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

Olga Kosheleva[1], Vladik Kreinovich[1], and Shahnaz Shahbazova[2]

[1]University of Texas at El Paso
El Paso, TX 79968, USA
olgak@utep.edu, vladik@utep.edu
[2]Azerbaijan Technical University
Baku, Azerbaijan
shahbazova@gmail.com

# 1. Need for Optimization Under Constraints

- In many practical problems:
  - we need to find an optimal alternative $a_{\text{opt}}$
  - among all alternatives from the set $P$ of all possible ones.
- Optimal means that the value of the corresponding objective function $f(x)$ is the largest possible:

$$f(a_{\text{opt}}) = \max_{a \in P} f(a).$$

## 2. Need for Fuzzy Constraints

- The above formulation works well if we know the set $P$.
- In practice, for some alternatives $a$, we are not 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.
- Zadeh invented fuzzy logic specifically:
  - to translate such imprecise natural-language knowledge
  - into precise computer-understandable form.
- E.g., we ask each expert to estimate, on a scale, say, 0 to 10, to what extend each alternative is possible.

## 3. Need for Fuzzy Constraints (cont-d)

- 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:
  - to 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 $\mu$ is known as a *membership function* or, alternatively, as a *fuzzy set*.

## 4. How to Optimize Under Fuzzy Constraints

- How to optimize a function $f(a)$ under fuzzy constraints – described by a membership function $\mu(a)$?
- This question was raised in a joint paper of L. Zadeh and Richard Bellman, a famous specialist in control.
- 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.

## 5. Optimizing Under Fuzzy Constraints (cont-d)

- Finally, we select an alternative $a_{\text{opt}}$ 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.
- So, we should consider only alternatives from the set

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

- 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.

# 6. Optimizing Under Fuzzy Constraints (cont-d)

- Thus, the degree $\mu_{\rm opt}(a)$ should only depend on the value $f(a)$.
- In other words, we should have $\mu_{\rm opt}(a) = F(f(a))$ for some function $F(x)$.
- Here:
  - when the value $f(a)$ is the smallest possible, i.e., when $f(a) = \underline{f} \stackrel{\rm def}{=} \min\limits_{a\in A} f(a)$,
  - then we are absolutely sure that this alternative is not optimal, i.e., that $\mu_{\rm opt}(a) = 0$.
- Thus, we should have $F(\underline{f}) = 0$.

# 7. Optimizing Under Fuzzy Constraints (cont-d)

- On the other hand:
  - if the value $f(a)$ is the largest possible: $f(a) = \overline{f} \stackrel{\text{def}}{=} \max\limits_{a \in A} f(a)$,
  - 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:

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

# 8. Optimizing Under Fuzzy Constraints (cont-d)

- However, non-linear functions are also possible.
- We can also have $F(f(a)) = S(L(F(a)))$ for some non-linear scaling f-n $S(x)$ for which $S(0) = 0$ and $S(1) = 1$.
- We need:
  - 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.
- For this, 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.

# 9. Optimizing Under Fuzzy Constraints (cont-d)

- If we use a linear scaling function $F(x)$, then we select $a$ for which the following value is the largest:

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

- 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

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

## 10. A Problem

- The problem with this definition is that it depends on the values $\underline{f}$ and $\overline{f}$.
- Thus, it depends on the exact shape of the set

$$A = \{a : \mu(a) > 0\}.$$

- 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 $A$.
- Thus, they can lead to huge changes in the values $\underline{f}$ and $\overline{f}$.

# 11. Case When This Problem Is Not So Crucial and Related Questions

- 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

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

- This new function does not depend on $\overline{f}$ at all.
- 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?

# 12. Our Answers

- In this talk, 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.

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

- Independence on $\overline{f}$ means, in particular, that:
  - if 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 replace $\overline{f}$ with $\overline{f}'$:

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

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

- This implication must be true for any $\mu(a)$, for any $f(a)$, and for any $\overline{f}$ and $\overline{f}'$.

# 14. Product is the Only t-Norm (cont-d)

- Let us denote $A \stackrel{\text{def}}{=} \mu(a)$, $A' \stackrel{\text{def}}{=} \mu(a')$,
$$b \stackrel{\text{def}}{=} \frac{f(a) - \underline{f}}{\overline{f} - \underline{f}}, \quad b' \stackrel{\text{def}}{=} \frac{f(a') - \underline{f}}{\overline{f} - \underline{f}}, \quad k \stackrel{\text{def}}{=} \frac{\overline{f} - \underline{f}}{\overline{f}' - \underline{f}}.$$
- In these terms, the desired implication takes the following form: for all $A$, $b$, $A'$, $b'$, and $k$:
$$\text{if } f_\&(A, b) = f_\&(A', b'), \text{ then } f_\&(A, k\cdot b) = f_\&(A', k\cdot b').$$
- Let us analyze which "and"-operations $f_\&(x, y)$ satisfy this property.
- 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 $b' = x$.

## 15. Product is the Only t-Norm (cont-d)

- Reminder: for for $A = x$, $b = 1$, $A' = 1$, and $b' = x$, we get $f_\&(x, 1) = f_\&(1, x)$.
- 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 $f_\&(1, z) = z$ for all $z$, we thus conclude that $f_\&(x, y) = x \cdot y$ for all $x$ and $y$.
- The statement is proven.

## 16. What If We Use a Non-Linear Scaling Function $S(x)$?

- Then, $d(a) = f_\&\left(\mu(a), S\left(\dfrac{f(a)-\underline{f}}{\overline{f}-\underline{f}}\right)\right).$
- Thus, the desired property 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 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
  $$f_\&(S(X), S(k\cdot b)) = f_\&(S(X'), S(k\cdot b')).$$

## 17. Second Result (cont-d)

- It is known that $f'_\&(x,y) \stackrel{\text{def}}{=} S^{-1}(f_\&(S(x),S(y))$ 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 $S(f'_\&(x,k\cdot b)) = S(f'_\&(x',k\cdot b'))$ for all $k>0$.
- 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 $f'_\&(x,k\cdot b) = f'_\&(x',k\cdot b')$ for all $k$.

## 18. Second Result (cont-d)

- We have proven that the only "and"-operation satisfying this condition is $f'_\&(x,y)=x\cdot y$.
- By definition of $f'_\&$, 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 $X\stackrel{\rm def}{=}S^{-1}(x)$ and $Y\stackrel{\rm def}{=}S^{-1}(y)$, we have $S(X)=x$, $S(y)=y$ and thus,

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

- So, the only "and"-operation for which the optimization does not depend on $\overline{f}$ is

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

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

- Independence on $\underline{f}$ means, in particular, that

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

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

- This implication must be true for any $\mu(a)$, for any $f(a)$, and for any values $\underline{f}$ and $\underline{f}'$.
- Let us take $\overline{f}=1$ and $\underline{f}=0$.
- Then, if $f_\&(\mu(a), f(a)) = f_\&(\mu(a'), f(a'))$, then

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

## 20. Third Result (cont-d)

- 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 $f_0 \in (0, 1)$:

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

- 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

$$f_\& \left( A, \frac{b - b}{1 - b} \right) = f_\& \left( A', \frac{1 - b}{1 - b} \right), \text{ i.e., } f_\&(A, 0) = f_\&(A', 1).$$

## 21. Third Result (cont-d)

- 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 $b<1$.
- On the other hand, $f_\&(A,1)=A>0$.
- This is not possible for a continuous "and"-operation.
- So, it is not possible to avoid the dependence of the optimization result on the value $\underline{f}$.

# 22. Acknowledgments

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