Low-Complexity Zonotopes Can Enhance Uncertainty Quantification (UQ)

Olga Kosheleva and Vladik Kreinovich
University of Texas at El Paso
500 W. University, El Paso, Texas 79968, USA
olgak@utep.edu, vladik@utep.edu

1. Main Objective of Science and Engineering

- What do we want?
- We want to predict what will happen in the future – this is what science does.
- We want to select the actions that will leads to the best possible future.
- This is, crudely speaking, what engineering is for.
- Both to predict the future state of the world and to select the best action, we must have information:
  - about the current state of the world,
  - i.e., about the values of all the quantities that characterize this state.
- This information mostly comes from measurements.

2. Main Objective of Science and Engineering (cont-d)

- We want to predict the future value y of a quantity or to describe each control parameter y.
- For this, we use the known relation $y = f(x_1, \ldots, x_N)$ between:
  - this future value (or control parameter) and
  - the current values of several related quantities

$$x_1, \ldots, x_N.$$

- So, we measure the values of the quantities $x_1, \ldots, x_N$; then we:
  - apply the algorithm $f(x_1, \ldots, x_N)$ to the results $\widetilde{x}_1, \ldots, \widetilde{x}_N$ of measuring the quantities $x_1, \ldots, x_N$,
  - and return the value $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_N)$.

3. Need for Uncertainty Quantification

- Measurements are never absolutely accurate.
- The result $\widetilde{x}$ of a measurement is, in general, different from the actual (unknown) value x of the quantity.
- In other words, the *measurement error* $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x$ is, in general, different from 0.
- In general, the measurement result $\widetilde{x}_i$ is, in general, different from x_i.
- So, our estimate $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_N)$ based on the measurement results:
  - is, in general, different
  - from the desired value $y = f(x_1, \ldots, x_N)$.
- How different can they be?

4. Need for Uncertainty Quantification (cont-d)

- What can we say about the estimation error

$$\Delta y \stackrel{\text{def}}{=} \widetilde{y} - y?$$

- This is very important to know in many practical situations.
- For example, suppose that we are prospecting for oil.
- We estimated that in some location, there is $\widetilde{y} = 150$ million tons.
- What shall we do? It depends on the accuracy of this estimate.
- If $y = 150 \pm 20$, this is very good news; we should dig a well and start producing oil.

5. Need for Uncertainty Quantification (cont-d)

- On the other hand, if $y = 150 \pm 200$, then maybe at this location, there is no oil at all.
- In this case, it is better to perform some additional measurements first.
- This will decrease the risk of wasting money on the expensive well.
- We need to estimate the approximation error Δy based on the known information about Δx_i.
- This is one of the main problems of uncertainty quantification.

6. Traditional Probability-Based Approach to Uncertainty Quantification and Its Limitations

- Traditional engineering approach to uncertainty quantification assumes that:
  - we know the probability distributions
  - of each measurement error Δx_i.
- And indeed, in many real-life situations we have this knowledge.
- However, there are many important practical situations when we do not know these probabilities.
- To explain why, let us recall where the information about the probabilities comes from.

7. Probability-Based Approach (cont-d)

- In the ideal world, for each measuring instrument, we should compare, several times:
  - the measurement result $\widetilde{x}$ with
  - the actual value x of the corresponding quantity.
- For each such comparison, we shall compute the measurement error $\Delta x = \widetilde{x} - x$.
- After a sufficient number of measurements, we would get a large sample of values Δx.
- Based on this sample, we will then be able to find the corresponding probability distribution.
- Of course, in reality, we never know the exact actual values of the physical quantities.

8. Probability-Based Approach (cont-d)

- However, in many cases:
  - there exists another – much more accurate – measuring instrument,
  - whose measurement error Δx_s is much smaller than the measuring error of the tested instrument.
- Such much-more-accurate measuring instruments are known as *standard* ones.
- In this case, with high accuracy:
  - the value $\widetilde{x}_s$ measured by the standard measuring instrument
  - is approximately equal to the actual value.
- The difference $\delta x \stackrel{\text{def}}{=} \widetilde{x} - \widetilde{x}_s$ is approximately equal to the desired measurement error Δx.

9. Probability-Based Approach (cont-d)

- Thus, we can measure, several times, the same quantities by both measuring instruments.
- Then we can use the resulting sample to find the probability distribution of the corr. measurement error.
- In many cases, such a calibration is indeed performed, and we get the corresponding probability distributions.
- However, in many other cases, such a calibration is not done.
- Thus, we do not know the corresponding probabilities.
- There are two main reasons why calibration is not done.

