5-2020

Optimization under Fuzzy Constraints: Need to Go Beyond Bellman-Zadeh Approach and How It Is Related to Skewed Distributions

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

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

Nguyen Hoang Phuong
Thang Long University, nhphuong2008@gmail.com

Comments:

Technical Report: UTEP-CS-20-46

Optimization under Fuzzy Constraints: Need to Go Beyond Bellman-Zadeh Approach and How It Is Related to Skewed Distributions

Olga Kosheleva[1], Vladik Kreinovich[2], and Hoang Phuong Nguyen[3]
[1]University of Texas at El Paso
500 W. University, El Paso, TX 79968, USA
olgak@utep.edu, vladik@utep.edu
[2]Division Informatics, Math-Informatics Faculty
Thang Long University, Nghiem Xuan Yem Road
Hoang Mai District, Hanoi, Vietnam
nhphuong2008@gmail.com

Abstract

In many practical situations, we need to optimize the objective function under fuzzy constraints. Formulas for such optimization are known since the 1970s paper by Richard Bellman and Lotfi Zadeh, but these formulas have a limitation: small changes in the corresponding degrees can lead to a drastic change in the resulting selection. In this paper, we propose a natural modification of this formula, a modification that no longer has this limitation. Interestingly, this formula turns out to be related for formulas for skewed (asymmetric) generalizations of the normal distribution.

1 Formulation of the Problem

Need for optimization under constraints. Whenever we have a choice, we want to select the alternative which is the best for us. The quality of each alternative a is usually described by a numerical value $f(a)$. In these terms, we want to select the alternative $a_{\rm opt}$ for which this numerical value is the largest possible:

$$f(a_{\rm opt}) = \max_a f(a). \tag{1}$$

Often, not all theoretically possible alternatives are actually possible, there are some constraints. For example, suppose we want to drive from point A to point B in the shortest possible time, so we plan the shortest path – but it may turn out that some of the roads are closed, e.g., due to an accident, or to extreme weather conditions, or to some public event. In such situations, we can only select an alternative that satisfies these constraints.

Let us describe this situation in precise terms. Let A denote the set of all actually available alternatives – i.e., all alternatives that satisfy all the given constraints. In this case, instead of the original unconstrained optimization problem (1), we have a modified problem: we need to select an alternative $a_{\rm opt} \in A$ for which the value of the objective function $f(a)$ attains its largest possible value on the set A:

$$f(a_{\rm opt}) = \max_{a \in A} f(a). \tag{2}$$

Need for optimization under fuzzy constraints. The above formulation assumes that we know exactly which alternatives are possible and which are not, i.e., that the set A of possible alternatives is crisp.

In practice, this knowledge may come in terms of words from natural language. For example, you may know that it is *highly probable* that a certain alternative a will be possible. A natural way to describe such knowledge in precise terms is to use fuzzy logic – technique specifically designed by Lotfi Zadeh to translate imprecise ("fuzzy") knowledge from natural language to numbers; see, e.g., [3, 6, 10, 11, 12, 15]. In this technique, to each alternative a, we assign the degree $\mu(a) \in [0, 1]$ to which this alternative is possible:

- degree $\mu(a) = 1$ means that we are absolutely sure that this alternative is possible,
- degree $\mu(a) = 0$ means that we are absolutely sure that this alternative is *not* possible,
- and degrees between 0 and 1 indicate that we have some – but not full – confidence that this alternative is possible.

How can we optimize the objective function $f(a)$ under such fuzzy constraints?

Bellman-Zadeh approach: a brief reminder. The most widely used approach to solving this problem was proposed in a joint paper [2] that Zadeh wrote in collaboration with Richard Bellman, one of the world's leading authorities in optimization.

Their main idea was to explicitly say that what we want is an alternative which is possible *and* optimal. We know the degree $\mu(a)$ to which each alternative is possible. To describe to the degree $\mu_{\rm opt}(a)$ to which an alternative a is optimal, Bellman and Zadeh proposed to use the following formula:

$$\mu_{\rm opt} = \frac{f(a) - m}{M - m}, \tag{3}$$

where m is the absolute minimum of the function $f(a)$ and M is its absolute maximum.

