Why Convex Optimization Is Ubiquitous and Why Pessimism Is Widely Spread

Angel F. Garcia Contreras, Martine Ceberio, and Vladik Kreinovich

Abstract In many practical applications, the objective function is convex. The use of convex objective functions makes optimization easier, but ubiquity of such objective function is a mystery: many practical optimization problems are not easy to solve, so it is not clear why the objective function – whose main goal is to describe our needs – would always describe easier-to-achieve goals. In this paper, we explain this ubiquity based on the fundamental ideas about human decision making. This explanation also helps us explain why in decision making under uncertainty, people often make pessimistic decisions, i.e.., decisions based more on the worst-case scenarios.

1 Why Convex Optimization Is Ubiquitous

Reasonable decision making means optimization. In many real life situations, we need to make a decision, i.e., we need to select an alternative x out of many possible alternatives.

Decision making theory has shown that the decision making of a rational person is equivalent to maximizing a special function $u(x)$ – known as *utility* – that describes this person's preferences; see, e.g., [1, 5, 6, 8]. Thus, maximization problems are very important for practical applications.

In many cases, the utility value is described by its monetary equivalent amount.

Small changes in an alternative should lead to small change in preferences, so the function $u(x)$ is usually continuous.

What if an optimization problem has several solutions? From the purely mathematical viewpoint, it is possible that an optimization problem has several solu-

Angel F. Garcia Contreras, Martine Ceberio, and Vladik Kreinovich
University of Texas at El Paso, El Paso, Texas 79968, USA
e-mail: afgarciacontreras@miners.utep.edu, mceberio@utep.edu, vladik@utep.edu

tions, i.e., several different alternatives $x^{(1)}, x^{(2)}, \ldots$ all maximize the objective function $u(x)$:

$$u(x^{(1)}) = u(x^{(2)}) = \ldots = \max_x u(x).$$

From the practical viewpoint, however, the fact that, by using some criterion, we get several possible solutions, means that we can use this non-uniqueness to optimize something else. For example, if a company selects a design x for a new plant, and several designs $x^{(1)}$, $x^{(2)}$, $\ldots$ are equally profitable, then a reasonable idea is to select, among these most-profitable solutions, the one which is, e.g., the most environmentally friendly. This will weed out some of the possible designs. If even after taking into account environmental impact, we still have several possible alternatives, we can use the remaining non-uniqueness to optimize something else – e.g., look for the most aesthetically pleasing design. This process continues until we end up with the single optimal alternative.

In other words, if the objective function $u(x)$ allows several optimal solutions, this means, from the practical viewpoint, that we need to modify our preferences – i.e., in effect, modify the corresponding objective function – until we end up with an objective function that attains its maximum at the unique point.

So, while, from the mathematical viewpoint, we can consider arbitrary objective functions $u(x)$ – and they can serve as good approximations to the way we make decisions – the *final* objective function, the function that describes exactly how we actually make decisions, should have the unique maximum.

How can we describe such final objective functions? In general, selecting a decision x involves selecting the values of many different parameters $x_1, \ldots, x_n$ that characterize this decision. For example, when we select a design of a plant, we must take into account the land area that we need to purchase, the amount of steel and concrete that goes into construction, the overall length of roads, pipes, etc. forming the supporting infrastructure, etc.

Our original decision x is based on known costs of all these attributes. However, costs can change. If the cost per unit of the i-th attribute changes by the value d_i, then the overall cost of an option x changes from the original value $u(x)$ to the new value

$$u'(x) = u(x) + \sum_{i=1}^{n} d_i \cdot x_i. \tag{1}$$

It is therefore reasonable to select an objective function $u(x)$ in such away that not only the original function $u(x)$ has the unique maximum, but that for all possible combinations of values d_i, the resulting combination (1) also has the unique maximum.

Need to consider constraints. From the purely mathematical viewpoint, we often consider *unconstrained* optimization, where we have no prior restrictions on the values of the parameters $x_1, \ldots, x_n$ that describe the desired solution $x = (x_1, \ldots, x_n)$. In practice, there are always physical and economical restrictions on the possible values of these parameters. As a result, in practice, for each parameter x_i, we always have bounds $\underline{x}_i$ and $\overline{x}_i$, and we only consider values x_i from the corresponding intervals $[\underline{x}_i, \overline{x}_i]$.

Once we take into account the existence of constraints, we can always guarantee that the corresponding optimization problem always has a solution: indeed, on a bounded closed set $B = [\underline{x}_1, \overline{x}_1] \times \ldots \times [\underline{x}_n, \overline{x}_n]$, every continuous function attaints its maximum at some point $x \in B$.