10. Probability-Based Approach (cont-d)

- The first reason is that sometimes:
  - we use state-of-the-art measuring instruments,
  - for which no other instrument is more accurate.
- This happens a lot in advanced science.
- E.g., it would be nice if near the Hubble telescope, we would have a 5 times more accurate instrument.
- However, the Hubble telescope is the best we have.
- This often happens in applications as well.
- For example, geophysical companies often use state-of-the-art measuring equipment.

11. Probability-Based Approach (cont-d)

- This equipment costs money; however:
  - it is still cheaper to use such expensive measuring instruments
  - than to risk wasting even more money
  - e.g., on drilling oil well where there is no oil at all.
- The second reason is more mundane.
- Yes, potentially, in a manufacturing plant, we can, in principle:
  - calibrate all the sensors, and
  - get the corresponding probability distributions.
- However, there is a problem.

12. Probability-Based Approach (cont-d)

- Many sensors are very cheap nowadays.
- Kids play with robotic toys that measure distances to the walls etc. as they go.
- These sensors can be bought for a few bucks.
- However, calibrating each sensor requires access to an expensive accurate measuring instrument.
- It would thus cost several orders of magnitude more than the sensor itself.
- This is too expensive for a manufacturing plant – which usually operates at a low profit margin.

13. Enter Interval Uncertainty

- If we do not know probabilities of different values of measurement error Δx, what do we know?
- For a device to be called a measuring instrument, we need to know:
  - at least some upper bound Δ
  - on the absolute value of the measurement error:

$$|\Delta x| \le \Delta.$$

- If we do not even know such an upper bound, this means that:
  - after a measurement by this instrument,
  - we cannot say anything about the actual value of the measured quantity,
  - it can be as far away from the measurement result as we can imagine.

14. Enter Interval Uncertainty (cont-d)

- In other words, what such a device would produce is a wild guess, not a measurement result.
- Thus, such a bound is always produced by the manufacturer of the measuring instrument.
- And if we cannot find the probabilities of different values Δx, this upper bound is all we know.
- In this case, once we know the measurement result $\widetilde{x}$:
  - the only information that we have about the actual value x of the measured quantity
  - is that this value is somewhere in the interval

$$[\widetilde{x} - \Delta, \widetilde{x} + \Delta].$$

- Such uncertainty is naturally called *interval uncertainty.*

15. Need for Interval Computations

- Let us go back to the situation when:
  - instead of the actual (ideal) value
$$y = f(x_1, \ldots, x_N),$$
  - we have an estimate $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_N)$ based on the measurement results $\widetilde{x}_1, \ldots, \widetilde{x}_N$.
- Suppose that for each of N measurements, we only know the upper bound Δ_i: $|\Delta x_i| \le \Delta_i$.
- Then, all we know about the actual value y is that it is equal to $f(x_1, \ldots, x_N)$ for some $x_i \in [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$.
- Thus, all we can say about the value y is that it belongs to the set
$$Y \stackrel{\text{def}}{=} \{f(x_1, \ldots, x_N) : x_i \in [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i] \text{ for all } i\}.$$
- For continuous functions $f(x_1, \ldots, x_N)$ this set is also an interval.

16. Need for Interval Computations (cont-d)

- The problem of computing the endpoint of this interval is known as the problem of *interval computations*.
- In general, the interval computation problem is NP-hard.
- This means that:
  - unless P = NP (which most computer scientists do not believe to be true),
  - it is not possible to have a feasible algorithm that solves all particular cases of this problems.
- However, in many practical situations, there exist efficient algorithms that:
  - either compute the desired range Y
  - or at least compute a good approximation to Y.

17. Possibility of Linearization

- A feasible algorithm for UQ is possible is when the measurement errors Δx_i are reasonably small.
- And usually, they are reasonable small.
- In this case, we can use one of the main ideas of computations in physics:
  - expand the corresponding expression in Taylor series in terms of the corresponding small quantities,
  - and keep only linear terms in this expansion.
- In our case, by definition of the measurement error $\Delta x_i = \widetilde{x}_i - x_i$, we have $x_i = \widetilde{x}_i - \Delta x_i$, thus:

$$\Delta y = f(\widetilde{x}_1, \ldots, \widetilde{x}_N) - f(x_1, \ldots, x_N) =$$

$$f(\widetilde{x}_1, \ldots, \widetilde{x}_N) - f(\widetilde{x}_1 - \Delta x_1, \ldots, \widetilde{x}_N - \Delta x_N).$$