For example, we can define m and M by considering all alternatives for which there is at least some degree of possibility, i.e., for which $\mu(a) > 0$:

$$m = \min_{a:\mu(a)>0} f(a), \quad M = \max_{a:\mu(a)>0} f(a). \tag{4}$$

Once we know the degrees $\mu(a)$ and $\mu_{\rm opt}(a)$ that the alternative a is possible and that this alternative is optimal, to find the degree $d(a)$ to which the alternative a is possible *and* optimal, we can use the usual idea of fuzzy logic – namely, apply an appropriate "and"-operation (t-norm) $f_\&(a,b)$ to these degrees, resulting in

$$d(a) = f_\&(\mu(a), \mu_{\rm opt}(a)). \tag{5}$$

In principle, we can use any "and"-operation: e.g., the operations $\min(a,b)$ and $a \cdot b$ proposed in the very first Zadeh's paper on fuzzy logic, or any more complex operation.

Once we have selected an "and"-operation and computed, for each alternative a, the degree $d(a)$ to which a is desired, a natural idea is to select the alternative for which this degree is the largest possible:

$$d(a_{\rm opt}) = \max_a d(a). \tag{5}$$

Comment. In formulating the formula (5), we do not need to explicitly restrict ourselves to alternatives a for which $\mu(a) > 0$: indeed, if $\mu(a) = 0$, then, by the properties of an "and"-operation, we have $d(a)$ equal to 0 – i.e., to the smallest possible value.

Limitations of the Bellman-Zadeh approach. Degrees $\mu(a)$ describing the person's degree characterize subjective feelings and are, thus, approximate; these values have some accuracy ε. This means that the same subjective feeling can be described by two different values μ and μ', as long as these values differ by no more than ε: $|\mu - \mu'| \le \varepsilon$. In particular, the same small degree of possibility can be characterized by 0 and by a small positive number ε.

It seems reasonable to expect that small – practically indistinguishable – changes in the value of the degrees would lead to small, practically indistinguishable, changes in the solution to the corresponding optimization problem. But, unfortunately, with the Zadeh-Bellman approach, this is not the case.

To show this, let us consider a very simple example when:

- each alternative is characterized by a single number,
- the objective function is simply $f(a) = a$,
- the membership function $\mu(a)$ – e.g., corresponding to "small positive" – is a triangular membership function $\mu(a)$ which is equal to $1 - a$ for $a \in [0, 1]$ and to 0 for all other values a, and
- the "and"-operation is $f_\&(a,b) = a \cdot b$.

In this case, the set $\{a : \mu(u) > 0\}$ is equal to $[0, 1)$, so $m = 0$, $M = 1$, and

$$\mu_{\rm opt}(a) = \frac{a - 0}{1 - 0} = \frac{a}{2}.$$

So,

$$d(a) = f_\&(\mu(a), \mu_{\rm opt}(a)) = (1 - a) \cdot \frac{a}{2} = \frac{a - a^2}{2}.$$

Differentiating this expression with respect to a and equating derivative to 0, we conclude that the maximum of this function is attained when $1 - 2a = 0$, i.e., for $a_{\rm opt} = 0.5$.

On the other hand, if we replace 0 values of the degree $\mu(a)$ for $a \in [-1, 0]$ with a small value $\mu(a) = \varepsilon > 0$, then we get $\{a : \mu(a) > 0\} = [-1, 1)$, so $m = -1$, thus

$$\mu_{\rm opt}(a) = \frac{a - (-1)}{1 - (-1)} = \frac{a + 1}{2}.$$

For $a \le 0$, the product $d(a)$ is increasing, so its maximum has to be attained for $a \ge 0$. For values $a \ge 0$, we have

$$d(a) = f_\&(\mu(a), \mu_{\rm opt}(a)) = (1 - a) \cdot \frac{a + 1}{2} = \frac{1 - a^2}{2}.$$

This is a decreasing function, so its maximum is attained when $a_{\rm opt} = 0$.

So, indeed, an arbitrarily small change in $\mu(a)$ can lead to a drastic change in the selected "optimal" alternative.

What is known about this problem. What we showed is that a change in m can lead to a drastic change in the selected alternative. Interestingly, a change in M is not that critical: for the product "and"-operation $f_\&(a, b) = a \cdot b$, we select an alternative that maximizes the expression

$$d(a) = \mu(a) \cdot \frac{f(a) - m}{M - m}.$$

If we multiply all the values of the maximized constant by the same positive constant $M - m$, its maximum remains attained for the same value a. Thus, it is sufficient to find the alternative that maximized the product $(M - m) \cdot d(a) = \mu(a) \cdot (f(a) - m)$. Good news is that this expression does not depend on M at all.

