Constraint Approach to Multi-objective Optimization

Martine Ceberio, Olga Kosheleva and Vladik Kreinovich

1 Formulation of the Problem

Multi-objective optimization: examples. In many practical situations, we would like to maximize several different criteria.

For example, in meteorology and *environmental research*, it is important to measure fluxes of heat, water, carbon dioxide, methane and other trace gases that are exchanged within the atmospheric boundary layer. To perform these measurements, researchers build up vertical towers equipped with sensors at different heights; these tower are called *Eddy flux* towers. When selecting a location for the Eddy flux tower, we have several criteria to satisfy; see, e.g., [1, 5]: The station should located as far away as possible from roads, so that the gas flux generated by the cars do not influence our measurements of atmospheric fluxes. On the other hand, the station should be located as close to the road as possible, so as to minimize the cost of carrying the heavy parts when building such a station. The inclination at the station location should be small, because otherwise, the flux will be mostly determined by this inclination and will not be reflective of the atmospheric processes, etc.

In *geophysics*, different type of data provide complementary information about the Earth structures. For example, information from the body waves (P-wave receiver functions) mostly covers deep areas, while the information about the Earth surface is mostly contained in surface waves. To get a good understanding of the Earth structure, it is therefore important to take into account data of different types; see, e.g., [3, 9].

If we had only one type of data, then we can use the usual Least Squares approach $f_i(x) \to$ min to find a model that best fits the data. If we knew the relative accuracy

M. Ceberio (✉) · O. Kosheleva · V. Kreinovich
University of Texas at El Paso, El Paso, TX 79968, USA
e-mail: mceberio@utep.edu

O. Kosheleva
e-mail: olgak@utep.edu

V. Kreinovich
e-mail: vladik@utep.edu

of different data types, we could apply the Least Squares approach to all the data. In practice, however, we do not have a good information about the relative accuracy of different data types. In this situation, all we can say that we want to minimize the errors $f_i(x)$ corresponding to all the observations i.

Multi-objective optimization is difficult. The difficulty with this problem is that, in contrast to a simple optimization, the problem of multi-objective optimization is not precisely defined. Indeed, if we want to minimize a single objective $f(x) \to$ min, this has a very precise meaning: we want to find an alternative x_0 for which $f(x_0) \leq f(x)$ for all other alternatives x. Similarly, if we want to maximize a single objective $f(x) \to$ max, this has a very precise meaning: we want to find an alternative x_0 for which $f(x_0) \geq f(x)$ for all other alternatives x.

In contrast, for a multi-objective optimization problem

$$f_1(x) \to \text{min}; \quad f_2(x) \to \text{min}; \quad \ldots; \quad f_n(x) \to \text{min} \tag{1}$$

or

$$f_1(x) \to \text{max}; \quad f_2(x) \to \text{max}; \quad \ldots; \quad f_n(x) \to \text{max}, \tag{2}$$

no such precise meaning is known.

Let us illustrate this ambiguity on the above trip example. In many cases, convenient direct flights which save on travel time are more expensive, while a cheaper trip may involve a long stay-over in between flights. So, if we find a trip that minimizes cost, the trip takes longer. Vice versa, if we minimize the travel time, the trip costs more.

It is therefore necessary to come up with a way to find an appropriate compromise between several objectives.

2   Analysis of the Problem and Two Main Ideas

Analysis of the problem. Without losing generality, let us consider a multi-objective maximization problem. In this problem, ideally, we would like to find an alternative x_0 that satisfies the constraints $f_i(x_0) \geq f_i(x)$ for all objectives i and for all alternatives x. In other words, in the ideal case, if we select an alternative x at random, then with probability 1, we satisfy the above constraint.

Main ideas. The problem is that we cannot satisfy all these constraints with probability 1. A natural idea is thus to find x_0 for which the probability of satisfying these constraints is as high as possible. Let us describe two approaches to formulating this idea (i.e., the corresponding probability) is precise terms.

First approach: probability to satisfy all n constraints. The first approach is to look for the probability that for a randomly selected alternative x, we have $f_i(x_0) \geq f_i(x)$ *for all i*.

Second approach: probability to satisfy a randomly selected constraint. An alternative approach is to look for the probability that for a randomly selected alternative x and *for a randomly selected objective i*, we have $f_i(x_0) \geq f_i(x)$.

