Why Cauchy Membership Functions

Javier Viaña[1], Stephan Ralescu[1],
Kelly Cohen[1], Anca Ralescu[1],
and Vladik Kreinovich[2]
[1]University of Cincinnati, Cincinnati, OH 45219, USA
[2]University of Texas at El Paso, El Paso, TX 79968, USA
vianajr@mail.uc.edu, ralescs@mail.uc.edu,
cohenky@ucmail.uc.edu, ralescal@ucmail.uc.edu,
vladik@utep.edu

Abstract

In many practical situations of fuzzy techniques, we can elicit membership functions from the experts, but what if we cannot do that? What functions should we then use? Experiments show that in many applications, what we call Cauchy membership functions – whose expressions are similar to Cauchy distributions – work the best. It this paper, we provide a theoretical explanation for this empirical fact.

1 Formulation of the Problem

In many practical applications of fuzzy techniques (see, e.g., [1, 2, 3, 4, 5, 9]), we can elicit membership functions from the experts, but what if we cannot do that? What functions should we then use? Experiments (see, e.g., [7, 8]) show that in many applications, the following membership functions work the best:

$$\mu_x(x) = \frac{1}{1 + \dfrac{(x-a)^2}{k^2}}. \tag{1.1}$$

The expression (1.1) describing these membership functions is similar to the known expression for the probability density function $f(x)$ of a Cauchy distribution (see, e.g., [6]):

$$f(x) = \text{const} \cdot \frac{1}{1 + \dfrac{(x-a)^2}{k^2}}. \tag{1.2}$$

Because of this similarity, membership functions (1.1) are known as *Cauchy* membership functions.

A natural question is: how can we explain this empirical fact – that Cauchy membership functions work better than other functions that we tried? In this paper, we provide two explanations:

- first, that these functions (as well as Gaussian membership functions) lead to the most efficient learning, and
- second, that these functions lead to the most reliable results.

2 Which Membership Functions Lead to the Most Efficient Learning

2.1 Formulation of the Problem

From expert rules to fuzzy learning. One of the main reasons why Lotfi Zadeh invented fuzzy techniques was to translate expert rules that use imprecise ("fuzzy") natural-language property like "small", "medium", etc., into a precise control strategy. For this purpose, to each such property P, Zadeh proposed to assign a function $\mu_P(x)$ (known as *membership function*) that describes, for each possible value x of the corresponding quantity, the degree to which, according to the expert, an object with this value satisfies the property P – e.g., to what extent the amount x is small. This degree is usually assumed to be from the interval $[0,1]$.

This is how first applications of fuzzy techniques emerged: researchers elicited rules and membership functions from the experts, and used fuzzy methodology to design a control strategy. The resulting control was often reasonably good, but not perfect. So, a natural idea was proposed: to use the original fuzzy control as a first approximation, and then to tune its parameters based on the practical behavior of the resulting system.

This "fuzzy learning" idea was first used in situations when we have expert rules that provide a reasonable first approximation. However, it turned out that this learning algorithm leads to a reasonable control even when we do not have any expert rules, i.e., when we only have data.

Natural question: which membership function should we use? When we start with expert knowledge, we elicit membership functions from the experts. But when we use fuzzy learning to situations when there is no expert knowledge, a natural question is: which membership functions should we use?

2.2 How to Select Membership Functions: Analysis and Conclusion

Main idea. A natural idea is to select a membership function that would make learning faster. How can we do that?

Need to compute derivatives. The main objective of any learning is to optimize the corresponding objective function – a function that describes which outputs are better and which are worse. For example, if we have examples of desired outputs, then the objective is to minimize the discrepancy between the values produced by the system and the values that we want to obtain.

Since the invention of calculus, the most efficient optimization techniques are based on computing the derivatives: one of the main objectives (and still one of main uses) of calculus is to identify points where a function attains its maximum or minimum as points where its derivative is 0, and the fastest ways to reach these points is to use the derivatives of the objective function. There are many optimization techniques, from the simplest gradient descent to more complex methods, all these techniques use differentiation.

The result of processing by several fuzzy layers is a composition of functions corresponding to each layer. So, to compute the derivative of the resulting transformation, we need to know the derivatives corresponding to each layer. From this viewpoint, since we want to find a membership function that will make learning faster, we need to find membership functions for which the computation of its derivatives is as easy as possible.