18. Possibility of Linearization (cont-d)

- Expanding the expression in the right-hand side in Taylor series in terms of Δx_i, we get

$$f(\widetilde{x}_1-\Delta x_1,\ldots,\widetilde{x}_N-\Delta x_N) = f(\widetilde{x}_1,\ldots,\widetilde{x}_N)-\sum_{i=1}^{N} c_i\cdot\Delta x_i,$$

- Here we denoted $c_i \stackrel{\text{def}}{=} \dfrac{\partial f}{\partial x_i}$.
- Thus, the above formula takes the following form:

$$\Delta y = \sum_{i=1}^{N} c_i \cdot \Delta x_i.$$

- In this linearized case, we can feasibly compute the bounds on Δy.
- Indeed, each measurement error Δx_i takes values from the interval $[-\Delta_i, \Delta_i]$.

19. Possibility of Linearization (cont-d)

- Different measurement errors do not depend on each other.
- So, the largest possible value of the sum is attained when each term $c_i \cdot \Delta x_i$ is the largest possible.
- The corresponding linear function $c_i \cdot \Delta x_i$ is increasing when $c_i > 0$ and decreasing when $c_i < 0$.
- When $c_i > 0$, the largest possible value of $c_i \cdot \Delta x_i$ is attained when Δx_i is the largest possible: $\Delta x_i = \Delta_i$.
- The resulting largest value of the quantity $c_i \cdot \Delta x_i$ is equal to $c_i \cdot \Delta_i$.
- When $c_i < 0$, the largest possible value of the quantity $c_i \cdot \Delta x_i$ is attained when Δx_i is the smallest possible:

$$\Delta x_i = -\Delta_i.$$

20. Possibility of Linearization (cont-d)

- The resulting largest value of the quantity $c_i \cdot \Delta x_i$ is equal to $-c_i \cdot \Delta_i$.
- In both cases, the largest possible value of the quantity $c_i \cdot \Delta x_i$ is equal to $|c_i| \cdot \Delta_i$.
- Thus, the largest possible value Δ of the sum Δx is:

$$\Delta = \sum_{i=1}^{N} |c_i| \cdot \Delta_i.$$

- By using this formula, we can explicitly compute Δ in N steps – i.e., in feasible time.
- Strictly speaking, this algorithm is feasible.
- Still, when we have a large number N of inputs, it requires a large amount of computation time.
- There exist more efficient algorithms for computing Δ.

21. Need to Estimate the Joint Uncertainty of Several Data Processing Results

- All the above discussions are about estimating a *single* quantity y.
- In reality, we usually estimate *several* different characteristics $y_1, \ldots, y_n$ based on the same data $\widetilde{x}_1, \ldots, \widetilde{x}_N$:

$$\widetilde{y}_1 = f_1(\widetilde{x}_1, \ldots, \widetilde{x}_N), \ldots, \widetilde{y}_n = f_n(\widetilde{x}_1, \ldots, \widetilde{x}_N).$$

- For example, when we predict weather:
  - we do not just predict temperature at one location,
  - we predict temperature, wind speed and direction, and humidity at several locations.
- What is the accuracy of the resulting estimations?
- In other words, what can we say about the corresponding approximation errors

$$\Delta y_j \stackrel{\text{def}}{=} \widetilde{y}_j - f_j(x_1, \ldots, x_N).$$

22. Joint Uncertainty (cont-d)

- In many practical situations, we only know the upper bounds on the measurement errors.
- So, we have interval uncertainty, for which:
  - the only information that we have about each measurement error Δx_i
  - is the upper bound Δ_i on its absolute value:

$$|\Delta x_i| \le \Delta_i.$$

- Also, in many practical situations, measurement errors are relatively small.
- So, we can ignore quadratic (and higher order) terms in the Taylor expansions.

23. Joint Uncertainty (cont-d)

- Then, we get the linearized formulas

$$\Delta y_1 = c_{1,1} \cdot \Delta x_1 + \ldots + c_{1,N} \cdot \Delta x_N;$$

$$\ldots$$

$$\Delta y_n = c_{n,1} \cdot \Delta x_1 + \ldots + c_{n,N} \cdot \Delta x_N.$$

- Here we denoted $c_{j,i} \stackrel{\text{def}}{=} \dfrac{\partial f_j}{\partial x_i}$.
- For each value y_j, we can use the above techniques and find the interval of possible values of Δy_j.
- However, this is not enough.
- We also need to also know:
  - what combinations of the values $(y_1, \ldots, y_n)$ are possible.
  - i.e., equivalently, of the approximation errors $(\Delta y_1, \ldots, \Delta y_n)$ are possible.

