2-2018

Which t-Norm Is Most Appropriate for Bellman-Zadeh Optimization

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

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

Shahnaz Shahbazova
Azerbaijan Technical University, shahbazova@gmail.com

Comments:
Technical Report: UTEP-CS-18-13
To appear in *Proceedings of the World Conference on Soft Computing*, Baku, Azerbaijan, May 29-31, 2018.

Which t-Norm Is Most Appropriate for Bellman-Zadeh Optimization

Vladik Kreinovich and Olga Kosheleva
University of Texas at El Paso
500 W. University
El Paso, Texas 79968, USA
vladik@utep.edu, olgak@utep.edu

Shahnaz Shahbazova
Azerbaijan Technical University
Baku, Azerbaijan
shahbazova@gmail.com

***Abstract*—In 1970, Richard Bellman and Lotfi Zadeh proposed a method for finding the maximum of a function under fuzzy constraints. The problem with this method is that it requires the knowledge of the minimum and the maximum of the objective function over the corresponding crisp set, and minor changes in this crisp set can lead to a drastic change in the resulting maximum. It is known that if we use a product "and"-operation (t-norm), the dependence on the maximum disappears. Natural questions are: what if we use other t-norms? Can we eliminate the dependence on the minimum? What if we use a different scaling in our derivation of the Bellman-Zadeh formula? In this paper, we provide answers to all these questions. It turns out that the product is the only t-norm for which there is no dependence on maximum, that it is impossible to eliminate the dependence on the minimum, and we also provide t-norms corresponding to the use of general scaling functions.**

I. Formulation of the Problem

Need for optimization under constraints. In many practical problems, we need to find an optimal alternative $a_{\rm opt}$, optimal in the sense that for this alternative, the value of the corresponding objective function $f(x)$ is the largest possible:

$$f(a_{\rm opt})=\max_{a\in P} f(a),$$

where P denotes the set of all possible alternatives.

Need for optimization under fuzzy constraints. The above formulation works well if we know the set P. In practice, however, for some alternatives a, we may not be absolute sure that these alternatives are possible. For such alternatives, an expert can describe to what extent these alternatives are possible. This description is often made in terms of imprecise ("fuzzy") words from natural language.

To describe such knowledge, it is therefore reasonable to use techniques that Zadeh invented specifically to translate such imprecise knowledge into precise computer-understandable form – namely, the technique of fuzzy logic; see, e.g., [2], [3], [4], [5], [6], [7].

Crudely speaking, we ask each expert to estimate, on a scale, say, from 0 to 10, to what extend each alternative is possible. If an expert marks 7 on a scale of 0 to 10, we say that the expert's degree of confidence that a is possible is $\mu(a)=7/10=0.7$.

This way, for each alternative a, we assign a degree $\mu(a)\in[0,1]$ to which, according to the experts, this alternative is possible. The corresponding function μ is known as a *membership function* or, alternatively, as a *fuzzy set*.

How to optimize under fuzzy constraints. So how can we optimize a function $f(a)$ under fuzzy constraints – described by a membership function $\mu(a)$? This question was raised in a joint paper that L. Zadeh wrote with Richard Bellman, a famous specialist in control [1].

Their main idea is to look for an alternative which is, to the largest extent, both possible *and* optimal. To be more precise:

- first, we need to describe the degree $\mu_{\rm opt}(a)$ to which an alternative is optimal,
- then, for each alternative a, we need to combine the degree $\mu(a)$ to which this alternative is possible and the degree $\mu_{\rm opt}(a)$ to which this alternative is optimal into a single degree to which a is possible *and* optimal;
- finally, we select an alternative $a_{\rm opt}$ for which the combined degree is the largest possible.

Let us start with the first step: finding out to what extent an alternative a is optimal.

Of course, if some alternative has 0 degree of possibility, this means that this alternative is not possible at all, so we should not consider it. So, we should consider only alternatives from the set

$$A\stackrel{\rm def}{=}\{a:\mu(a)>0\}$$

of all alternatives for which there is a non-zero degree of possibility.

If two alternatives a and a' have the same value of the objective function $f(a)=f(a')$, then, intuitively, our degree of confidence that the alternative a is optimal should be the same as our degree of confidence that the alternative a' is possible. Thus, the degree $\mu_{\rm opt}(a)$ should only depend on the value $f(a)$, i.e., we should have $\mu_{\rm opt}(a)=F(f(a))$ for some function $F(x)$.

When the value $f(a)$ is the smallest possible, i.e., when

$$f(a)=\underline{f}\stackrel{\rm def}{=}\min_{a\in A} f(a),$$

then we are absolutely sure that this alternative is not optimal, i.e., that $\mu_{\rm opt}(a)=0$. Thus, we should have $F(\underline{f})=0$.