Thus, we arrive at the following definition.

Definition 1. *A continuous function $u(x) = u(x_1, \ldots, x_n)$ is called a* final objective function *if for every combination of tuples $d = (d_1, \ldots, d_n)$, $\underline{x} = (\underline{x}_1, \ldots, \underline{x}_n)$, and $\overline{x} = (\overline{x}_1, \ldots, \overline{x}_n)$ the following constrained optimization problem has the unique solution:*

$$\textit{Maximize } u(x) + \sum_{i=1}^{n} d_i \cdot x_i \textit{ under constraints } \underline{x}_i \le x_i \le \overline{x}_i.$$

Discussion. There is a class of functions which are realistic objective functions in the sense of the above definition – namely, the class of *strictly convex* functions $u(x)$, i.e., functions for which $u\left(\frac{x+x'}{2}\right) > \frac{u(x)+u(x')}{2}$ for all $x \ne x'$; see, e.g., [9]. Indeed, it is easy to prove that for a strictly convex function, maximum is attained at a unique point: if we have two different points $x \ne x'$ at which $u(x) = u(x') = \max\limits_x u(x)$, then, due to strong convexity, for the midpoint $x'' \stackrel{\text{def}}{=} \frac{x+x'}{2}$, we would have $u(x'') > u(x) = u(x')$, i.e., we would have $u(x'') > \max\limits_x u(x)$, which is not possible.

One can also easily check that if a function $u(x)$ is strictly convex, and if we add a linear expression $\sum\limits_{i=1}^{n} d_i \cdot x_i$ to this function, then the resulting sum $u'(x)$ is also strictly convex. Thus, strictly convex functions are indeed final objective functions in the sense of Definition 1.

Interestingly, if we restrict ourselves to smooth (at least three times differentiable) functions, the opposite is also true: only convex objective functions are final in the sense of the above definition.

Proposition 1. *Every smooth final objective function $u(x)$ is convex.*

Comments.

- This result explains why convex objective functions are ubiquitous in practical applications; see, e.g. [9].
- This result is also good for practical applications since, while optimization in general is NP-hard, feasible algorithms are known for solving convex optimization problem; see, e.g., [4, 7].

Proof of Proposition 1. Let us prove this by contradiction. Let us assume that there exists a smooth final objective function $u(x)$ which is not convex. A smooth function is convex if and only if at all points, its matrix of second derivatives is non-positive definite [9]. Since $u(x)$ is not convex, there exists a point p at which this matrix is not non-negative definite. At this point, the Taylor expansion of the function $u(x)$ has the form

$$u(x) = u(p) + \sum_{i=1}^{n} u_{,i} \cdot (x_i - p_i) + \frac{1}{2} \cdot \sum_{i=1}^{n} \sum_{j=1}^{n} u_{,ij} \cdot (x_i - p_i) \cdot (x_j - p_j) + o((x-p)^2),$$

where $u_{,i} \stackrel{\text{def}}{=} \dfrac{\partial u}{\partial x_i}$ and $u_{,ij} \stackrel{\text{def}}{=} \dfrac{\partial^2 u}{\partial x_i \partial x_j}$. Thus, the function $u'(x) = u(x) - \sum\limits_{i=1}^{n} u_{,i} \cdot x_i$ has the form $u'(x) = q(x) + o((x-p)^2)$, where

$$q(x) \stackrel{\text{def}}{=} u'(p) + \frac{1}{2} \cdot \sum_{i=1}^{n} \sum_{j=1}^{n} u_{,ij} \cdot (x_i - p_i) \cdot (x_j - p_j).$$

Let us take $\underline{x}_i = x_i^{(0)} - \varepsilon$ and $\overline{x}_i = x_i^{(0)} + \varepsilon$ for some small $\varepsilon > 0$. Then, for small $\varepsilon > 0$, $u(x)$ is very close to $q(x)$.

Non-negative definite would mean that $\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{n} u_{,ij} \cdot (x_i - p_i) \cdot (x_j - p_j) \le 0$ for all x_i. The fact that the matrix $u_{,ij}$ is not non-negative definite means that there exists a vector $x_i - p_i$ for which $\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{n} u_{,ij} \cdot (x_i - p_i) \cdot (x_j - p_j) > 0$. So, for a vector proportional to $x_i - p_i$ and which is within the box B, we have $q(x) > q(p)$. Thus, the maximum of the function $q(x)$ on the box B is *not* attained at p. Since the function $q(x)$ does not change if we reverse the sign of all the differences $x_i - p_i$, with each point $x = p + (x - p)$, the same maximum is attained at a different point $p - (x - p)$. So, for the function $q(x)$, the maximum is attained in at least two different points.