How to formulate these two ideas in precise terms. To formulate the above two ideas in precise terms, we need to estimate two probabilities:

- the probability $p_I(x_0)$ that for a randomly selected x, we have $f_i(x_0) \geq f_i(x)$ for all i, and
- the probability $p_{II}(x_0)$ that for a randomly selected x and a randomly selected i, we have $f_i(x_0) \geq f_i(x)$.

Let us estimate the first probability. Since we do not have any prior information about the dependence between different objective functions $f_i(x)$ and $f_j(x)$, $i \neq j$, it is reasonable to assume that the events $f_i(x_0) \geq f_i(x)$ and $f_j(x_0) \geq f_j(x)$ are independent for different i and j. Thus, the desired probability $p_I(x_0)$ that all n such inequalities are satisfied can be estimated as the product $p_I(x_0) = \prod_{i=1}^{n} p_i(x_0)$ of n probabilities p_i of satisfying the corresponding inequalities.

So, to estimate p, it is sufficient to estimate, for every i, the probability $p_i(x_0)$ that $f_i(x_0) \geq f_i(x)$ for a randomly selected alternative x.

How can we estimate this probability $p_i(x_0)$? Again, in general, we do not have much prior knowledge of the i-th objective function $f_i(x)$. What do we know? Before starting to solve this problem as a multi-objective optimization problem, we probably tried to simply optimize each of the objective functions—hoping that the corresponding solution would also optimize all other objective functions. Since we are interesting in maximizing, this means that we know the largest possible value M_i of each of the objective functions: $M_i = \max_x f_i(x)$.

In many practical cases, the optimum can be attained by differentiating the objective function and equating all its derivatives to 0. This is, for example, how the Least Squares method works: to optimize the quadratic function that describes how well the model fits the data, we solve the system of linear equations obtained by equating all partial derivatives to 0. It is important to mention that when we consider the points where all the partial derivatives are equal to 0, we find not only maxima but also minima of the objective function. Thus, it is reasonable to assume that in the process of maximizing each objective function $f_i(x)$, in addition to this function's maximum, we also compute its minimum $m_i = \min_x f_i(x)$.

Since we know the smallest possible value m_i of the objective function $f_i(x)$, and we know its largest possible value M_i, we thus know that the value $f_i(x)$ corresponding to a randomly selected alternative x must lie inside the interval $[m_i, M_i]$.

In effect, this is all the information that we have: that the random value $f_i(x)$ is somewhere in the interval $[m_i, M_i]$. Since we do not have any reason to believe that some values from this interval are more probable and some values are less probable, it is reasonable to assume that all the values from this interval are equally probable, i.e., that we have a uniform distribution on the interval $[m_i, M_i]$.

This argument—known as Laplace Indeterminacy Principle—can be formalized as selecting the distribution with the probability density $\rho(x)$ for which the entropy $S = -\int \rho(x) \cdot \ln(\rho(x)) \, dx$ is the largest possible. One can check that for distributions on the given interval, the uniform distribution is the one with the largest entropy [6].

For the uniform distribution on the values $f_i(x) \in [m_i, M_i]$, the probability $p_i(x_0)$ that the random value $f_i(x)$ does not exceed $f_i(x_0)$, i.e., belongs to the subinterval $[m_i, f_i(x_0)]$, is equal to the ratio of the corresponding intervals, i.e., to $p_i(x_0) = \dfrac{f_i(x_0) - m_i}{M_i - m_i}$. Thus, the desired probability $p_I(x_0)$ is equal to the product of such probabilities. So, we arrive at the following precise formulation of the first idea:

Precise formulation of the first idea. To solve a multi-objective optimization problem (2), we find a value x_0 for which the product $p_I(x_0) = \prod\limits_{i=1}^{n} \dfrac{f_i(x_0) - m_i}{M_i - m_i}$ attains the largest possible value, where $m_i \stackrel{\text{def}}{=} \min\limits_{x} f_i(x)$ and $M_i \stackrel{\text{def}}{=} \max\limits_{x} f_i(x)$.

