Why Convex Optimization Is Ubiquitous and Why Pessimism Is Widely Spread

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

Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, Texas 79968, USA
afgarciacontreras@miners.utep.edu,
mceberio@utep.edu@utep.edu, vladik@utep.edu

1. Decision Making Means Optimization

- In many real life situations, we need to make a decision, i.e., select an alternative x out of many.
- Decision making theory has shown that:
  - the decision making of a rational person
  - is equivalent to maximizing a special function $u(x)$ (*utility*) that describes this person's preferences.
- 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 $u(x)$ is continuous.

2. What If the Problem has Several Solutions?

- The optimization problem can have several solutions:

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

- From the practical viewpoint, we can use this non-uniqueness to optimize something else.
- E.g., if several designs $x^{(1)}$, $x^{(2)}$, ... are equally profitable, we select the most environmentally friendly one.
- If we still have several possible alternatives, we can, e.g., look for the most aesthetically pleasing design.
- This process continues until we end up with the single optimal alternative.
- So, the *final* objective function should have the unique maximum.

3. How to Describe Final Objective Functions?

- In general, selecting a decision x involves selecting the values of many different parameters $x_1, \ldots, x_n$.
- 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.

4. Describing Final Objective Functions (cont-d)

- If the cost per unit of the i-th attribute changes by the value d_i, then the overall cost of x changes to

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

- It is therefore reasonable to select an objective function $u(x)$ in such away that:
  - for all possible combinations of values d_i,
  - the resulting combination also has the unique maximum.

5. Need to Consider Constraints

- In practice, there are always physical and economical restrictions on the possible values of these parameters.
- As a result, for each parameter x_i, we always have bounds $\underline{x}_i$ and $\overline{x}_i$, so $x_i \in [\underline{x}_i, \overline{x}_i]$.
- Under such constraints, the 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 $u(x)$ attaints its maximum.

6. Definition and Discussion

- 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:

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

- This is true for *strictly convex* functions $u(x)$, for which $u\left(\frac{x + x'}{2}\right) > \frac{u(x) + u(x')}{2}$ for all $x \neq x'$.

7. Discussion (cont-d)

- 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 $x \neq x'$ at which $u(x) = u(x') = \max\limits_x u(x)$,
  - then, due to strong convexity, for the midpoint $x'' \stackrel{\text{def}}{=} \dfrac{x + x'}{2}$, we would have $u(x'') > u(x) = u(x')$;
  - this would imply $u(x'') > \max\limits_x u(x)$, which is not possible.
- If $u(x)$ is strictly convex, it remains strictly convex after adding $\sum\limits_{i=1}^{n} d_i \cdot x_i$.
- Thus, strictly convex functions are indeed final objective functions.
- Interestingly, they are the only ones.

8. Main Result

- **Proposition.** *Every smooth final objective function $u(x)$ is convex.*
- This result explains why convex objective functions are ubiquitous in practical applications.
- 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.

9. Decision Making Under Uncertainty

- In many practical situations, we do not know the exact consequences of different actions.
- So, for each alternative x, we have several different values $u(x,s)$ depending on the situation s.
- According to decision theory, a reasonable idea is to optimize the so-called Hurwicz criterion

$$U(x) = \alpha \cdot \max_s u(x,s) + (1-\alpha) \cdot \min_s u(x,s) \text{ for some } \alpha \in [0,1].$$

- 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.
- $\alpha \in (0,1)$ means that we consider both the best and the worst cases.

10. When Is This Convex?

- We showed that we should 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.
- For $\alpha = 0.5$, we get arithmetic average – also convex.
- Case $\alpha < 0.5$ is a convex combination of $\alpha = 0$ and $\alpha = 0.5$, so also convex.
- However, for $\alpha > 0.5$, this is no longer true:
- E.g., for $u(x,+) = |x-1|$ and $u(x,-) = |x+1|$, the function $U(x)$ attains maximum for two different x.
- Thus, $U(x)$ is not convex.

11. This Explains Why Pessimism Is Widely Spread

- We showed that:
  - only in the pessimistic approach ($\alpha \le 0.5$)
  - we can guaranteed that the resulting objective function is final.
- This explains why the pessimistic approach is widely spread.
- I.e., why in many real-life situations, decision makers make decisions based on the worst-case scenarios.

12. 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 from Prudential Foundation.

13. Proof of Proposition

- 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.
- Since $u(x)$ is not convex, there exists a point p at which this matrix is not non-positive definite.
- At p, the Taylor expansion of $u(x)$ has the form

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

- Here, we denoted $u_{,i} \stackrel{\text{def}}{=} \dfrac{\partial u}{\partial x_i}$ and $u_{,ij} \stackrel{\text{def}}{=} \dfrac{\partial^2 u}{\partial x_i \partial x_j}$.

14. Proof (cont-d)

- Thus, the function $u'(x) = u(x) - \sum_{i=1}^{n} u_{,i} \cdot x_i$ 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 $\underline{x}_i = p_i - \varepsilon$ 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

$$\sum_{i=1}^{n} \sum_{j=1}^{n} u_{,ij} \cdot (x_i - p_i) \cdot (x_j - p_j) \le 0 \text{ for all } x_i.$$

15. Proof (cont-d)

- The fact that the matrix $u_{,ij}$ is not non-negative definite means that there exists a vector $x_i - p_i$ for which
$$\sum_{i=1}^{n}\sum_{j=1}^{n} u_{,ij}\cdot(x_i-p_i)\cdot(x_j-p_j) > 0.$$
- 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.
- The function $q(x)$ does not change if we reverse the sign of all the differences $x_i - p_i$.
- So, with each point $x = p+(x-p)$, the same maximum is attained at a point $p-(x-p)\neq x$.
- So, for the function $q(x)$, the maximum is attained in at least two different points.

16. Proof (cont-d)

- 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 $q(x)$.
- 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 $m = u'(y)$.

17. Proof (cont-d)

- 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, it is possible 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.
- $u'(x)$ is itself a sum of $u(x)$ and a linear function.
- Thus, $u''(x)$ is also a sum of $u(x)$ and a linear function.

18. Proof (final)

- So, a linear combination of $u(x)$ and a linear function has two maxima.
- Thus, we get a contradiction with our assumption that the function $u(x)$ is a final objective function.
- The proposition is proven.