frac{m}{n}$ to either $\frac{m}{n}$ or $\frac{m+1}{n+1}$. In both cases, the difference between the new and the old ratios does not exceed $\frac{1}{n}$. In our description of the limit process, we assumed that the number of labelled values n in the interval $[q-\varepsilon, q+\varepsilon]$ tends to ∞, so this difference between the two sequential elements of the sequence tends to 0.

One can prove that if we have a sequence $r_n \in [0,1]$ for which $|r_n - r_{n+1}| \to 0$, then the set of limit points of this sequence forms a subinterval of the interval $[0,1]$. Thus, we arrive at the following conclusion.

Need to consider interval-valued fuzzy sets. It is reasonable to consider fuzzy sets in which, for each value q, the expert's degree of confidence $\mu(q)$ is not a number, but an interval – a subinterval of the interval $[0,1]$.

Interval-valued fuzzy sets have indeed been efficiently used. This is the least controversial of our predictions, since interval-valued fuzzy sets have been indeed successfully used in many practical applications; see, e.g., [20,21].

6. Second Promising Future Direction: Computing with Words

Need for computing with words. The main idea behind fuzzy techniques is that to describe a quantity such as temperature, height, etc., we use words such as "small", "medium", "high", etc. However, if we only use the selected words, we get a rather crude description of the quantity.

To get a more accurate description, we can say, e.g., that someone is rather short, but closer to medium height. In this case, to a large extent, this person is short, but to some degree, this person is of medium height.

In general, an accurate description of the quantity may include not just one word, but several words, with degrees associated with different words. A similar description can be used to describe situations when we are uncertain: e.g., if we do not know whether a person is short, medium, or tall, we assign, e.g., the same degree 1 to the possible degree to which this person is short, medium, and tall.

Once we fix the words $w_1, \ldots, w_n$, each quantity is then represented by the corresponding tuple of degrees $d = (d_1, \ldots, d_n)$.

Need for data processing. In many practical situations, we are interested in the value of a real-valued quantity y which is difficult (or even impossible) to measure or estimate directly. For example, we are often interested in the future value y of some quantity (e.g., tomorrow's weather), and it is not possible to directly measure a future value of a quantity.

In such situations, a usual approach is:

- to estimate easier-to-estimate auxiliary real-valued quantities $x_1, \ldots, x_m$ which are related to y by a known dependence $y = f(x_1, \ldots, x_m)$, and
- to use the estimates of x_i to compute the estimate for y.

This computation of y based on $x_1, \ldots, x_m$ (or, to be more precise, computation of an estimate for y from the estimates for x_i) is known as *data processing*.

When the estimates for x_i are given in the form of tuples d, we face the following problem:

- we know the tuples $d^{(j)} = \left(d_1^{(j)}, \ldots, d_n^{(j)}\right)$ which describes our knowledge about each input x_j, $1 \le j \le m$;
- we want to describe the resulting knowledge about y in a similar tuple form.

In particular, in the simplest case when we have only two inputs x_1 and x_2, and data processing consists of applying a simple arithmetic operation, e.g., $f(x_1, x_2) = x_1 + x_2$, $f(x_1, x_2) = x_1 - x_2$, or $f(x_1, x_2) = x_1 \cdot x_2$, we face the following problem:

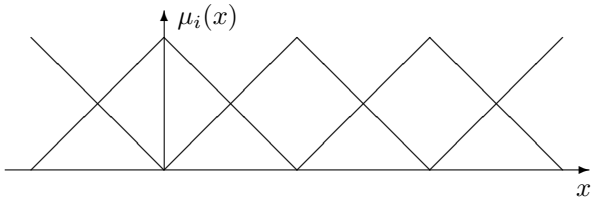
- if we have one quantity x_1 which is characterized by the tuple $d^{(1)}$ and
- we have another quantity x_2 which is characterized by the tuple $d^{(2)}$,
- then we should be able to produce a tuple characterizing the sum $x_1 + x_2$ of these quantities, a tuple characterizing their difference $x_1 - x_2$, their product $x_1 \cdot x_2$, etc.

In general, instead of computing with numbers, we should be able to compute with words. The need for such computing with words was first emphasized by L. Zadeh; see, e.g., [42].

How to represent the original words. Usually, the membership functions are triangular, i.e., first linearly increase from 0 to 1, then linearly decrease from 1 to

0 at the same rate. Different functions $\mu_i(x)$ usually differ by a shift, so for some some starting point s and step h, we have

$$\mu_i(x) = \max\left(0, 1 - \frac{|x - (s + i \cdot h)|}{h}\right). \quad (6.1)$$



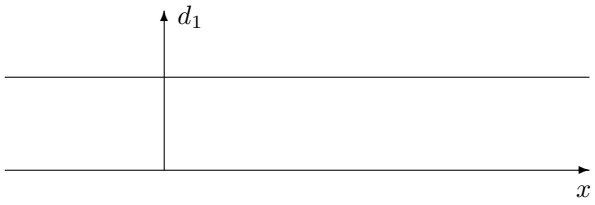
Comment. For example, these membership functions can describe the usual terms "negligible", "small", "medium", etc.

From a words-related tuple representation to a membership function. A tuple $d = (d_1, \ldots, d_n)$ represents a value x if one of the following conditions hold:

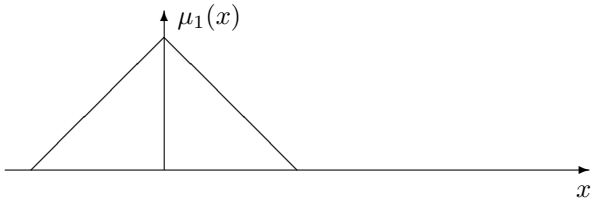
- the corresponding quantity q is characterized by the first word w_1, and x satisfies the property described by this word;
- the corresponding quantity q is characterized by the second word w_2, and x satisfies the property described by this word;
- ...
- the corresponding quantity q is characterized by the n-th word w_n, and x satisfies the property described by this word.

Here:

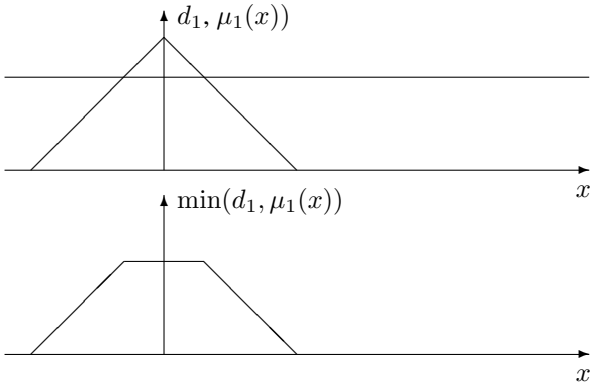
- the degree to which q is characterized by the first word is d_1, and



- the degree to which x satisfies the property described by this word is $\mu_1(x)$.

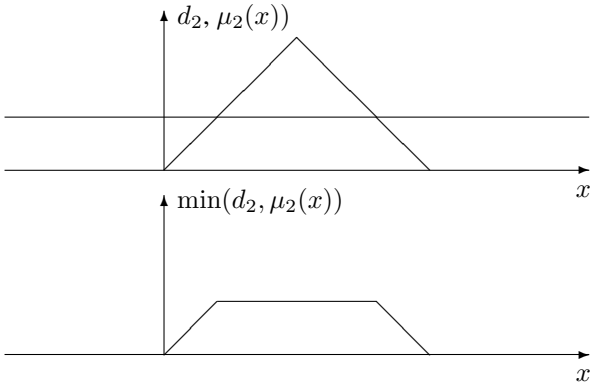


From the computational viewpoint, the simplest way to describe "and" in fuzzy logic is to use minimum: out of all t-norms, only $a \cdot b$ and $\min$ can be implemented by a single arithmetic operation, and computing $\min$ requires fewer bit operations and is, thus, faster. So, the degree to which q is characterized by the first word w_1 *and* x satisfies the property described by this word can be described as $\min(d_1, \mu_1(x))$.



Similarly:

- the degree to which q is characterized by the second word w_2 *and* x satisfies the property described by this word can be described as $\min(d_2, \mu_2(x))$;

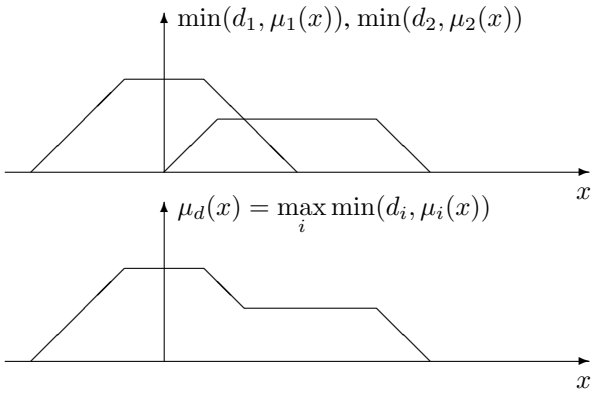


- ...
- the degree to which q is characterized by the n-th word w_n *and* x satisfies the property described by this word can be described as $\min(d_n, \mu_n(x))$.

The simplest way to describe "or" in fuzzy logic is to use maximum, so the degree to which one of these conditions is satisfied is equal to

$$\mu_d(x) =$$

$$\max(\min(d_1, \mu_1(x)), \ldots, \min(d_n, \mu_n(x))). \quad (6.2)$$



Resulting fuzzy-based formalization of computing with words: a natural idea. We know the tuples $d^{(1)}$ and $d^{(2)}$ describing the two quantities x_1 and x_2, and we want to find a tuple d corresponding to the value $y = f(x_1, x_2)$ for some known function $f(x_1, x_2)$. A natural idea to do it is as follows:

- first, we use the formula (1) to generate membership functions $\mu^{(1)}(x_1)$ and $\mu^{(2)}(x_2)$ corresponding to the tuples $d^{(1)}$ and $d^{(2)}$;
- then, by applying Zadeh's extension principle to these membership functions, we compute the membership function $\mu(x)$ corresponding to $y = f(x_1, x_2)$;
- finally, we need to generate the tuple d corresponding to the resulting membership function $\mu(x)$.