It turns out (see, e.g., [8]) that $f_\&(a, b)$ is the only "and"-operation for which there is no such dependence. Thus, in the following text, we will use this "and"-operation. On the other hand, in [8], it was also shown that no matter what "and"-operation we select, the result will always depend on m – and thus, will always have the same problem as we described above.

Remaining problem. So, to make sure that the selection does not change much if we make a small change to the membership function $\mu(a)$, we cannot just change the "and"-operation, we need to change the formulas (3) and (4).

What we do in this paper. In this paper, we propose an alternative to the formulas (3) and (4), under which small changes in the degree $\mu(a)$ lead to small changes in the resulting selection.

2 Main Idea and the Resulting Definition

Main idea. We want to use the fact – mentioned several times by Zadeh himself – that the same uncertainty can be described both in terms of the probability density function $\rho(x)$ and in terms of the membership function $\mu(x)$. In both cases, we start with the observed number of cases $N(x)$ corresponding to different values x, but then the procedure differs:

- to get a probability density function, we need to appropriately normalize the values $N(x)$, i.e., take $\rho(x) = c \cdot N(x)$, where the constant c must be determined from the condition that the overall probability is 1:

$$\int \rho(x)\,dx = 1; \tag{6}$$

- to get a membership function, we also need to appropriately normalize the values $N(x)$, i.e., take $\mu(x) = c \cdot N(x)$, where the constant c must be determined from the condition that the largest value of the membership function is 1: $\max\limits_x \mu(x) = 1$.

Because of this possibility, if we start with a membership function, we can normalize it into a probability density function $\rho(x) = c \cdot \mu(x)$ by multiplying all the degrees $\mu(x)$ by an appropriate constant c. One can easily find this constant by substituting $\rho(x) = c \cdot \mu(x)$ into the formula (6). As a result, we get

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

How to use this idea: analysis. Based on the known membership function $\mu(a)$, we can use the usual Zadeh extension principle (see, e.g., [3, 6, 10, 11, 12]) to find the membership function $\nu(x)$ corresponding to the value $x = f(a)$:

$$\nu(x) = \sup_{a: f(a)=x} \mu(a). \tag{7}$$

Based on this membership function, we can find the corresponding probability density function $\rho(x)$ on the set of all the value of the objective function:

$$\rho_X(x) = \frac{\nu(x)}{\int \nu(y)\,dy}. \tag{8}$$

In these terms, a reasonable way to gauge how optimal is an alternative a with the value $X = f(a)$ is by the probability $F(X)$ that a randomly selected value x will be smaller than or equal to X. If this probability is equal to 1, this means that almost all values $f(a')$ are smaller than or equal to $f(a)$ – i.e., that we are practically certain that this alternative a is optimal. The smaller this probability, the less sure we are that this alternative is optimal.

In probability and statistics, the probability $F(X)$ is known as the cumulative distribution function (see, e.g., [13]); it is determined by the formula

$$F(X) = \int_{-\infty}^{X} \rho_X(x)\,dx. \tag{9}$$

Substituting the expression (8) into this formula, we can express $F(X)$ in terms of the membership function $\nu(x)$:

$$F(X) = \frac{\int_{-\infty}^{X} \nu(x)\,dx}{\int \nu(x)\,dx}. \tag{10}$$

The probability $\rho(a)$ that a is possible is also proportional to $\mu(a)$: $\rho(a) = c \cdot \mu(a)$ for an appropriate coefficient c. The probability that an alternative a is possible *and* optimal can be estimated as the product $\rho(a) \cdot F(f(a))$ of the corresponding probabilities. It is therefore reasonable to select an alternative for which this probability is the largest possible. Since c is a positive constant, maximizing the product $\rho(a) \cdot F(f(a)) = c \cdot \mu(a) \cdot F(f(a))$ is equivalent to maximizing a simpler expression $\mu(a) \cdot F(f(a))$. Thus, we arrive at the following idea.