On the other hand, if for the alternative a, the value of the objective function is the largest possible:

$$f(a) = \overline{f} \stackrel{\text{def}}{=} \max_{a \in A} f(a),$$

then we are absolutely sure that this alternative is optimal, i.e., that $\mu_{\text{opt}}(a) = 1$. Thus, we should have $F(\overline{f}) = 1$.

So, we need to select a function $F(x)$ for which $F(\underline{f}) = 0$ and $F(\overline{f}) = 1$. It is also reasonable to require that the function $F(f)$ increases with f. The simplest such function is linear:

$$F(f(a)) = L(f(a)) \stackrel{\text{def}}{=} \frac{f(a) - \underline{f}}{\overline{f} - \underline{f}},$$

but non-linear functions are also possible. Alternatively, we can have non-linear scaling functions

$$F(f(a)) = S(L(F(a)))$$

for some nonlinear function $S(x)$ for which $S(0) = 0$ and $S(1) = 1$.

To combine the degrees $\mu(a)$ and $F(f(a))$ of the statements "a is possible" and "a is optimal" into a single degree describing to what extent a is both possible *and* optimal, we can use an "and"-operation (t-norm) $f_\&(x, y)$. The most widely used "and"-operations are $\min(x, y)$ and $x \cdot y$.

Thus, we find the alternative a for which the value

$$d(a) = f_\&(\mu(a), F(f(a)))$$

is the largest possible. In particular, if we use a linear scaling function $F(x)$, then we select an alternative a for which the value

$$d(a) = f_\& \left(\mu(a), \frac{f(a) - \underline{f}}{\overline{f} - \underline{f}} \right)$$

is the largest possible.

When $f_\&(x, y) = \min(x, y)$, then we get

$$d(a) = \min \left(\mu(a), \frac{f(a) - \underline{f}}{\overline{f} - \underline{f}} \right).$$

When $f_\&(x, y) = x \cdot y$, then we get

$$d(a) = \mu(a) \cdot \frac{f(a) - \underline{f}}{\overline{f} - \underline{f}}.$$

Problem. The problem with this definition is that it depends on the values $\underline{f}$ and $\overline{f}$ and thus, on the exact shape of the set $A = \{\mu(a) > 0\}$.

In practice, experts have only approximate idea of the corresponding degrees $\mu(a)$, so when $\mu(a)$ is very small, it could be 0, or vice versa. These seemingly minor changes in the membership function can lead to huge changes in the set A and thus, to huge changes in the values $\underline{f}$ and $\overline{f}$.

There is one case when this problem is not so crucial. There is one case when the problem stops being dependent on $\overline{f}$: namely, the case of the product t-norm. Indeed, in this case, maximizing the function $d(a)$ is equivalent to maximizing the function

$$D(a) \stackrel{\text{def}}{=} (\overline{f} - \underline{f}) \cdot d(a)$$

which is equal to

$$D(a) = \mu(a) \cdot (f(a) - \underline{f}),$$

and thus, does not depend on $\overline{f}$ at all.

Natural questions and what we do in this paper. Natural questions are:

- What if we use other t-norms?
- Can we eliminate the dependence on the minimum?
- What if we use a different scaling in our derivation of the Bellman-Zadeh formula?

In this paper, we provide answers to all these questions. It turns out:

- that the product is the only t-norm for which there is no dependence on maximum,
- that it is impossible to eliminate the dependence on the minimum, and
- we also provide t-norms corresponding to the use of general scaling functions.

II. First Result: Product is the Only t-Norm for Which Optimization under Fuzzy Constraints Does Not Depend on $\overline{f}$

Analysis of the problem. Independence on $\overline{f}$ means, in particular, that two alternatives a and a' have the same value of $d(a)$, i.e., that $d(a) = d(a')$, then the same equality holds if we change the value $\overline{f}$. In other words, we want to make sure that if

$$f_\& \left(\mu(a), \frac{f(a) - \underline{f}}{\overline{f} - \underline{f}} \right) = f_\& \left(\mu(a'), \frac{f(a') - \underline{f}}{\overline{f} - \underline{f}} \right),$$

then for a new value $\overline{f}'$, we will also have

$$f_\& \left(\mu(a), \frac{f(a) - \underline{f}}{\overline{f}' - \underline{f}} \right) = f_\& \left(\mu(a'), \frac{f(a') - \underline{f}}{\overline{f}' - \underline{f}} \right).$$

This implication must be true for any membership function $\mu(a)$, for any objective function $f(a)$, and for any values $\overline{f}$ and $\overline{f}'$. Let us denote $A \stackrel{\text{def}}{=} \mu(a)$, $A' \stackrel{\text{def}}{=} \mu(a')$,

$$b \stackrel{\text{def}}{=} \frac{f(a) - \underline{f}}{\overline{f} - \underline{f}}, \quad b' \stackrel{\text{def}}{=} \frac{f(a') - \underline{f}}{\overline{f} - \underline{f}},$$

and

$$k \stackrel{\text{def}}{=} \frac{\overline{f} - \underline{f}}{\overline{f}' - \underline{f}}.$$

In these terms, the desired implication takes the following form:

- if $f_\&(A, b) = f_\&(A', b')$,
- then for every $k > 0$, we have

$$f_\&(A, k \cdot b) = f_\&(A', k \cdot b').$$

Let us analyze which "and"-operations $f_\&(x, y)$ satisfy this property.