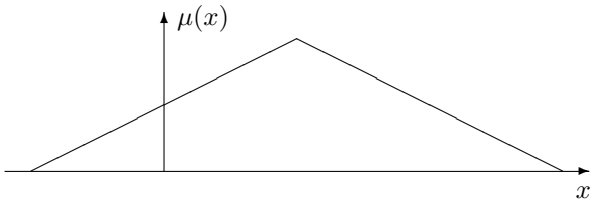
What is necessary to implement the above idea. To implement the above idea, we need to be able to generate a tuple corresponding to a given membership function.

How to transform a membership function into a tuple: a seemingly natural idea. At first glance, we have a natural way of computing the degree d_i to which it is possible that a quantity described by the new membership function $\mu(x)$ satisfies the property described by the word w_i. There are several possible value x, so the quantity corresponds to w_i if one of the following statements hold:

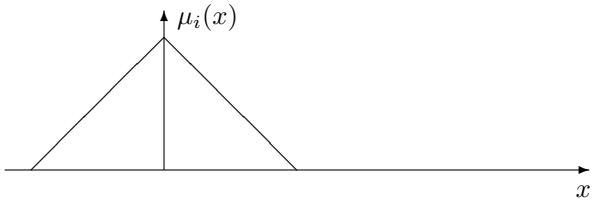
- for one possible value x, this value is in agreement with the new membership function and with the word w_i;
- for another possible value x, this value is in agreement with the new membership function and with the word w_i;
- ...

For each number x:

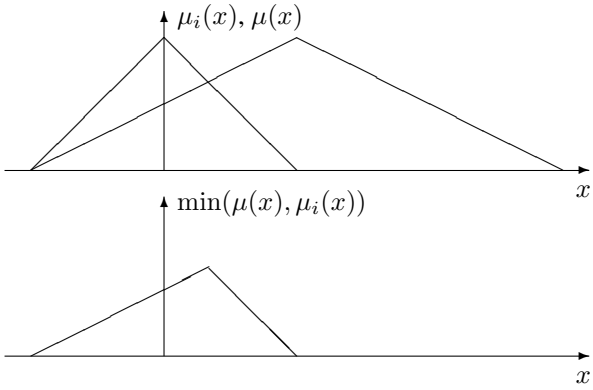
- the degree to which this number is in agreement with the new membership function is equal to $\mu(x)$,



- the degree to which this number is consistent with the word w_i is equal to $\mu_i(x)$.



Thus, the degree to which this number x is in agreement with the new membership function *and* in agreement with the word w_i is equal to $\min(\mu(x), \mu_i(x))$:



The degree to which one of the conditions corresponding to different value x hold is equal to the maximum of all such values, i.e., to

$$A_i \stackrel{\text{def}}{=} \max_x(\min(\mu(x), \mu_i(x))). \qquad (6.3)$$

Comment. In deriving the formula (6.3), we used $\max$ for "or" and $\min$ for "and". If we use sum for "or" and product for "and", we get F-transform instead of the formula (2); see, e.g., [11,26,27,29,30].

How to transform a membership function into a tuple: a reasonable requirement. It is reasonable to require that:

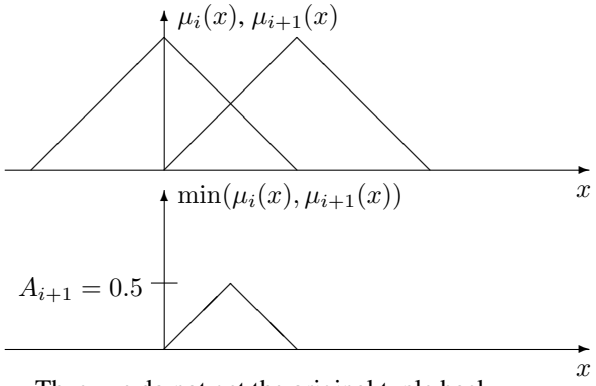
- if we start with the word w_i itself, i.e., with the tuple $d = (0, \ldots, 0, 1, 0, \ldots, 0)$ in which $d_i = 1$ and $d_j = 0$ for all $j \neq i$, and

– we perform no operations at all, i.e., use the function $f(x) = x$,
– then we would like to get back the same tuple

$$d = (0, \ldots, 0, 1, 0, \ldots, 0).$$

How to transform a membership function into a tuple: limitations of a seemingly natural idea. Unfortunately, the above seemingly reasonable approach does not satisfy this requirement. Specifically, when we apply this approach to the word w_i, then:

– for this index i, we do get $A_i = 1 = d_i$;
– however, instead of the desired $d_{i-1} = d_{i+1} = 0$, we get $A_{i-1} = A_{i+1} = 0.5$.



Thus, we do not get the original tuple back.

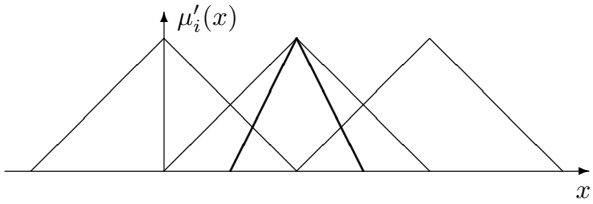
What is needed. We therefore need to modify the above seemingly natural procedure, to make sure that it returns the original tuple if no operation is performed.

How can we do this? The main problem with the above idea is that the neighboring membership functions overlap. This leads us to the following natural idea.

Since intersection is a problem, let us remove the intersecting parts from the membership function before applying a formula of the above type. We still want to compute the i-th coefficient d_i corresponding to the membership function $\mu(x)$ by comparing this membership function with the membership function $\mu_i(x)$ representing the i-th word w_i. However, instead of directly comparing these functions, let us first cut off, from both of them, parts intersecting with the neighboring membership functions. In other words, we first compute "reduced" functions

$$\mu_i'(x) =$$

$$\max(0, \mu_i(x) - \max(\mu_{i-1}(x), \mu_{i+1}(x))) \qquad (6.4)$$



and

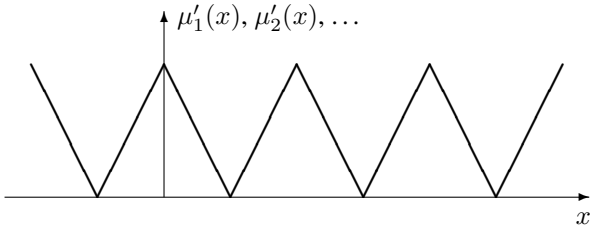
$$\mu'(x) =$$

$$\max(0, \mu(x) - \max(\mu_{i-1}(x), \mu_{i+1}(x))), \qquad (6.5)$$

and then compute the degrees based on these reduced functions, as

$$A_i = \max_x(\min(\mu'(x), \mu_i'(x))). \qquad (6.6)$$

Here, the reduced functions $\mu_i'(x)$ no longer overlap:



Thus, we arrive at the following definitions.

Definition 6.1.

– *Let $\mu_i(x)$ be a sequence of membership functions described by the formula* (1)*, and*
– *let $\mu(x)$ be a membership function.*

By a tuple $d = (d_1, \ldots, d_n)$ *corresponding to the membership function $\mu(x)$, means a sequence of values obtained by using formulas* (6.4)–(6.6).

The following easy-to-prove result shows that this modification of the original formula (6.3) indeed enables us to reconstruct the original degrees d_i:

Proposition 6.1.

– *Let $\mu_i(x)$ be a sequence of triangular functions* (6.1)*,*
– *let $d = (d_1, \ldots, d_n)$ be a tuple of a numbers $d_i \in [0, 1]$,*
– *and let $\mu_d(x)$ be described by the formula* (6.2)*.*

For this function $\mu_d(x)$, formulas (6.4)–(6.6) *lead to $A_i = d_i$ for all i.*

Comment. As we can see from the proof, Proposition 6.1 is valid not only for the triangular functions, but also for any set of membership functions $\mu_i(x)$ for which, for some sequence of values t_i:

- $\mu_i(t_i) = 1$, and
- $\mu_i(x)$ is only different from 0 for $x \in [t_{i-1}, t_{i+1}]$.

Resulting definition of an operation with tuples. Now that we know that our idea enables us to recover the original degrees, we can formalize what it means to perform computations with words.

Definition 6.2.

- *Let $\mu_i(x)$ be a family of membership functions described by the formula* (1)*,*
- *let $y = f(x_1, \ldots, x_m)$ be a function of n real numbers,*
- *and let $d^{(j)} = (d_1^{(j)}, \ldots, d_n^{(j)})$, $j = 1, \ldots, m$, be tuples.*

By the result $f\left(d^{(1)}, \ldots, d^{(m)}\right)$ of applying the function $f(x_1, \ldots, x_m)$ to the tuples $d^{(1)}, \ldots, d^{(m)}$, *we mean the tuple obtained by using the following three-stage procedure:*

- *first, we use the formula* (6.2) *to compute the membership functions $\mu^{(j)}(x_j)$ corresponding to the given tuples*

$$d^{(1)}, \ldots, d^{(m)};$$

- *then, we apply Zadeh's extension principle to the membership functions $\mu^{(j)}(x_j)$, producing a new membership function $\mu(y)$ defined as*

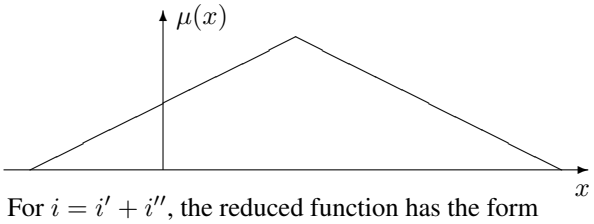
$$\sup_{x_i : f(x_1, \ldots, x_m) = y} \min\left(\mu^{(1)}(x_1), \ldots, \mu^{(m)}(x_m)\right);$$

- *finally, we use the formulas* (6.4)–(6.6) *to transform the resulting membership function $\mu(y)$ into a tuple d.*

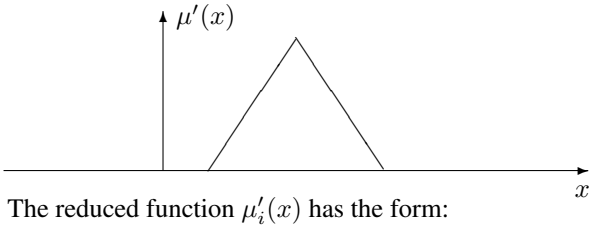
Examples. Let us consider triangular membership functions starting with $s = 0$. Each original word w_i is described by a tuple $d = (d_1, \ldots, d_n)$ in which $d_i = 1$ and $d_j = 0$ for all $j \neq i$.