Let us now consider the original function $u'(x)$. If its maximum is attained at two different points, we get our contradiction. Let us now assume that its maximum m is attained at a single point y. This maximum is close to a maximum of the function $q(x)$. The fact that this function has only one maximum means that the value of $u'(x)$ at the point $p - (y - p)$ is slightly smaller than the value $m = u'(y)$. We can then take the plane (linear function) $u = m$, and, keeping its value to be m at the point y, we slightly rotate it and lower it until we touch some other point on the graph – close to $p - (y - p)$. This is possible for $q(x)$, thus it is possible for any function which is sufficiently close to $q(x)$ – in particular, for a function $u'(x)$ corresponding to a sufficiently small value $\varepsilon > 0$. Thus, we get a sum $u''(x)$ of $u'(x)$ and a linear function that has at least two maxima. Since $u'(x)$ is itself a sum of $u(x)$ and a linear function, this means that $u''(x)$ is also a sum of $u(x)$ and a linear function – so we get a contradiction with our assumption that the function $u(x)$ is a final objective function.

The proposition is proven.

2 Why Pessimism Is Widely Spread

Decision making under uncertainty. In many practical situations, we do not know the exact consequences of different actions. In other words, for each alternative x, instead of a single value $u(x)$, we have several different values $u(x,s)$ depending on the situation s. According to decision theory, in such situation, a reasonable idea is to optimize the so-called Hurwicz criterion

$$U(x) = \alpha \cdot \max_s u(x,s) + (1-\alpha) \cdot \min_s u(x,s)$$

for some $\alpha \in [0,1]$; see, e.g., [2, 3, 5]. Here, $\alpha = 1$ corresponds to the optimistic approach, when we only consider the best-case scenarios, $\alpha = 0$ is pessimistic approach, when we only consider the worst cases, and $\alpha \in (0,1)$ means that we consider both the best and the worst cases.

When is this convex? From the viewpoint described in the previous section, it is reasonable to consider situations in which $u(x,s)$ is convex for every s *and* the objective function $U(x)$ is also convex.

For $\alpha = 0$, it is easy to show that the minimum of convex function is always convex; see, e.g., [9]. For $\alpha = 0.5$, we get the arithmetic average which is also convex. For $\alpha < 0.5$, we get a convex combination of cases $\alpha = 0$ and $\alpha = 0.5$, so we also get a convex functions.

However, for any $\alpha > 0.5$, this is no longer true. For example, let us take $s \in \{-,+\}$, with

$$u(x,+) = |x-1| \text{ and } u(x,-) = |x+1|.$$

Then, for every $\alpha > 0.5$, the function $U(x)$ attains its maximum value α at two difference points. Thus, $U(x)$ is not convex.

This explains why pessimism is widely spread. The fact that only in the pessimistic approach we can guaranteed that the resulting objective function is final explains why the pessimistic approach ($\alpha \le 0.5$) is widely spread.

Acknowledgments. This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721, and by an award "UTEP and Prudential Actuarial Science Academy and Pipeline Initiative" from Prudential Foundation.

References

1. P. C. Fishburn, *Utility Theory for Decision Making*, John Wiley & Sons Inc., New York, 1969.
2. L. Hurwicz, *Optimality Criteria for Decision Making Under Ignorance*, Cowles Commission Discussion Paper, Statistics, No. 370, 1951.
3. V. Kreinovich, "Decision making under interval uncertainty (and beyond)", In: P. Guo and W. Pedrycz (eds.), *Human-Centric Decision-Making Models for Social Sciences*, Springer Verlag, 2014, pp. 163–193.
4. V. Kreinovich, A. Lakeyev, J. Rohn, and P. Kahl, *Computational Complexity and Feasibility of Data Processing and Interval Computations*, Kluwer, Dordrecht, 1998.
5. R. D. Luce and R. Raiffa, *Games and Decisions: Introduction and Critical Survey*, Dover, New York, 1989.

6. H. T. Nguyen, O. Kosheleva, and V. Kreinovich, "Decision making beyond Arrow's 'impossibility theorem', with the analysis of effects of collusion and mutual attraction", *International Journal of Intelligent Systems*, 2009, Vol. 24, No. 1, pp. 27–47.
7. P. Pardalos, *Complexity in Numerical Optimization*, World Scientific, Singapore, 1993.
8. H. Raiffa, *Decision Analysis*, Addison-Wesley, Reading, Massachusetts, 1970.
9. R. T. Rockafeller, *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1997.