Main result of this section: the product $f_\&(x, y) = x \cdot y$ is the only "and"-operation that satisfies the desired implication. We want to prove that the algebraic product $f_\&(x, y)$ is the only "and"-operation that satisfies the desired implication.

Indeed, by the general properties of the "and"-operation, we have

$$f_\&(x, 1) = f_\&(1, x) = x$$

for all x. Thus, the condition $f_\&(A, b) = f_\&(A', b')$ is satisfied for $A = x$, $b = 1$, $A' = 1$, and $b' = x$. So, if the desired implication holds, then, for $k = y$, we get

$$f_\&(x, y \cdot 1) = f_\&(1, y \cdot x),$$

i.e., that $f_\&(x, y) = f_\&(1, y \cdot x)$. Since $f_\&(z) = z$ for all z, we thus conclude that $f_\&(x, y) = x \cdot y$ for all x and y. The statement is proven.

III. Second Result: What If We Use a Non-linear Scaling Function $S(x)$?

Analysis of the problem. What if, instead of a linear scaling function, we use a non-linear function $S(x)$?

In this case, we have

$$d(a) = f_\& \left(\mu(a), \left(\frac{f(a) - \underline{f}}{\overline{f} - \underline{f}} \right) \right).$$

Thus, the desired property takes the following form: if

$$f_\& \left(\mu(a), S\left(\frac{f(a) - \underline{f}}{\overline{f} - \underline{f}} \right) \right) =$$

$$f_\& \left(\mu(a'), S\left(\frac{f(a') - \underline{f}}{\overline{f} - \underline{f}} \right) \right),$$

then for a new value $\overline{f}'$, we will also have

$$f_\& \left(\mu(a), S\left(\frac{f(a) - \underline{f}}{\overline{f}' - \underline{f}} \right) \right) =$$

$$f_\& \left(\mu(a'), S\left(\frac{f(a') - \underline{f}}{\overline{f}' - \underline{f}} \right) \right).$$

If we use the above notations A, a', b, b', and k, then the desired implication takes the following form:

- if $f_\&(A, S(b)) = f_\&(A', S(b'))$,
- then for every $k > 0$, we have

$$f_\&(A, S(k \cdot b)) = f_\&(A', S(k \cdot b')).$$

Let us analyze which "and"-operations $f_\&(x, y)$ satisfy this property.

Let us denote $X \stackrel{\text{def}}{=} S^{-1}(A)$ and $X' \stackrel{\text{def}}{=} S^{-1}(A')$. Then $A = S(X)$, $A' = S(X')$, and the above implication takes the following form:

- if $f_\&(S(X), S(b)) = f_\&(S(X'), S(b'))$,
- then for every $k > 0$, we have

$$f_\&(S(X), S(k \cdot b)) = f_\&(S(X'), S(k \cdot b')).$$

It is known that for every 1-1 continuous monotonic function $S(x) : [0, 1] \to [0, 1]$ and for every "and"-operation $f_\&(x, y)$, the re-scaled function

$$f'_\&(x, y) \stackrel{\text{def}}{=} S^{-1}(f_\&(S(x), S(y))$$

is also an "and"-operation. In terms of this new "and"-operation,

$$f_\&(S(x), S(y)) = S(f'_\&(x, y)).$$

Thus, the desired implication takes the form:

- if $S(f'_\&(x, b)) = S(f'_\&(x', b'))$,
- then for every $k > 0$, we have

$$S(f'_\&(x, k \cdot b)) = S(f'_\&(x', k \cdot b')).$$

Since the scaling function $S(x)$ is increasing, $S(x) = S(y)$ is equivalent to $x = y$. Thus, the desired condition can be further simplified into the following form:

- if $f'_\&(x, b) = f'_\&(x', b')$,
- then for every $k > 0$, we have

$$f'_\&(x, k \cdot b) = f'_\&(x', k \cdot b').$$

In the previous section, we have proven that the only "and"-operation satisfying this condition is $f'_\&(x, y) = x \cdot y$. By definition of the re-scaled function $f'_\&$, this means that

$$S^{-1}(f_\&(S(x), S(y)) = x \cdot y.$$

Applying $S(x)$ to both sides, we conclude that

$$f_\&(S(x), S(y)) = S(x \cdot y).$$

Thus, for any $X \stackrel{\text{def}}{=} S^{-1}(x)$ and $Y \stackrel{\text{def}}{=} S^{-1}(y)$, we have $S(X) = x$, $S(y) = y$ and thus,

$$f_\&(X, Y) = S(x \cdot y) = S(S^{-1}(X) \cdot S^{-1}(Y)).$$

Thus, we arrive at the following conclusion.