Example 6.1. Let us first consider the case when we add two words $w_{i'}$ and $w_{i''}$ corresponding to tuples $(0, \ldots, 0, 1, 0, \ldots, 0)$ (with 1 on the i-th place) and $(0, \ldots, 0, 1, 0, \ldots, 0)$ (with 1 on the i'-th place). In this case, we will show that we get a tuple with $d_{i'+i''} = 1$, $d_{(i'+i'')-1} = d_{(i'+i'')+1} = 0.5$, and $d_j = 0$ for all other j, i.e., a tuple $(0, \ldots, 0, 0.5, 1, 0.5, 0, \ldots, 0)$.

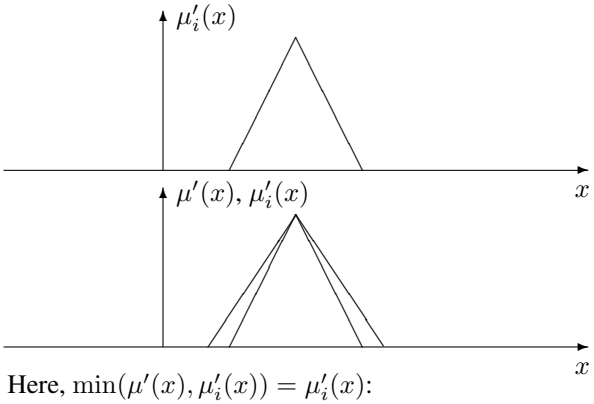
Indeed, here, Zadeh's extension principle leads to the following membership function:



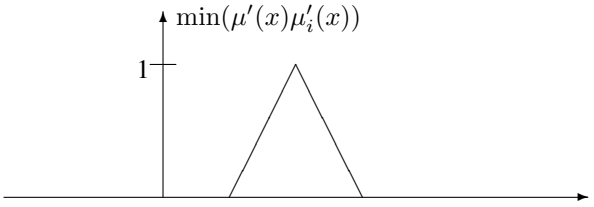
For $i = i' + i''$, the reduced function has the form



The reduced function $\mu_i'(x)$ has the form:



Here, $\min(\mu'(x), \mu_i'(x)) = \mu_i'(x)$:



Thus, the maximum $d_i = \max\limits_x \min(\mu'(x), \mu_i'(x))$ is equal to 1.

Similarly, for $i = i'+i''+1$ and for for $i = i'+i''-1$, we get $d_i = \max\limits_x \min(\mu'(x), \mu_i'(x)) = 0.5$, and we get $d_i = 0$ for all other i.

Example 6.2. For subtracting two words $w_{i'}$ (corresponding to the tuple $(0, \ldots, 0, 1, 0, \ldots, 0)$) and $w_{i''}$ (corresponding to the tuple $(0, \ldots, 0, 1, 0, \ldots, 0)$), we similarly get a tuple with $d_{i'-i''} = 1$ and $d_{(i'-i'')-1} = d_{(i'-i'')+1} = 0.5$, i.e., a tuple

$$(0, \ldots, 0, 0.5, 1, 0.5, 0, \ldots, 0).$$

Example 6.3. A shift of a word w_i as described by the tuple $(0, \ldots, 0, 1, \ldots, 0)$, i.e., the result of applying a function $f(x) = x + a \cdot h$ with $0 < a < 1$ to the word w_i, leads to $d_i = 1 - a$ and $d_{i+1} = a$, i.e., to the tuple $(0, \ldots, 0, 1 - a, a, 0, \ldots, 0)$.

Need for extending these results. In all three examples, we get reasonable results. The fact that we get reasonable results for simple examples shows that this approach is worth pursuing. To make this approach useful, we need to come up with similar explicit formulas for the result of applying other functions $f(x_1, \ldots, x_m)$ to tuples.

7. Third Promising Future Direction: Complex-Valued Fuzzy Sets

Ideal case. In the ideal case, we know the expert's degrees of belief $d_1, \ldots, d_n$ in the basic statements $S_1, \ldots, S_n$, and we use "and"- and "or"-operations to estimate the expert's degree of confidence in different propositional combinations of the basic statements.

In practice, the situation may be somewhat more complicated. Sometimes, instead of knowing the expert's degree of belief in the basic statements, we only know the expert's degree of belief in some propositional combinations of the basic statements. In this case:

- first, we need to recover the degrees $d_1, \ldots, d_n$ from the available information;
- then, we use the recovered values $d_1, \ldots, d_n$ to estimate the expert's degree of belief in other propositional combinations.

If "and"- and "or"-operations were exact, this procedure would always succeed. In the ideal case, when the expert's degree of belief in $A \,\&\, B$ is exactly equal to $f_{\&}(d(A), d(B))$ and the expert's degree of belief in $A \vee B$ is exactly equal to $f_{\vee}(d(A), d(B))$, we can indeed recover the desired degrees by solving the corresponding system of equations.

Example. Suppose that we use the algebraic product $f_{\&}(a, b) = a \cdot b$ as an "and"-operation and

$$f_{\vee}(a, b) = a + b - a \cdot b$$

as an "or"-operation. Suppose that instead of the actual values $d_1 = d(S_1)$ and $d_2 = d(S_2)$ we only know the degrees

$$d(S_1 \,\&\, S_2) = f_{\&}(d_1, d_2) = d_1 \cdot d_2$$

and

$$d(S_1 \vee S_2) = f_{\vee}(d_1, d_2) = d_1 + d_2 - d_1 \cdot d_2.$$

In particular, if we actual (unknown) values of d_1 and d_2 are $d_1 = 0.4$ and $d_2 = 0.6$, then

$$d(S_1 \,\&\, S_2) = 0.4 \cdot 0.6 = 0.24$$

and

$$d(S_1 \vee S_2) = 0.6 + 0.4 - 0.6 \cdot 0.4 = 0.76.$$

These two numbers

$$d(S_1 \,\&\, S_2) = 0.24 \text{ and } d(S_1 \vee S_2) = 0.76$$

are the only information that we have about the expert's degrees d_1 and d_2. Based on these numbers, we want to recover the values d_1 and d_2.

Since we assumed that the t-norm $a \cdot b$ and the t-conorm $a + b - a \cdot b$ describe the expert's belief in composite statements, we form two equations for the two unknowns d_1 and d_2: $d_1 \cdot d_2 = 0.24$ and $d_1 + d_2 - d_1 \cdot d_2 = 0.76$. After adding these two equations, we get $d_1 + d_2 = 1$, hence $d_2 = 1 - d_1$. Substituting $d_2 = 1 - d_1$ into the first equation, we get

$$d_1 \cdot (1 - d_1) = 0.24.$$

After opening parentheses and moving all the terms to the right-hand side, we get the equation

$$d_1^2 - d_1 + 0.24 = 0.$$

By using the known formula for solving quadratic equations, we get

$$d_1 = \frac{1}{2} \pm \sqrt{\left(\frac{1}{2}\right)^2 - 0.24} =$$

$$0.5 \pm \sqrt{0.25 - 0.24} = 0.5 \pm \sqrt{0.01} = 0.5 \pm 0.1.$$

Thus, $d_1 = 0.4$ or $d_1 = 0.6$, i.e., (almost) exactly the expert's original estimates.

In this case, due to symmetry, we cannot distinguish between d_1 and d_2, but we can make this distinction if we have additional information.

What happens in practical cases, when the "and"- and "or"-operations are only approximate? Let us now analyze what will happen if we take into account that in reality, "and"- and "or"-operations provide only an approximate description of the expert's degrees of belief. As an example, let us assume that in general, the expert's reasoning is best described by the same "and"- and "or"-operations $f_{\&}(a,b) = a \cdot b$ and $f_{\vee}(a,b) = a + b - a \cdot b$. The fact that these operations are the best "on average" does not necessarily mean that these operations always exactly describe the expert's degree of belief in composite statements.

For example, as we have mentioned earlier, if the statements S_1 and S_2 coincide, then

$$d(S_1 \,\&\, S_2) = d(S_1 \vee S_2) = d(S_1).$$

For such two statements with $d(S_1) = d(S_2) = 0.5$, we will get $d(S_1 \,\&\, S_2) = 0.5$ and $d(S_1 \vee S_2) = 0.5$.

Let us see what happens if we try to apply, to these two values $d(A \,\&\, B) = 0.5$ and $d(A \vee B) = 0.5$, the above procedure of reconstructing d_1 and d_2. Specifically, we form two equations: $d_1 \cdot d_2 = 0.5$ and

$$d_1 + d_2 - d_1 \cdot d_2 = 0.5,$$

and we try to find d_1 and d_2 by solving this system of two equations. After adding the two equations, we get $d_1 + d_2 = 1$ and thus, $d_2 = 1 - d_1$. Substituting $d_2 = 1 - d_1$ into the first equation, we get

$$d_1 \cdot (1 - d_1) = 0.5.$$

After opening parentheses and moving all the terms to the right-hand side, we get the equation

$$d_1^2 - d_1 + 0.5 = 0.$$

The determinant of this equation is negative

$$(-1)^2 - 4 \cdot 1 \cdot 0.5 = 1 - 2 = -1 < 0$$

and thus, this equation does not have any real solution – and hence, no solutions with $d_1 \in [0,1]$.

Natural idea leads to complex-valued degrees. Since we cannot get the degrees from the interval $[0,1]$, a natural idea is to extend real numbers so that the corresponding equation (or system of equations) has a solution. In principle, we could get any solutions, so it is desirable to make sure that all (or at least almost all) equations (and systems of equations) have a solution. The need to consider quadratic equations immediately leads to the appearance of the imaginary unit $\mathrm{i} = \sqrt{-1}$, which is a solution of the equation $x^2 + 1 = 0$, and to the appearance of general complex numbers as solutions of generic quadratic equations.

Good news is that nothing else needs to be added to take care of cubic and higher order equations: a so-called main theorem of algebra states that every polynomial equation has a complex-valued solution (unless this equation has the form $c = 0$ with a constant c which is different from 0).

Thus, we arrive at the need to use complex-valued degrees.