Let us describe it in precise terms. When we compute the value of the derivative $\mu'(x)$ for some input x, we can use the fact that we have already computed the output signal and thus, we have already computed the value $\mu(x)$. Thus, in computing the value $\mu'(x)$, we can use not only the input x, but also the value $\mu(x)$. In other words, we are looking for an expression $\mu'(x) = f(\mu(x), x)$ for the simplest possible function $f(a, x)$ of two variables.

What does "simplest" mean? In a computer, the only hardware supported operations with numbers are arithmetic operations: addition, subtraction (which, for the computer, is, in effect, the same as addition), multiplication, and taking an inverse (division is implemented as $a/b = a \cdot (1/b)$). To be more precise, computing an inverse is also implemented as a sequence of additions, subtractions, and multiplications, so each computation actually consists of additions, subtractions, and multiplications – and thus, computes a polynomial, since a polynomial can be defined as any function that can be obtained from variables and constants by using addition, subtraction, and multiplication. For example, when we ask a computer to compute $\exp(x)$ or $\sin(x)$, what most compilers do is compute the value of a polynomial that approximates the desired function – usually this polynomial is simply the sum of the first few terms of this function's Taylor expansion.

From this viewpoint, looking for the simplest function $f(a, x)$ means looking for a polynomial $f(a, x)$ that can be obtained by using the smallest possible number of arithmetic operations. (In a computer, unary minus is easy, so we do not count unary minuses.)

Additional idea: asymptotic behavior. A typical membership function corresponding to notions like "small" and "medium" is only satisfied, with a reasonable degree, for a bounded set of values. Thus, in the limits, when $x \to \infty$ or $x \to -\infty$, we should have $\mu(x) \to 0$. Thus, it makes sense to consider membership functions with this asymptotic property.

Most membership functions do not just asymptotically tend to 0, they are equal to 0 outside some intervals. For such function, in the areas where $\mu(x) = 0$, we expect $\mu'(x) = 0$, i.e., we have $f(0, x) = 0$ for all x. Since the function $f(a, x)$ is a polynomial, this means that all its monomials must be proportional to a, i.e., we must have $f(a, x) = a \cdot g(a, x)$ for some function $g(a, x)$. Thus, looking for the simplest function $f(a, x)$ means looking for the simplest functions $g(a, x)$.

We will consider the cases when computing $g(a, x)$ requires 0 or 1 arithmetic operation.

Case when computing $g(a, x)$ does not require any arithmetic operations at all. This means that the value $g(a, x)$ is equal to one of the given values, i.e., either to a or to x or to some constant c.

In the first case, when $g(a, x) = a$, we have $\mu' = f(\mu, x) = \mu \cdot g(\mu, x) = \mu \cdot \mu = \mu^2$, i.e., $\dfrac{d\mu}{dx} = \mu^2$ hence $\dfrac{d\mu}{\mu^2} = dx$. Integrating, we get $-\dfrac{1}{\mu} = x + C$, hence $\mu(x) = -\dfrac{1}{x + C}$. This function is unbounded, so it cannot serve as a membership function. In this case, adding unary minus, i.e., considering $g(a, x) = -a$, does not help.

In the second case, when $g(a, x) = x$, we have $\mu' = \mu \cdot x$, i.e., $\dfrac{d\mu}{dx} = \mu \cdot x$ hence $\dfrac{d\mu}{\mu} = x \cdot dx$. Integrating, we get $\ln(\mu(x)) = \dfrac{x^2}{2} + C$, i.e., $\mu(x) = A \exp\left(\dfrac{x^2}{2}\right)$ for some constant $A = \exp(C)$. This is not a membership function, but by adding unary negation, i.e., by considering $g(a, x) = -x$, we get $\mu(x) = \exp\left(-\dfrac{x^2}{2}\right)$ – a very reasonable case of Gaussian membership functions.

In the third case, when $g(a, x) = c$, we have $\mu' = c \cdot \mu$, i.e., $\dfrac{d\mu}{dx} = c \cdot \mu$ hence $\dfrac{d\mu}{\mu} = c \cdot dx$. Integrating, we get $\ln(\mu(x)) = c \cdot x + C$, i.e., $\mu(x) = A \exp(c \cdot x)$ – also not membership functions.