Let us estimate the second probability. In the second approach, we select the objective function f_i at random. Since we have no reason to prefer one of the n objective functions, it makes sense to select each of these n functions with equal probability $\dfrac{1}{n}$.

For each selection of the objective function i, we know the probability $p_i(x_0) = \dfrac{f_i(x_0) - m_i}{M_i - m_i}$ that we will have $f_i(x_0) \geq f_i(x)$ for a randomly selected alternative x. The probability of selecting each objective function $f_i(x)$ is equal to $\dfrac{1}{n}$. Thus, we can use the complete probability formula to compute the desired probability $p_{II}(x_0)$:

Precise formulation of the second idea. To solve a multi-objective optimization problem (2), we find a value x_0 for which the expression $p_{II}(x_0) = \sum\limits_{i=1}^{n} \dfrac{1}{n} \cdot \dfrac{f_i(x_0) - m_i}{M_i - m_i}$ attains the largest possible value.

Discussion. Let us show that both ideas lead to known (and widely used) methods for solving multi-objective optimization problems.

The second idea leads to optimizing a linear combination of objective functions. Let us start with analyzing the second idea, since the resulting formula with the sum looks somewhat simpler than the product-based formula corresponding to the first idea.

In the case of the second idea, the optimized value $p_{II}(x_0)$ is a linear combination of n objective functions—to be more precise, it is an arithmetic average of the objective functions normalized in such a way that their values are within the interval $[0, 1]$: $p_{II}(x_0) = \dfrac{1}{n} \cdot \sum\limits_{i=1}^{n} f_i'(x_0)$, where $f_i'(x_0) \stackrel{\text{def}}{=} \dfrac{f_i(x_0) - m_i}{M_i - m_i}$.

Maximizing a linear combination of the objective functions is indeed the most widely used approach to solving multi-objective optimization problems; see, e.g., [4].

The first idea leads to maximizing a product of (normalized) objective functions. One can easily see that the first idea leads to maximizing a product of normalized objective functions: $p_I(x_0) = \prod_{i=1}^{n} f_i'(x_0)$.

Maximizing such a product is exactly what Bellman-Zadeh fuzzy approach recommends (if we use the product as an "and" operation); see, e.g., [2, 8]. It fits will with our own proposal for such a situation; see, e.g., [5].

This is also exactly what the the Nobelist John Nash recommended for a similar situation of making a group decision when each participant would like to optimize his/her own utility $f_i(x) \to$ max; see, e.g., [7].

References

1. Aubinet, M., Vesala, T., Papale, D. (eds.): Eddy Covariance - A Practical Guide to Measurement and Data Analysis. Springer, Dordrecht (2012)
2. Bellman, R.E., Zadeh, L.A.: "Decision making in a fuzzy environment". Manag. Sci. **17**(4), B141–B164 (1970)
3. Bodin, T., Sambridge, M., Tkalcic, H., Arroucau, P., Gallagher, K., Rawlinson, N.: "Transdimensional inversion of receiver functions and surface wave dispersion". J. Geophys. Res. Vol. **117**, doi:10.1029/2011JB008560 (2012)
4. Figueira, J., Greco, S., Ehrgott, M. (eds.): Multiple Criteria Decision Analysis: State of the Art Surveys. Kluwer, Dordrecht (2004)
5. Jaimes, A., Tweedie, C., Kreinovich, V., Ceberio, M.: Scale-invariant approach to multi-criterion optimization under uncertainty, with applications to optimal sensor placement, in particular, to sensor placement in environmental research. Int. J. Reliab. Saf. **6**(1–3), 188–203 (2012)
6. Jaynes, E.T., Bretthorst, G.L.: Probability Theory: The Logic of Science. Cambridge University Press, Cambridge, UK (2003)
7. Nash, J.: "Two-Person Cooperative Games". Econometrica **21**, 128–140 (1953)
8. Nguyen, H.T., Walker, E.A.: A First Course in Fuzzy Logic. Chapman and Hall/CRC, Boca Raton, Florida (2006)
9. Thompson, L., Velasco, A., Kreinovich, V.: "Construction of ShearWave models by applying multi-objective optimization to multiple geophysical data sets". In: Tost, G.O., Vasilieva, O. (eds.) Analysis, Modelling, Optimization, and Numerical Techniques. Springer, Berlin (2015, to appear)