How good are complex-valued degrees in practice? Complex-valued fuzzy sets have been indeed successfully used in practical applications; see, e.g., [1,4,6,13, 24] – and our analysis explains why. Let us show, on several examples, that they indeed help to reconstruct the membership values d_i.

Example 7.1. Let us check which complex numbers appear in the above example. By using the known formula for solving quadratic equations, we get

$$d_1 = \frac{1}{2} \pm \sqrt{\left(\frac{1}{2}\right)^2 - 0.5} =$$

$$0.5 \pm \sqrt{0.25 - 0.5} = 0.5 \pm \sqrt{-0.25} = 0.5 \pm 0.5 \cdot \mathrm{i}.$$

Of course, it is difficult to interpret complex-valued degrees (or, for that purpose, any degrees outside the interval $[0,1]$). So, it is natural, for each such complex-valued degree, to take the closest value from the interval $[0,1]$.

For complex numbers, the natural distance is Euclidean distance

$$d(a_1 + a_2 \cdot \mathrm{i}, b_1 + b_2 \cdot \mathrm{i}) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}.$$

It is easy to see that for a complex number $a_1 + a_2 \cdot \mathrm{i}$, the closest point on the real line is its real part a_1, and the closest point on $[0,1]$ is:

- the same value a_1 is $a_1 \in [0,1]$;
- the value 0 is $a_1 < 0$, and
- the value 1 if $a_1 > 1$.

Thus, for the complex numbers $0.5 + 0.5 \cdot \mathrm{i}$ and $0.5 - 0.5 \cdot \mathrm{i}$, the closest numbers from the interval $[0,1]$ are 0.5 and 0.5 – exactly the values that the expert assigned!

Example 7.2. Let us consider a slightly more general example, with the same "and"- and "or"-operations and with $S_1 = S_2$, but this time, with an arbitrary value $d \in [0,1]$ for which $d(S_1) = d(S_2) = d$. In this case, we get $d(S_1 \,\&\, S_2) = d$ and $d(S_1 \vee S_2) = d$. These two values $d(S_1 \,\&\, S_2) = d$ and $d(S_1 \vee S_2) = d$ are all we get from the expert. Based on these two values, we want to reconstruct d_1 and d_2.

In this example, we get a system of equations

$$d_1 \cdot d_2 = d \text{ and } d_1 + d_2 - d_1 \cdot d_2 = d.$$

After adding these two equations, we get $d_1 + d_2 = 2d$, hence $d_2 = 2d - d_1$. Substituting $d_2 = 2d - d_1$ into the first equation, we get $d_1 \cdot (2d - d_1) = d$. After opening parentheses and moving all the terms to the right-hand side, we get the equation $d_1^2 - 2d \cdot d_1 + d = 0$. By using the known formula for solving quadratic equations, we get $d_1 = d \pm \sqrt{d^2 - d}$. Here, $d^2 - d = (d-1) \cdot d \le 0$, so $d_1 = d \pm \sqrt{d - d^2} \cdot \mathrm{i}$. For both complex values $d + \sqrt{d - d^2} \cdot \mathrm{i}$ and $d - \sqrt{d - d^2} \cdot \mathrm{i}$, the closest number from the interval $[0,1]$ is the value d – also exactly what the experts assigned.

Example 7.3: complex numbers are not a panacea. To avoid a false impression that complex numbers also lead to perfect results, let us consider another example in which general "and"- and "or"-operations may not be applicable: an example when S_2 implies S_1. In this case, $S_1 \,\&\, S_2$ is simply equivalent to S_2, and $S_1 \vee S_2$ is equivalent to S_1. So, for example, for $d_1 = 0.6$ and $d_2 = 0.4$, we get $d(S_1 \,\&\, S_2) = 0.4$ and $d(S_1 \vee S_2) = 0.6$. These two values $d(S_1 \,\&\, S_2) = 0.4$ and $d(S_1 \vee S_2) = 0.6$ are all we get from the expert. Based on these two values, we want to reconstruct d_1 and d_2.

In this example, we get a system of equations

$$d_1 \cdot d_2 = 0.4 \text{ and } d_1 + d_2 - d_1 \cdot d_2 = 0.6.$$

After adding these two equations, we get $d_1 + d_2 = 1$, hence $d_2 = 1 - d_1$. Substituting $d_2 = 1 - d_1$ into the first equation, we get $d_1 \cdot (1 - d_1) = 0.4$. After opening parentheses and moving all the terms to the right-hand side, we get the equation $d_1^2 - d_1 + 0.4 = 0$. By using the known formula for solving quadratic equations, we get

$$d_1 = 0.5 \pm \sqrt{0.25 - 0.4} = 0.5 \pm \sqrt{-0.15} =$$

$$0.5 \pm \sqrt{0.15} \cdot \mathrm{i}.$$

For both complex values $0.5 + \sqrt{0.15} \cdot \mathrm{i}$ and $0.5 - \sqrt{0.15} \cdot \mathrm{i}$, the closest number from the interval $[0,1]$ is the value 0.5, which is somewhat different from the original expert values 0.4 and 0.6 (but still rather close to these values).

8. Fourth Promising Future Direction: Dynamic Fuzzy Sets

Variety of t-norms and t-conorms. In fuzzy logic, there are numerous t-norms and t-conorms. Which one to apply depends on the relation between the statements A and B. This dependence can be illustrated in the probabilistic approaches, when degree a represents the probability that A is true (or the probability that a randomly selected expert considers A to be true).

If A and B are independent, then the probability $f_{\&}(a,b)$ of $A \,\&\, B$ is equal to the product

$$a \cdot b = P(A) \cdot P(B)$$

of the corresponding probabilities. In this case, the most adequate t-norm is a product $f_{\&}(a,b) = a \cdot b$.

Of the other hand, if we know that A and B are strongly correlated, then a t-norm $f_{\&}(a,b) = \min(a,b)$ which leads to $P(A \,\&\, B) = P(A) = P(B)$ if $A = B$ is more adequate.

The problem is that in many cases, we do not know whether A and B are correlated or not. In such cases, we select *some* t-norm. The selected t-norm may not necessarily coincide with the ideal one; hence, the resulting recommendations may not be always adequate. This "truth-functionality", the fact that the degree of confidence in $A \,\&\, B$ depends only on the degrees of confidence in A and B – without fully adequately taking into account the possibility of different correlations – if often cited as one of the main limitations of fuzzy techniques.

Towards dynamic fuzzy logic. As we have mentioned, one of the origins of fuzzy logic is a description of processes that changes with time. However, the traditional fuzzy logic assumes that the expert's degrees of confidence do not change. In reality, the expert's opinions often change with time. Thus, to get a more adequate description of the expert opinions and rules, it is necessary to take these changes into account, i.e., to take into account that the expert's degree of confidence in each statement A changes with time. In other words, to describe the expert's opinion about a state-

ment A, instead of a single value $a \in [0,1]$, we need to use a function $a(t)$ that describes how this degree changes with time t. Such *dynamic fuzzy logic* was proposed in [5,31,32,33].

What we do in this section. In this section, we show that, if we take this dynamics into consideration, then we can get a more adequate description of "and" and "or" operations, a description in which it is possible to distinguish between the cases when the statements are independent and when they are strongly dependent.

This possibility will be illustrated on the example when the fuzzy degrees have a probabilistic meaning.

Relation between correlation and the probabilities $P(A\,\&\,B)$ and $P(A\vee B)$: reminder. In statistics, the most frequent way to describe correlation between two random variables x and y is to use the correlation coefficient

$$\rho = \frac{E[x\cdot y] - E[x]\cdot E[y]}{\sqrt{V[x]\cdot V[y]}},$$

where $E[x]$ denote the mean (expected value) of the variable x and the variance $V[x]$ is defined as

$$V[x] \stackrel{\rm def}{=} E[(x-E[x])^2] = E[x^2] - (E[x])^2;$$

see, e.g., [35].

A statement A which is true with probability a and false with the remaining probability $1-a$ can be viewed as a random variable that takes the value 1 (= "true") with probability a and 0 (= "false") with probability $1-a$. For this variable,

$$E[A] = 1\cdot a + 0\cdot(1-a) = a$$

and similarly, $E[B] = b$. Similarly, $E[A\,\&\,B] = P(A\,\&\,B)$.

Here, $A = 0$ or $A = 1$, hence $A^2 = A$, $E[A^2] = E[A]$ and thus, $V[A] = E[A^2] - (E[A])^2 = a - a^2 = a\cdot(1-a)$. Similarly, we can conclude that $V[B] = b\cdot(1-b)$.

For true and false statements, "and" is simply a product, so $A\,\&\,B = A\cdot B$ and thus, $E[A\,\&\,B] = P(A\,\&\,B) = E[A\cdot B]$. Thus, the above formula for the correlation takes the following form:

$$\rho = \frac{P(A\,\&\,B) - a\cdot b}{\sqrt{a\cdot(1-a)\cdot b\cdot(1-b)}}.$$

Once we know the probabilities $P(A) = a$ and $P(B) = b$ and the correlation coefficient, we can uniquely reconstruct the probabilities $P(A\,\&\,B)$ and $P(A\vee B)$. From the above formula, we can conclude that

$$P(A\,\&\,B) =$$

$$a\cdot b + \rho\cdot\sqrt{a\cdot(1-a)\cdot b\cdot(1-b)}. \qquad (8.1)$$

The expression for $P(A\vee B)$ can be found if we take into account the known property

$$P(A\,\&\,B) + P(A\vee B) = P(A) + P(B),$$

from which we conclude that

$$P(A\vee B) = P(A) + P(B) - P(A\,\&\,B) =$$

$$a + b - P(A\,\&\,B),$$

i.e.,

$$P(A\vee B) =$$

$$a + b - a\cdot b - \rho\cdot\sqrt{a\cdot(1-a)\cdot b\cdot(1-b)}. \qquad (8.2)$$

How do we find the correlation coefficient? In the dynamic case, we not only know the current expert's degrees of confidence a and b in statements A and B, we also know the past degrees $a(t)$ and $b(t)$ which were, in general, different from a and b.