Case when computing $g(a, x)$ requires a single arithmetic operation. This operation can be addition/substraction or multiplication. For addition, we can have $g(a, x) = a + a$, $g(a, x) = a + c$, $g(a, x) = x + x$, $g(a, x) = x + c$, or $g(a, x) = a + x$. In the first case, we get an unbounded function. In the second case, we get a sigmoid function – that does not have the right asymptotics for $x \to \pm\infty$. In the third and fourth cases, we get Gaussian functions – re-scaled in the third case and shifted in the fourth case. In the last case, we have a reasonable differential equation $\mu' = \mu \cdot (\mu + x)$, but the problem is that this equation does not have an explicit solution, so while computing $\mu'(x)$ is easy when we know $\mu(x)$, computing $\mu(x)$ will be difficult – so this case should also be dismissed.

For multiplication, we can similarly have $g(a, x) = a \cdot c$, $g(a, x) = x \cdot c$, $g(a, x) = a \cdot a$, $g(a, x) = x \cdot x$, or $g(a, x) = a \cdot x$. In the first case, we get an unbounded function. In the second case, we get a re-scaled Gaussian function. For $g(a, x) = a \cdot a$, we have $\dfrac{d\mu}{dx} = \mu^3$ hence $\dfrac{d\mu}{\mu^3} = dx$. Integrating, we get

$-\dfrac{1}{2\mu^2} = x + C$, i.e., $\mu(x) = \sqrt{-2(x+C)}$. This expression is not defined for large positive x, so it should also be dismissed.

For $g(a,x) = x \cdot x$, we get $\dfrac{d\mu}{dx} = \mu \cdot x^2$ hence $\dfrac{d\mu}{\mu} = x^2 \cdot dx$. Integrating, we get $\ln(\mu) = \dfrac{1}{3} \cdot x^3 + C$, hence $\mu(x) = \exp\left(\dfrac{1}{3} \cdot x^3 + C\right)$. This function is also not bounded, so it has to be dismissed.

Finally, for $g(a,x) = a \cdot x$, we get $\dfrac{d\mu}{dx} = \mu^2 \cdot x$ hence $\dfrac{d\mu}{\mu^2} = x \cdot dx$. Integrating, we get $-\dfrac{1}{\mu(x)} = \dfrac{1}{2} \cdot x^2 + C$, hence $\mu(x) = -\dfrac{1}{\dfrac{1}{2} \cdot x^2 + C}$. This is not a membership function, but if we add unary minus, i.e., consider $g(a,x) = -a \cdot x$, we get $\mu(x) = \dfrac{1}{\dfrac{1}{2} \cdot x^2 + C}$, i.e., what we called a Cauchy membership function.

Resulting membership functions. A membership function is usually defined in such a way that its largest value is 1. For the function $1/(x^2/2+C)$, the largest possible value is $1/C$, so we should take $C = 1$ and consider the membership function

$$\mu(x) = \frac{1}{1 + \dfrac{x^2}{2}}. \tag{2.1}$$

We also need to take into account that the numerical value of a physical quantity depends on the choice of the measuring unit and on the choice of the starting point. If we change a measuring unit and/or a starting point, then we get new numerical values X which can be obtained from previous values x by a linear transformation $X = k \cdot x + a$, where k is the ratio of the measuring units and a is the difference in starting points. A classical example is the relation between temperature t_C in Celsius and temperature t_F in Fahrenheit: $t_F = 1.8 \cdot t_C + 32$.

When the original values x are described by the membership function (2.1), then, to get the membership function for the new numerical values X, we need to substitute, into the formula (2.1), the expression $x = \dfrac{X - a}{k}$ that describes the old value x in terms of the new value X. As a result, for the new values, we get the following membership function

$$\mu_X(X) = \frac{1}{1 + \dfrac{(X-a)^2}{2k^2}}. \tag{2.2}$$

This expression can be somewhat simplified if we define a new parameter $K \stackrel{\text{def}}{=} \sqrt{2} \cdot k$ for which $2k^2 = K^2$, then we get:

$$\mu_X(X) = \frac{1}{1 + \dfrac{(X-a)^2}{K^2}}. \tag{2.3}$$

Similarly, substituting $x = \dfrac{X-a}{k}$ into the expression $\mu(x) = \exp\left(-\dfrac{x^2}{2}\right)$, we get

$$\mu_X(X) = \exp\left(-\frac{(X-a)^2}{2k^2}\right), \tag{2.4}$$

i.e., in terms of the new parameter K:

$$\mu_X(X) = \exp\left(-\frac{(X-a)^2}{K^2}\right). \tag{2.5}$$

Conclusion to this section. The membership functions for which fuzzy learning is the simplest are Gaussian and Cauchy membership functions (2.3) and (2.5).