24. Joint Uncertainty (cont-d)

- For example, in some cases, the future temperature in two nearby locations can range from 15 to 25 degrees.
- However:
  - unless these two locations are separated by a mountain – as we have in our city of El Paso,
  - the temperatures at these two locations cannot differ too much.
- We can have $(15, 16)$ and even, probably, $(15, 17)$.
- However, we cannot have $(15, 25)$.
- How can we take this into account? How can we describe the corresponding set of tuples $(y_1, \ldots, y_n)$?
- This is the problem that we analyze in this talk.

25. General Approach to Solving Problems: Reduce And/Or Reformulate

- A usual approach to solving a new problem is to try to find similar problems:
  - that have been already solved,
  - or at least for which there are some partial solutions.
- Sometimes, we cannot immediately come up with such a similar somewhat-solved problem.
- Then, a natural idea is to try to reformulate our problems in equivalent terms.
- This often makes it easier to find a similar problem.

26. Enter Zonotopes

- For our problem, this reformulation becomes possible if we reformulate the formulas in vector terms, as

$$\Delta y = c_1 \cdot \Delta x_1 + \ldots + c_N \cdot \Delta x_N,$$

- Here $\Delta x_i \in [-\Delta_i, \Delta_i]$, and we denoted

$$\Delta y \stackrel{\text{def}}{=} (\Delta y_1, \ldots, \Delta y_n) \text{ and } c_i \stackrel{\text{def}}{=} (c_{1,i}, \ldots, c_{n,i}).$$

- For each i, the set S_i of all the vectors $\Delta x_i \cdot c_i$ for $\Delta x_i \in [-\Delta_i, \Delta_i]$ forms a straight line segment.
- This segment connects the points $\Delta_i \cdot c_i$ and $-\Delta_i \cdot c_i$.
- We want to find the set S of all possible values of the sum.
- This set is thus to the set of all possible sums of vectors from the corresponding sets S_i:

$$S = \{s_1 + \ldots + s_N : s_1 \in S_1, \ldots, s_N \in S_N\}.$$

27. Enter Zonotopes (cont-d)

- In geometry, this construction is known as the *Minkowski sum* of the sets $S_1, \ldots, S_N$.
- This sum is denoted by $S = S_1 + \ldots + S_N$.
- The Minkowski sum of several straight line segments is known as a *zonotope*.
- Thus, our conclusion is that the desired set of possible values of the tuple $\Delta y = (\Delta y_1, \ldots, \Delta y_n)$ is a zonotope.
- So:
  - to solve our main problem – of estimating the joint uncertainty of several data processing results
  - we need to be able to deal with zonotopes.

28. An Interesting Observation: Every Zonotope Can Be Thus Represented

- We have shown that every set of possible values of the tuple $(\Delta y_1, \ldots, \Delta y_n)$ is a zonotope.
- Thus, the set of possible values of the tuple $(y_1, \ldots, y_n)$ is also a zonotope.
- Let us show that, vice versa, every zonotope can be thus represented.
- Indeed, in the above representation, we use straight-line segments centered at 0.
- Every straight-line segment T_i can be represented as the sum $T_i = m_i + S_i$ of:
  - its midpoint m_i, and
  - a segment $S_i \stackrel{\text{def}}{=} T_i - m_i$ centered at 0.

29. Every Zonotope Can Be Thus Represented (cont-d)

- This segment connects its endpoints $c_i = (c_{i,1}, \ldots, c_{i,N})$ and $-c_i = (-c_{i,1}, \ldots, -c_{i,N})$.
- Thus, each zonotope $T = T_1 + \ldots + T_N$ can be represented as $T = (m_1 + \ldots + m_N) + (S_1 + \ldots + S_N)$.
- The set $S_1 + \ldots + S_n$ can be interpreted as the set of possible approximation errors for the algorithm

$$f_i(x_1, \ldots, x_N) = c_{i,1} \cdot x_1 + \ldots + c_{i,N} \cdot x_n, \text{ for } \Delta_1 = \ldots = \Delta_N = 1.$$

- Thus indeed:
  - every zonotope can be represented
  - as the set of possible tuples $(y_1, \ldots, y_n)$ for some data processing algorithm.

30. Historical Comment

- The idea of using zonotopes was known already in 2005.
- In 2005, it was shown that:
  - for a *specific* data processing algorithm.
  - namely, for the least square estimation under interval uncertainty,
  - the resulting set of possible tuples is a zonotope.
- In this talk, we show that this is true for *all* data processing algorithms.
- We also show that, vice versa, every zonotope can be thus represented.