When the statements A and B are strongly correlated, then it is reasonable to expect that the corresponding changes $a(t)$ and $b(t)$ are also correlated. If the statements A and B are independent, then it is reasonable to expect that the changes $a(t)$ and $b(t)$ are also independent. In general, to find the correlation coefficient between A and B, we can use, as random variables, the values $a(t)$ and $b(t)$ corresponding to T known moments of time. Under this idea,

$$E[A] = \frac{1}{T}\cdot\sum_t a(t),\ \ E[B] = \frac{1}{T}\cdot\sum_t b(t),$$

$$V[A] = \frac{1}{T}\cdot\sum_t a^2(t) - \left(\frac{1}{T}\cdot\sum_t a(t)\right)^2,$$

$$V[B] = \frac{1}{T} \cdot \sum_t b^2(t) - \left(\frac{1}{T} \cdot \sum_t b(t) \right)^2,$$

$$E[A \cdot B] = \frac{1}{T} \cdot \sum_t a(t) \cdot b(t),$$

and thus,

$$\rho = \frac{E[A \cdot B] - E[A] \cdot E[B]}{\sqrt{V[A] \cdot V[B]}}.$$

Substituting this value ρ into (8.1) and (8.2), we get the desired estimates for $P(A \,\&\, B)$ and $P(A \vee B)$.

Mathematical comment. In producing these estimates, we implicitly assumed that for the desired statistical characteristic (in our case, correlation), averaging over time leads to the same result as averaging over a sample. This property is called *ergodicity*; it is often assumed and/or proved in statistical physics and in statistical data analysis; see, e.g., [3,36].

Computational comment. In the above formulas, we implicitly assumed that the correlation between different expert estimates does not change in time. In reality, just like the expert degrees change with time, the correlation between these degrees may also change. It is therefore necessary to take this change into account when estimating correlation. One way to do that is to consider the recent values with higher weights, and the past values with lower weights. In other words, to each of T moments of time, we assign a weight $w(t) \geq 0$ such that $\sum\limits_t w(t) = 1$, and then consider the modified formulas

$$E[A] = \sum_t w(t) \cdot a(t), \quad E[B] = \sum_t w(t) \cdot b(t),$$

$$V[A] = \sum_t w(t) \cdot a^2(t) - \left(\sum_t w(t) \cdot a(t) \right)^2,$$

$$V[B] = \sum_t w(t) \cdot b^2(t) - \left(\sum_t w(t) \cdot b(t) \right)^2,$$

$$E[A \cdot B] = \sum_t w(t) \cdot a(t) \cdot b(t).$$

The above case corresponds to $w(t) = \dfrac{1}{T}$. A usual selection of "discount" weights is $w(t) = C \cdot q^t$ for some $q < 1$. In this case, the sum $\sum w(t) = \sum C \cdot q^t$ is the sum of a geometric progression:

$$\sum_{t=1}^{T} C \cdot q^t = C \cdot \sum_{t=1}^{T} q^t = C \cdot \frac{1 - q^{T+1}}{1 - q}.$$

Thus, once q is selected, the value C is determined from the condition that $\sum\limits_t w(t) = 1$, as

$$C = \frac{1 - q}{1 - q^{T+1}}.$$

Computational complexity: description. What are the limitations of this approach? An obvious limitation is that to find the degree of confidence in $A \,\&\, B$ or in $A \vee B$, we now need to perform a large number of computations – instead of simply applying a t-norm or a t-conorm to two numbers.

Computational complexity is unavoidable. This limitation is unavoidable: in the dynamic fuzzy logic, we have more values for representing the expert's degree of confidence in each statement, so processing these degrees takes more computation time.

Non-associativity: description. Another limitation is that, in contrast to the usual (static) fuzzy logic, dynamic logic operations are not necessarily associative, i.e., the estimates for $(A \vee B) \vee C$ and for $A \vee (B \vee C)$ are, in general, different.

Non-associativity is unavoidable. Let us show that this non-associativity is also a limitation not of a specific *method* of extending "and"- and "or"-operations to dynamic fuzzy logic, but a limitation of the very *dynamic character* of these logics.

Let us show that non-associativity occurs even if we restrict ourselves to linear operations. This possibility comes from the fact that one of the most frequently used probability-related fuzzy "or"-operations $f_\vee(a, b) = a + b - a \cdot b$ is approximately linear for small a and b, and that it is isomorphic to $a + b$ if we appropriately re-scale the values from the interval $[0, 1]$ to the set $\mathbb{R}_0^+$ of all non-negative numbers.

Definition 8.1.

– *For every integer t, by a* dynamical fuzzy value *corresponding to time t, we mean a sequence of values $a = \{a_s\}_{s \leq t}$, where each value a_s belongs to the set $\mathbb{R}_0^+$.*

- *For every integer t_0 and for each dynamic fuzzy value a, by a* shift $S_{t_0}(a)$*, we mean a sequence $a' = \{a'_s\}_{s \le t+t_0}$ for which $a'_s = a_{s-t_0}$.*
- *By a* aggregation operation*, we mean an operation f that transforms two sequences $a = \{a_s\}_{s \le t}$ and $b = \{b_s\}_{s \le t}$ into a value $c_t \in \mathbb{R}_0^+$.*
- *An operation f is called* shift-invariant *if for every a, b, and t, whenever it transforms a and b into a value c_t, it transforms shifted values $S_{t_0}(a)$ and $S_{t_0}(b)$ into the same value c_{t+t_0}.*
- *We say that an aggregation operation f is* linear *if it is a linear function of all its variables a_s and b_s, i.e.,*

$$c_t = Z_t + \sum_{s \le t} A_{t,s} \cdot a_s + \sum_{s \le t} B_{t,s} \cdot b_s.$$

- *For any aggregation operation f, by the* result $c = \{c_s\}_{s \le t} = f(a, b)$ *of applying this operation to sequences $a = \{a_s\}_{s \le t}$ and $b = \{b_s\}_{s \le t}$ we mean a sequence for which, for every $s \le t$, $c_s = f(\{a_u\}_{u \le s}, \{b_u\}_{u \le s})$.*
- *We say that an operation is* commutative *if $f(a, b) = f(b, a)$ for all a and b, and* associative *if $f(f(a, b), c) = f(a, f(b, c))$.

Proposition 8.1. *If $c = f(a, b)$ is a shift-invariant linear commutative and associative operation, then the value c_t depends only on a_t and b_t and does not depend on the values a_s and b_s for $s < t$.*

Discussion. In other words, any commutative linear operation that takes into account previous fuzzy estimates is not associative.

Comment. The fact that not all algebraic properties can be satisfied in the dynamical case is known in other similar situations: e.g., in [18], it is proven that if we formulate natural requirements for a reasonable next step in a bargaining process, then every function satisfying these requirements does not depend on the bargaining pre-history.

Acknowledgments

This work is supported by a grant from Chiang Mai University, Thailand, and also by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721.

References

[1] S. Aghakhani and S. Dick, An on-line learning algorithm for complex fuzzy logic, *Proceedings of the 2010 International IEEE Conference on Fuzzy Sets and Systems FUZZ-IEEE'2010*, pp. 1–7.

[2] G. Beliakov, A. Pradera, and T. Calvo, *Aggregation Functions: A Guide for Practitioners*, Springer, New York, 2007.

[3] M. Brin and S. Garrett, *Introduction to Dynamical Systems*, Cambridge University Press, 2002.

[4] Z. Chen, S. Aghakhani, J. Man, and S. Dick, ANCFIS: A Neurofuzzy Architecture Employing Complex Fuzzy Sets, *IEEE Transactions on Fuzzy Systems* **19**(2) (2011), 305–322.

[5] R. W. Deming and L. I. Perlovsky, "Using Fuzzy Dynamic Logic to Fuse Information from Multiple Platforms", *Proceedings of the Second Annual Integrated Sensing and Decision Support Workshop ISDS'2005*, MIT Lincoln Laboratory, Lexington, MA, Apr. 12–13, 2005.

[6] S. Dick, Towards complex fuzzy logic, *IEEE Transactions on Fuzzy Systems* **13**(3) (2005), 405–414.

[7] D. Dubois and H. Prade, Possibility theory, probability theory and multiple-valued logics: a clarification, *Annals of Mathematics and Artificial Intelligence* **32**(2001), 35–66.

[8] R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Addison Wesley, Boston, Massachusetts, 2005.

[9] M. E. Garey and D. S. Johnson, *Computers and intractability: a guide to the theory of NP-completeness*, Freeman, San Francisco, 1979.

[10] G. Gerla, *Fuzzy logic: Mathematical Tools for Approximate Reasoning*, Kluwer Academic Publishers, Dordrecht, 2001.

[11] M. Holčapek and T. Tichý, A smoothing filter based on fuzzy transform, *Fuzzy Sets and Systems* **180**(2011), 69–97.

[12] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.

[13] O. Kosheleva, V. Kreinovich, and T. Ngamsantivong, Why complex-valued fuzzy? Why complex values in general? A computational explanation, *Proceedings of the Joint World Congress of the International Fuzzy Systems Association and Annual Conference of the North American Fuzzy Information Processing Society IFSA/NAFIPS'2013*, Edmonton, Canada, June 24–28, 2013, pp. 1233–1236.

[14] O. Kosheleva and V. Kreinovich, Approximate nature of traditional fuzzy methodology naturally leads to complex-valued fuzzy degrees, *Proceedings of the IEEE World Congress on Computational Intelligence WCCI'2014*, Beijing, China, July 6–11, 2014.

[15] O. Kosheleva, V. Kreinovich, A. Garcia, F. Jovel, L. T. Escobedo, and T. Ngamsantivong, Computing with words: towards a new tuple-based formalization, *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics IEEE SMC'2013*, Manchester, UK, October 13–16, 2013, pp. 344–349.

[16] O. Kosheleva, V. Kreinovich, and T. Ngamsantivong, Fuzzy sets can be interpreted as limits of crisp sets, and this can help to fuzzify crisp notions, *Proceedings of 3rd World Conference on Soft Computing*, San Antonio, December 15–18, 2013, pp. 327–337.

