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

Olga Kosheleva[1], Vladik Kreinovich[1],
and Hoang Phuong Nguyen[2]
[1]University of Texas at El Paso, 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

1. 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:

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

- 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.

2. Optimization Under Constraints (cont-d)

- However, 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 the alternatives that satisfy all the given constraints.

3. Optimization Under Constraints (cont-d)

- In this case, instead of the original unconstrained optimization problem, we have a modified problem.
- We need to select an alternative $a_{\rm opt} \in A$ for which the value of the objective function $f(a)$ is the largest:

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

4. 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 Lotfi Zadeh's fuzzy logic.
- This technique was designed to translate imprecise ("fuzzy") knowledge from natural language to numbers.

5. Need for Fuzzy Constraints (cont-d)

- 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 confidence in this alternative.
- How can we optimize the objective function $f(a)$ under such fuzzy constraints?

6. Bellman-Zadeh Approach: A Brief Reminder

- The most widely used approach to solving this problem was proposed in a joint paper:
  - by Lotfi Zadeh, founder of fuzzy logic, and
  - by 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.
- Bellman and Zadeh proposed $\mu_{\rm opt} = \dfrac{f(a) - m}{M - m}$.
- Here m is the absolute minimum of the function $f(a)$ and M is its absolute maximum.

7. Bellman-Zadeh Approach (cont-d)

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

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

- To find the degree $d(a)$ to which a is possible *and* optimal, we can use an "and"-operation (t-norm) $f_\&(a, b)$:

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

- 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.

8. Bellman-Zadeh Approach (cont-d)

- Then we select the alternative for which $d(a)$ is the largest possible: $d(a_{\rm opt}) = \max\limits_{a} d(a)$.
- 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., equal to the smallest possible value.

9. 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:

$$|\mu - \mu'| \le \varepsilon.$$

- For example, 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 optimization problem.

10. Limitations (cont-d)

- But, unfortunately, with the Zadeh-Bellman approach, this is not the case.
- To show this, let us consider a very simple example.
- 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 triangular:
  - with $\mu(a) = 1 - a$ for $a \in [0, 1]$ and
  - with $\mu(a) = 0$ for all other values a.
- 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 $\mu_{\text{opt}}(a) = \dfrac{a - 0}{1 - 0} = a$.

11. Limitations (cont-d)

- So, $d(a) = f_{\&}(\mu(a), \mu_{\rm opt}(a)) = (1-a)\cdot a = a - a^2$.
- Differentiating this expression w.r.t. a and equating derivative to 0, we get $1-2a=0$, i.e., $a_{\rm opt} = 0.5$.
- If we replace 0 values of the degree $\mu(a)$ for $a \in [-1,0]$ with a small value $\mu(a)=\varepsilon>0$, then:

$$\{a : \mu(a) > 0\} = [-1, 1).$$

- So, $m=-1$, thus $\mu_{\rm opt}(a) = \dfrac{a-(-1)}{1-(-1)} = \dfrac{a+1}{2}$.
- For $a \le 0$, the product $d(a)$ is increasing, so its maximum has to be attained for $a \ge 0$.

12. Limitations (cont-d)

- For values $a \geq 0$, we have

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

- This is a decreasing function, so its maximum is attained when $a_{\text{opt}} = 0$.
- So, indeed, an arbitrarily small change in $\mu(a)$ can lead to a drastic change in the "optimal" alternative.

13. 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.
- Indeed, 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 $d(a)$ by a constant $M - m > 0$, its maximum is 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.

14. What Is Known (cont-d)

- It turns out 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, 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.

15. Remaining Problem

- So, to make sure that the selection does not change much if we make a small change to $\mu(a)$:
  - we cannot just change the "and"-operation,
  - we need to change the formula
$$d(a) = f_\& \left(\mu(a), \frac{f(a) - m}{M - m} \right).$$
- In this talk, we propose an alternative formula.
- In this formula, small changes in the degree $\mu(a)$ lead to small changes in the resulting selection.

16. Main Idea

- Zadeh mentioned several times 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

$$\rho(x) = c \cdot N(x).$$

- The constant c must be determined from the condition that the overall probability is 1: $\int \rho(x)\,dx = 1$.

17. Main Idea (cont-d)

- To get a membership function, we also need to appropriately normalize the values $N(x)$, i.e., take

$$\mu(x) = c \cdot N(x).$$

- Here c must be determined from the condition that the largest value of the membership function is 1:

$$\max_x \mu(x) = 1.$$

- Because of this possibility:
  - if we start with a membership function,
  - we can normalize it into a probability density function $\rho(x) = c \cdot \dfrac{\mu(x)}{\int \mu(y)\,dy}$.

18. How to Use This Idea: Analysis

- We know the membership function $\mu(a)$.
- We can use Zadeh extension principle to find the membership function for $x = f(a)$: $\nu(x) = \sup\limits_{a:f(a)=x} \mu(a)$.
- Based on this membership function, we can find the corresponding probability density function:

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

- 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.

19. How to Use This Idea (cont-d)

- If this probability is equal to 1, this means that almost all values $f(a')$ are smaller than or equal to $f(a)$.
- So, 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.
- It is determined by the formula $F(X) = \int_{-\infty}^{X} \rho_X(x)\,dx$.
- Substituting the expression for $\rho_X(x)$ into this formula, we can express $F(X)$ in terms of $\nu(x)$:

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

20. How to Use This Idea (cont-d)

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

21. Resulting Idea

- We want to select an alternative under fuzzy constraints.
- We suggest to find the alternative that maximizes the product $\mu(a) \cdot F(f(a))$, where $F(X) = \dfrac{\int_{-\infty}^{X} \nu(x)\,dx}{\int \nu(x)\,dx}$.
- The corresponding function $\nu(x)$ is determined by the formula $\nu(x) = \sup\limits_{a: f(a) = x} \mu(a)$.
- One can see that:
  - if we make minor changes to the degrees $\mu(a)$,
  - we will get only minor changes to the resulting selection.

22. Discussion

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

23. Simplest 1-D Case

- In the 1-D case, when $f(a) = a$, we have $\nu(x) = \mu(x)$.
- Thus, maximizing $\nu(a) \cdot F(f(a)) = \mu(a) \cdot F(a)$ is equivalent to maximizing $\rho(a) \cdot F(a)$.
- Interestingly, the probability density function of the skew-normal distribution is $\rho(a) \cdot F(a)$, where:
  - $\rho(a)$ is the probability density function of the normal distribution and
  - $F(a)$ is the corresponding cumulative distribution function.
- It is also worth mentioning that, vice versa, fuzzy ideas can be used to explain the skew-normal formulas.

24. Example

- In the above example, $F(X) = \int_0^X (1-x)\,dx = X - \dfrac{X^2}{2}$.
- So we need to find the value $a_{\rm opt}$ for which the product $(1-a)\cdot\left(a - \dfrac{a^2}{2}\right)$ 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, $-a + \dfrac{a^2}{2} + 1 - 2a + a^2 = 0$, thus $\dfrac{3}{2}\cdot a^2 - 3a + 1 = 0$, and $a_{\rm opt} = \dfrac{3 \pm \sqrt{9-6}}{3}$.
- Since $a \le 1$, we get $a_{\rm opt} = \dfrac{3-\sqrt{3}}{3} = 1 - \dfrac{\sqrt{3}}{3} \approx 0.42$.

25. 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),
- HRD-1242122 (Cyber-ShARE Center of Excellence).