3 Which Membership Functions Lead to the Most Reliable Results

3.1 Idea

General idea. We want to select membership functions for which we will be most confident in the results of the corresponding data processing. What often makes us more confident is when two different techniques lead to the same result – just like:

- when we have two experts making the same statement, it makes us more confident that this statement is true, and
- when two different measurements of the same quantity agree, this makes more confident that both measurement results are correct.

Specific idea. As Zadeh himself mentioned several times, whatever we can describe by using a membership function $\mu(x)$ – which is usually normalized by taking $\max\limits_x \mu(x) = 1$ – can also, from the mathematical viewpoint, be described by using subjective probabilities, with the probability density

$$f(x) = \frac{\mu(x)}{\int \mu(y)\,dy}.$$

Vice versa, the probability density distribution can be transformed into a membership function if we normalize it by dividing by its largest value

$$\mu(x) = \frac{f(x)}{\max\limits_y f(y)}.$$

So, it is reasonable to select a membership function $\mu(x)$ for which fuzzy data processing will lead to the same result as using the corresponding subjective probabilities.

3.2 Data Processing: Reminder and the Resulting Explanation

What is data processing.

- Whether we are using the known current values $\widetilde{x}_1, \ldots, \widetilde{x}_n$ of different quantities $x_1, \ldots, x_n$ to predict the future value of some physical quantity y,
- whether we are reconstructing the current value of some difficult-to-measure quantity y from the results $\widetilde{x}_1, \ldots, \widetilde{x}_n$ of measuring related easier-to-measure quantities $x_1, \ldots, x_n$,
- whether we are finding the best control y based on the known values $x_1, \ldots, x_n$ of the related quantities,

in all these cases we have an algorithm f that transforms the known values $\widetilde{x}_1, \ldots, \widetilde{x}_n$ into the desired estimate $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$.

Need to take uncertainty into account. The values $\widetilde{x}_i$ come from measurements or from expert estimates. Both measurement and expert estimates are never absolutely accurate: in general, each measurement result $\widetilde{x}_i$ is different from the actual (unknown) value x_i, i.e., there is a non-zero approximation error $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$. Because of this, the estimate $\widetilde{y}$ is, in general, different from the value $y = f(x_1, \ldots, x_n)$ that we would have obtained if we used the actual values $x_i = \widetilde{x}_i - \Delta x_i$. From the practical viewpoint, an important question is: how big is this difference $\Delta y \stackrel{\text{def}}{=} \widetilde{y} - y$?

In this section, we consider the case when our information about possible values of Δx_i is characterized in fuzzy terms, by a membership function.

Linearization. In many practical situations, the approximation errors are relatively small. So, we can expand the expression for Δy:

$$\Delta y = \widetilde{y} - y = f(\widetilde{x}_1, \ldots, \widetilde{x}_n) - f(\widetilde{x}_1 - \Delta x_1, \ldots, \widetilde{x}_n - \Delta x_n)$$

in Taylor series in terms of Δx_i, and ignore terms which are quadratic (or of higher order) in terms of Δx_i. In this approximation:

$$f(\widetilde{x}_1 - \Delta x_1, \ldots, \widetilde{x}_n - \Delta x_n) = f(\widetilde{x}_1, \ldots, \widetilde{x}_n) - \sum_{i=1}^n c_i \cdot \Delta x_i,$$

where we denoted $c_i \stackrel{\text{def}}{=} \dfrac{\partial f}{\partial x_i}\,(\widetilde{x}_1, \ldots, \widetilde{x}_n)$. In this case, we get

$$\Delta y = \sum_{i=1}^n c_i \cdot \Delta x_i. \tag{3.1}$$

This is the case we consider in this section.

How to describe and process fuzzy uncertainty. We assume that for each estimate $\widetilde{x}_i$, we have a numerical estimate Δ_i of the corresponding approximation error. This assumption is in good accordance with the usual practice, according to which we say something like

"x_i is approximately 1.0, with an error of about 0.1";

in this example, the estimate $\widetilde{x}_i$ is equal to 1.0, and $\Delta_i = 0.1$.