[17] V. Kreinovich, Dynamic fuzzy logic leads to more adequate "and" and "or" operations, *Proceedings of the 2011 Sixth International Conference on Soft Computing, Computing with Words and Perceptions in System Analysis, Decision and Control IC-*

SCCW'2011, Antalya, Turkey, September 1–2, 2011, pp. 21–29.
[18] V. Kreinovich, H. T. Nguyen, and S. Sriboonchitta, How to bargain: an interval approach", *International Journal of Intelligent Technologies and Applied Statistics (IJITAS)* **4**(2) (2011), 147–164.
[19] V. Kreinovich, O. Kosheleva, J. Y. Cabrera, M. Gutierrez, and T. Ngamsantivong, A symmetry-based approach to selecting membership functions and its relation to chemical kinetics, *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics IEEE SMC'2013*, Manchester, UK, October 13–16, 2013, pp. 339–343.
[20] J. M. Mendel, *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*, Prentice-Hall, Upper Saddle River, 2001.
[21] J. M. Mendel and D. Wu, *Perceptual Computing: Aiding People in Making Subjective Judgments*, IEEE Press and Wiley, New York, 2010.
[22] I. S. Molchanov, *Theory of Random Sets*, Springer-Verlag, London, 2005.
[23] H. T. Nguyen and V. Kreinovich, *Applications of Continuous Mathematics to Computer Science*, Kluwer, Dordrecht, 1997.
[24] 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* **13**(8) (1998), 683–714.
[25] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2006.
[26] V. Novák, M. Štěpnička, A. Dvořák, I. Perfilieva, V. Pavliska, and L. Vavřícková, Analysis of seasonal time series using fuzzy approach, *International Journal of General Systems* **39**(2010), 305–328.
[27] V. Novák, M. Štěpnička, I. Perfilieva, and V. Pavliska, Analysis of periodical time series using soft computing techniques, *Proceedings of the 8th International FLINS Conference on Computational Intelligence in Decision and Control FLINS'2008*, Madrid, Spain, September 21–24, 2008.
[28] C. H. Papadimitriou, *Computational Complexity*, Addison Wesley, San Diego, 1994.
[29] I. Perfilieva, Fuzzy transforms: theory and applications, *Fuzzy Sets and Systems* **157**(2006), 993–1023.
[30] I. Perfilieva, V. Novák, V. Pavliska, A. Dvořák, and M. Štěpnička, Analysis and prediction of time series using fuzzy transform, *Proceedings of IEEE World Congress on Computational Intelligence WCCI'2008*, 2008, pp. 3875–3879.
[31] L. I. Perlovsky, Neural network with fuzzy dynamic logic, In: *Proceedings of the International IEEE and INNS Joint Conference on Neural Networks IJCNNŠ05*, Montreal, Quebec, Canada, 2005.
[32] L. I. Perlovsky, Fuzzy dynamic logic, *New Math. and Natural Computation* **2**(1) (2006), 43–55.
[33] L. I. Perlovsky, Neural networks, fuzzy models and dynamic logic, In: R. Köhler and A. Mehler, eds., *Aspects of Automatic Text Analysis: Festschrift in Honor of Burghard Rieger*, Springer, Germany, 2007, pp. 363–386.
[34] K. Richart, O. Kosheleva, and V. Kreinovich, r-bounded fuzzy measures are equivalent to ε-possibility measures, *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics SMC'2014*, San Diego, California, October 5–8, 2014, pp. 1229–1234.
[35] D. J. Sheskin, Handbook of Parametric and Nonparametric Statistical Procedures, Chapman & Hall/CRC, Boca Raton, Florida, 2011.
[36] P. Walters, *An Introduction to Ergodic Theory*, Springer Verlag, 1992.
[37] Z. Wang and G. J. Klir, *Fuzzy Measure Theory*, Plenum Press, New York, 1991.
[38] G. S. Yablonsky, D. Constales, and G. B. Marin, Coincidences in chemical kinetics: surprising news about simple reactions, *Chemical Engineering Science* **65** (2010), 6065–6076.
[39] G. S. Yablonsky, D. Constales, and G. B. Marin, Equilibrium relationships for non-equilibrium chemical dependencies, *Chemical Engineering Science* **66** (2011), 111–114.
[40] L. A. Zadeh, Fuzzy sets, *Information and Control* **8** (1965), 338–353.
[41] L. A. Zadeh, Fuzzy sets as the basis for a theory of possibility, *Fuzzy Sets and Systems* **1**(1978), 3–28.
[42] L. A. Zadeh, From computing with numbers to computing with words – from manipulation of measurements to manipulation of perceptions, *Int. J. Appl. Math. Comput. Sci.* **12**(3) (2002), Vol. 12, 307–324.

Appendix

A. Proof of Proposition 2.1.

1°. Let $L = \{\sum c_i \cdot \mu_i(q)\}$ be a shift-invariant linear space. By definition of shift-invariance, this mean that for each function $f(q)$ from this space (in particular, for each function $f(q) = \mu_i(q)$), the shifted function $f(q+q_0)$ also belongs to this space. For $f(q) = \mu_i(q)$, this means that the function $\mu_i(q+q_0)$ also belongs to the space L. By definition of the space L, this means that the shifted function $\mu_i(q+q_0)$ is a linear combination of the original functions $\mu_1(q), \ldots, \mu_n(q)$, i.e., that

$$\mu_i(q+q_0) = \sum_{j=1}^{n} c_{ij}(q_0) \cdot \mu_j(q), \qquad (A1.1)$$

for some real numbers values $c_{ij}(q_0)$.

The equality (A1.1) holds for all $i = 1, \ldots, n$, so we arrive at the following system of equalities:

$$\mu_i(q+q_0) = \sum_{j=1}^{n} c_{ij}(q_0) \cdot \mu_j(q), \qquad (A1.2)$$

2°. By Definition 1, each function $\mu_i(q)$ is differentiable. Let us use the formula (A1.1) to prove that the functions $c_{ij}(q_0)$ are also differentiable. Indeed, let us select n different values $q_1, \ldots, q_k, \ldots, q_n$ of the quantity q, and let us repeat the formula (A1.1) for

each of these values. We then get the following system of equalities:

$$\mu_i(q_k + q_0) = \sum_{j=1}^{n} c_{ij}(q_0) \cdot \mu_j(q_k), \qquad (A1.3)$$

The system (A1.3) is a linear system of equations with n unknowns $c_{i1}(q_0)$, ..., $c_{in}(q_0)$. It is known that in general, each element of the solution to a system of linear equations can be described – via the so-called *Cramer rule* – as a ratio of two determinants, i.e., as a smooth function of the coefficients and the free terms. In our case, the coefficients $\mu_j(q_k)$ are constants (hence, are differentiable), and the free terms $\mu_i(q_k + q_0)$ are also differentiable functions of q_0. Thus, each element $c_{ij}(q_0)$ is a result of applying a smooth function to smooth functions and is, therefore, a differentiable function of q_0.

3°. Now that we know that both the functions $\mu_i(q)$ and the functions $c_{ij}(q)$ are differentiable, we can differentiate both sides of the equations (A1.2) and set q_0 to 0. As a result, we get the following system of differential equations:

$$\mu_i'(q) = \sum_{j=1}^{n} C_{ij} \cdot \mu_j(q), \qquad (A1.4)$$

where we denoted $C_{ij} \stackrel{\text{def}}{=} c_{ij}'(0)$. Thus, for the functions $\mu_1(q)$, ..., $\mu_n(q)$, we have a system of linear differential equations with constant coefficients. Solutions to such systems are well-known: they have the form $x^k \cdot \exp(-\lambda \cdot q)$, where $-\lambda$ is an eigenvalue of the matrix C_{ij}, and k is an integer corresponding to degenerate eigenvalues, i.e., eigenvalues for which the linear space of the corresponding eigenvectors is more than 1-dimensional:

- if we have only one linear independent eigenvector corresponding to the eigenvalue λ, we only get the term corresponding to $k = 0$;
- if we have two linear independent eigenvectors corresponding to the eigenvalue λ, we get terms corresponding to $k = 0$ and $k = 1$;
- ...

4°. Terms corresponding to the same eigenvalue λ form a shift-invariant linear subspace; thus, from the fact that the linear space L is basic, it follows that all the functions from this space correspond to the same eigenvalue λ.

5°. In general, for a linear system of differential equations with constant coefficients, we can have positive, zero, and complex eigenvalues, corresponding to negative, zero, or complex values λ.

5.1°. In our cases, negative values λ are not possible, since then we will have $\mu_i(q) = q^{i-1} \cdot \exp(-\lambda \cdot q)$ tend to infinity for $q \to \infty$, which contradicts to our assumption that the linear space is fuzzy-related, i.e., that we can select a basis of functions whose values are, for all $q \geq 0$, bounded to the interval $[0, 1]$.

5.2°. Zero values λ are also not possible:

- for $i = 1$, $\mu_i(q) = q^{i-1} \cdot \exp(-\lambda \cdot q)$ with $\lambda = 0$ will be a constant function, which contradicts Definition 3, and
- for $i > 1$, this expression tend to infinity for $q \to \infty$, which contradicts to our assumption that the linear space is fuzzy-related, i.e., that we can select a basis of functions whose values are, for all $q \geq 0$, bounded to the interval $[0, 1]$.

5.3°. Similarly, complex values $\lambda = a + \mathrm{i} \cdot b$ are impossible, since then terms $\mu_i(q)$ are then proportional to $q^{i-1} \cdot \exp(-a \cdot q) \cdot \sin(b \cdot q + \varphi)$ for some φ, and thus, cannot be non-negative for all $q \geq 0$.

5.4°. Since negative, zero, or complex values λ are not possible, we conclude that the value λ must be positive. The proposition is proven.

B. Proof of Proposition 4.1

1°. Let us first define a doubly infinite sequence $\ldots < c_{-(k+1)} < c_k < \ldots < c_{-1} < c_0 < c_1 < \ldots < c_k < c_{k+1} < \ldots$ as follows.

We take $c_0 = 1$.

Once we have defined the value c_k for some $k \geq 0$, we define c_{k+1} as follows. By definition of an r-bounded measure, there exists a value $\Delta_k > 0$ such that if $\mu(A) \leq c_k$ and $\mu(B) \leq c_k$, then $\mu(A \cup B) \leq \Delta_k$. We then take $c_{k+1} \stackrel{\text{def}}{=} (1 + \varepsilon) \cdot \max(c_k, \Delta_k)$.

Here, $c_0 = 1$ and $c_{k+1} \geq (1 + \varepsilon) \cdot c_k$. By induction over k, we can prove that $c_k \geq (1 + \varepsilon)^k$ and thus, $c_k \to \infty$ when k increases.