Resulting idea. To select an alternative under fuzzy constraints, we suggest to find the alternative that maximizes the product $\mu(a) \cdot F(f(a))$, where the function $F(X)$ is determined by the formula

$$F(X) = \frac{\int_{-\infty}^{X} \nu(x)\,dx}{\int \nu(x)\,dx}, \tag{10}$$

and the corresponding function $\nu(x)$ is determined by the formula

$$\nu(x) = \sup_{a: f(a) = x} \mu(a). \tag{7}$$

Discussion. One can see that if we make minor changes to the degrees $\mu(a)$, we will get only minor changes to the resulting selection.

Simplest 1-D case. In the 1-D case, when $f(a) = a$, we have $\nu(x) = \mu(x)$ and thus, maximizing the product $\mu(a) \cdot F(f(a))$ – or, equivalently, the product $\rho(a) \cdot F(f(a))$ is equivalent to maximizing the product $\rho(a) \cdot F(a)$.

Interestingly, the standard formula for the probability density function of the skewed generalization of normal distribution – skew-normal distribution – has exactly this form $\rho(a) \cdot F(a)$, where $\rho(a)$ is the probability density function of the normal distribution and $F(a)$ is the corresponding cumulative distribution function; see, e.g., [1, 9].

It is also worth mentioning that, vice versa, fuzzy ideas can be used to explain the formulas for the skew-normal distribution; see, e.g., [5].

Example. In the above example,

$$F(X) = \int_0^X (1-x)\,dx = X - \frac{X^2}{2},$$

so we need to find the value $a_{\rm opt}$ for which the product $(1-a)\cdot\left(a-\frac{a^2}{2}\right)$ attains the largest possible value. Differentiating this expression with respect to a and equating the derivative to 0, we get

$$-\left(a-\frac{a^2}{2}\right)+(1-a)\cdot(1-a)=0,$$

so

$$-a+\frac{a^2}{2}+1-2a+a^2=0,$$

and

$$\frac{3}{2}\cdot a^2-3a+1=0.$$

Thus,

$$a_{\rm opt}=\frac{3\pm\sqrt{9-6}}{3},$$

i.e., taking into account that $a\le 1$, we take

$$a_{\rm opt}=\frac{3-\sqrt{3}}{3}=1-\frac{\sqrt{3}}{3}\approx 0.42.$$

One can see that for small $\varepsilon>0$ we get very close values.

Comment. The original Bellman-Zadeh formula can be described in the same way, but with the cumulative distribution function $F(X)$ corresponding to the uniform distribution on the interval $[m,M]$; see, e.g., [7]. From this viewpoint, our proposal can be viewed as a natural generalization of the original formula, a generalization that takes into account that not all the values from the interval $[m,M]$ are equally possible.

Acknowledgments

This work was supported in part by the National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science) and HRD-1242122 (Cyber-ShARE Center of Excellence).

References

[1] A. Azzalini and A. Capitanio, *The Skew-Normal and Related Families*, Cambridge University Press, Cambridge, Massachusetts, 2013.

[2] R. E. Bellman and L. A. Zadeh, "Decision making in a fuzzy environment", *Management Science*, 1970, Vol. 17, No. 4, pp. B 141–B 164.

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

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

[5] J. G. Flores Muñiz, V. V. Kalashnikov, N. Kalashnykova, O. Kosheleva, and V. Kreinovich, "Why skew normal: a simple pedagogical explanation", *International Journal of Intelligent Technologies and Applied Statistics*, 2018, Vol. 11, No. 2, pp. 113–120.

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

[7] O. Kosheleva and V. Kreinovich, "Why Bellman-Zadeh approach to fuzzy optimization", *Applied Mathematical Sciences*, 2018, Vol. 12, No. 11, pp. 517–522.

[8] V. Kreinovich, O. Kosheleva, and S. Shahbazova, "Which t-norm is most appropriate for Bellman-Zadeh optimization", In: S. N. Shahbazova, J. Kacprzyk, V. E. Balas, and V. Kreinovich (eds.), *Proceedings of the World Conference on Soft Computing*, Baku, Azerbaijan, May 29–31, 2018.

[9] B. Li, D. Shi, and T. Wang, "Some applications of one-sided skew distributions", *International Journal Intelligent Technologies and Applied Statistics*, 2009. Vol. 2, No. 1.

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

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

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

[13] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC, Boca Raton, Florida, 2011.

[14] K. S. Thorne and R. D. Blandford, *Modern Classical Physics: Optics, Fluids, Plasmas, Elasticity, Relativity, and Statistical Physics*, Princeton University Press, Princeton, New Jersey, 2017.

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