If we select a membership function $\mu(x)$ corresponding to the case $\Delta_i = 1$, then for each i for which $\Delta_i \neq 1$, as a membership function for Δx_i, it s reasonable to take

$$\mu_i(\Delta x_i) = \mu\left(\frac{\Delta x_i}{\Delta_i}\right) \tag{3.2}$$

To process this fuzzy uncertainty, we can use Zadeh's extension principle, according to which the resulting membership function $\mu_y(\Delta y)$ has the form

$$\mu_y(\Delta y) = \max\left\{\min(\mu_1(\Delta x_1), \ldots, \mu_n(\Delta x_n)) : \sum_{i=1}^{n} c_i \cdot \Delta x_i = \Delta y\right\}.$$

Since we have no information about the membership function $\mu(x)$, we have no reason to conclude that positive or negative values of x are more possible. Thus, it makes sense to assume that such values are equally possible, i.e., that $\mu(x) = \mu(-x)$ for all x. It is known for such even functions $\mu(x)$, when all the membership function have the same shape – i.e., have the form (3.2) – then the resulting membership function also has the same form $\mu_y(\Delta y) = \mu\left(\frac{\Delta y}{\Delta}\right)$, where we denoted

$$\Delta = \sum_{i=1}^{n} |c_i| \cdot \Delta_i. \tag{3.3}$$

How to process the corresponding subjective probabilities. Based on each membership function (3.2), we form the corresponding probability density functions

$$f_i(\Delta x_i) = \text{const} \cdot \mu_i(\Delta x_i) = \text{const} \cdot \mu\left(\frac{\Delta x_i}{\Delta_i}\right).$$

One can easily check that if by ξ we denote a random variable corresponding to $\Delta_i = 1$, with probability density $f(x)$, then the distribution of the random variable ξ_i corresponding to $\Delta_i \neq 1$ is equivalent to the distribution of $\Delta_i \cdot \xi$. We therefore write that $\xi_i = \Delta_i \cdot \xi^{(i)}$, where $\xi^{(i)}$ is distributed according to the distribution $f(x)$ (corresponding to $\Delta_i = 1$).

Since we have no reason to expect positive or negative correlation between these random variables, it makes sense to assume that they are independent. Thus, due to the formula (3.1), the random variable ξ_y corresponding to Δy has the form

$$\xi_y = \sum_{i=1}^{n} c_i \cdot \Delta_i \cdot \xi^{(i)}. \tag{3.4}$$

So, the condition that the resulting probability density will lead, after renormalization, to the membership function $\mu_y(\Delta y) = \mu\left(\frac{\Delta y}{\Delta}\right)$, with the value Δ described by the formula (3.3), is equivalent to requiring that:

- for n independent identically distributed random variables $\xi^{(i)}$, with common probability density $f(x)$,
- the distribution of their linear combination (3.4) is equivalent to the distribution of $\Delta \cdot \xi$, where Δ is determined by the formula (3.3).

This condition can be described in terms of the characteristic functions $\chi_\alpha(\omega) \stackrel{\text{def}}{=} E[\exp(\mathrm{i} \cdot \omega \cdot \alpha)]$, were $E[\cdot]$ denotes the mean value and $\mathrm{i} \stackrel{\text{def}}{=} \sqrt{-1}$. Indeed, from (3.4), we conclude that for

$$E[\exp(\mathrm{i} \cdot \omega \cdot \xi_y)] = E[\exp(\mathrm{i} \cdot \omega \cdot \Delta \cdot \xi)] = \chi_0(\Delta \cdot \omega), \tag{3.5}$$

where χ_0 denotes the characteristic function of the random variable ξ, we have

$$E[\exp(\mathrm{i} \cdot \omega \cdot \xi_y)] = E\left[\exp\left(\mathrm{i} \cdot \omega \cdot \sum_{i=1}^{n} c_i \cdot \Delta_i \cdot \xi^{(i)}\right)\right] =$$

$$E\left[\prod_{i=1}^{n} \exp\left(\mathrm{i} \cdot \omega \cdot c_i \cdot \Delta_i \cdot \xi^{(i)}\right)\right].$$

Since the variables $\xi^{(i)}$ are independent, the expected value of the product is equal to the product of expected values, i.e.,

$$E[\exp(\mathrm{i} \cdot \omega \cdot \xi_y)] = \prod_{i=1}^{n} E\left[\exp\left(\mathrm{i} \cdot \omega \cdot c_i \cdot \Delta_i \cdot \xi^{(i)}\right)\right] = \prod_{i=1}^{n} \chi_0(c_i \cdot \Delta_i \cdot \omega). \tag{3.6}$$

Comparing the expression (3.5) and (3.6), we conclude that

$$\chi_0\left(\left(\sum_{i=1} |c_i| \cdot \Delta_i\right) \cdot \omega\right) = \prod_{i=1}^{n} \chi_0(c_i \cdot \Delta_i \cdot \omega). \tag{3.7}$$

For any $a > 0$, for $\omega = 1$, $\Delta_1 = a$, and $c_1 = -1$, we get $\chi_0(a) = \chi_0(-a)$, so the function $\chi_0(a)$ is even.