Similarly, once we have defined the value c_{-k} for some $k \geq 0$, we define $c_{-(k+1)}$ as follows. By definition of an r-bounded measure, there exists a value $\nu_k > 0$ such that if $\mu(A) \leq \nu_k$ and $\mu(B) \leq \nu_k$, then $\mu(A \cup B) \leq c_{-k}$. We then take

$$c_{-(k+1)} \stackrel{\text{def}}{=} (1-\varepsilon) \cdot \min(c_{-k}, \nu_k).$$

Here, $c_0 = 1$ and $0 < c_{-(k+1)} \leq (1-\varepsilon) \cdot c_{-k}$. By induction over k, we can prove that $0 < c_{-k} \leq (1-\varepsilon)^k$ and thus, $c_{-k} \to 0$ when $k \to \infty$.

2°. Let us now define the desired function $f(x)$. Since the sequence c_k is strictly increasing, $c_k \to \infty$ when $k \to +\infty$, and $c_k \to 0$ when $k \to -\infty$, for every positive number $x > 0$, there exists an integer k for which $c_{k-1} < x \leq c_k$. We can then define $f(x)$ as follows:

- for each integer k, we take $f(c_k) = (1+\varepsilon)^{k/2}$ and
- for each value x between c_{k-1} and c_k, we define $f(x)$ by linear interpolation: if $c_{k-1} < x \leq c_k$, then

$$f(x) = f(c_{k-1}) + \frac{x - c_{k-1}}{c_k - c_{k-1}} \cdot (f(c_k) - f(c_{k-1})).$$

Since the sequence c_k is strictly increasing, the resulting function $f(x)$ is also strictly increasing.

3°. Let us now prove that for for the new measure $\mu'(A) \stackrel{\text{def}}{=} f(\mu(A))$ (which is equivalent to $\mu(A)$), for every two sets A and B, we have

$$\mu'(A \cup B) \leq (1+\varepsilon) \cdot \max(\mu'(A), \mu'(B)).$$

Without losing generality, let us assume that $\mu(A) \geq \mu(B)$. As we have mentioned in Part 2 of this proof, there exist integers k and ℓ for which $c_{k-1} < \mu(A) \leq c_{k+1}$ and $c_{\ell-1} < \mu(B) \leq c_\ell$. Since $\mu(A) \geq \mu(B)$ and c_k is an increasing sequence, we cannot have $k < \ell$, so $k \geq \ell$ and thus, $c_\ell \leq c_k$.

Hence, we have $\mu(A) \leq c_k$ and $\mu(B) \leq c_k$. By definition of Δ_k, we therefore have $\mu(A \cup B) \leq \Delta_k$. By definition of c_{k+1}, this value is always great than Δ_k, thence we have $\mu(A \cup B) \leq c_{k+1}$.

Since the function $f(x)$ is increasing, we get

$$\mu'(A \cup B) = f(\mu(A \cup B)) \leq f(c_{k+1}) = (1+\varepsilon)^{(k+1)/2}.$$

On the other hand, here, $\max(\mu(A), \mu(B)) = \mu(A) > c_{k-1}$. Due to monotonicity, we have

$$\max(\mu'(A), \mu'(B)) = \mu'(A) = f(\mu(A)) >$$

$$f(c_{k-1}) = (1+\varepsilon)^{(k-1)/2}.$$

In other words, we have

$$(1+\varepsilon)^{(k-1)/2} < \max(\mu'(A), \mu'(B)).$$

Multiplying both sides of this inequality by $1+\varepsilon$, we get

$$(1+\varepsilon)^{(k+1)/2} < (1+\varepsilon) \cdot \max(\mu'(A), \mu'(B)).$$

We already know that $\mu'(A \cup B) \leq (1+\varepsilon)^{(k+1)/2}$. Thus, we conclude that

$$\mu'(A \cup B) \leq (1+\varepsilon) \cdot \max(\mu'(A), \mu'(B)).$$

The proposition is proven.

C. Proof of Proposition 4.2

This proposition can be proven in a way which is similar to the proof of Proposition 4.1, the only difference is how the sequence c_k is built.

We still take $c_0 = 1$.

Once we have defined the value c_k for some $k \geq 0$, we define c_{k+1} as follows. By definition of an r-bounded measure, for each i from 1 to n, there exists a value $\Delta_{ki} > 0$ such that if $\mu_i(A) \leq c_k$ and $\mu_i(B) \leq c_k$, then $\mu_i(A \cup B) \leq \Delta_{ki}$. We then take

$$c_{k+1} = (1+\varepsilon) \cdot \max(c_k, \Delta_{k1}, \ldots, \Delta_{kn}).$$

Similarly, once we have defined the value c_{-k} for some $k \geq 0$, we define $c_{-(k+1)}$ as follows. By definition of an r-bounded measure, for each i from 1 to n, there exists a value $\nu_{ki} > 0$ such that if $\mu_i(A) \leq \nu_{ki}$ and $\mu_i(B) \leq \nu_{ki}$, then $\mu_i(A \cup B) \leq c_{-k}$. We then take

$$c_{-(k+1)} = (1-\varepsilon) \cdot \min(c_{-k}, \nu_{k1}, \ldots, \nu_{kn}).$$

The rest of the proof is the same as for Proposition 4.1.

D. Proof of Proposition 6.1

1°. We are interested in the maximum of the function $\min(\mu'_d(x), \mu'_i(x))$. Let us first show that this maximum is always attained on the interval $[t_{i-1}, t_{i+1}]$, where we denoted $t_i \stackrel{\text{def}}{=} s + i \cdot h$, because outside this interval, the above function is equal to 0.

Indeed, the minimum function is always non-negative. The function $\mu_i(x)$ (as defined by the formula (6.1)) is only different from 0 on the interval $[t_{i-1}, t_{i+1}]$. By definition of a reduced function (formula (6.4)), $\mu'_i(x) \le \mu_i(x)$ and thus, the reduced function $\mu'_i(x)$ can only be different from 0 on the interval $[t_{i-1}, t_{i+1}]$. By definition (6.4) of the reduced function, we have $\mu'_i(x) \le \mu_i(x)$, hence $\min(\mu'_d(x), \mu'_i(x)) \le \mu'_i(x) \le \mu_i(x)$.

Since outside the interval $[t_{i-1}, t_{i+1}]$, we have $\mu_i(x) = 0$, the minimum $\min(\mu'_d(x), \mu'_i(x))$ is also equal to 0 for $x \notin [t_{i-1}, t_{i+1}]$.

2°. We want to prove that the largest value of

$$\min(\mu'_d(x), \mu'_i(x))$$

is equal to d_i. To prove this, we will prove two auxiliary statements:

- that $\min(\mu'_d(t_i), \mu'_i(t_i)) = d_i$, and
- that $\min(\mu'_d(x), \mu'_i(x)) \le d_i$ for all other values

$$x \in [t_{i-1}, t_{i+1}].$$

2.1°. Let us first prove that $\min(\mu'_d(t_i), \mu'_i(t_i)) = d_i$. To prove this equality, we will:

- first compute $\mu'_d(t_i)$,
- then compute $\mu'_i(t_i)$,
- and finally compute the minimum of these two values.

2.1.1°. Let us first compute $\mu'_i(t_i)$.

From the formula (1), we can conclude that for $x = t_i$, we have $\mu_i(t_i) = 1$ and $\mu_j(t_i) = 0$ for all $j \ne i$. Thus, by definition (4) of the reduced function $\mu'_i(x)$, we have

$$\mu'_i(t_i) = \max(0, \mu_i(t_i) - \max(\mu_{i-1}(t_i), \mu_{i+1}(t_i))) =$$

$$\max(0, 1 - \max(0, 0)) = \max(0, 1) = 1.$$

2.1.2°. Let us now compute $\mu'_d(t_i)$.

From the fact that $\mu_i(t_i) = 1$ and $\mu_j(t_i) = 0$ for all $j \ne i$, we conclude that $\min(d_i, \mu_i(t_i)) = \min(d_i, 1) = d_i$ and $\min(d_j, \mu_j(t_i)) = \min(d_j, 0) = 0$ for all $j \ne i$. Thus, the value $\mu_d(t_i)$ (as defined by the formula (6.2)) is equal to $\mu_d(t_i) = \max(d_i, 0, \ldots, 0) = d_i$. Hence,

$$\mu'_d(t_i) = \max(0, \mu_d(t_i) - \max(\mu_{i-1}(t_i), \mu_{i+1}(t_i))) =$$

$$\max(0, d_i - \max(0, 0)) = \max(0, d_i) = d_i.$$

2.1.3°. We have computed $\mu'_i(t_i) = 1$ and $\mu'_d(t_i) = d_i$; thus, $\min(\mu'_d(t_i), \mu'_i(t_i)) = \min(d_i, 1) = d_i$.

The first auxiliary statement is proven.

2.2°. Let us now prove that $\min(\mu'_d(x), \mu'_i(x)) \le d_i$ for all $x \in [t_{i-1}, t_{i+1}]$.

On this interval, only the function $\mu_i(x)$ and two neighboring membership functions $\mu_{i-1}(x)$ and $\mu_{i+1}(x)$ are different from 0, all the other are equal to 0. Thus, for these x, the value $\mu_d(x)$ (as defined by the formula (6.2)) is equal to the largest of the three values:

- the value $\min(\mu_{i-1}(x), d_{i-1})$,
- the value $\min(\mu_i(x), d_i)$, and
- the value $\min(\mu_{i+1}(x), d_{i+1}))$.

The maximum of the three numbers is equal to one of them. Let us consider these three cases one by one and show that in all three cases, we have $\min(\mu'_d(x), \mu'_i(x)) \le d_i$.

2.2.1°. When $\mu_d(x) = \min(\mu_{i-1}(x), d_{i-1})$, then $\mu_d(x) \le \mu_{i-1}(x)$ and thus,

$$\mu_d(x) \le \max(\mu_{i-1}(x), \mu_{i+1}(x)).$$

Hence, $\mu'_d(x)$ is equal to

$$\max(0, \mu_d(x) - \max \mu_{i-1}(x), \mu_{i+1}(x))) = 0,$$

and so, $\min(\mu'_d(x), \mu'_i(x)) = 0 \le d_i$.