Main result of this section: for which "and"-operation is the optimization independent on $\overline{f}$. For each scaling function $S(x)$, the only "and"-operation for which the optimization does not depend on $\overline{f}$ is the operation

$$f_\&(x, y) = S(S^{-1}(x) \cdot S^{-1}(y)).$$

IV. Third Result: It Is Not Possible to Avoid the Dependence on $\underline{f}$

Analysis of the problem. Independence on $\underline{f}$ means, in particular, that two alternatives a and a' have the same value of $d(a)$, i.e., that $d(a) = d(a')$, then the same equality holds if we change the value $\underline{f}$. In other words, we want to make sure that if

$$f_\& \left(\mu(a), \frac{f(a) - \underline{f}}{\overline{f} - \underline{f}} \right) = f_\& \left(\mu(a'), \frac{f(a') - \underline{f}}{\overline{f} - \underline{f}} \right),$$

then for a new value $\underline{f}'$, we will also have

$$f_{\&}\left(\mu(a), \frac{f(a)-\underline{f}'}{\overline{f}-\underline{f}'}\right) = f_{\&}\left(\mu(a'), \frac{f(a')-\underline{f}'}{\overline{f}-\underline{f}'}\right).$$

This implication must be true for any membership function $\mu(a)$, for any objective function $f(a)$, and for any values $\underline{f}$ and $\underline{f}'$. Let us take $\overline{f}=1$ and $\underline{f}=0$, in this case the desired condition takes the following form: if

$$f_{\&}(\mu(a), f(a)) = f_{\&}(\mu(a'), f(a')),$$

then for a new value $\underline{f}'$, we will also have

$$f_{\&}\left(\mu(a), \frac{f(a)-\underline{f}'}{1-\underline{f}'}\right) = f_{\&}\left(\mu(a'), \frac{f(a')-\underline{f}'}{1-\underline{f}'}\right).$$

Let us denote $A \stackrel{\text{def}}{=} \mu(a)$, $A' \stackrel{\text{def}}{=} \mu(a')$, $b \stackrel{\text{def}}{=} f(a)$, $b' \stackrel{\text{def}}{=} f(a')$, and $f_0 \stackrel{\text{def}}{=} \underline{f}'$. In these terms, the desired implication takes the following form:

- if $f_{\&}(A, b) = f_{\&}(A', b')$,
- then for every $f_0 \in (0,1)$, we have

$$f_{\&}\left(A, \frac{b-f_0}{1-f_0}\right) = f_{\&}\left(A', \frac{b'-f_0}{1-f_0}\right).$$

Let us take any A and any $b<1$. Then, for $A' = f_{\&}(A.b)$ and for $b'=1$, we have

$$f_{\&}(A', b') = f_{\&}(A', 1) = A' = f_{\&}(A, b).$$

Thus, due to the desired property, for $f_0 = b$, we have

$$f_{\&}\left(A, \frac{b-b}{1-b}\right) = f_{\&}\left(A', \frac{1-b}{1-b}\right),$$

i.e.,

$$f_{\&}(A, 0) = f_{\&}(A', 1).$$

By the properties of the "and"-operation, we have $f_{\&}(A,0) = 0$ and $f_{\&}(A',1) = A'$, thus we conclude that $A'=0$. But A' is equal to $f_{\&}(A.b)$, so we get $f_{\&}(A,b) = 0$ for all A and $b<1$ – which is impossible for a continuous "and"-operation. So, we arrive at the following conclusion.

Main result of this section: for fuzzy optimization, it is not possible to get rid of the dependence on $\underline{f}$. No matter what "and"-operation we use, it is not possible to avoid the dependence of the optimization result on the value $\underline{f}$.

Acknowledgments

This work was supported in part by the US National Science Foundation grant HRD-1242122.

References

[1] 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.
[2] R. Belohlavek, J. W. Dauben, and G. J. Klir, *Fuzzy Logic and Mathematics: A Historical Perspective*, Oxford University Press, New York, 2017.
[3] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.
[4] J. M. Mendel, *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*, Springer, Cham, Switzerland, 2017.
[5] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2006.
[6] V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer, Boston, Dordrecht, 1999.
[7] L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.