For any $a > 0$ and $b > 0$, for $n = 2$, $\omega = 1$, $\Delta_1 = a$, and $\Delta_2 = b$, we conclude that

$$\chi_0(a + b) = \chi_0(a) \cdot \chi_0(b). \tag{3.8}$$

Taking logarithms of both sides, we get Cauchy's functional equation $\ell(a + b) = \ell(a) + \ell(b)$, where we denoted $\ell(a) \stackrel{\text{def}}{=} \ln(\chi_0(a))$. The function $\ell(a)$ is measurable, and it is known that the only measurable solutions of Cauchy's functional equation are linear functions, so $\ell(a) = k \cdot a$ for some constant k, and thus, $\chi_0(a) = \exp(k \cdot a)$ Since the function $\chi_0(a)$ is even, we have $\chi_0(a) = \exp(k \cdot |a|)$.

The characteristic function is a Fourier transform of the probability density function. So, by applying the inverse Fourier transform to the characteristic function, we can reconstruct the probability density function. For the above expression, we get $f(x) = \dfrac{\text{const}}{1 + \dfrac{x^2}{k^2}}$. So, after normalizing it back to the membership function, we get

$$\mu(x) = \frac{1}{1 + \dfrac{x^2}{k^2}}, \tag{3.9}$$

which is exactly what we called Cauchy membership function.

From the membership function for the approximation error to the membership function for the actual quantity. According to the formula (3.9), the membership function for each approximation error Δx should have the form

$$\mu_{\Delta x}(\Delta x) = \frac{1}{1 + \dfrac{(\Delta x)^2}{k^2}}. \tag{3.10}$$

Substituting the expression $\Delta x = \widetilde{x} - x$ into the formula (3.10), we get the membership function corresponding to each quantity x:

$$\mu_x(x) = \frac{1}{1 + \dfrac{(x-a)^2}{k^2}}, \tag{3.11}$$

for a constant $a \stackrel{\text{def}}{=} \widetilde{x}$.

Conclusion to this section. For each membership function, we can process the corresponding uncertainty in two different ways. First, we can apply Zadeh's extension principle. Alternatively, we can:

- transform the corresponding membership functions into probability density functions,
- process the corresponding random variable, and then
- transform the probability density function for the result back into a membership function.

The only case when these two results coincide – and thus, when we have additional confidence in this joint result – is when we use Cauchy membership functions (3.11).

Acknowledgments

This work was supported in part by a grant from the "la Caixa" Banking Foundation (ID 100010434), whose code is LCF / BQ / AA19 / 11720045, by the

National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science), and HRD-1834620 and HRD-2034030 (CAHSI Includes), and by the AT&T Fellowship in Information Technology. It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478.

References

[1] R. Belohlavek, J. W. Dauben, and G. J. Klir, *Fuzzy Logic and Mathematics: A Historical Perspective*, Oxford University Press, New York, 2017.

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

[3] J. M. Mendel, *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*, Springer, Cham, Switzerland, 2017.

[4] H. T. Nguyen, C. L. Walker, and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2019.

[5] V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer, Boston, Dordrecht, 1999.

[6] D. J. Sheskin, *Handbook of Parametric and Non-Parametric Statistical Procedures*, Chapman & Hall/CRC, London, UK, 2011.

[7] J. Viaña and K. Cohen, "Fuzzy-based, noise-resilient, explainable algorithm for regression", *Proceedings of the Annual Conference of the North American Fuzzy Information Processing Society NAFIPS'2021*, West Lafayette, Indiana, June 7–9, 2021.

[8] J. Viaña, S. Ralescu, K. Cohen, A. Ralescu, and V. Kreinovich, "Extension to multi-dimensional problems of a fuzzy-based explainable and noise-resilient algorithm", *Proceedings of the 14th International Workshop on Constraint Programming and Decision Making CoProd'2021*, Szeged, Hungary, September 12, 2021.

[9] L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.