2.2.2°. When $\mu_d(x) = \min(\mu_i(x), d_i)$, then $\mu_d(x) \le d_i$ and thus,

$$\mu'_d(x) = \max(0, \mu_d(x) - \max\mu_{i-1}(x), \mu_{i+1}(x))) \le$$

$$\mu_d(x) \le d_i.$$

Hence, $\min(\mu'_d(x), \mu'_i(x)) \le \mu'_d(x) \le d_i$.

2.2.3°. When $\mu_d(x) = \min(\mu_{i+1}(x), d_{i+1})$, then $\mu_d(x) \le \mu_{i+1}(x)$ and thus,

$$\mu_d(x) \le \max(\mu_{i-1}(x), \mu_{i+1}(x).$$

Hence, $\mu'_d(x)$ is equal to

$$\max(0, \mu_d(x) - \max\mu_{i-1}(x), \mu_{i+1}(x))) = 0,$$

and so, $\min(\mu'_d(x), \mu'_i(x)) = 0 \le d_i$.

2.2.4°. In all three cases, we have the desired inequality. Thus, the inequality always holds, and the proposition is proven.

E. Proof of Proposition 8.1

1°. Let us first use the fact that our linear aggregation operation is shift-invariance.

By definition, shift-invariance means that for every two sequences a and b, if

$$c_t = Z_t + \sum_{s\le t} A_{t,s}\cdot a_s + \sum_{s\le t} B_{t,s}\cdot b_s,$$

and we combine the shifted sequences $a' = S_{t_0}(a)$ and $b' = S_{t_0}(b)$:

$$c'_{t+t_0} = Z_{t+t_0} + \sum_{s\le t+t_0} A_{t+t_0,s}\cdot a'_s + \sum_{s\le t} B_{t+t_0,s}\cdot b'_s,$$

then we should get the same result: $c_t = c'_{t+t_0}$. Substituting $a'_s = a_{s-t_0}$ and $b'_s = b_{s-t_0}$ into the formula for c'_{t+t_0}, we conclude that

$$c'_{t+t_0} = Z_{t+t_0} + \sum_{s\le t+t_0} A_{t+t_0,s}\cdot a_{s-t_0} +$$

$$\sum_{s\le t} B_{t+t_0,s}\cdot b_{s-t_0}.$$

Introducing a new variable $s' \stackrel{\text{def}}{=} s - t_0$ for which $s = s' + t_0$, we get

$$c'_{t+t_0} = Z_{t+t_0} + \sum_{s'\le t} A_{t+t_0,s'+t_0}\cdot a_{s'} +$$

$$\sum_{s\le t} B_{t+t_0,s'+t_0}\cdot b_{s'}.$$

For the following arguments, it is convenient to rename s' into s, as a result, we conclude that

$$c'_{t+t_0} = Z_{t+t_0} + \sum_{s\le t} A_{t+t_0,s+t_0}\cdot a_s + \sum_{s\le t} B_{t+t_0,s+t_0}\cdot b_s.$$

The fact that c_t and c'_{t+t_0} are equal means that the following equality holds for all possible sequences a and b:

$$Z_t + \sum_{s\le t} A_{t,s}\cdot a_s + \sum_{s\le t} B_{t,s}\cdot b_s =$$

$$Z_{t+t_0} + \sum_{s\le t} A_{t+t_0,s+t_0}\cdot a_s + \sum_{s\le t} B_{t+t_0,s+t_0}\cdot b_s.$$

Two linear functions coincide if and only if all their coefficients coincide. Thus, for every t and t_0, we have $Z_t = Z_{t+t_0}$, $A_{t,s} = A_{t+t_0,s+t_0}$, and $B_{t,s} = B_{t+t_0,s+t_0}$.

1.1°. Let us first use the first equality $Z_t = Z_{t+t_0}$.

For every two values t and t', we can take $t_0 = t' - t$, then $t + t_0 = t'$ hence $Z_t = Z_{t'}$. Thus, every two values Z_t coincide, so the value Z_t does not depend on t. We will denote this common value by Z.

1.2°. From $A_{t,s} = A_{t+t_0,s+t_0}$, by taking $t_0 = -s$, we conclude that $A_{t,s} = A_{t-s,0}$. By denoting $A_t \stackrel{\text{def}}{=} A_{t,0}$, we can describe this as $A_{t,s} = A_{t-s}$.

1.3°. Similarly, we conclude that $B_{t,s} = B_{t-s}$.

Thus, a shift-invariant linear operation has the form

$$c_t = Z + \sum_{s\le t} A_{t-s}\cdot a_s + \sum_{s\le t} B_{t-s}\cdot b_s.$$

2°. Let us now use commutativity.

Commutativity means that the result of applying this operation to a and b is the same as the result of applying it to b and a, i.e., that

$$Z + \sum_{s \le t} A_{t-s} \cdot a_s + \sum_{s \le t} B_{t-s} \cdot b_s =$$

$$Z + \sum_{s \le t} A_{t-s} \cdot b_s + \sum_{s \le t} B_{t-s} \cdot a_s.$$

Here again, the fact that the two linear functions coincide means that all their coefficients must coincide, i.e., that $A_t = B_t$ for all t. Thus, the above formula for c_t takes the form

$$c_t = Z + \sum_{s \le t} A_{t-s} \cdot (a_s + b_s).$$

3°. Let us now use associativity $f(f(a,b),c) = f(a,f(b,c))$.

Associativity means that the two aggregation expressions $f(f(a,b),c)$ and $f(a,f(b,c))$ coincide.

3.1°. In the expression $f(f(a,b),c)$, we first combine sequences a and b into a new sequence $d = f(a,b)$, and then combine d and c into a new sequence $e = f(d,c)$. For the t-th component of these two new sequences d and e, if we keep track only of the dependence on a_t, b_t, and c_t, we get $d_t = A_0 \cdot (a_t + b_t) + \ldots$ and thus,

$$e_t = A_0 \cdot (d_t + c_t) + \ldots = A_0^2 \cdot (a_t + b_t) + A_0 \cdot c_t + \ldots$$

A similar expression for $f(a,f(b,c))$ takes the form

$$A_0^2 \cdot (b_t + c_t) + A_0 \cdot a_t + \ldots$$

The fact that the two expression coincide means that for all possible values a_t, b_t, and c_t, we have

$$A_0^2 \cdot (a_t + b_t) + A_0 \cdot c_t = A_0^2 \cdot (b_t + c_t) + A_0 \cdot a_t.$$

Since the two linear functions coincide, their coefficients must coincide, i.e., we must have $A_0 = A_0^2$. Thus, we have $A_0 = 0$ or $A_0 = 1$.

3.2°. Let us show that in both cases $A_0 = 0$ and $A_0 = 1$, we have $A_1 = A_2 = \ldots = 0$, i.e., the value c_t depends only on a_t and b_t and does not depend on the previous values a_s and b_s.

In both cases, we will prove it by contradiction. Indeed, let us assume that $A_j \neq 0$ for some $j \ge 1$; let k denote the smallest index $k \ge 0$ for which $A_k \neq 0$.

$3.2.1^\circ$. When $A_0 = 0$, this means that the aggregation operation $f(a,b)$ leads to

$$d_t = Z + A_k \cdot (a_{t-k} + b_{t-k}) + \ldots$$

and

$$e_t = Z + A_k \cdot (d_{t-k} + c_{t-k}) + \ldots$$

Here,

$$d_{t-k} = Z + A_k \cdot (a_{t-2k} + b_{t-2k}) + \ldots$$

Thus, we have

$$e_t = Z + A_k \cdot Z + A_k^2 \cdot (a_{t-2k} + b_{t-2k}) + A_k \cdot c_{t-k} + \ldots$$

Similarly, the second expression $f(a,f(b,c))$ leads to

$$e_t = Z + A_k \cdot Z + A_k^2 \cdot (b_{t-2k} + c_{t-2k}) + A_k \cdot a_{t-k} + \ldots,$$

thus

$$Z + A_k \cdot Z + A_k^2 \cdot (a_{t-2k} + b_{t-2k}) + A_k \cdot c_{t-k} + \ldots =$$

$$Z + A_k \cdot Z + A_k^2 \cdot (b_{t-2k} + c_{t-2k}) + A_k \cdot a_{t-k} + \ldots$$

The left-hand side of this equality does not depend on the value a_{t-k}, it only depends on the previous values, while the right-hand side explicitly depends on a_{t-k}: this term enters with a coefficient $A_k \neq 0$.

Thus, the equality is indeed impossible.

$3.2.2^\circ$. When $A_0 = 1$, the aggregation operation $f(a,b)$ leads to

$$d_t = Z + a_t + b_t + A_k \cdot (a_{t-k} + b_{t-k}) + \ldots$$

and

$$e_t = Z + d_t + c_k + A_k \cdot (d_{t-k} + c_{t-k}) + \ldots$$

Here,

$$d_{t-k} = Z + a_{t-k} + b_{t-k} + \ldots$$

Thus, we have

$$e_t = Z + (Z + a_t + b_t + A_k \cdot (a_{t-k} + b_{t-k}) + \ldots) + c_t +$$

$$A_k \cdot ((Z + a_{t-k} + b_{t-k} + \ldots) + c_{t-k}) + \ldots =$$

$$2Z + a_t + b_t + c_t + A_k \cdot (2a_{t-k} + 2b_{t-k} + c_{t-k}) + \ldots$$

Similarly, the second expression $f(a, f(b, c))$ leads to

$$e_t = 2Z + a_t + b_t + c_t + A_k \cdot (2b_{t-k} + 2c_{t-k} + a_{t-k}) + \ldots,$$

thus

$$2Z + a_t + b_t + c_t + A_k \cdot (2a_{t-k} + 2b_{t-k} + c_{t-k}) + \ldots =$$

$$2Z + a_t + b_t + c_t + A_k \cdot (2b_{t-k} + 2c_{t-k} + a_{t-k}) + \ldots$$

Since the two linear functions coincide, all corresponding coefficients must coincide. The left-hand side of this equality contains a_{t-k} with a coefficient $2A_k$, while the right-hand side has this variable with a different coefficient $A_k \neq 2A_k$.

Thus, the equality is impossible in this case as well. The proposition is proven.