31. How To Deal with Zonotopes: What Is Known

- In computational geometry, there are several efficient algorithms for dealing with zonotopes.
- The main problem with these algorithms is that:
  - the exact description of the uncertainty-related zonotope in an n-dimensional space
  - requires as many n-dimensional parameters c_i as there are measured quantities $x_1, \ldots, x_N$.
- In many practical problems, e.g., in seismology, N is in thousands.
- So this description becomes difficult to process.

32. What We Propose: General Idea

- The possibility to make computations easier comes from the fact that:
  - the number n of desired properties $y_1, \ldots, y_n$
  - is much smaller than the number N of inputs.
- So, to speed up computations, we propose to use the known results about the possibility of approximating zonotopes:
  - with "low-complexity" sets,
  - i.e., with sets determined by a much smaller number of n-dimensional parameters.
- By approximating a set S, we mean, as usual, producing a set A for which, for some small number δ:
  - each element $s \in S$ is δ-close to some $a \in A$, and
  - each element $a \in A$ is δ-close to some $s \in S$.

33. General Idea (cont-d)

- In mathematics, this closeness is usually described by saying that:
  - the Hausdorff distance $d_H(A, S)$ between the sets S and A
  - is smaller than or equal to δ.
- $d_H(A, S)$ is defined as the small distance for which the above two conditions are true.
- We can also say that the set A approximates the set S with *relative accuracy* ε if $d_H(A, S) \leq \delta \cdot \text{diam}(S)$.
- Here the diameter $\text{diam}(S)$ is defined as the largest distance between two points from the set S.
- This is a natural generalization of the width of an interval and the diameter of a sphere to general sets.

34. First Simplifying Result

- Each n-dimensional zonotope S can be approximated:
  - with any given relative accuracy $\varepsilon > 0$,
  - by a "low-complexity" zonotope, which is the sum of $N' = c(\varepsilon) \cdot n \cdot (\log(n))^3$ segments;
  - here, c is a constant depending on ε.
- Since $n \ll N$, the new number of n-dimensional vector parameters is much smaller than N.
- Thus, the problem becomes easier to handle.

35. Second Simplifying Result

- Each symmetric convex polyhedron – in particular, each zonotope – can be approximated:
  - with any given relative accuracy $\varepsilon > 0$,
  - by a convex polyhedron with "low" number of vertices $v_1, \ldots, v_{N'}$.
- Namely, it can be approximated by a polyhedron for which:
  - the number N' is also bounded, from above,
  - by the value $c(\varepsilon) \cdot n \cdot (\log(n))^d \ll N$, for some small constant d.

36. Second Simplifying Result (cont-d)

- In this case, the approximating set A is the convex combination of these vertices v_j.
- In other words, all elements $a = (a_1, \ldots, a_n)$ from the approximating set A have the form

$$a = c_1 \cdot v_1 + \ldots + c_{N'} \cdot v_{N'}, \text{ where}$$

$$c_i \geq 0 \text{ and } c_1 + \ldots + c_{N'} = 1.$$

37. Remaining Open Problems

- The results that we propose to use are effective.
- They drastically reduce the complexity of the corresponding problem.
- However, at present, they are not supported:
  - by efficient computational algorithms
  - for producing the corresponding approximations.
- These results are still useful:
  - we spend time on computing the approximation only once, and
  - then, we can enjoy the benefits of this reduction for every single way we want to process this set.
- However, it would be nice to have efficient algorithms for producing the corresponding approximations.

38. Remaining Open Problems (cont-d)

- Designing such efficient algorithms is the main open problem that we want to emphasize.
- We have shown – such algorithms will be very helpful:
  - not just in somewhat obscure computational geometry problems,
  - but also in generic problems of uncertainty quantification
- We hope that this will hopefully encourage researchers to design such algorithms.

39. 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-1834620 and HRD-2034030 (CAHSI Includes).

It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478.

40. References

- J. Bourgain, J. Lindenstrauss, and V. Milman, Approximation of zonoids by zonotopes, *Acta Mathematica*, **162**(1–2), 73–141, 1989.
- J.E. Goodman and J. O'Rourke (eds.), *Handbook of Discrete and Computational Geometry*, CRC Press, Boca Raton, Florida, 1997.
- P.M. Gruber and J.M. Willis (eds.), *Handbook of Convex Geometry*, Elsevier, Amsterdam, 1993.
- S. Schön and H. Kurtterer, Using zonotopes for overestimation-free interval least squares – some geodetic applications, *Reliable Computing*, **11**, 137–155, 2004.
- T. Tao, Exploring the toolkit of Jean Bourgain, *Bulletin on the American Mathematical Society*, **58**, 155–171, 2021.